

**THERMOSTABLE DNA POLYMERASES
IN REPLICATION, REPAIR AND
BIOTECHNOLOGY**



Louise Gilroy

A thesis submitted for the Degree of Doctor of Philosophy

Institute of Cell and Molecular Biosciences

November 2013

Abstract

Many archaea contain a unique DNA polymerase, DNA Pol D. This enzyme is a heterodimer composed of a large subunit (polymerase) and a small subunit (3'-5' proof reading exonuclease). The enzyme from *Pyrococcus furiosus* is inhibited by the presence of uracil in template strands. This research has shown that a single uracil located as far as 134 bases ahead of the primer-template junction causes inhibition of replication. Further, using replication fork mimics, it is shown that, as expected, uracil on a template strand being copied by Pol D causes inhibition. Surprisingly, though, the presence of uracil on a complementary non-copied strand is also inhibitory. A model for uracil recognition by Pol D is proposed.

The biochemical properties of the individual, large and small, subunits of the Pol D heterodimer were analysed. Both subunits were found to possess activity when expressed alone although the activity was greatly reduced compared to the Pol D heterodimer. It was not possible to regain the level of activity observed in the Pol D holoenzyme by mixing the two subunits *in vitro*. This finding contributed to the hypothesis that the carboxyl-terminal region of the large subunit contains an Fe-S cluster that is lost when the protein is purified aerobically. Attempts were made to express Pol D in archaeal hosts and purify the protein with the correct metallo-status; regrettably, these were not successful.

Two thermostable bacterial family-B (pol II) DNA polymerases were cloned and expressed in *E.coli* and their biochemical properties analysed. The enzymes were found to possess many properties that make them amenable to biotechnology: polymerase activity, 3'-5' proofreading activity, high fidelity rates and the ability to bypass uracil located in template strand DNA. Unfortunately, thermostability assays revealed that the polymerases denatured on exposure to temperatures ~85°C, making them unsuitable in the PCR. Thus, further manipulation is required to determine whether the polymerases have applications in biotechnology.

Acknowledgments

I would like to take the opportunity to thank all members of lab M3008, both past and present, for their help and support throughout my PhD. Specifically, I would like to thank Pauline Heslop for her endless support, for always taking the time to help and for sharing her seemingly limitless scientific knowledge. I would also like to thank members of the Chong lab at York University and the Allers lab at the University of Nottingham for their willingness to share their time, expertise and equipment to facilitate my research. I would like to thank Professor Bernard Connolly for the opportunity to work with him and for the academic encouragement and support I have received. I have learned a huge amount under his guidance for which I shall be forever grateful.

I would also like to thank my friends and family in Glasgow for their continuous support and love throughout my 3 years in Newcastle. Without their help I would have been unable to complete this PhD. I would like to specifically mention Robert Harvey, who always took the time to call me simply to ensure I was happy- a true friend. Finally, I would like to thank David Gallagher for his love, support and encouragement over the past three years. David's ability to make me smile, no matter how badly my experiments were going, our weekends together and his perpetual belief in my abilities gave me the necessary motivation and confidence to complete my PhD. Thank you.

Key words and Abbreviations

A	Adenine
A ₂₆₀	Absorbance at wavelength 260 nanometres
A ₂₈₀	Absorbance at wavelength 280 nanometres
AAA+	ATPase associated with diverse cellular activities
Amp ₁₀₀	Ampicillin (concentration 100mg/ml)
ANS	1-anilino 9-naphthalene sulfonic acid
AP	Abasic (AP) lesions
APS	Ammonium persulfate
BCM	Barycentric wavelength mean
Bp	Base pair
C	Cytosine
Cam ₃₄	Chloramphenicol (concentration 34mg/ml)
CSR	Compartmentalized self replication
CTD	C terminal domain
Cy5	Cyanine5
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DESERVED	In depth Simple rapid small volume detection analysis
dGTP	Deoxyguanosine triphosphate
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DP1	Family D polymerase small subunit
DP2	Family D polymerase large subunit
DSF	Differential scanning fluorimetry
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuracil triphosphate
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylene diamine tetra-acetic acid
EMSA	Electrophoretic mobility shift assay
Exo	Exonuclease
FEN1	Flap structure-specific endonuclease 1
Fluor	Fluorescein
G	Guanine
Hex	Hexachlorofluorescein
His	Histidine
HPLC	High performance liquid chromatography
<i>Hvo</i>	<i>Haloferax volcanii</i>
<i>Hvo</i> YPC	<i>Haloferax volcanii</i> growth media
Kan ₅₀	Kanamycin (concentration 50mg/ml)
K _D	Dissociation constant
LB	Luria-Bertani growth medium
Ld ^{Cy5}	Leading strand primer with Cy5 primer annealed at the 5' end
Lg ^{Fluor}	Lagging strand primer with fluor annealed at the 5' end
LIC	Ligase independent cloning
<i>McCas</i>	<i>Methanococcus maripaludis</i> growth media
MCM	Mini-chromosome maintenance
<i>Mja</i>	<i>Methanocaldococcus jannaschii</i>

<i>Mma</i>	<i>Methanococcus maripaludis</i>
<i>Mth</i>	<i>Methanobacter thermoautotrophicum</i>
Neo	Neomycin
NTP	Nucleotide triphosphate
O.D. ₆₀₀	Optical density observed at a wavelength of 600 nm
ORC	Origin recognition complex
p/t	Primer-template
PCR	Polymerase chain reaction
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
<i>Pfu</i> -Pol B	<i>Pyrococcus furiosus</i> family B polymerase
<i>Pfu</i> -Pol D	<i>Pyrococcus furiosus</i> family D polymerase
Pol B	Family B polymerase
Pol D	Family D polymerase
Pol II	DNA polymerase II
Pol α	DNA polymerase alpha
Pol β	DNA polymerase beta
Pol γ	DNA polymerase gamma
Pol δ	DNA polymerase delta
Pol ε	DNA polymerase epsilon
Pol ζ	DNA polymerase zeta
Pol θ	DNA polymerase theta
Pol ι	DNA polymerase iota
Pol κ	DNA polymerase kappa
Pol σ	DNA polymerase sigma
Pur	Puromycin
<i>R.marinus</i>	<i>Rhodothermus marinus</i>
RDC	Restriction digest cloning
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase PCR
RT PCR	Real-time PCR
<i>S. YO3</i>	<i>Sulfurihydrogenibium</i> sp. <i>YO3AOP1</i>
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SDS-PAGE.	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ss	Single stranded
SSB	Single stranded binding proteins
SW	Salt water
T	Thymidine
<i>Taq</i>	<i>Thermus aquaticus</i>
TBE	Tris/Borate/EDTA buffer
TdT	Deoxynucleotidyl transferase
TEMED	Tetramethylethylenediamine
TLS	Translesion synthesis
T _m	Melting temperature
U	Uracil
UDG	Uracil-DNA glycosylase
UV	Ultra violet
WT	Wild-type

Table of Contents

Abstract.....	i
Acknowledgments.....	ii
Key words and abbreviations.....	iii
List of figures.....	ix
List of Tables.....	xv
Chapter 1 Introduction to DNA polymerases.....	1-45
1.1 Discovery of Deoxyribonucleic acid (DNA).....	2
1.2 DNA structure.....	2
1.3 Domains of life	4
1.3.1 Eukaryial domain.....	5
1.3.2 Bacterial domain.....	5
1.3.3 Archaeal domain.....	6
1.4 DNA Replication in the three domains of life.....	7
1.5 Variations in the mechanism of DNA replication among the three domains of life.....	10
1.6 Archaea as a model organism.....	18
1.7 DNA polymerases	19
1.7.1 DNA polymerase classification	20
1.7.2 DNA polymerase structure.....	21
1.7.3 DNA polymerase function	22
1.7.4 Two metal iron requirement for polymerases.....	23
1.7.5 Fidelity of polymerases and 3'-5' exonuclease activity	26
1.8 3'-5' exonuclease activity	29
1.9 Family-B polymerases.....	30
1.9.1 Family B DNA polymerases in eukaryotes.....	31
1.9.2 Family B DNA polymerases in bacteria	32
1.9.3 Family B DNA polymerases in archaea	33
1.9.4 Recognition of uracil and hypoxanthine by archaeal Pol-B.....	34
1.10 Family-D polymerases	36
1.10.1 Structure and function of archaeal Pol-D	36
1.11 Replicative polymerases in the archaea.....	37
1.12 Recognition of uracil by archaeal Pol D	38
1.13 DNA deamination.....	39
1.13.1 Repair of deaminated bases.....	42
1.14 DNA polymerases as PCR reagents.....	44
Chapter 2 Materials and methods	446-79
2.1 Oligodeoxynucleotide design, synthesis and purification	47
2.1.1 Oligodeoxynucleotide design and synthesis.....	47
2.1.2 Synthesis of a 134 base oligodeoxynucleotide	47
2.1.3 Synthesis of RNA oligonucleotides and primers for PCR	48
2.1.4 Oligodeoxynucleotide purification.....	48
2.1.5 Oligodeoxynucleotide concentration calculation	50
2.2 PCR, Real time (RT)-PCR and site-directed mutagenesis.....	51
2.2.1 Polymerase Chain Reaction (PCR)	51

2.2.2	RT-PCR.....	52
2.2.3	Site-directed mutagenesis	53
2.3	Primer list.....	54
2.4	Agarose gel electrophoresis, DNA extraction, DNA purification and DNA quantification.....	55
2.4.1	Agarose gel electrophoresis	55
2.4.2	Extraction of DNA bands	55
2.4.3	DNA purification.....	55
2.4.4	DNA quantification.....	55
2.5	Bacterial, <i>Haloferax volcanii</i> (<i>Hvo</i>) and <i>Methanococcus maripaludis</i> (<i>Mma</i>) growth media and agar	56
2.5.1	Bacterial growth media and agar	56
2.5.2	<i>Haloferax volcanii</i> (<i>Hvo</i>) growth media and agar	56
2.5.3	<i>Methanococcus maripaludis</i> (<i>Mma</i>) growth media and agar	57
2.6	Competent cell preparation, vector transformations and vector preparation.....	58
2.6.1	Competent cell preparation.....	58
2.6.2	Bacterial transformation.....	58
2.6.3	<i>Haloferax volcanii</i> (<i>Hvo</i>) transformation.....	59
2.6.4	<i>Methanococcus maripaludis</i> (<i>Mma</i>) transformation	60
2.6.5	Bacterial and <i>Haloferax volcanii</i> (<i>Hvo</i>) vector preparation.....	61
2.7	Restriction digest cloning	61
2.7.1	Vector and insert design.....	61
2.7.2	Restriction endonuclease digestion	61
2.7.3	Vector dephosphorylation	62
2.7.4	Ligation and harvesting	62
2.8	Construct preparation and gene insertion for ligase independent cloning (LIC).....	64
2.8.1	Vector preparation	64
2.8.2	LIC vector and DNA insert T4 polymerase reaction.....	64
2.8.3	LIC annealing	64
2.9	TOPO Cloning.....	65
2.10	Screening for successful cloning: restriction digest, analytical PCR, and DNA sequencing.....	66
2.10.1	Analytical restriction digests	66
2.10.2	Analytical PCR.....	66
2.10.3	DNA sequencing.....	66
2.11	Protein expression, harvesting and purification.....	67
2.11.1	Protein expression.....	67
2.11.2	Protein expression and harvesting in <i>E.coli</i>	68
2.11.3	Protein expression in <i>Haloferax volcanii</i> (<i>Hvo</i>)	68
2.11.4	Protein expression and harvesting in <i>Methanococcus maripaludis</i> (<i>Mma</i>).....	69
2.11.5	Protein purification from <i>E.coli</i> , <i>Haloferax volcanii</i> (<i>Hvo</i>) and <i>Methanococcus maripaludis</i> (<i>Mma</i>).....	69

2.12	Protein storage, analysis, identification and concentration determination	72
12.2.1	Protein storage	72
12.2.2	Protein analysis - Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	73
12.2.3	Protein identification	73
12.2.4	Protein concentration determination	74
12.2.5	Extinction coefficients of purified proteins	74
2.13	Oligodeoxynucleotide hybridisation, primer-template extension and exonuclease assays	75
2.13.1	Oligodeoxynucleotide hybridisation assays	75
2.13.2	Primer-template extension assays	75
2.13.3	Exonuclease reactions	76
2.14	Determination of KD - fluorescence anisotropy	76
2.15	Protein thermostability assays	77
2.15.1	Differential scanning flurimetry	77
2.15.2	DESERVED analysis	77
2.16	Plasmid based fidelity assay	78
2.17	DNA and protein ladders	79
	Chapter 3 Inhibition of archaeal Pol D by uracil	80-109
3.1	Background	81
3.2	Purification of <i>Pyrococcus furiosus</i> Pol D	83
3.3	Design of uracil-containing oligodeoxynucleotides	84
3.4	Synthesis of uracil-containing, long oligodeoxynucleotides	85
3.5	Primer-template annealing assays	88
3.6	Extension of uracil containing primer-templates by <i>Pfu</i> -Pol D	89
3.7	Copying of replication forks by DNA Pol D: archaeal DNA replication <i>in vivo</i>	92
3.8	Replication fork mimic design	94
3.9	Attempted "click" chemistry synthesis of replication fork mimics	95
3.10	Replication fork mimics nomenclature	96
3.11	Annealing of primers to replication fork mimic	96
3.12	Extension of replication fork mimic by <i>Pfu</i> -Pol D	97
3.13	Rate of exonucleolysis in the presence of uracil	104
3.14	Discussion	106
	Chapter 4 Family D polymerases: characterisation of the individual subunits	110-144
4.1	Background	111
4.2	Cloning <i>Mja</i> and <i>Mth</i> DP2 and DP1 subunits into expression vectors	114
4.3	Expression and purification of <i>Mja</i> and <i>Mth</i> Pol Ds as a holoenzyme and individual subunits	116
4.4	Extension of DNA primer-templates by <i>Mja</i> -Pol D, <i>Mja</i> DP2 and <i>Mja</i> -DP1	119
4.5	Extension of uracil containing DNA primer-templates by <i>Mja</i> -Pol D and <i>Mja</i> -DP2	121
4.6	Extension of uracil containing DNA primer-templates by <i>Mja</i> -Pol D and <i>Mja</i> -DP2	122
4.7	3'-5' exonucleolysis of template DNA by <i>Mja</i> -Pol D, <i>Mja</i> -DP2/ <i>Mth</i> DP1, <i>Mja</i> -DP1 and <i>Mth</i> DP1	123

4.8	Assembling Pol D by mixing <i>Mja</i> -DP2 and <i>Mja</i> -DP1 subunits in vitro...	126
4.9	Binding of Pol D and its individual subunits to DNA.....	128
4.10	Interaction of Pol D and its individual subunits to DNA.....	129
4.11	Preparation of <i>Mja</i> -Pol D lacking 3'-5' proofreading exonuclease activity.....	133
4.12	Fidelity of Pol D.....	134
4.12.1	Sequencing of mutant colonies obtained during the plasmid based fidelity assay.....	138
4.13	Discussion.....	142
Chapter 5 Overexpressing family D polymerases in archaea.....		145-171
5.1	Background.....	147
5.2	<i>Methanococcus maripaludis</i> as a model organism.....	152
5.3	Overexpressing recombinant proteins in the Euryarchaea.....	153
5.4	Cloning of <i>Mja</i> -DP2 subunit into pAW42.....	153
5.5	Expression and purification of <i>Mja</i> -DP2 from <i>Mma</i> strain S0001	154
5.6	Thermostability of <i>Mja</i> -DP2	157
5.7	Purification of <i>Mja</i> -DP2 using the N-terminal His tag	158
5.8	Purification of <i>Mja</i> -DP2 via gel filtration.....	158
5.9	Cloning of <i>Mja</i> -DP1 into pLW40	159
5.10	Expression and purification of <i>Mja</i> -Pol D from <i>Mma</i> strain S0001	161
5.11	<i>Haloferax volcanii</i> as a model organism.....	162
5.12	Overexpression of <i>Haloferax volcanii</i> Pol D in <i>Haloferax volcanii</i>	163
5.12.1	Purification of <i>Haloferax volcanii</i> (<i>Hvo</i>) Pol D	165
5.13	Cloning of <i>Haloferax volcanii</i> Pol D into pTA1392.....	166
5.14	Discussion.....	170
Chapter 6 Characterisation of bacterial DNA Polymerase II from Rhodothermus marinus and Sulfurihydrogenibium sp. YO3AOP1.....		172-215
6.1	Background.....	174
6.2	Sequence of <i>R. marinus</i> and S.YO3 DNA Pol II	175
6.3	Cloning of <i>R. marinus</i> and S.YO3 Pol II	180
6.4	<i>R. marinus</i> and S.YO3 Pol II protein purification.....	181
6.5	Extension of DNA primer-templates by DNA Pol II enzymes	185
6.6	DNA Pol II 3'-5' exonuclease assays	187
6.7	RNA polymerase activity of DNA Pol IIs.....	191
6.8	Reverse transcriptase activity of DNA Pol II's	193
6.9	DNA Pol II exonucleolysis of DNA/RNA primer-templates	195
6.10	Thermostability of <i>R. marinus</i> and S.YO3 DNA Pol II	197
6.10.1	DESERVED analysis.....	197
6.10.2	Differential scanning fluorimetry (DSF).....	201
6.11	Attempted PCR with <i>R. marinus</i> and S.YO3 DNA Pol II	203
6.11.1	Realtime PCR	207
6.12	Fidelity of <i>E. coli</i> , <i>R. marinus</i> and S.YO3 DNA Pol IIs.....	209
6.13	Discussion.....	210
6.14	Future Work	214
Chapter 7 Conclusion.....		216
Chapter 8 References.....		220

List of figures

Chapter 1

Figure 1. 1 Structure of DNA.....	3
Figure 1. 2 The 3 domains of life as determined by a 16S ribosomal sequence.	5
Figure 1. 3 The simultaneous synthesis of leading and lagging DNA strands at the replication fork.....	8
Figure 1. 4 Semi-conservative DNA replication.....	10
Figure 1. 5 DNA replication proteins and their functions in the three domains of life.	16
Figure 1. 6 Organisation of bacterial, eukaryotic and archaeal replisome machines.	17
Figure 1. 7 DNA replication reaction catalysed by DNA polymerases.....	19
Figure 1. 8 Structure of <i>E.coli</i> DNA polymerase.....	21
Figure 1. 9 A summary of the 5 main steps in the kinetic pathway of nucleotide incorporation	22
Figure 1. 10 The two metal ion mechanism of polymerisation in the context of the T7 DNA polymerase-substrate.....	25
Figure 1. 11 The proposed transition state of the two metal ion enzymatic mechanisms for the 3'-5' exonuclease reaction.....	30
Figure 1. 12 Recognition of uracil, located 4 bases ahead of the primer template junction, by <i>Pfu</i> -Pol B.....	35
Figure 1. 13 Structure of cytosine, adenine, guanine, 5-methylcytosine and their deaminated products uracil, hypoxanthine, xanthine and thymine respectively. Deamination occurs where an exocyclic amino group is lost.	39
Figure 1. 14 Deamination of adenine, cytosine and guanine.....	41
Figure 1. 15 Base excision repair pathway initiated by uracil-DNA glycosylase.	43
Figure 1. 16 Image of the polymerase chain reaction.....	45

Chapter 2

Figure 2.1 DNA and protein ladders.....	79
---	----

Chapter 3

Figure 3. 1 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing purified <i>Pfu</i> -Pol D.....	83
Figure 3. 2 Primer-templates containing template-strand uracil at a defined position.	84
Figure 3.3 Synthesis of the T134/U134 single stranded long oligodeoxynucleotides with uracil (thymine in controls) located at a defined position.	86
Figure 3. 4 PCR based method for synthesis of long oligodeoxynucleotides..	87
Figure 3. 5 Hybridisation of a Cy5 labelled primer to the T70/U70, T102/U102, and T134/U134 DNA templates.....	88
Figure 3. 6 Summary of data from Figure 3.7-Figure 3.10 showing remaining primer-template over time	90
Figure 3. 7 Primer extensino reaction with T42/U42. The percentage of starting primer extended was calculated with ImageQuant software.	90
Figure 3. 8 Primer extensino reaction with T70/U70.	91
Figure 3. 9 Primer extension reaction with T102/U102..	91
Figure 3. 10 Primer extension reaction with T134/U134..	92
Figure 3. 11 Components of the archaeal replisome.	93
Figure 3.12 106 base DNA fork mimics synthesised as a single oligodeoxynucleotide strand.	94
Figure 3. 13 Materials used for click chemistry synthesis of replication fork mimics.....	95
Figure 3. 14 Nomenclature of replication fork mimics.....	96
Figure 3. 15 TTCF replication fork mimic annealing assays visualised using a Typhoon scanner.	97
Figure 3.16 TTCF and TUCF replication fork mimic primer extension reactions.	98
Figure 3.17 TTCF and UTCF replication fork mimic primer extension reactions.	99
Figure 3.18 TUCF and UUCF replication fork mimic primer extension reactions.	100
Figure 3. 19 UUCF and UTCF replication fork mimic primer extension reactions.	101

Figure 3.20 Replication fork primer-extension reactions.....	103
Figure 3. 21 Summary of data from Figure 3.20 showing remaining primer- template against time.....	104
Figure 3. 22 Cy5 labelled primer and complementary template used in single turnover assays.....	104
Figure 3. 23 Proof reading exonucleolysis of primer-templates containing uracil (thymine in control) initiated by <i>Pfu</i> -Pol D	105
Figure 3.24 Model for the interaction of <i>Pfu</i> -Pol D with uracil.....	106

Chapter 4

Figure 4. 1 Unrooted Bayesian tree of the archaeal Euryarchaeota based on a concatenation of 57 ribosomal proteins.....	113
Figure 4. 2 Ligase independent cloning.	115
Figure 4. 3 Denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (12 %) showing the purified <i>Mja</i> and <i>Mth</i> proteins.....	117
Figure 4. 4 Oligodeoxynucleotides used in primer-extension reactions....	119
Figure 4. 5 Primer extension reactions initiated by A) 150 nM <i>Mja</i> -Pol D B) 150 nM <i>Mja</i> -DP2 C) <i>Mja</i> -DP1.	120
Figure 4. 6 Primer extension reactions initiated by A) 150 nM <i>Mja</i> -Pol D B) 150 nM <i>Mja</i> -DP2.....	122
Figure 4. 7 Primer extension reactions initiated with 150 nM <i>Mja</i> -DP2/ <i>Mth</i> -DP1.	123
Figure 4. 8 Hex labelled oligodeoxynucleotide used in exonuclease reactions.	123
Figure 4. 9 Exonuclease reactions.....	124
Figure 4. 10 Primer extension reactions.	127
Figure 4. 11 Illustration of fluorescence anisotropy.....	128
Figure 4. 12 Interaction of polymerases with single stranded control (thymine containing) and uracil containing DNA.	130
Figure 4. 13 Interaction of polymerases with single stranded control (thymine containing) and uracil containing DNA.	131
Figure 4. 14 Exonuclease reaction initiated by 80 nM <i>Mja</i> -Pol D Exo ⁻	133
Figure 4. 15 Gapped plasmid used in the fidelity assay and principle of the assay.	134

Figure 4. 16 Agarose gel electrophoresis showing filled Psj3 plasmid.....	135
--	------------

Figure 4. 17 Mutations detected, via sequencing, in 50 white colonies from the DNA polymerases during the plasmid-based fidelity assay..	139-140
--	---------

Figure 4. 18 Analysis of mutation type obtained during the plasmid-based fidelity assay.	141
--	-----

Chapter 5

Figure 5.1 Amino acid sequence conservation of the large subunit C-terminus of family D DNA polymerases.	148
--	-----

Figure 5.2 Structures of common iron-sulphur clusters. A) Rubredoxin-style iron centre B) [2Fe-2S] cluster C) [3Fe-4S] cluster D)[4Fe-4S] cluster.....	150
--	-----

Figure 5.3 Fe-S cluster destruction by oxidation.....	151
---	-----

Figure 5.4 pAW42 shuttle plasmid for protein expression in <i>Mma</i>	154
---	-----

Figure 5.5 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing heated and centrifuged <i>Mma</i> cell extract.....	155
---	-----

Figure 5.6 Mass spectrometry results identifying the large protein band highlighted on figure 5.5 as <i>Mja</i> -DP2.....	156
---	-----

Figure 5.7 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing elute samples containing <i>Mja</i> -DP2 purified from <i>Mma</i>	157
---	-----

Figure 5.8 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %).	159
---	-----

Figure 5.9 pLW40neo shuttle plasmid for protein expression in <i>Mma</i>	160
--	-----

Figure 5.10 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing elute samples containing <i>Mja</i> -DP2 and <i>Mja</i> -DP1 purified from <i>Mma</i> using heat (60°C).	161
--	-----

Figure 5.11 The pop-in/pop-out cloning system.	164
---	-----

Figure 5.12 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing samples purified from SMH610 and SMH622.....	165
---	-----

Figure 5.13 pTA1392 vector.....	167
---------------------------------	-----

Figure 5.14 Agarose gel (1 %) electrophoresis showing restriction digested plasmid following transformation of Top10 cells with ligated pTA1392 [<i>Hvo</i> -DP2] and pTA1392 [<i>Hvo</i> -DP1].	169
---	-----

Chapter 6

Figure 6.1 Figure 6.1 Alignment of DNA Pol II sequences identified in BLAST searches using <i>E.coli</i> DNA Pol II.....	176-178
Figure 6. 2 Thermocycler conditions used in touchdown PCR.....	180
Figure 6.3 Alignment of DNA Pol II sequences identified in BLAST searches using <i>E.coli</i> DNA Pol II.....	1790
Figure 6.4 Denaturing sodium dodecyl sulphate polyacrylamide gel (10 %) showing purified <i>R.marinus</i> DNA Pol II.....	181
Figure 6.5 Denaturing sodium dodecyl sulphate polyacrylamide gel (10 %) showing the purified <i>S.YO3</i> DNA Pol II protein.....	182
Figure 6.6 Mass spectrometry results identifying the large protein band highlighted on Figure 6.3 as <i>R.marinus</i> DNA Pol II.....	183
Figure 6.7 Mass spectrometry results identifying the large protein band highlighted on Figure 6.4 as <i>S.YO3</i> DNA Pol II.....	184
Figure 6.8 Primer-template extensions with bacterial DNA Pol II observed. .	186
Figure 6. 9 Oligodeoxynucleotides used in 3'-5' exonuclease reactions. "ss" represents single stranded and "Fluor" represents fluorescein.	187
Figure 6.10 3'-5' exonuclease reactions initiated with 200 nM <i>Pfu</i> -Pol B (control) or 200nM <i>E.coli</i> DNA Pol II.	188
Figure 6.11 3'-5' exonuclease reactions initiated with 200 nM <i>Pfu</i> -Pol B (control) and 200nM <i>R.marinus</i> DNA Pol II.	189
Figure 6.12 3'-5' exonuclease reactions initiated with 200 nM <i>Pfu</i> -Pol B (control) and 400 nM <i>S.YO3</i> DNA Pol II.....	190
Figure 6. 13 Fluorescent labelled DNA primer-template used in RNA polymerase assays.	191
Figure 6.14 RNA polymerase reactions.....	192
Figure 6.15 Oligonucleotide sequences used in reverse transcriptase assay.	192
Figure 6.16 Reverse transcriptase assays initiated with A) 200 nM <i>E.coli</i> Pol II B) 200 nM <i>R.marinus</i> Pol II or C) 400 nM <i>S.YO3</i> Pol II	194
Figure 6.17 3'-5' exonuclease reactions containing RNA primer-template.....	196
Figure 6.18 Total fluorescence emitted between 395 and 751 nm plotted against heating time.	198
Figure 6. 19 A graphical representation of the ANS spectra between 395 and 751 nm.	200
Figure 6.20 Barycentric wavelength mean.....	201

Figure 6.21 Differential scanning fluorimetry of DNA Pol II.	202
Figure 6.22 Differential plot of DSF data shown in Figure 6.21.....	203
Figure 6.23 Products from PCRs, targeting a ~ 200 bp region of pET28a, visualised on a 1 % agarose gel.	205
Figure 6. 24 Products from PCRs, targeting a ~ 200 bp region of pET28a, visualised on a 1 % agarose gel..	206
Figure 6.25 RT-PCR targeting a 250 bp region of the <i>S. cerevisiae</i> Pol 2 gene.	207
Figure 6. 26 Analysis of the melting temperatures of the DNA products produced during RT-PCR..	208
Figure 6.27 Primer extension and exonuclease reactions performed using modified <i>S.YO3</i> DNA Pol II.	214
Figure 6. 28 Compartmentalised self-replication.....	216

List of Tables

Chapter 1

Table 1. 1 DNA polymerase families and the roles which they each play within cells.....	20
Table 1. 2 Fidelity rate of some well characterised DNA polymerases.	28
Table 1. 3 A summary of half lives of common chemical bonds contained within normal double-stranded DNA at 25°C and 92°C.....	40

Chapter 2

Table 2. 1 Sequence of oligodeoxynucleotides used to synthesise single stranded T134/U134 templates.....	48
Table 2. 2 Buffer gradients used in HPLC purification of oligodeoxynucleotides based on the 3' purification handle.	49
Table 2. 3 Extinction coefficients (260 nm) of DNA bases and fluorophores used in oligodeoxynucleotide synthesis.	50
Table 2. 4 Oligodeoxynucleotide sequence of primers used in PCR amplification, DNA sequencing, and site-directed mutagenesis reactions.....	54
Table 2. 5 Summary of the vectors used to clones DNA inserts and the cloning technique that was utilised.....	63
Table 2. 6 Summary of expression vectors used in this thesisd.	67
Table 2. 7 Buffers used in the first column chromatography purification step of proteins over-expressed in <i>E.coli</i> , <i>Hvo</i> and <i>Mma</i>	70
Table 2. 8 Buffers used for the long term storage of purified proteins from <i>E.coli</i> , <i>Hvo</i> and <i>Mma</i>	72
Table 2. 9 Extinction coefficients of proteins purified at 280 nm in $\text{mM}^{-1}\text{cm}^{-1}$...74	

Chapter 3

Table 3. 1 DNA sequence of oligodeoxynucleotides used to synthesise single stranded T134/U134 templates.....	85
--	----

Chapter 4

Table 4. 1 Summary of vectors used to clone the <i>Mja</i> and <i>Mth</i> DP2 and DP1 subunits.	114
Table 4. 2 Summary of vectors used to express Pol D and the proteins that were successfully purified.	116
Table 4. 3 Reaction buffers used in exonuclease reactions.	125
Table 4. 4 Oligodeoxynucleotide used in fluorescence anisotropy.	129
Table 4.5 The K_D values for the binding of protein (mean \pm standard deviation from at least three determinations) to oligodeoxynucleotides containing uracil (thymine in controls).....	131
Table 4. 6 Error rate of DNA polymerases determined using pSJ3 in the plasmid based fidelity assay.	136

Chapter 5

Table 5.1 All eukaryotic family-B polymerases contain two cysteine motifs, designated CysA (binds Zn^{2+}) and CysB (4Fe-4S centre).	149
Table 5.2 <i>Haloferax volcanii</i> strains and their associated properties.	163

Chapter 6

Table 6. 1 Buffers used in PCRs by Phusion, <i>R.marinus</i> DNA Pol II and S.YO3 DNA Pol II.	205
Table 6. 2 Error rates of DNA polymerases determined using the gapped, pSJ3.	209

Chapter 1

Introduction to DNA polymerases

1.1 Discovery of Deoxyribonucleic acid (DNA)

In 1865, Gregor Mendel's experiments with peas showed that certain traits were inherited in "packages", now referred to as genes. Shortly after Mendel's discovery, Friedrich Miescher isolated a "compound" from the nucleus of a white blood cell which he referred to as "nuclein" (Dahm, 2008). However, the relationship between nuclein (now referred to as deoxyribonucleic acids, DNA) and genes remained unknown until 1944 when Oswald Avery transferred the ability to cause disease from one strain of bacteria to another (Avery *et al.*, 1995). During his research, Avery had moved DNA from one bacterium to another, and noticed that the "genes" and the associated traits were transferred. This finding combined with experiments using bacteriophage conducted by Hershey and Chase (Hershey and Chase, 1952) led to the conclusion that genes are made of DNA.

By the mid 1940's the importance of DNA in inheritance was widely accepted, however the structure remained unknown until 1953 when James Watson and Francis Crick proposed the double-helix model (**Figure 1. 1**) (Watson and Crick, 1953). Watson and Crick's double-helix model was based on a photograph (photograph 51) of DNA produced via X-ray crystallography by Raymond Gosling and Rosalind Franklin (Franklin and Gosling, 1953). Since Watson and Crick's original discovery scientists have made some minor changes to the double-helix model (Richard R. Sinden *et al.*, 1998), however, the model's major features remain the same today.

1.2 DNA structure

DNA exists within cells as a helical polymer made of 2 anti-parallel strands that are connected by hydrogen bonds (**Figure 1. 1**). Each DNA strand is composed of deoxyribonucleotides, compounds that consist of a base, a pentose sugar and a phosphate residues (Watson and Crick, 1953). The backbone of each DNA strand is made of pentose sugars, joined together by phosphate residues, which form phosphodiester bonds between the fifth and third carbon atoms of adjacent sugar rings (**Figure 1. 1**). Each DNA backbone has bases attached to the 1' carbon atom of the sugar rings that interact, via hydrogen bonding, with

complementary bases (adenine: thymine and guanine: cytosine) on the opposite DNA strand (Figure 1. 1); this interaction is commonly referred to as “base-pairing”. It is this sequence of complementary bases that is responsible for encoding genetic information.

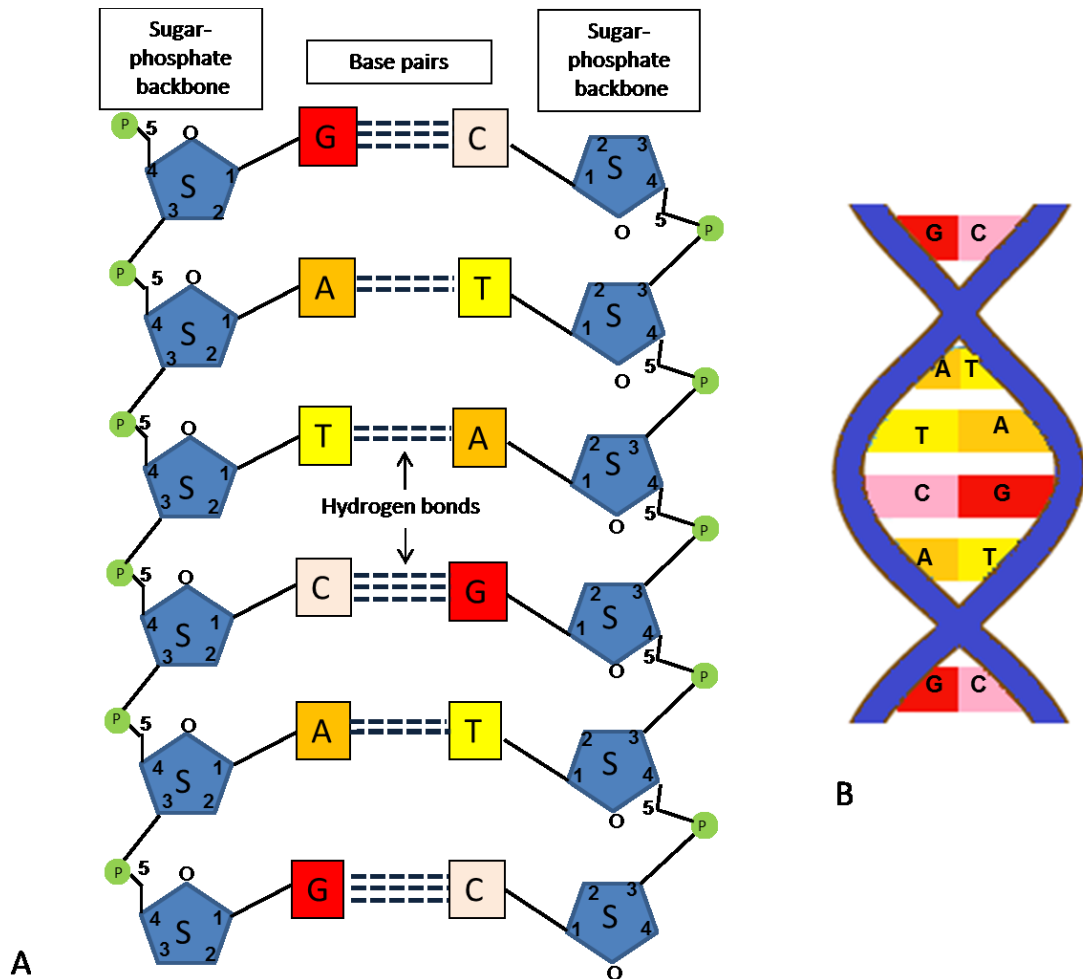


Figure 1. 1 Structure of DNA. A) The sugar-phosphate backbone held together by phosphodiester bonds and the complementary base-pairs held together by hydrogen bonds. B) The double-helix conformation.

1.3 Domains of life

The “three domains of life” classification system, which splits cellular organisms into three distinct categories (domains) based on DNA sequence and molecular structures was proposed by Carol Woese in 1990 (Woese *et al.*, 1990). The three domains: Archaea, Bacteria and Eukaryotes, are believed to have evolved separately from a single ancestor known as a progenitor (Woese *et al.*, 1990) (Figure 1. 2). Evolution from the progenitor split in two directions and formed the bacterial domain and a further lineage which later diverged to form the archaeal and eukaryal domains (Woese *et al.*, 1990; Leipe *et al.*, 1999) (Figure 1. 2). Thus the archaea and eukarya domains share a more recent common ancestor than the bacteria.

The three domains are distinguished based on differences such as variation in ribosomal RNAs; archaeal 16S rRNAs can be identified by a unique structure found in the regions between position 180 and/or between positions 405 and 498 (Woese *et al.*, 1983). Molecular similarities within each domain can also be used. Thus bacteria possess highly similar subunit patterns (in terms of numbers and sizes) in their RNA polymerases, which is unlike that seen in either the archaea or eukaryotes (Huet *et al.*, 1983).

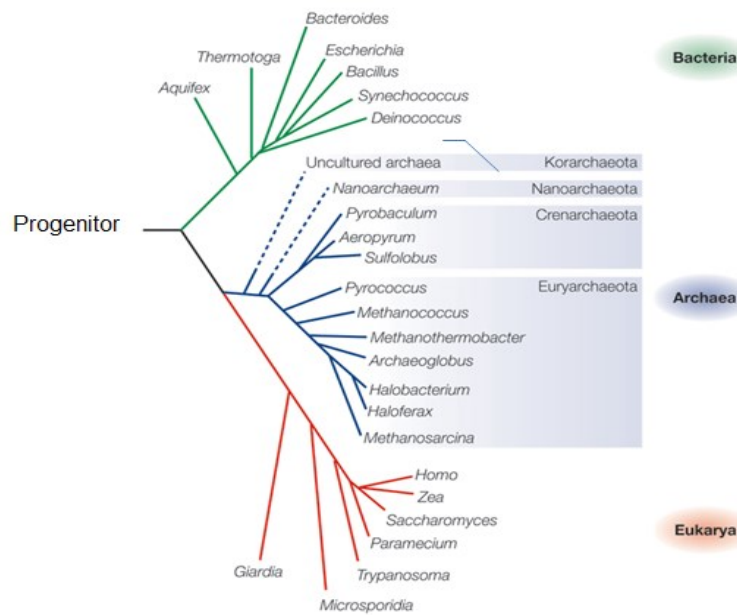


Figure 1. 2 The 3 domains of life as determined by a 16S ribosomal sequence.
 Taken from Allers and Mevarech, 2005.

1.3.1 Eukaryal domain

The eukaryal domain contains all multicellular organisms and many unicellular organisms (e.g. protozoa) and incorporates the Animalia, Plantae, Fungi and Protista kingdoms (Margulis, 1996). All members of the eukarya possess a membrane bound nucleus which contains DNA stored on linear chromosomes (Alberts *et al.*, 2002). The majority of eukaryotes also possess other membrane-bound organelles that are responsible for a wide range of cell functions including: energy production (mitochondria), processing and modification of proteins (Golgi apparatus) and photosynthesis (chloroplasts). The presence of membrane bound organelles allows different biochemical environments to exist within one eukaryotic cell (Bock *et al.*, 2001).

1.3.2 Bacterial domain

All members of the bacterial domain are prokaryotic and as such they are all unicellular organisms (White, 2007). Unlike eukaryotes, bacteria do not possess a nucleus or other membrane-bound organelles. Instead, their

intracellular (water-soluble) apparatuses are located together within the cytoplasm and surrounded by a single, lipid cell membrane (Gitai, 2005). This cell membrane acts as a barrier and holds the essential components of the cytoplasm within the cell (e.g. proteins and nutrients). The most important cellular component, the genome, is normally stored on a circular double-stranded piece of DNA held in a region of the cytoplasm, within the cell membrane, called the nucleoid (Shih and Rothfield, 2006).

1.3.3 Archaeal domain

Despite being unicellular prokaryotes and appearing morphologically similar to bacteria; archaea share a more recent common ancestor with eukarya (Figure 1. 2) (Woese *et al.*, 1990; Leipe *et al.*, 1999). Thus, even though they do not possess a nucleus or any membrane bound organelles, the information forming processes and functions within archaea (including DNA replication, translation, and transcription) are more closely related to those found in eukaryotes than bacteria (Barns *et al.*, 1996; Ishino and Ishino, 2012).

The archaeal domain is currently subdivided into 6 distinct phyla: Crenarchaeota, Euryarchaeota, Korarchaeota, Nanoarchaeota, Thaumarchaeota and Aigarchaeota (Brochier-Armanet *et al.*, 2011; Ishino and Ishino, 2012). The largest and best characterized phyla are the Crenarchaeota and Euryarchaeota. The Crenarchaeota consists of hyperthermophilic and thermophilic aerobic species, whereas the Euryarchaeota includes all known methanogens and halophiles as well as some thermophilic species. Many Crenarchaeota and Euryarchaeota species have had their genome sequenced and are used as model organisms to study eukaryotic DNA replication and repair pathways (Leigh *et al.*, 2011).

The other 4 phyla, Korarchaeota, Nanoarchaeota, Thaumarchaeota and Aigarchaeota, have been proposed recently and, so far, contain few members (Ishino and Ishino, 2012). The Korarchaeota phylum consists of mainly uncultivated organisms that have been found in low abundance in high temperature hydrothermal environments (Elkins *et al.*, 2008; Miller-Coleman *et al.*, 2012). The Nanoarchaeota phylum, introduced in 2002, only has one

member, *Nanoarchaeum equitans* (Huber *et al.*, 2002). *Nanoarchaeum equitans* has the second smallest, non-viral, cellular genome ever sequenced (490,855 bases) and grows and divides on the surface of other archaea (Huber *et al.*, 2002).

The Thaumarchaeota currently consists of four members, all of which are chemolithoautotrophic ammonia-oxidizers (Hallam *et al.*, 2006) and believed to play an important role in biogeochemical cycles (e.g. nitrogen and carbon cycles). Members of this phylum contain a form of type I topoisomerase that is commonly found in eukaryotes and that had never before been identified in archaea (Forterre *et al.*, 2007; Brochier-Armanet *et al.*, 2008). The Aigarchaeota, the most recently identified phyla, possess genes encoding Euryarchaeota Pol D and crenarchaeotic Pol BII (Brochier-Armanet *et al.*, 2008; Nunoura *et al.*, 2011). This phylum shows strong sequence similarity with the Thaumarchaeota, and thus strong debate exists over whether it should be classified as a distinct phyla or whether its members should be included within the Thaumarchaeota (Nunoura *et al.*, 2011).

1.4 DNA Replication in the three domains of life

DNA replication is the essential process in which cells replicate their entire genome and is the foundation of biological inheritance. DNA replication is an ordered, multifaceted process that occurs at the replisome (a multi-protein complex) and consists of three defined steps: initiation, elongation and termination. Across the three domains of life, the overall mechanism of DNA replication is highly conserved (Stillman, 2005), however, important variations in the process exist (Robinson and Bell, 2005; Dahm, 2008; O'Donnell *et al.*, 2013)

Initiation of DNA replication occurs at specific sites within the genome referred to as the “origins of replication”. The origins of replication are sequences of the genome where origin binding proteins bind. These origin binding proteins initiate limited unwinding of the DNA duplex at the runs of A-T base-pairs. Following initial unwinding a DNA helicase encircles each of the strands and

continues the process of strand separation by breaking the hydrogen bonds between complementary bases (Yao and O'Donnell, 2009). The helicase moves incrementally and directionally along the DNA strand, unzipping the double helix and forming a replication fork (Figure 1. 3). All single strands exposed by helicase are bound by a single stranded binding protein (Meyer and Laine, 1990; Dickey *et al.*, 2013).

The exposed bases in the leading and lagging strands of the replication fork then serve as a template for DNA synthesis (Figure 1. 3). However, as DNA polymerases require a free 3' hydroxyl group to initiate DNA synthesis; a primase is employed, prior to elongation by the DNA polymerase, to synthesise a short RNA primer with a free 3' OH group (Figure 1. 3) (Griep, 1995).

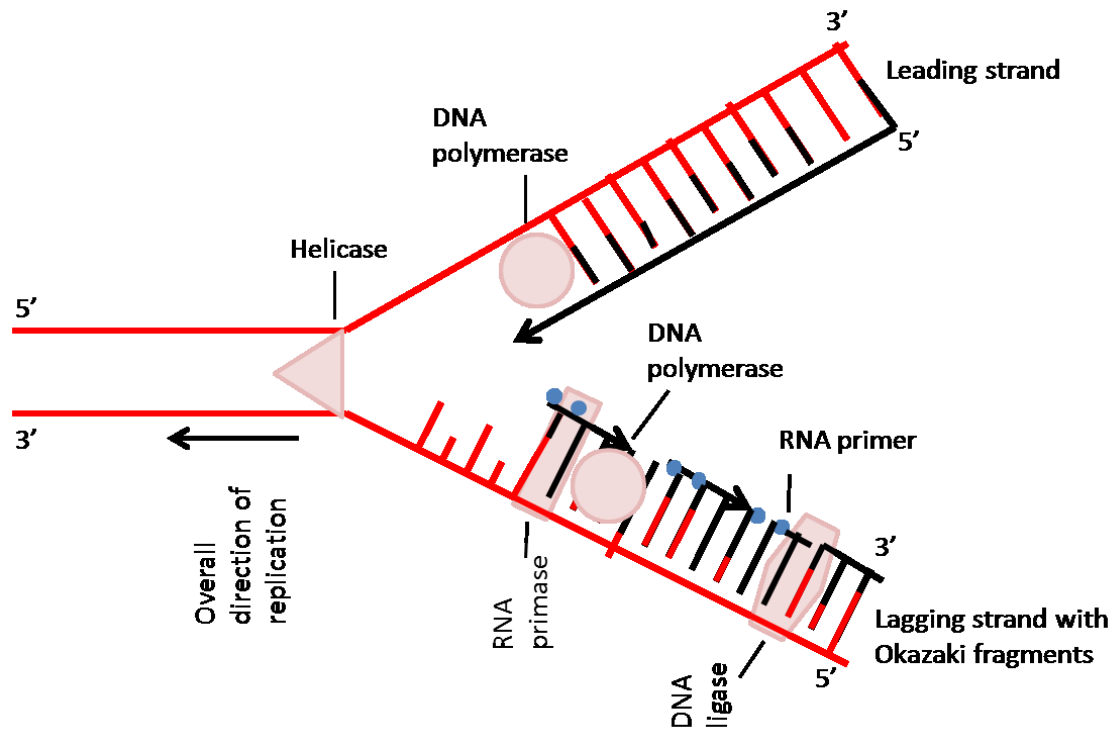


Figure 1. 3 The simultaneous synthesis of leading and lagging DNA strands at the replication fork. The leading strand is formed through continuous replication while the lagging strand is synthesised in short 5'-3' Okazaki fragments.

As DNA replication moves along the parent strand in the 3'-5' direction (with the polymerase adding bases in 5'-3' direction), replication of the leading strand is continuous (Figure 1. 3). However, as polymerases are unable to synthesise DNA in the 3'-5' direction, replication of the anti-parallel, lagging strand is more complicated. Rather than the continuous replication observed in the leading strand, the lagging strand is replicated in small segments called Okazaki fragments (Okazaki *et al.*, 1968; Lehman, 1974). Primases synthesise short RNA primers on the lagging strand which act as a template for the DNA polymerases to initiate polymerisation and create the Okazaki fragments. After elongation by the DNA polymerase, the RNA primers are removed using either the 5'-3' exonuclease activity of a DNA polymerase or by a flap endonuclease (Qiu *et al.*, 1999; Liu *et al.*, 2004). The separate Okazaki fragments are joined together by DNA ligase to produce a single strand of DNA. When both the leading and lagging strands have been copied, the two new duplex DNA strands separate and DNA replication is terminated (Figure 1. 3). However, many other proteins are essential for efficient DNA replication. These include sliding clamps (proteins that encircle the DNA and interact with the polymerase conferring processivity) and "clamp loaders" which form multiple contacts to other replisome proteins holding the entire ensemble together (Jeruzalmi *et al.*, 2002; O'Donnell *et al.*, 2013).

Thus, during DNA replication, each new daughter cell receives an exact copy of the genetic material from the original parent cell (Hanawalt, 2004). This form of replication is referred to as "semi-conservative" as the two daughter cells inherit a DNA double-helix that contains one DNA strand from the parent cell and one newly copied DNA strand synthesised during DNA replication (Meselson and Stahl, 1958) (Figure 1. 4).

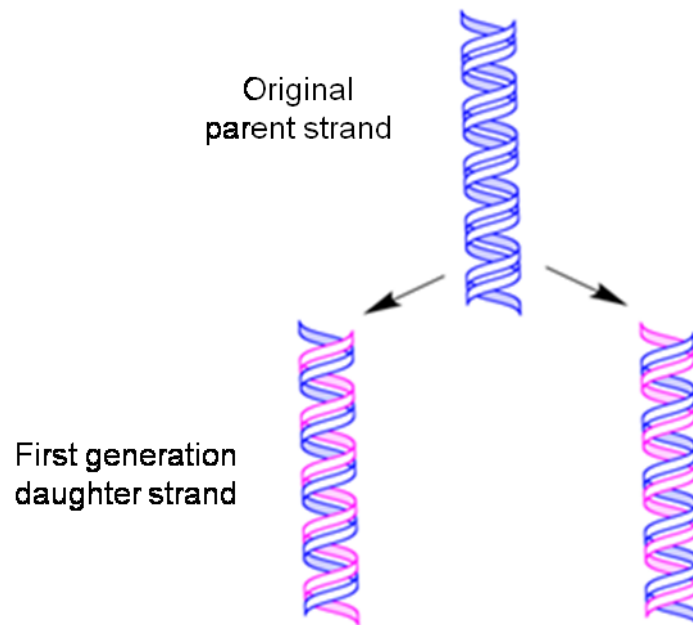


Figure 1. 4 Semi-conservative DNA replication as proposed by Meselson and Stahl in 1958. The daughter cells receive one parental DNA strand and one newly synthesised DNA strand.

1.5 Variations in the mechanism of DNA replication among the three domains of life

Despite similarities in the overall mechanism of DNA replication among the three domains of life, many variations exist (**Figure 1. 5** & **Figure 1. 6**) (Dahm, 2008). For example, in all three domains, DNA replication is initiated at the origins of replication. However, while bacteria have one origin of replication located on a single chromosome (Mott and Berger, 2007); eukaryotes possess multiple chromosomes with many origins of replication (Nasheuer *et al.*, 2002). This is in contrast to archaea where variations exist between species, some species possess a single chromosome with a single origin and others possess multiple chromosomes with many origins of replication (Kelman and Kelman, 2004; Kelman and White, 2005).

In all organisms origins of replication are activated by origin binding proteins, normally members of the AAA+ family that function as multimeric machines

(Erzberger and Berger, 2006). These proteins assist in the melting of double stranded DNA and the initiation of DNA replication. However, the origin binding proteins and the manner in which DNA replication is initiated differs between the three domains (Robinson and Bell, 2005).

In bacteria, multiple copies of the origin binding protein, DnaA, form a helical filament that binds to the origin of replication (Kaguni, 2011). This DnaA filament binds ATP and unwinds the A/T rich regions of the origin resulting in a single stranded DNA bubble onto which the replicative helicase is loaded. Unlike bacteria, eukaryotes contain six origin binding proteins (5 of which are related to AAA+ proteins) referred to as the ORC (origin recognition complex) (Stillman, 2005). The ORC subunits together with another AAA+ family protein, Cdc6, form a ring shaped hexamer that binds DNA (Sun *et al.*, 2012). Unlike bacterial DnaA, the ORC does not unwind the DNA directly; rather it requires further protein interactions (Bell, 2002). In archaea, the initiation of replication is believed to be similar to that observed in eukaryotes as archaea contain homologues of eukaryotic origin binding proteins. Almost all sequenced archaea have been found to possess a homologue of the largest member of the ORC, Orc1 and the Cdc6 protein (Myllykallio and Forterre, 2000). It is likely that these proteins bind to the origins of replication, which causes a distortion and localized melting in the DNA. However the number of origin binding proteins appears to differ between archaeal species (Barry and Bell, 2006), and the precise functional mechanism of these proteins is currently unknown and requires further investigation (Kelman and Kelman, 2004).

In all organisms, after the origin binding proteins have bound at the origin of replication, helicase is loaded at the replisome prior to the binding of DNA polymerase. Helicase is a six subunit complex, found in all 3 domains that unwinds double stranded DNA by encircling one strand of the parental DNA duplex and moving along to unzip the helix and form the replication fork (Gai *et al.*, 2010; Wu, 2012). Despite the similarities in the functioning of helicase, there are variations in helicase structure across the three domains (Kelman and Kelman, 2004).

In bacteria the DnaC protein assists with the loading of the DnaB helicase, a homohexamer, based on a Rec-A protein, that travels along single stranded DNA in the 5'-3' direction (Kelman and Kelman, 2004). However, in eukaryotes, MCM, a family of six proteins (MCM₂₋₇) with highly conserved amino acid sequences, is responsible for helicase activity. The MCM_{4, 6, 7} heterotrimer possesses 3'-5' DNA helicase activity, single-stranded DNA binding and ATPase activities and is believed to be the eukaryotic helicase (Tye, 1999; Tye and Sawyer, 2000). Evidence suggests that the MCM_{2, 3 & 5} assist in stabilization of this complex. Interestingly, at least one MCM homologue has been found in all sequenced archaeal species (Myllykallio and Forterre, 2000). Archaeal helicase has similar biochemical properties as those identified in eukaryotic helicase: 3'-5' helicase activity which is dependent on ATP, single stranded binding properties and DNA dependent ATPase activity (Chong *et al.*, 2000; Shechter *et al.*, 2000; Carpentieri *et al.*, 2002). However, archaeal MCM helicase interacts with a bacterial-like DnaG primase to form a unique primosome complex involved in synthesising primers on the leading and lagging strand (You *et al.*, 2013). The exact mechanism of this primosome complex and the impact on helicase formation is currently unknown and requires further investigation.

Single stranded binding (SSB) proteins play essential roles in DNA replication in all three domains of life. They have been identified in all organisms except the Thermoproteales, a group of extremophile archaea, in which the SSB proteins have been displaced by the ThermoDBP protein (Paytubi *et al.*, 2012). SSB proteins bind to single-stranded (ss) regions of DNA. They serve to prevent premature annealing, protect ssDNA from being digested by nucleases and to remove secondary structures to allow other enzymes to function effectively (Curth *et al.*, 1996; Dickey *et al.*, 2013). SSB proteins also bind to and regulate the function of other proteins involved in DNA replication (Dickey *et al.*, 2013).

Most bacterial SSB proteins work as homo-tetramers (Mijakovic *et al.*, 2006); the monomers are built from two fragments (Genschel *et al.*, 1996; Purnapatre and Varshney, 1999). The N terminal fragment is ~ 120 amino acids and possesses conserved residues that are responsible for binding to ssDNA,

tetramerization and stabilization of the monomer fold (Williams *et al.*, 1983; Carlini *et al.*, 1998). The C-terminal fragment is responsible for interactions with other proteins (Handa *et al.*, 2001). Despite low sequence homology with bacterial SSB proteins, eukaryotic SSB proteins, display similar biochemical properties. They are normally hetero-tetramers and crystal structures have revealed that many form an almost identical fold as in the structure of bacterial SSB proteins (Curth *et al.*, 1994). Replication protein A (RPA), identified as a eukaryotic nuclear ssDNA-binding protein (Smith *et al.*, 1997), is a hetero-trimer composed of subunits of 70, 32 and 14 kDa, each of which is conserved in all eukaryotes (Brill and Stillman, 1991; Wold, 1997). The whole RPA protein possesses four ssDNA-binding domains which share structural similarities (Bochkareva *et al.*, 2000). A special conserved feature of the RPA70 protein not observed in bacterial SSBs is the presence of a C-4 zinc-finger motif in the C-termini, which is required for effective function of RPA proteins (Lin *et al.*, 1998; Bochkareva *et al.*, 2000).

The first archaeal SSB to be identified was a monomer, built from four tandem repeats found in the genome sequence of *Methanococcus jannaschii* (Kelly *et al.*, 1998). *Mja*-SSB protein was found to have a high level of sequence similarity to that of eukaryotic RPA and found to have similar biochemical properties (e.g. binding affinity) (Kelly *et al.*, 1998). Homologues have since been found in other archaeal species and they have been found to contain the conserved zinc finger domains also found in eukaryotes (Kelman *et al.*, 1999). Additionally, unique SSB proteins have been identified in thermophilic archaea that are found as homodimers (Dabrowski *et al.*, 2002). It is believed that these unique SSB proteins arose from the fusion of two ssDNA binding domains into one polypeptide (Dabrowski *et al.*, 2002).

Despite differences in SSBs from the three domains of life, analysis of structure and sequence has shown regions of homology indicating that they all evolved from a single, common ancestor (Dabrowski *et al.*, 2002). The ancestral SSB protein probably possessed four ssDNA-binding domains with the features identified in the core of all currently known SSB proteins with an OB fold (Murzin, 1993).

DNA polymerases are responsible for DNA replication and repair pathways in all organisms (Rothwell and Waksman, 2005). However, they are incapable of *de novo* DNA synthesis and thus DNA primases exist to synthesise short regions of RNA to act as primers (Frick and Richardson, 2001). Eukaryotic primases, are comprised of two subunits of ~ 58 kDa and 48 kDa - and are normally found in a complex with DNA polymerase α (Arezi and Kuchta, 2000). The pol α -primase complex contains an additional, fourth subunit with a molecular weight of approximately ~ 70-90 kDa (Collins *et al.*, 1993). This subunit does not bind to the primase subunits, but binds tightly to the pol α and is believed to play a role in tethering the pol α -complex to other proteins within the replication fork. The 48 kDa subunit of the pol α -complex synthesises a short RNA primer of ~ 12 nucleotides and pol α extends the RNA primer with ~ 25 nucleotides of DNA making an RNA/DNA hybrid primer (Muzi-Falconi *et al.*, 2003).

Archaeal DNA primase is believed to be similar to eukaryotic primase; however, some variations do exist. *Thermococcus kodakaraensis* DNA primase complex is a heterodimer containing p41 and p46 subunits with the catalytic activity located within the p41 subunit (Galal *et al.*, 2012). Similar to the eukaryotic complex, the archaeal primase-complex is able to synthesise both DNA and RNA. However, the *T.kodakaraensis* primase complex preferentially interacts with dNTP's rather than the ribonucleoside triphosphates required by the eukaryotic homologue and it is able to initiate RNA as well as DNA chains *de novo* (Galal *et al.*, 2012).

DNA primase within bacteria differs from the priming complex identified in eukaryotes and archaea (Corn and Berger, 2006). The bacterial primase, DnaG, transiently binds DnaB helicase to synthesise an RNA primer ~ 12 nucleotides long (Kaguni, 2011). Binding of DnaG primase to DnaB stimulates the release of the regulatory protein DnaC from DnaB indicating that initial priming and unwinding are tightly coordinated (Kaguni, 2011).

Following the creation of RNA or RNA/DNA primers, DNA polymerases bind and replicate the single stranded DNA at the replication fork. Within eukaryotes family B polymerases, Pol α , Pol δ , and Pol ϵ are responsible for DNA

replication (Rothwell and Wakemen, 2005). Pol α (family-B) works as a primase to synthesise short RNA/DNA primer while Pol ϵ and Pol δ are responsible for leading and lagging strand synthesis respectively (Jin *et al.*, 2003; Muzi-Falconi *et al.*, 2003; Shikata *et al.*, 2006; Eckardt, 2009). The family B polymerases contain strong, 3'-5' proofreading exonuclease activity and possess high fidelity rates and thus ensure accurate replication of the genome (Capson *et al.*, 1992; Lin *et al.*, 1994).

It was originally believed that family B polymerases were also responsible for DNA replication within all archaeal species (Grabowski and Kelman, 2003). However, a hypothesis now exists that suggests Pol D, a recently discovered, novel polymerase, found in all archaea except the Crenarchaea phyla, is involved in DNA replication in these archaea (Cann *et al.*, 1998; Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013). Studies of the biochemical properties of family D polymerases have identified many properties associated with replicative polymerases (Cann *et al.*, 1998; Henneke *et al.*, 2005; Tori *et al.*, 2007). However, all species of archaea possess at least one copy of a family B polymerase, and it is currently unknown whether the family B and D polymerases work together to replicate the genome, or if one polymerase is responsible. However, as family B polymerases are the only replicative polymerases that have been identified in Crenarchaea; they are believed to be responsible for Crenarchaeal replication.

Within bacteria, family C polymerases are responsible for DNA replication (Kronberg and Baker, 1992; Rothwell and Wakemen, 2005). The family C polymerase holoenzyme interacts with other proteins and forms a large multi-subunit complex consisting of at least 10 subunits. The α -subunit of the holoenzyme contains the DNA polymerase activity that is tightly associated with the subunit which contains a 3'-5' exonuclease activity (Rothwell and Wakemen, 2005). Thus, despite all three domains employing polymerases to extend DNA, the mechanism differs between domains. As such, replication speed, and fidelity varies between the three domains of life.

Thus, despite the same core components (origins of replication, helicases, primases and DNA polymerases) being located in all cells, the manner in which

these proteins are arranged and interact differs among the three domains of life (Figure 1.5, Figure 1.6) (Avery *et al.*, 1995; Kelman and White, 2005; Stillman, 2005; Barry and Bell, 2006). Due to the large range of variations in the DNA replication mechanism between domains, they are not all discussed in great detail here. However, it should be highlighted that archaeal DNA replication appears to be more similar to that observed in eukaryotes, than bacteria (Figure 1.5) (Leipe *et al.*, 1999). Figure 1.5 & 1.6 briefly summarise some key differences in proteins involved in DNA replication between the three domains of life.






Stage of DNA replication	Bacteria	Archaea	Eukaryote
Initiation			
 Origin recognition	DnaA	Cdc6/Orc1	ORC (Orc1, 2, 3, 4, 5, 6)
 DNA unwinding	DnaC DnaB	Cdc6/Orc1 MCM GIN5 (Gins23, Gins51)	Cdc6 Cdt1 MCM (Mcm2, 3, 4, 5, 6, 7) GIN5 (Sld5, Psf1, 2, 3) Cdc45
 Primer synthesis	DnaG	DNA primase	Pol α / primase
Elongation			
 DNA synthesis	Family C DNA polymerase (Pol III) Clamp loader (γ -complex) Clamp (β -clamp)	Family B DNA polymerase (Pol B) Family D DNA polymerase (Pol D) Clamp loader (RFC) Clamp (PCNA)	Family B DNA polymerase (Pol δ) (Pol ϵ) Clamp loader (RFC) Clamp (PCNA)
 Maturation	Pol I RNaseH DNA ligase	Fen1 Dna2 DNA ligase	Fen1 Dna2 DNA ligase

Figure 1.5 DNA replication proteins and their functions in the three domains of life. Most of the archaeal proteins are designated as homologues of the proteins identified from eukaryotic DNA replication studies. Taken from Ishino and Ishino., 2012.

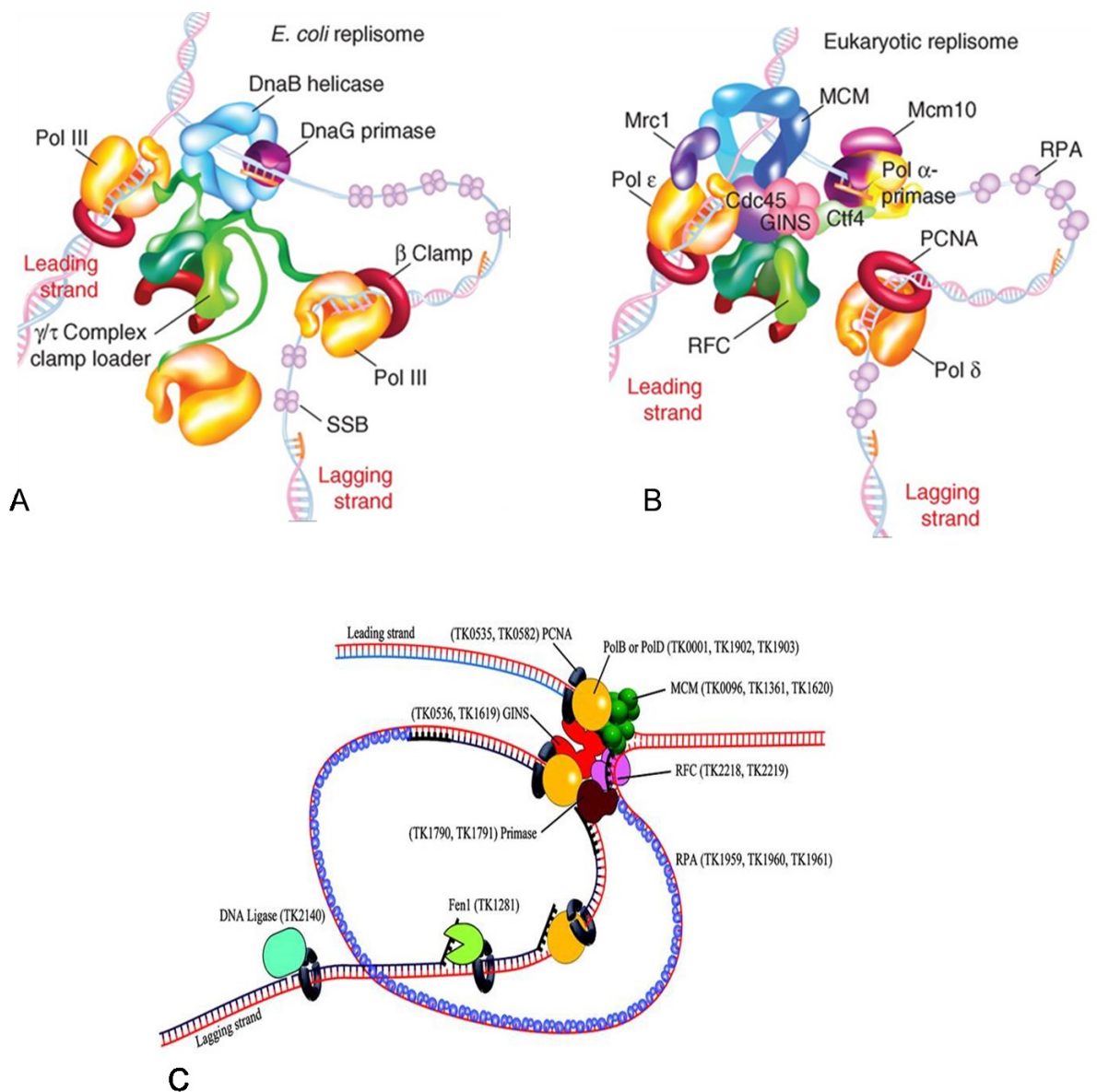


Figure 1. 6 Organisation of bacterial, eukaryotic and archaeal replisome machines. A) Replisome architecture in *E.coli*. B) Proposed architecture of a eukaryotic replisome. C) Components of the archaeal replisome with the *T. kodakaraensis* numerical gene designations listed adjacent to the protein sub complexes. Image A and B taken from O' Donnell *et al.*, 2013 and image C taken from Li *et al.*, 2010.

1.6 Archaea as a model organism

Despite being prokaryotic, archaeal cells contain many proteins that are highly homologous to those involved in eukaryotic genetic information processing pathways including: DNA replication, transcription and translation (Yutin *et al.*, 2008). Thus, due to the greater simplicity of archaea and the inherent difficulties of working with eukaryotes, some archaeal species are utilised as model organisms (Leigh *et al.*, 2011).

The natural properties of many archaeal species including ease of growth on solid and liquid media, short generation time, and thermostability have made them ideal target model organisms. However, the need for specialist growth media and strict anaerobic conditions for some species can be disadvantageous as specialist equipment is required (Hendrickson *et al.*, 2004; Leigh *et al.*, 2011). The development of genetic systems (including methods for transformation, and gene-knock-out), selectable markers and plasmid vectors for some archaeal species have enabled *in vitro* and *in vivo* biochemical and genetic analysis of the DNA replication, DNA repair, transcription and translation pathways (Leigh *et al.*, 2011).

The majority of archaeal model organisms are members of the Crenarchaea and Euryarchaea phyla. These phyla contain many hyperthermophilic species that are often used to isolate and analyse the structure and function of replisome components (Ishino and Ishino, 2012). These hyperthermophilic species are targeted as the structural integrity of thermophilic proteins makes them easier to study than their mesophilic counterparts (Razvi and Scholtz., 2006). Other well developed archaeal model systems include methanogens (e.g. *Methanococcus maripaludis*) and halophiles (e.g. *Haloferax volcanii*). The methanogens are often used to gain understanding of anaerobic cellular processing while excellent genetic tools, available for the halophiles, enable complex processes including genetic recombination and structural genomics to be analysed (Leigh *et al.*, 2011).

1.7 DNA polymerases

Although the replication and repair of DNA utilises many enzymes and cellular regulators (Hubscher *et al.*, 2002; Barry and Bell, 2006), DNA polymerases are the essential enzymes responsible for DNA replication within living cells (Rothwell and Waksman, 2005). They function by catalysing the addition of deoxyribonucleotide triphosphates (dNTPs) to the 3' OH- group of growing DNA strands (Lehman *et al.*, 1958) (Figure 1.7). The process of deoxyribonucleotide incorporation is usually carried out by a polymerase while reading from a template polynucleotide during semi-conservative replication (Figure 1. 4) (Meselson and Stahl, 1958).

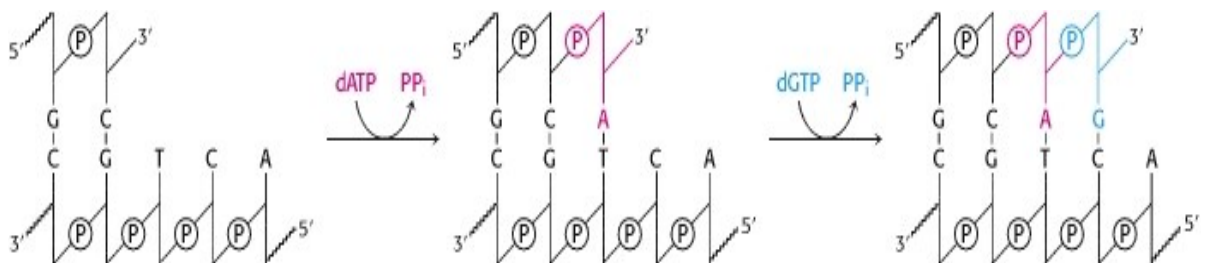


Figure 1. 7 DNA replication reaction catalysed by DNA polymerases. The DNA polymerase catalyses the addition of deoxyribonucleotide triphosphates (dNTPs) to the 3' OH- group of the growing DNA strands. Image taken from Berg JM, 2002.

In addition to their intrinsic polymerase activity, many DNA polymerases boast additional properties such as 3'-5' exonuclease activity, 5'-3' exonuclease activity, reverse transcriptase activity, DNA repair activities, and the ability to replicate mismatched DNA (translesion synthesis) (Bebenek and Kunkel, 2004; Joyce and Benkovic, 2004; Rothwell and Waksman, 2005). These properties assist in many cellular processes, including DNA replication, DNA repair, genetic recombination and reverse transcription, to ensure the high fidelity replication of the genome (Steitz, 1999).

1.7.1 DNA polymerase classification

DNA polymerases found in the 3 domains can be categorized into 7 polymerase families: A, B, C, D, X, Y, and RT (Table 1. 1) based on sequence homology (Braithwaite and Ito, 1993; Rothwell and Waksman, 2005). Some polymerase families are well conserved and found in all 3 domains of life, (e.g. family B) while others are believed to be specific to individual domains (e.g. family D).

Family	Description
A	Found in eukaryotes, bacteria and archaea, family A polymerases are involved in the DNA replication and repair processes. The replicative T7 DNA polymerase is the best characterized member (Franklin <i>et al.</i> , 2001). <i>Escherichia coli</i> DNA Pol I, is an example of a repair enzyme and is involved in the excision repair process and also helps process Okazaki fragments during lagging strand synthesis (Rothwell and Waksman, 2005). Most Pol I enzymes contain a 5'-3' exonuclease activity and a 3'-5' proofreading activity. The 5'-3' exonuclease activity is required for viability as it is essential for the removal of RNA primers from Okazaki fragments generated during DNA replication.
B	Family B polymerases are widespread across all 3 domains of life. This family includes the main eukaryotic replicative DNA polymerases: α , δ and ϵ . Pol B is the main replicative polymerase in the Crenarchaeal kingdom (Henneke <i>et al.</i> , 2005). This family is predominantly involved in DNA replication and usually possesses strong 3'-5' exonuclease activity (Jokela <i>et al.</i> , 2004b; Rothwell and Waksman, 2005).
C	Only found in bacteria, this family contains the major chromosomal replicative enzymes in bacteria (Kornberg and Baker, 1992). This family can be divided into classes (I, II and III) based on primary sequences and domain structures.
D	Found in all archaeal phyla except the Crenarchaea (Yulong Shen, 2001; Brochier-Armanet <i>et al.</i> , 2011). The members of this family have strong polymerase activity and 3'-5' exonuclease activity and are thought to be involved in DNA replication (Cann <i>et al.</i> , 1999; Henneke <i>et al</i> 2005; Castrec <i>et al</i> , 2009). It is believed that all members are hetero-dimeric (Cann and Ishino, 1999), with the small subunit, DP1, believed to be responsible for exonuclease activity (Jokela <i>et al.</i> , 2004) and the large, DP2, subunit responsible for polymerase activity.
X	Only found within eukaryotic cells and have roles in DNA replication and repair pathways. This family includes pol β , pol λ , pol μ , and terminal deoxynucleotidyl transferase (TdT). Pol β , the smallest eukaryotic polymerase, is required for short-patch base excision repair (Matsumoto and Kim, 1995) while pol λ and pol μ are involved in non-homologous end-joining. None of these polymerases are believed to have a proofreading domain thus have low intrinsic fidelity of DNA synthesis (Yamasaki <i>et al.</i> , 2010; Yamtich and Sweasy, 2010).
Y	Found in all 3 domains of life, the family Y DNA polymerases are often referred to as translesion synthesis (TLS) polymerases due to their ability to replicate regions of damaged DNA. They are insensitive to geometric distortions in DNA (Ohmori <i>et al.</i> , 2001) and so exhibit low fidelity. Polymerases within this family have no intrinsic 3'-5' proofreading activity making them advantageous for strand termination and apoptosis (Zhou <i>et al.</i> , 2001).
RT	The reverse transcriptase polymerases are found in both retroviruses and eukaryotes. They use single stranded RNA templates to synthesize double stranded pro-viral DNA strands within viruses (Goff, 1990). The eukaryotic RT polymerases are usually restricted to telomerases (Greider and Blackburn, 1985; Greider and Blackburn, 1987).

Table 1. 1 DNA polymerase families and the roles which they each play within cells.

Within families, the polymerase structure and mechanism is usually well conserved, whereas significant differences are often observed between different families (Rothwell and Waksman, 2005).

1.7.2 DNA polymerase structure

Although the sequence homology of all DNA polymerase is relatively low, their overall structure shows a high degree of conservation across all domains (Rothwell and Waksman, 2005). The structures of all DNA polymerases that have been identified via crystallography appear to share common architectural features within their three core domains: finger, thumb and palm. The three core polymerase domains are named as the overall structure of polymerases tends to resemble that of an open right hand (Figure 1. 8).

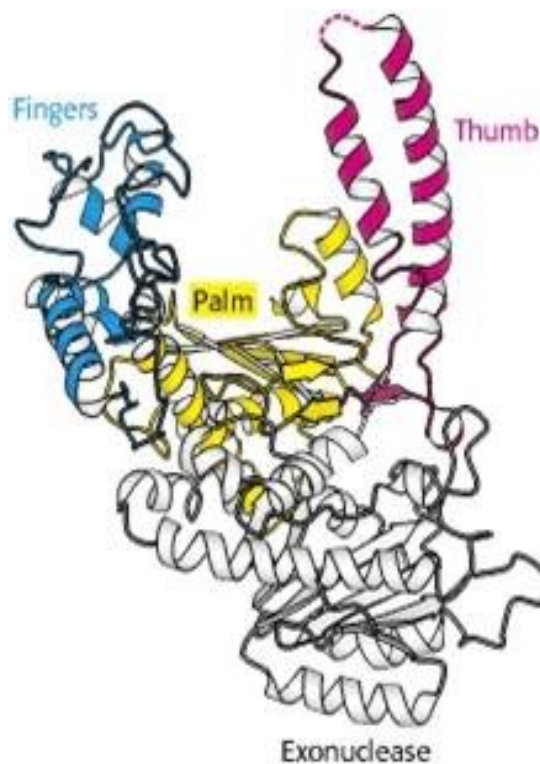


Figure 1. 8 Structure of *E.coli* DNA polymerase 1. The palm (yellow), thumb (red), and finger (blue) domains are well conserved between the three domains of life. Taken from Berg *et al.*, 2002.

The functional role of each domain appears to be consistent across almost all known polymerases (Rothwell and Waksman, 2005). The fingers domain forms important interactions with the incoming deoxynucleoside triphosphates to ensure they are in the correct orientation and is thus important for nucleotide recognition and binding. The thumb domain assists with positioning the duplex DNA and in processivity and translocation. The palm domain contains catalytically essential amino acids and the polymerases' active site and is responsible for catalysis of the phosphoryl transfer reaction (Braithwaite and Ito, 1993; Steitz, 1999; Hubscher *et al.*, 2002). The high degree of conservation of the 3 core domains between the eukarya, bacteria and archaea highlights the irreplaceable functions of the DNA polymerases.

1.7.3 DNA polymerase function

The mechanism of polymerisation is highly conserved among most polymerases and can be distinguished by the basic model of nucleotide incorporation that divides the process into five main steps (Figure 1. 9) (Rothwell and Waksman, 2005). The process is initiated as the unwound DNA primer-template (p/t) binds to the unliganded enzyme (E) to form the enzyme p/t complex (E:p/t) (Step 1, Figure 1. 9). The second step, nucleotide incorporation into the enzyme-p/t complex, is initiated by the binding of dNTPs to form the enzyme-p/t-dNTP complex (E:p/t:dNTP) (Step 2 Figure 1. 9).

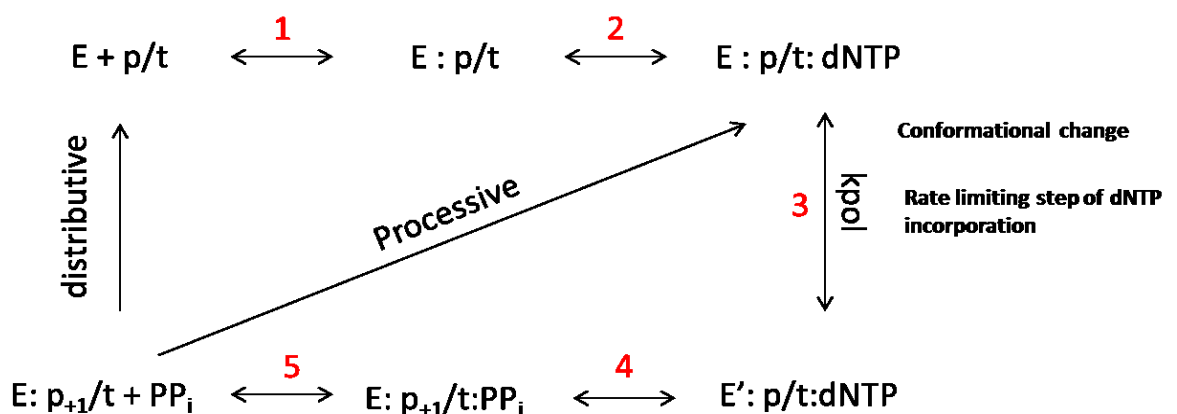


Figure 1. 9 A summary of the 5 main steps in the kinetic pathway of nucleotide incorporation. The rate constant of the conformational change, k_{pol} , which forms the rate-limiting step in the nucleotide incorporation, is indicated in step 3. Image taken from Rothwell and Wakeman, 2005.

The third step, believed to be the rate-limiting step of DNA polymerisation, involves the E:p/t:dNTP complex undergoing a conformational change. This change forms the active complex, E':p/t:dNTP, in which all components of the active site are assembled to facilitate polymerisation (Step 3 Figure 1. 9). The process is continued by the nucleophilic attack by the 3'-OH primer terminus on the α -phosphate of the dNTP resulting in the formation of a phosphodiester bond (E:p₊₁/t:PP_i) (Step 4, Figure 1. 9). Another conformational change then occurs which allows the release of the pyrophosphate (PP_i) product (Step 5, Figure 1. 9). The resulting enzyme complex, E:p₊₁/t can then act in a processive manner, translocating directly to stage 2 of the polymerisation process, and partake in further rounds of nucleotide incorporation. Alternatively, the E:p₊₁/t complex can act in a distributive manner and dissociate from the p/t and return to the start of the process (Rothwell and Waksman, 2005).

While most polymerases conform to the basic mechanism of polymerisation (Figure 1. 9), some polymerases possess different mechanisms which vary in regards to primer-template selection and binding. For example, pol β makes use of a 3 step process and is able to bind to gapped DNA (Jezewska *et al.*, 2002) while HIV-1 RT utilises a 2 step process, that facilitates the generation of complementary DNA (cDNA) from an RNA template (Rittinger *et al.*, 1995; Wohrl *et al.*, 1999)

1.7.4 Two metal ion requirement for polymerases

The mechanism of nucleotide addition by DNA polymerases, initially proposed in 1991 (Beese and Steitz, 1991), was based on an almost identical mechanism that had already been observed during 3'-5' exonucleolysis (Freemont *et al.*, 1988; Beese and Steitz, 1991). The mechanism relies on two divalent metal ions (Mg²⁺), that accompany incoming dNTPs during polymerisation, and two aspartic acid residues that are functionally conserved within the palm domain of all polymerases (Sawaya *et al.*, 1994).

During polymerization, the two divalent metal ions that are bound to the phosphates of the incoming nucleotide promote the stability of the

polymerase:p/t:dNTP complex in the transition state (Pelletier *et al.*, 1994). The two metal ions bind simultaneously to the two conserved aspartic acid residues within the palm domain of the polymerase (Figure 1. 10). Metal ion A interacts with the 3'-hydroxyl of the primer strand and is proposed to lower the dissociation constant (pK_a) of the hydroxyl, facilitating its attack on the α -phosphate of the incoming dNTP. Metal ion A is ligated by an oxygen atom from one conserved aspartic acid and by the 3'-OH group of the ribose moiety on the primer strand (Figure 1. 10). In addition to promoting the stability of the structure and charge during the transition state, metal ion B binds to the β - and γ -phosphates creating a better leaving group (Steitz, 1999).

The first observation of a polymerase complex with both p/t DNA and dNTP:Mg²⁺ bound to the polymerase active site was with rat pol β . This observation directly showed the structural basis of the two metal ion mechanism (Pelletier *et al.*, 1994), which has since been observed in numerous polymerases across the three domains of life (Steitz, 1999).

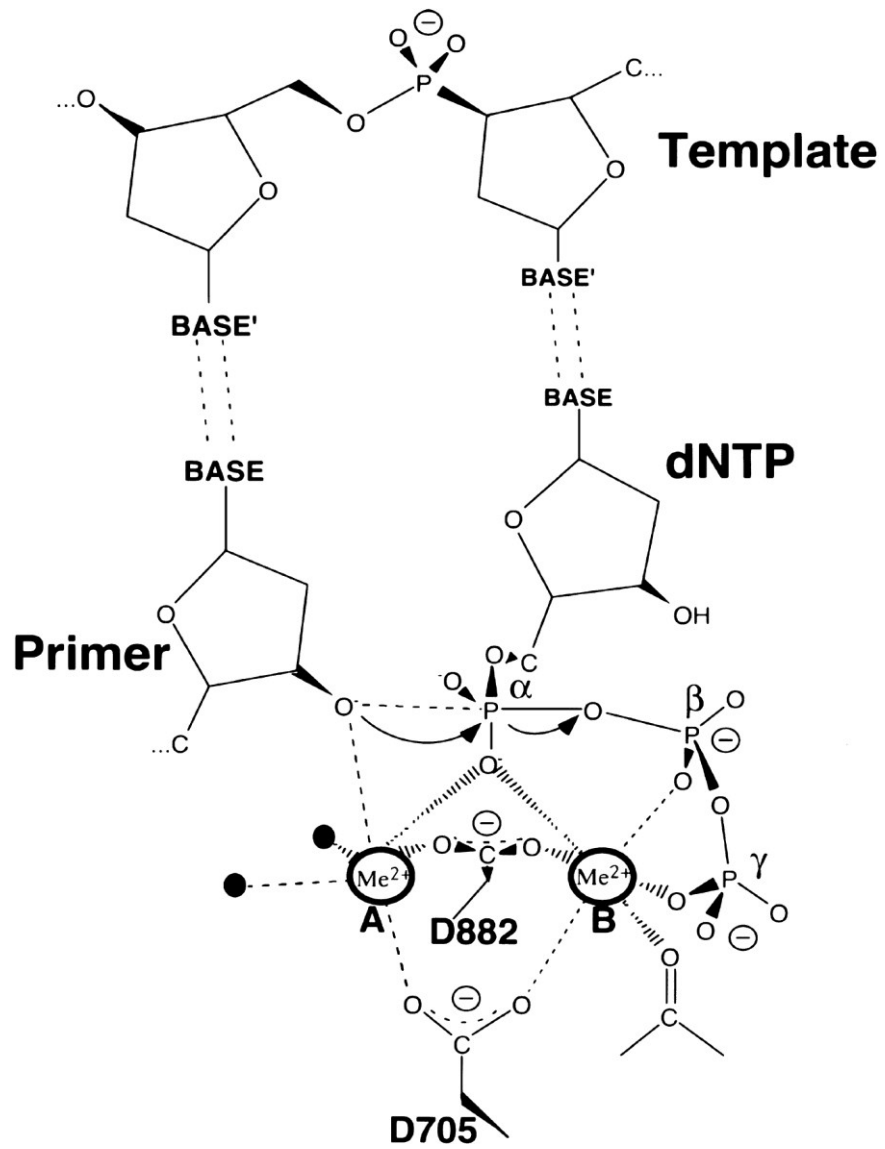


Figure 1. 10 The two metal ion mechanism of polymerisation in the context of the T7 DNA polymerase-substrate. Two divalent metal ions, A and B, are ligated to enzymes of the *E.coli* DNA polymerase I family by aspartic acid residues 705 and 882. The black spots are water molecules. Taken from Steitz, 1999.

1.7.5 Fidelity of polymerases and 3'-5' exonuclease activity

DNA replication fidelity is a key determinant of genome stability and is central to the evolution of species and the origins of human disease. High fidelity genome replication maintains genetic information over many generations with a low mutation rate and thus prevents changes which may lead to the onset of disease (Kunkel and Bebenek, 2000; Bebenek and Kunkel, 2004). Low fidelity rates lead to the incorporation of mutations within the replicated genome which increases the genetic diversity of the daughter cells (Kunkel and Bebenek, 2000; Bebenek and Kunkel, 2004). Low fidelity DNA replication may, therefore, be beneficial when organisms are subjected to changing environments as the increased genetic diversity promotes selection which may be evolutionarily beneficial.

In general, genome replication results in only one error for every 10^9 - 10^{10} bases copied (Echols and Goodman, 1991; Kunkel, 2004). This high fidelity rate is due to multiple kinetic and steric nucleotide incorporation mechanisms in the replicative polymerases combined with effective 3'-5' proofreading activity and DNA repair processes (de Laat., *et al.*, 1999; Kunkel and Bebenek, 2000; Leclere *et al.*, 2013). The nucleotide incorporation mechanisms utilised to promote high fidelity include: base-base hydrogen bonding, water exclusion and enthalpy-entropy compensation and geometric selection for correct shape and size (Bebenek and Kunkel, 2004).

Base-base hydrogen bonding contributes to the specificity of DNA base pairing and thus influences fidelity. However, the contribution, of base-base hydrogen bonding, varies between polymerases and appears to be small (Kool, 2001; Kool, 2002). Thus, the high fidelity of many polymerases is attributed to other nucleotide incorporation mechanisms. One explanation for the increased fidelity rates is that the polymerases are able to amplify free energy differences between correct and incorrect base pairs by partially excluding water from the active site, thus increasing enthalpy differences and reducing entropy differences (Petruska and Goodman, 1995; Bebenek and Kunkel, 2004). This, water exclusion and enthalpy-entropy compensation, hypothesis is supported by observations in the crystal structure of Y family polymerases. These crystal

structures show that the active site of the Y family polymerases is more accessible to solvents than the active sites of more accurate polymerases (Ling *et al.*, 2001; Ling *et al.*, 2003). Thus it is believed that the inability to amplify free energy differences between correct and incorrect base pairs, by partially excluding water from the active site, contributes to the Y family polymerases exceptionally low fidelity (Table 1. 2).

Another factor that affects nucleotide incorporation fidelity is geometric selection for the shape and size of correct Watson and Crick base-pairs. The geometries of A: T and G:C base pairs are remarkably similar to each other but differ from the geometries of mismatched base pairs (Echols and Goodman, 1991; Kool, 2001). Abnormal geometry is thought to result in steric clashes in and around the active site that preclude efficient catalysis. This hypothesis is supported by numerous studies with base analogs (Kool *et al.*, 1998; Kool *et al.*, 2002) and is believed to have a large effect on fidelity rate.

Many other molecular factors are likely to contribute to correct nucleotide incorporation during DNA synthesis (e.g. dNTP binding affinities). The exact contribution of each molecular event is unknown; however, as the fidelity rates of polymerases vary drastically (Table 1. 2), it is likely that the correct nucleotide insertion may depend on the polymerase structure, the base pair, and the DNA sequence context.

In general, the replicative polymerases have a higher fidelity rate than polymerases that have roles in DNA repair and translesion synthesis (Table 1. 2). However, within the cell, all polymerases work together to ensure the accurate replication of the genome.

Polymerase	Polymerase family	Role	Domain	3'-5' activity	Error rate
<i>Taq</i> -Pol	A	DNA repair,	Bacteria	No	1.1×10^{-5}
<i>Pfu</i> -Pol B	B	DNA replication	Archaea	Yes	1.6×10^{-6}
Pol ϵ	B	DNA replication	Eukarya	Yes	1.1×10^{-5}
Pol δ	B	DNA replication	Eukarya	Yes	1.3×10^{-5}
<i>Pfu</i> -Pol D	D	DNA replication	Archaea	Yes	1.7×10^{-5}
DNA Pol β	X	DNA repair	Eukarya	No	6.7×10^{-4}
Pol η	Y	TLS	Eukarya	No	2.1×10^{-3}
M-MLV	RT	RT	Bacteria	No	3.3×10^{-5}

Table 1. 2 Fidelity rate of some well characterised DNA polymerases. “TLS” represents translesion synthesis “RT” represents reverse transcriptase. (Kunkel and Alexander, 1986; Washington *et al.*, 1999; Arezi and Kuchta, 2000; Shimizu *et al.*, 2002; Haracska *et al.*, 2003; Acharya *et al.*, 2006; Zhong *et al.*, 2006; McCulloch and Kunkel, 2008; Keith *et al.*, 2013).

In addition to the kinetic and steric incorporation mechanisms, some DNA polymerases (family A and B) possess 3'-5' exonuclease, proofreading activity. Exonuclease activity excises mis-incorporated nucleotides from the 3' end of the DNA, giving the polymerase a second chance to incorporate the correct dNTP, leading to a reduced error rate (Shevelev and Hubsecher, 2002). Exonuclease activity can improve fidelity rates 3–100 fold (Drake, 1991; Igor V. Shevelev1, 2002) and has been found to be essential in many cells as its loss can create a strong mutator phenotype which can be lethal (Hubscher *et al.*, 2002; Igor V. Shevelev1, 2002; Morita *et al.*, 2004).

1.8 3'-5' exonuclease activity

The mechanism of nucleotide excision via 3'-5' exonuclease activity was described using the refined crystal structures of the Klenow fragment of *E.coli* Pol I (Freemont *et al.*, 1988; Beese and Steitz, 1991). Crystallography structures showed that the polymerase and exonuclease sites within the Klenow fragment are spatially separated and able to function independently (Beese and Steitz, 1991). For exonuclease activity about four bases of DNA are melted and the resulting single stranded primer threads into the active site for mismatched base removal.

Similar to polymerisation, two divalent metal ions (Mg^{2+}) and two highly conserved aspartic acid residues are essential for 3'-5' exonuclease activity (Figure 1. 11). During exonucleolysis, protein residues bind and orientate the two metal ions, the single stranded DNA and an attacking water molecule (Beese and Steitz, 1991). The catalysis of the hydrolytic phosphoryl transfer reaction is strictly dependent on the correct orientation of these components.

Metal ion A facilitates hydroxide ion formation and stabilization of the transition state of the reaction (Figure 1.11). The carboxylate of the glutamic acid functions to bind and orient metal ion A, the 3' OH of the terminal nucleotide and the attacking hydroxide ion (Figure 1. 11). The tyrosine residue also orients the attacking water molecule and provides hydrophobic stabilisation of the single stranded substrate. The second metal ion, metal ion B, functions to stabilise the transient penta-covalent species and facilitates the leaving of the 3' oxyanion from an apical position (Pelletier *et al.*, 1994).

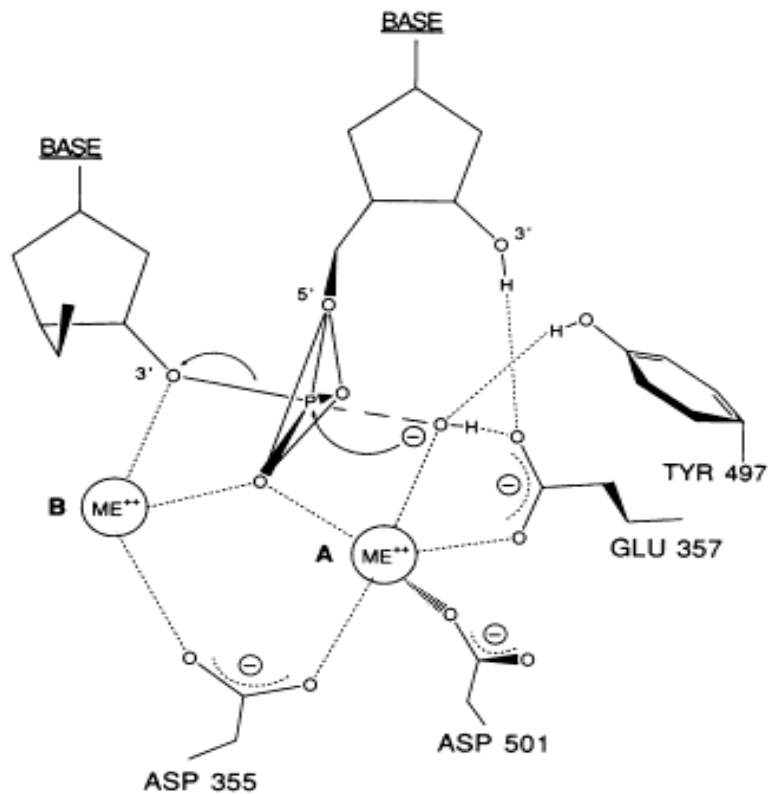


Figure 1. 11 The proposed transition state of the two metal ion enzymatic mechanisms for the 3'-5' exonuclease reaction. Taken from Pelletier *et al.*, 1994.

1.9 Family-B polymerases

Family-B DNA polymerases are found in all 3 domains of life and play key roles in DNA replication and repair (Rothwell and Wakeman, 2005). Many family B polymerases possess a strong, intrinsic 3'-5' exonuclease activity (over a 1000 times higher than that of *E.coli* pol I) (Capson *et al.*, 1992; Lin *et al.*, 1994) which contributes to the high fidelity rate associated with these polymerases (Table 1. 2). The family B polymerases have been well characterised and crystal structures are available from members from all three domains of life (Wang *et al.*, 1997; Hopfner *et al.*, 1999).

1.9.1 Family B DNA polymerases in eukaryotes

The three main replicative polymerases in eukaryotes, DNA polymerases α , δ , and ϵ , are structurally and functionally distinct from each other and are well conserved throughout eukaryotes (Hubscher *et al.*, 2000; Miyabe *et al.*, 2011). Pol α forms part of the heterotetrameric DNA pol α -prim complex involved in RNA primer synthesis. A 48 kDa subunit from the pol α -prim complex is able to start *de novo* synthesis and is responsible for the formation of an RNA primer (Muzi-Falconi *et al.*, 2003). The eukaryotic family B polymerase, pol α , then extends this short RNA primer using its intrinsic DNA polymerase activity to form an RNA/DNA primer \sim 35 nucleotides (Muzi-Falconi *et al.*, 2003; Klinge *et al.*, 2009).

Pol ϵ and Pol δ (family-B) are then responsible for the extension of the leading and lagging strands from the short RNA/DNA primers. It has been proposed that Pol ϵ is primarily responsible for leading strand synthesis from the DNA/RNA primer (Eckardt, 2009) while pol δ is responsible for lagging strand synthesis (Pursell *et al.*, 2007). DNA Pol ϵ has high processivity, without the need for PCNA, and possesses 3'5' proofreading exonuclease activity (Shikata *et al.*, 2006). Deletion of pol ϵ leads to a deficiency in DNA elongation and causes premature senescence (Shikata *et al.*, 2006) and thus is vital for the integrity of the genome. Pol ϵ also plays important roles in lagging strand repair, double strand break repair, base excision repair and cell cycle regulation (Edgell and Doolittle, 1997; Kawasaki and Sugino, 2002; Pospiech and Syvaaja, 2003; Shikata *et al.*, 2006).

DNA Pol δ is responsible for elongation of the lagging strand creating mature Okazaki fragments during DNA replication (Pursell *et al.*, 2007). Pol δ interacts with the eukaryotic DNA clamp, PCNA, and its loader, replication factor C (RFC), to carry out processive synthesis of the lagging strand *in vitro* (Garg and Burgers, 2005). Many other polymerases are required for the high fidelity replication of the eukaryotic genome including DNA polymerases ι , η and ζ (Hubscher *et al.*, 2000). DNA Pol ζ (family B polymerase) is made of two subunits Rav3 and Rev 7, and is involved in translesion synthesis (Gan *et al.*, 2008). It lacks 3'-5' exonuclease activity and is able to extend primer past

terminal mismatches which would stall the replicative polymerases Pol δ and Pol ϵ . Thus, although it is not essential for DNA replication, the ability of DNA Pol ζ to carry out synthesis downstream of DNA lesions is important to maintain the integrity of the genome.

Recent studies have shown that all four yeast family B polymerases (Pols α , ϵ , δ and ζ) possess eight conserved cysteine residues at their CTD (Netz *et al.*, 2012). Two distinct metal binding motifs (CysA and CysB) have been identified. CysA binds Zn^+ ions while CysB is an essential Fe-S centre that co-ordinates the formation of a [4Fe-4S] cluster *in vivo*. The Fe-S cluster has been characterised as playing a role in stabilising the CTD, maintaining the catalytic polymerase subunit with its respective accessory proteins, with loss of the Fe-S cluster resulting in reduced processivity (Netz *et al.*, 2012).

1.9.2 Family B DNA polymerases in bacteria

A few bacteria possess a family B polymerase, Pol II, which has the ability to bypass damaged DNA (translesion synthesis, TLS). TLS is normally associated with family Y polymerases which lack intrinsic 3'-5' proofreading activity (Bienko *et al.*, 2005). The best characterised DNA Pol II enzyme, *E.coli* DNA Pol II, is able to extend primers past mutagenic DNA, and is involved in nucleotide excision repair (Berardini *et al.*, 1999) and replication restart following UV exposure (Rangarajan *et al.*, 1999). However, unlike the family Y polymerases, *E.coli* DNA Pol II maintains 3'-5' exonuclease activity and the efficient DNA polymerase activity normally associated with family B polymerases (Wang and Yang, 2009). Additionally, *E.coli* Pol II possesses a high fidelity rate (Banach-Orlowska *et al.*, 2005) normally associated with replicative polymerases (Table 1.2).

Unlike family Y polymerases, the catalytic site of DNA Pol II translesion synthesis properties is distant from its polymerase active site (Wang and Yang, 2009). Thus, it is hypothesised that DNA Pol IIs evolved as gradual changes distant to the active site were selected for due to their beneficial TLS properties.

As this function was not detrimental to the main function of the replicative polymerase it offered a selective advantage.

DNA Pol IIs appear to be rare in nature with few homologues of the *E.coli* DNA Pol II identified in bacteria via BLAST searches (Figure 6.1). However, due to the unique properties of *E.coli* DNA Pol II, there is a likelihood that characterisation of homologues would identify polymerases with potential applications in biotechnology. For example, a DNA Pol II with similar properties to *E.coli* Pol II that was also thermostable may prove useful in PCR of damaged or low quality DNA.

1.9.3 Family B DNA polymerases in archaea

Family B polymerases are found in all archaea with Crenarchaea possessing multiple (up to three) slightly different proteins and all other phyla possessing a single exemplar (Edgell *et al.*, 1997; Henneke *et al.*, 2005). Biochemical properties of archaeal pol B including rapid synthesis of DNA, 5'-3' proofreading exonuclease activity, high fidelity rates and interaction with PCNA processivity factor combined with their strong sequence homology with the replicative eukaryotic pols δ and ϵ , suggests that they are involved in DNA replication (Edgell *et al.*, 1997). Thus, it was originally believed that Pol B was responsible for DNA replication within all archaea (Grabowski and Kelman, 2003).

However, a recently identified polymerase, Pol D, has been discovered in all archaea phyla except Crenarchaea (Uemori *et al.*, 1997a; Cann *et al.*, 1998; Cann and Ishino, 1999). Pol D has been found to possess many of the properties associated with replicative polymerases and thus is believed to be involved in archaeal DNA replication (Isaac K. O. Cann, 1999; Henneke *et al.*, 2005; Castrec *et al.*, 2009; Castrec *et al.*, 2010). However, as Pol B is the only polymerase in Crenarchaea that possess properties associated with a replicative polymerase, it is still believed that Pol B is solely responsible for DNA replication in Crenarchaea (Barry and Bell, 2006). However many hypotheses exist that suggest that both Pol B and Pol D or that Pol D alone is responsible

for DNA replication in the other archaea phyla (Henneke *et al.*, 2005; Rouillon *et al.*, 2007). These hypotheses stem from findings that show that Pol B can be deleted from *T.kodakarensis* and *M.maripaludis* with no change in growth rates and no increase in mutation rates while Pol D has been shown to be essential (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013).

Archaeal Pol B possesses a unique ability to specifically recognise uracil and hypoxanthine located in template strand DNA (Greagg *et al.*, 1999; Shuttleworth *et al.*, 2004; Firbank *et al.*, 2008b). Archaeal Pol B stalls the replication process when uracil is identified. It is believed that archaea have developed this unique property to improve replication fidelity due to exposure at high temperatures leading to increased rates of deamination (Wardle *et al.*, 2008).

1.9.4 Recognition of uracil and hypoxanthine by archaeal Pol-B

The “read-ahead” uracil recognition mechanism observed in archaeal Pol B has been well characterised (Firbank *et al.*, 2008). Pol B scans ahead of the replication fork and is able to identify uracil or hypoxanthine located in the template strand DNA (Gill *et al.*, 2007). If either of these deaminated bases is detected, DNA replication is stalled. The most profound stalling, when the polymerase binds uracil with the highest binding affinity, occurs when uracil is located at the +4 base pair position from the primer-template junction (Figure 1.12) (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Gouge *et al.*, 2012). This is the only known example of DNA repair based on template strand proofreading (Greagg *et al.*, 1999).

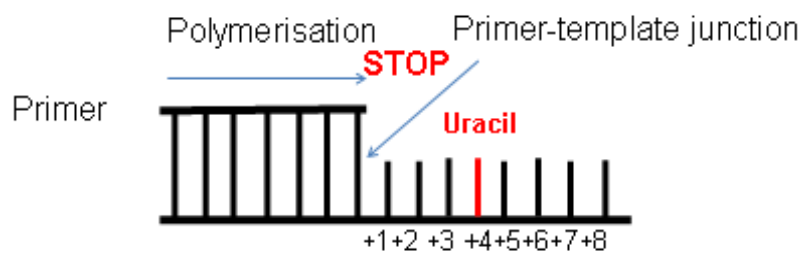


Figure 1. 12 Recognition of uracil, located 4 bases ahead of the primer template junction, by *Pfu*-Pol B. The presence of uracil at the +4 position results in the termination of polymerisation.

The uracil recognition properties of *Thermococcus gorgonarius* Pol B have been captured by X-ray crystallography (Firbank *et al.*, 2008b). This research identified a specialised uracil-binding pocket at the amino-terminal domain of the polymerase which is able to recognise and bind tightly with uracil in single stranded DNA (Fogg *et al.*, 2002). Binding of the uracil is stabilised by hydrogen bonds between the O2 and O4 atoms of the uracil and the amide nitrogen atoms of Ile114 and Tyr37 within the backbone of the polymerase (Firbank *et al.*, 2008).

However, despite the uracil recognition mechanism being well characterised; the physiological role of this novel property is not fully understood. It was initially believed that the uracil recognition mechanisms of archaeal Pol B served to prevent the replication of pro-mutagenic uracil and hypoxanthine that occur at an increased rate in archaea due to exposure to high temperatures. However, more recent evidence has identified uracil recognition properties in all archaeal Pol Bs including those from mesophilic archaea (including *Methanosarcina acetivorans*) (Firbank *et al.*, 2008b; Wardle *et al.*, 2008). This evidence combined with findings that eukaryotic and bacterial polymerases do not possess uracil or hypoxanthine recognition properties suggests that uracil recognition is a peculiarity of archaeal life (Wardle *et al.*, 2008).

At present the events that follow uracil/hypoxanthine induced stalling in archaea are unknown. However, when replication forks stall, they are normally repaired by damage tolerant recombination pathways (Michel *et al.*, 2004; Heller and

Marians, 2006; Lecointe *et al.*, 2007). Thus, it is likely that polymerase mediated uracil stalling of archaeal Pol B is the first step in an additional DNA repair pathway that prevents the copying of G: U mismatches and the permanent fixation of transition mutations. Thus DNA replication might offer a final opportunity for mutation avoidance within archaeal cells (Emptage *et al.*, 2008).

1.10 Family-D polymerases

It was originally believed that all archaeal replicative polymerases were members of the B family DNA polymerases. However, in 1997, Ishino *et al.*, identified novel DNA polymerase genes in the hyperthermophilic archaeon *Pyrococcus furiosus* (Uemori *et al.*, 1997a; Cann *et al.*, 1998). These genes were found to encode a protein that was distinct from any other DNA polymerases that had previously been characterised and are now categorised as a novel family of polymerase, family D (Pol D).

Since the original discovery, the family D polymerase have been identified in all Euryarchaeota, Thaumarchaeota, Korarchaeota, and Aigarchaeota (Cann *et al.*, 1998; Jokela *et al.*, 2004a; Brochier-Armanet *et al.*, 2011; Ishino and Ishino, 2012) . These phyla usually possess one Pol B and one Pol D enzyme while the Crenarchaeota possess multiple Pol Bs (up to 3) and lack Pol D (Edgell *et al.*, 1997; Cann *et al.*, 1999).

1.10.1 Structure and function of archaeal Pol-D

Family D polymerases are composed of a small, DP1, and a large, DP2, subunit and have been reported to be active as both hetero-dimers and L_2S_2 hetero-tetramers $(L_2S_2)_2$ (Uemori *et al.*, 1997b; Cann *et al.*, 1998; Cann and Ishino, 1999; Yulong Shen, 2001). The DP1 subunit shows homology to the non-catalytic B-subunits of the eukaryotic replicative pols α , δ and ϵ and is believed to work as an unidirectional, non-processive proofreading exonuclease (prefers mis-paired DNA and single stranded DNA) (Jokela *et al.*, 2004a). The DP2 subunit possesses polymerase activity and despite being highly conserved within Euryarchaeota, DP2 has no amino acid sequence homology to other known DNA polymerases (Cann *et al.*, 1998; Yulong Shen, 2001). Both the

DP1 and DP2 subunits have been reported to possess activity when expressed alone, however activity is much stronger when both subunits are present (Cann and Ishino, 1999; Jokela *et al.*, 2004b).

The Pol D hetero-dimer has been shown to interact with proliferating cell nuclear antigen (PCNA) in the replisome. Interaction with PCNA allows long strands of DNA to be copied as the DNA polymerases' ability to slide along the DNA template increases processivity (Castrec *et al.*, 2009). Additionally, it has been shown that Pol D possesses strand displacement activity and is able to elongate RNA primers *in vivo* (Henneke *et al.*, 2005)

This ability to interact with PCNA, combined with the strong polymerase, 3'-5' exonuclease activity, RNA elongation properties and evidence that Pol D possesses strand displacement activity *in vivo* (Henneke *et al.*, 2005) supports the hypothesis that Pol D is a replicative polymerase. Additional evidence comprises the ability of Pol D to interact with other replisome components (Li *et al.*, 2010) (Figure 1.6C) and that polymerase D genes adjacent to a cluster of genetically essential genes involved in DNA repair, replication and recombination (Yulong Shen, 2001; Tang *et al.*, 2004) also implies that Pol D may be a main replicative polymerase. Since the original discovery of archaeal Pol D, characterisation has revealed that it is likely to be a replicative polymerase. Despite this characterisation, there is still much debate regarding the precise roles of Pol B and Pol D in Euryarchaeota (Cubonova *et al.*, 2013). Experiments using targeted gene deletion have shown that the family D polymerases are essential for viability in Euryarchaeota (family B is not essential) (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013).

1.11 Replicative polymerases in the archaea

At present, the exact roles of Pol B and Pol D in DNA replication are unknown, however, based on biochemical properties of family B and D polymerases from *P.abyssii* it has been hypothesised that Pol B copies the leading strand while Pol D copies the lagging (Henneke *et al.*, 2005; Rouillon *et al.*, 2007; Castrec *et al.*, 2009). However, other hypotheses have since been proposed including a

scenario in which Pol D initially elongates RNA primers before a switch to Pol B catalysed synthesis. In this hypothesis Pol B synthesises almost all the leading strand and some of lagging strand (Rouillon *et al.*, 2007). The most recent hypothesis is that Pol D is the main replicative polymerase in archaea rather than Pol B (Richardson *et al.*, 2013). This hypothesis has been strongly supported by evidence that Pol B is not essential in *T.kodakarensis* and *M.maripaludis* whereas Pol D is essential (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013). The role of Pol D is of great interest to elucidate the exact replication mechanism in archaea and to understand the evolutionary relationship of DNA polymerases (Tahirov *et al.*, 2009).

1.12 Recognition of uracil by archaeal Pol D

The ability to recognise uracil and hypoxanthine located in template strand DNA was originally believed to be unique to archaeal Pol-B (Gill *et al.*, 2007; Wardle *et al.*, 2008). However, a brief report has suggested that Pol D is also able to recognise uracil (Sawai *et al.*, 2007). However, the manner of uracil recognition appears to be markedly different to the well-characterised read-ahead recognition described for archaeal Pol B (Greagg *et al.*, 1999; Firbank *et al.*, 2008a).

1.13 DNA deamination

Deamination is a term that refers to the removal of an amino group from a molecule. Spontaneous deamination occurs in DNA via the removal of the exocyclic amino groups in cytosine, adenine, guanine and 5-methylcytosine (a methylated form of cytosine). The deamination of these bases produces uracil, hypoxanthine, xanthine and thymine respectively (Figure 1. 13). Deamination rates are increased by high salt concentrations, low pH and high temperatures (Table 1. 3) (Schroeder and Wolfenden, 2007). The deaminated bases are referred to as “pro-mutagenic” as if they are left unrepaired prior to DNA replication; they result in transition mutations in 50 % of the progeny cells (Figure 1. 14).

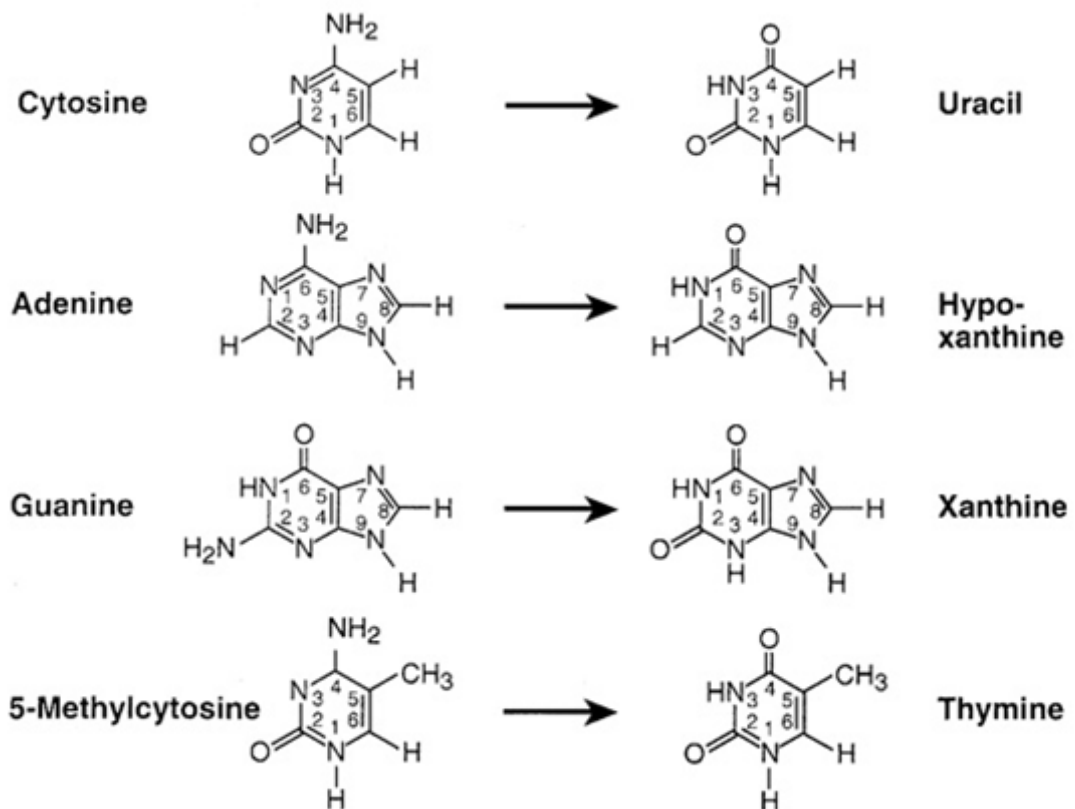


Figure 1. 13 Structure of cytosine, adenine, guanine, 5-methylcytosine and their deaminated products uracil, hypoxanthine, xanthine and thymine respectively. Deamination occurs where an exocyclic amino group is lost.

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

Table 1. 3 A summary of half lives of common chemical bonds contained within normal double-stranded DNA at 25°C and 92°C. Data taken from Schroeder and Wolfenden, 2007.

Transition mutations occur when the deaminated nucleotides “base pair” during DNA replication. For example, cytosine normally pairs with guanine during DNA replication, however, its deaminated product, uracil, pairs with adenine. Thus if uracil is not repaired prior to DNA replication, uracil will base pair with adenine causing a GC – AT transition mutation in 50 % of the progeny (Figure 1. 14). Similarly, deaminated guanine, adenine and 5-methylcytosine will result in 50 % of the progeny containing transition mutations. Thus, it is of utmost importance that the cell is able to repair deaminated bases.

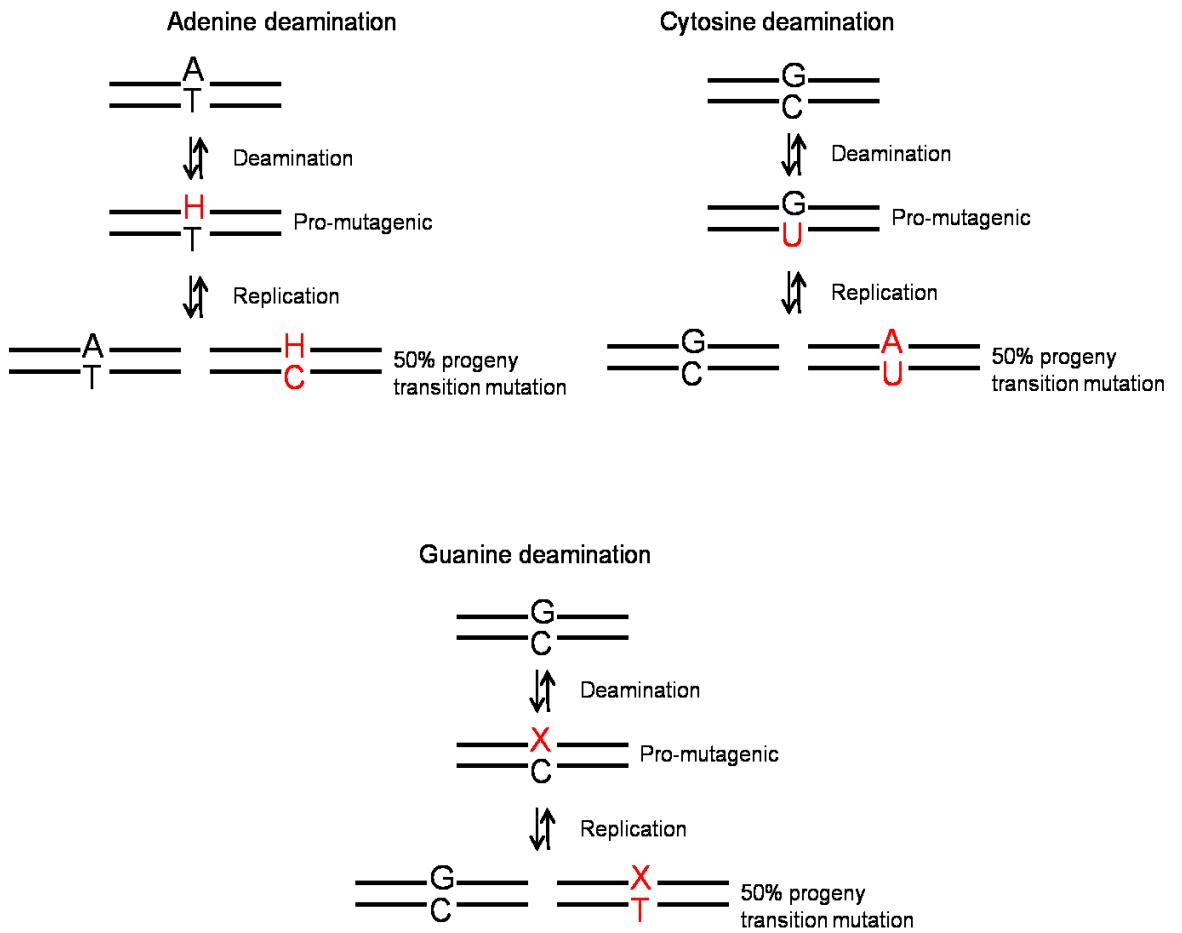


Figure 1. 14 Deamination of adenine, cytosine and guanine. Deamination results in a transition mutation in 50 % of the progeny if it is not repaired prior to DNA replication.

1.13.1 Repair of deaminated bases

The repair of G: U mis-pairs within DNA is usually initiated by uracil-DNA glycosylases (UDGs). UDGs are found in nearly all organisms and are highly conserved. They catalyse the efficient removal of uracil from both single and double stranded DNA to generate AP sites and do not require Mg^{2+} for activity (Pearl, 2000b). UDGs are also able to recognise uracil located opposite all DNA bases and various oxidised cytosine derivatives e.g. 5-fluorouracil, 5-OH-Ura, and 5,6-dihydroxyuracil. However, UDGs are unable to recognise uracil located in RNA (Pearl, 2000).

UDGs remove uracil from DNA via glycosidic bond hydrolysis which initiates a base excision pathway to restore the G-C base pair (Figure 1. 15) (Pearl, 2000; Barnes and Lindahl, 2004; Firbank *et al.*, 2008). Hydrolysis of the glycosidic bond results in an abasic site that is later repaired during the BER pathway (Figure 1. 15). The BER pathway removes the pro-mutagenic uracil from the DNA and thus reduces the frequencies of C-T/G-A transition mutations in the DNA.

Uracil can also be recognised by double-stranded DNA specific UDGs. These enzymes, referred to as mismatch specific uracil DNA N-glycosylase (MUG) in bacteria and thymine DNA N-glycosylase (TDG) in eukaryotes; only recognise uracil located in duplex DNA (Moe *et al.*, 2006). They are structurally related to UDGs despite little sequence homology and differ in uracil recognition mechanism and substrate specificity.

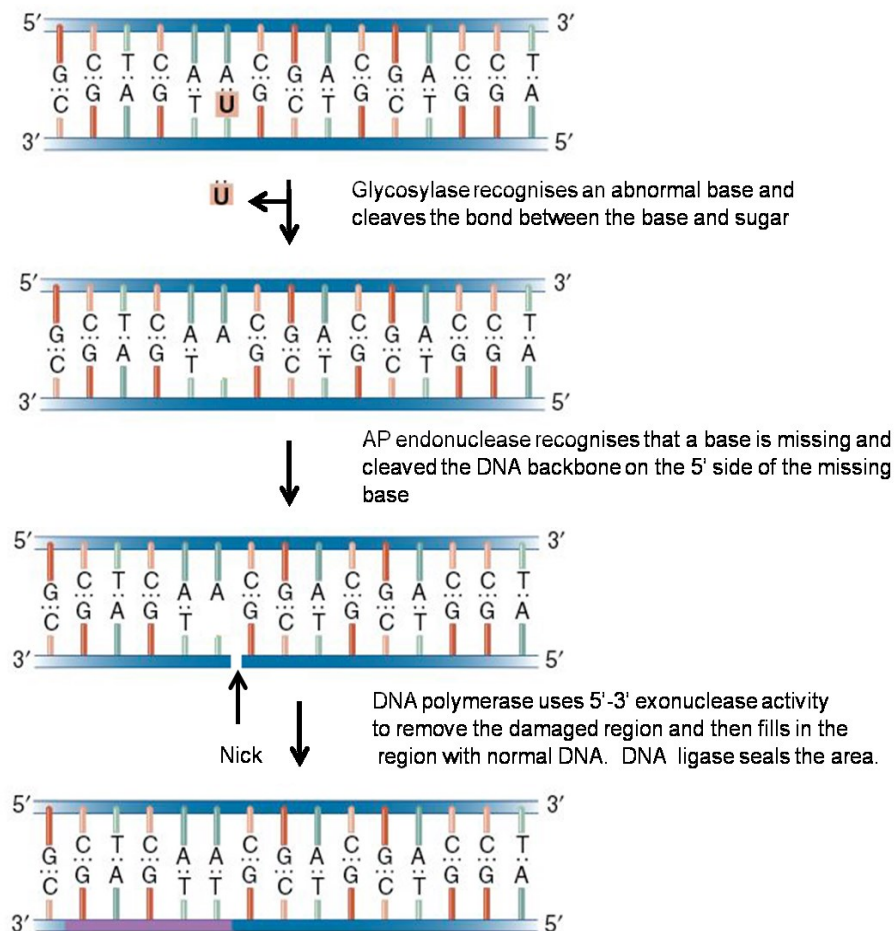


Figure 1. 15 Base excision repair pathway initiated by uracil-DNA glycosylase.

Another enzyme involved in the removal of deaminated bases is Endonuclease V, an enzyme encoded by the *nFi* gene. EndoV is highly conserved throughout the three domains of life and is able to recognise all deaminated bases (Lopez-Olmos *et al.*, 2012). EndoV functions by hydrolysing the second phosphodiester bond 3' of a deaminated base using Mg^{2+} as a cofactor (Feng *et al.*, 2006). Currently, the downstream processing for the EndoV pathway remain unknown, however, it is hypothesised that 3'-5' exonuclease activity may generate a DNA-repair patch spanning only 2–3 nucleotides (nt) to either side of a hypoxanthine base (Mi *et al.*, 2011).

Although EndoV cleaves a spectrum of DNA lesions, genetic analysis of *E.coli* NFI insertion mutations and overproducing strains suggests a major role for EndoV in the *in vivo* repair of deaminated purine bases (Dalhus *et al.*, 2009).

1.14 DNA polymerases as PCR reagents

PCR is a technique by which a specific region of DNA is exponentially amplified through successive rounds of cycling (Figure 1. 16) (Bartlett and Stirling, 2003). The process requires a pair of oligodeoxynucleotides complementary to either side of the target DNA, dNTPs to copy the target DNA and a processive, replicative polymerase. During the PCR, several rounds of thermo-cycling facilitate denaturation of the DNA, annealing of the specific primers and extension of the primers using polymerase and dNTPs. Due to the high temperatures used to denature the double stranded DNA, ~ 95°C, thermostable polymerases that can withstand high temperatures are required.

In 1988, the discovery of *Taq* polymerase was a landmark in DNA research as its ability to withstand extreme temperatures, facilitated PCR, and improved the technique (Saiki *et al.*, 1988). Since the original discovery of *Taq*, considerable research has been invested into identifying new thermostable polymerases with desirable properties for use in PCR such as high fidelity rates and high processivity (Kranaster and Marx, 2010).

However, as the applications of the PCR have evolved to incorporate error prone PCR, real-time PCR, and reverse-transcriptase PCR, there is a need for new thermostable polymerases with additional desirable properties. For example, the ability to bypass damaged DNA yet possess a high fidelity rate and high processivity would aid the replication of low quality or degraded DNA. Furthermore, a thermostable polymerase with DNA polymerase and reverse transcriptase activity would be highly desirable for applications in reverse transcriptase PCR. Thus, research into discovering and developing polymerases with desirable properties for PCR are on-going (Holland *et al.*, 1991; Hamilton *et al.*, 2001; Biles and Connolly, 2004; Eid *et al.*, 2009).

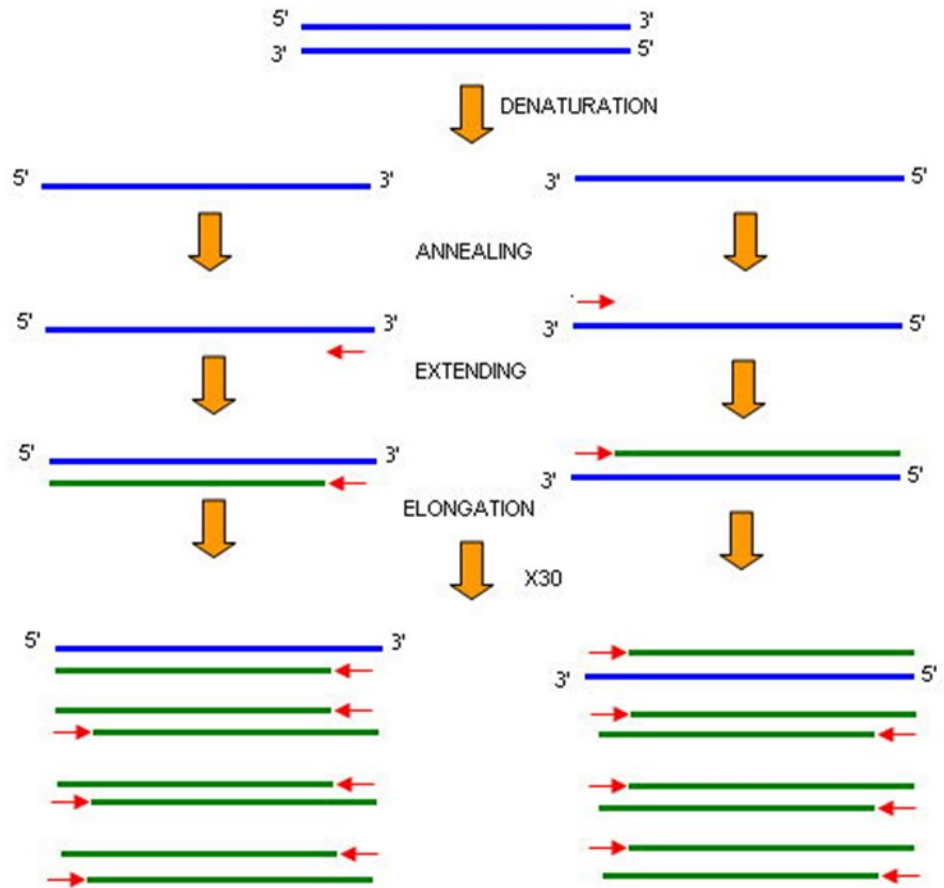


Figure 1. 16 Image of the polymerase chain reaction. → Represents a specific primer.

Chapter 2

Materials and methods

2.1 Oligodeoxynucleotide design, synthesis and purification

2.1.1 Oligodeoxynucleotide design and synthesis

Oligodeoxynucleotides were designed using Clone Manager Professional Suite version 8.0 (Scientific & Educational Software, Cary, NC, USA) and synthesised using an Applied Biosystems 392 DNA/RNA Synthesizer using phosphoramidite chemistry. Standard phosphoramidites and reagents used in the synthesis were purchased from Sigma-Aldrich (Dorset, England) and Proligo (Boulder, CO, USA). The fluorescent phosphoramidites including hexachlorofluorescein (Hex), fluorescein (Fluor) and cyanine5 (Cy5) were purchased from Cambio (Cambridge, UK). Synthesised oligodeoxynucleotides were removed from the glass beads by incubation in 35 % aqueous ammonia at 50°C for 5-15 hours. Samples were then filtered through 0.22 µm Millipore Millex filters (Millipore, Watford, UK) to remove all solid products. Ammonia was evaporated using a Savant Speedvac SC100 (Savant Instruments, Holbrook, NY, USA) and the remaining sample was de-salted using a Nap-25 column (GE Healthcare, Little Chalfont, UK) and concentrated with a Savant Speed Vac SC100. Samples were diluted by the addition of nanopure water to a final volume of 1 ml. Oligodeoxynucleotides containing Hex, Cy5 and Fluor were synthesised using ammonia-labile bases and the ammonia de-blocking step carried out overnight at room temperature.

2.1.2 Synthesis of a 134 base oligodeoxynucleotide

An oligodeoxynucleotide that contained uracil (thymine in controls) 134 bases ahead of a primer-template junction was assembled using three different oligodeoxynucleotides (Table 2. 1). A uracil-containing 3'OH oligodeoxynucleotide (oligodeoxynucleotide A) was ligated to an oligodeoxynucleotide that contained a 5' phosphate group (oligodeoxynucleotide B) using a complementary "splint" oligodeoxynucleotide (oligodeoxynucleotide C). The splint anneals to both the 3'OH and the 5'phosphate containing oligodeoxynucleotides enabling their ligation by mixing the three oligodeoxynucleotides. Oligodeoxynucleotides A, B and C (800 nM of each) were heated at 95°C for 10 minutes and cooled slowly to room

temperature. DNA ligase (1 µl) and 1 x DNA ligase reaction buffer were added and the mixture was left at room temperature for 1 hour. The product was purified by reverse phase HPLC (page 48).

Oligodeoxynucleotide	Sequence
A	5'ATGATTC <u>U</u> ACAGCTTGAGGCCAGTGAAGTAAGGCCTAAATGACCTGCTATACGCTACACGTTAT <u>CATTA</u> ACTAGATTGAC-OH-3'
B	5' <u>p</u> GCATGACGTTACCATTGCAATGCTCTAAGCTCAGTACGGTCTTAACGTGAACGTGAACGTCCGGCGTTGATCAGATCAGCGC-3
C	3'GTAATTGATCTAACTGCGTACTGCAATGGTA-5'

Table 2. 1 Sequence of oligodeoxynucleotides used to synthesise single stranded T134/U134 templates. Underlined region represents DNA complementary to the “splint”, C, oligodeoxynucleotide. “p” represents a phosphate group and “OH” represents a hydroxide group. Uracil is highlighted in red.

2.1.3 Synthesis of RNA oligonucleotides and primers for PCR

Primers, used in PCR, and RNA oligonucleotides were designed using Clone Manager Professional Suite version 8.0 and synthesised by Sigma-Aldrich Ltd (Dorset, UK). Primers were desalted by the company and delivered in dry form. Water was added to obtain the desired concentration. RNA oligonucleotides were supplied HPLC purified and as 100 µM solutions.

2.1.4 Oligodeoxynucleotide purification

Oligodeoxynucleotides that contained a purification handle were purified using reverse phase HPLC at 65°C. Purification was performed using an Apex C18 octadecylsilyl 0.5 micron column (Jones Chromatography, Llanbradach, Wales) and two buffers: buffer A (0.6 % acetic acid, 5 % acetonitrile) and buffer B (0.6 % acetic acid, 65 % acetonitrile), both adjusted to pH 6.5 using triethylamine. The buffers were de-gassed and the column was pre-equilibrated with the

buffers at 1 ml/min for 20 minutes. All purifications were performed at a flow rate of 1 ml/min for 30 minutes and the oligodeoxynucleotides were detected based on UV absorbance at 260 nm. The buffer gradient used was determined by the purification handle attached to the oligodeoxynucleotide (Table 2. 2).

Purification handle	Gradient (buffer B)
None	0-20 %
DMT	25-75 %
Cy5	10-65%
Hex	10-50 %
Fluor	10-60 %

Table 2. 2 Buffer gradients used in HPLC purification of oligodeoxynucleotides based on the 3' purification handle.

After purification, oligodeoxynucleotides containing a DMT group were placed in a rotary evaporator and concentrated at 40°C until oil was produced. The DMT group was then removed by incubation in 80 % acetic acid for 1 hour at room temperature. Acetic acid was removed by rotary evaporation followed by re-suspension in 15 ml of nanopure water. The water was then evaporated to 0.5 ml; this washing procedure was repeated 3 times and the purified oligodeoxynucleotides were re-suspended in 1 ml of nanopure water.

The 134 base oligodeoxynucleotides were subjected to a second round of purification using HPLC ion exchange chromatography. Purification was performed using two buffers: buffer A (25mM Tris-HCl [pH7.5], 10% Acetonitrile) buffer B (25mM Tris-HCl [pH7.5], 10% Acetonitrile, 1 M NaCl). Purifications were performed at a flow rate of 1 ml/min for 30 minutes using a gradient of 35 % to 75 % buffer B at 55°C. The oligodeoxynucleotides were detected by UV absorbance at 260 nm and separated on the basis of charge. Following HPLC purification, DNA samples were desalted using a NAP-25 column (GE Healthcare). Samples were then concentrated in a Speedvac SC100 (Savant) and stored at -20°C.

2.1.5 Oligodeoxynucleotide concentration calculation

The Beer-Lambert Law was used to determine the concentration of oligodeoxynucleotides:

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

Where C is the concentration of the oligodeoxynucleotide (mM); A₂₆₀ is the absorbance of light measured at 260 nm wavelength using a NanopDrop™ spectrophotometer (Thermo Scientific, Loughborough, UK), l is the path length and ϵ is the extinction coefficient of the oligodeoxynucleotide sequence at 260 nm ($\text{mM}^{-1} \text{cm}^{-1}$). The extinction coefficient of each oligodeoxynucleotide was determined by adding the extinction coefficients of the individual bases involved in the oligodeoxynucleotides composition (Table 2. 3).

Oligodeoxynucleotide components	Extinction coefficient ($\text{mM}^{-1}\text{cm}^{-1}$)
dA	14.7
dC	6.1
dG	11.8
dT	8.7
dU	9.3
Hex*	31.6
Cy5*	10.0
Fluor*	39.8

Table 2. 3 Extinction coefficients (260 nm) of DNA bases and fluorophores used in oligodeoxynucleotide synthesis.

* values taken from: <http://glenresearch.com/Reference/Extinctions.html>

2.2 PCR, Real time (RT)-PCR and site-directed mutagenesis

2.2.1 Polymerase Chain Reaction (PCR)

Reaction mixtures for use in PCR contained: 10-100 ng template DNA, 1 μ M of each oligodeoxynucleotide primer, 400 μ M of each dNTP, 1 U of *Taq* (Thermo Scientific, Hempstead, UK) or 1 U Phusion (New England Biolabs, Hitchin, UK), 1 x manufacturer's reaction buffer (supplemented with 1-3 mM MgCl₂ where optimisation was required) and water to a final volume of 50 μ l. PCRs used to amplify GC rich (> 60 %) DNA was supplemented with 2 μ l of dimethyl sulfoxide (DMSO) to reduce the melting temperature of the DNA and increase the efficiency of the reaction. Thermocycling was performed in a Biometra TPersonal Combi thermocycler (Thistle scientific, Glasgow, UK) using the following conditions:

95°C	2 minutes	
95°C	35 seconds	} X30
*T _m -5°C	35 seconds	
72°C	1 minute per kb extended	
*T _m -5°C	1 minute	
72°C	3 minutes per kb extended	
4°C	Pause	

*Represents temperature 5°C below the lowest melting temperatures of the primers used in PCR as determined by Clone Manager software.

2.2.2 RT-PCR

RT-PCR was performed using a Roto-Gene 6000 thermocycler (Corbett Life Science, Crawly, UK). Primers were designed to amplify a 250 base region of the *Saccharomyces cerevisiae* Pol II gene. Reaction mixtures for use in RT-PCR contained: 30 ng *S. cerevesiae* genomic DNA (Novagen, Darmstadt, Germany), 1 x Pol B reaction buffer (20 mM Tris-HCl [pH 8.0], 10 mM KCl, 2 mM MgSO₄, 10 mM (NH₄)₂SO₄, 0.1 % Triton X-100, 1 mg/ml BSA), 1 μM of each primer, 400 μM of each dNTP, 1 in 10 dilution of SYBR green (Invitrogen, Paisley, UK) and water to 25 μl. Thermocycling was performed using the following cycling conditions:

95°C	2 minutes	
95°C	10 seconds	} X 40
58°C	20 seconds	
72°C	30 seconds	
72°C	2 minutes	

After PCR cycling a pre-melt step was performed for melt curve analysis. Samples were heated for 90 seconds at 67°C. The temperature was then increased by 1°C every 5 seconds until 95°C was reached.

After PCR amplification the reaction mixtures were loaded onto a 1 % agarose gel, run at 100 V, and visualised under a UV light to confirm amplification of the correct product.

2.2.3 Site-directed mutagenesis

Site-directed mutagenesis was used to introduce specific mutations at a desired location into DNA. Site-directed mutagenesis was performed using a Quickchange site-directed mutagenesis kit (Agilent Technologies, Berkshire, UK) according to the manufacturer's protocol. The digested PCR product was transformed into Top10 competent *E.coli* cells (Invitrogen, Paisley, UK) and the mutagenized plasmid was isolated. Mutagenized plasmids were sent to GATC biotech (Constance, Germany) to be sequenced to confirm the desired mutation had been produced.

2.3 Primer list

All primers used in this thesis for PCR amplification, DNA sequencing, and site-directed mutagenesis reactions are described in Table 2. 4.

Primer	Use	Sequence (5'-3')
<i>Mth</i> -DP1 FP Nde1	Cloning	GTTTCTT CATATG AACGAGATAATCGGAAAATTTGCAAGGGAAGGGATC
<i>Mth</i> -DP1 RP Sal1	Cloning	GTTTCTT GTCGACT CAGCTGAATCAAGGAGTTTCATCACACCCTGTTAAG
<i>Mth</i> -DP2 FP Nde1	Cloning	CAC CATATG ATGGACTACTTCAATGAGCTTGAAAGGGAGACAGAAC
<i>Mth</i> -DP2 RP Sal1	Cloning	GTTTCTT GTCGACT TAAAGGAATACGTCGAGTGAGCTCTGTTTTG
<i>Mja</i> -DP1 FP Nde1	Cloning	GTTTCTT CATATG GAGATAATAAATAAATTCCTAGATTTAGAGGCTTTATTATCAC
<i>MjaMja</i> -DP1 RP Sal1	Cloning	GTTTCTT GTCGACT TAAATATCTAACCTCTAAAACCTCCCTATCCCATTC
<i>MjaMja</i> -DP2 FP Nde1	Cloning	CAC CATATG ATTGTTATGGTTCATGTTGCATGCTCCGAAAATATG
<i>MjaMja</i> -DP2 RP Sal1	Cloning	GTTTCTT GTCGACT TATCCTATCTTAAAGAAGTCACTCAACTTAAC
<i>Mth</i> -DP1700+	Sequencing	GCCAGACCTTCCTTGAGGACGCCTTC
<i>MjaMja</i> -DP1 700+	Sequencing	AGGATATTTTTGTCGTAGGAATCGTTAGTG
<i>Mth</i> -DP2 FP LIC	Cloning	CACCACCACCAC ATGGACTACTTCAATGAGCTTGAAAGGGAGACAGAAC
<i>Mth</i> -RP DP2 LIC	Cloning	GAGGAGAAGGCGCGTTA AAGGAATACGTCGAGTGAGCTCTGTTTTG
<i>MjaMja</i> -DP2 FP LIC	Cloning	CACCACCACCAC ATTGTTATGGTTCATGTTGCATGCTCCGAAAATATG
<i>MjaMja</i> -DP2 RP LIC	Cloning	GAGGAGAAGGCGCGTTA TCTATCTTAAAGAAGTCACTCAACTTAAC
<i>R.marinus</i> Pol II FP Nde1	Cloning	GTTCTTT CATATG GCGGAGACGACGCACCTGTAC
<i>R.marinus</i> Pol II RP Sal1	Cloning	GTTCTTT GTCGAC AGGGCGCCCGGATTTTTTCAAACG
<i>S.YO3</i> Pol II FP Nhe1	Cloning	GTTTCTT GCTAGC ATGAAAATATATCAAATCCTTGATGTATAC
<i>S.YO3</i> Pol II RP Sal1	Cloning	GTTTCTT GTCGACT TACCTAATACATTACCAAAAAGCTTTTTGGTAATAATC
<i>Hvo</i> -DP2 FP EcoR1	Cloning	GTTTCTT GAATTC AAATGCGCGAGGAGGAAACCCGGTACTTCC
<i>Hvo</i> -DP2 RP Not1	Cloning	GTTTCTT GCGGCCG CCTACATGAAGTCCGCGATGCCGACTGTTTG
<i>MjaMja</i> -DP2 FP Asc1 (pAW42)	Cloning	GTTCTTT GCGGCCG CGTGATTGTTATGGTTCATGTTGCATGCTCCGAAAATATG
<i>MjaMja</i> -DP2 RP BglII (pAW42)	Cloning	GTTCTTT AGATCT TTATCCTATCTTAAAGAAGTCACTCAACTTAAC
<i>MjaMja</i> -DP1 FP NsiI (pLW40)	Cloning	GTTCTTT ATGCAT GGAGATAATAAATAAATTCCTAGATTTAGAG
<i>MjaMja</i> -DP1 RP BglII (pLW40)	Cloning	GTTCTTT AGATCT TTAATATCTAACCTCTAAAACCTCCCTATCC
<i>Hvo</i> -DP1 FP Nhe1	Cloning	GTTCTTT GCTAGC CCGGGTATGCCACTGGAGAC
<i>Hvo</i> -DP1 RP Nhe1	Cloning	GTTCTTT GCTAGC GAACTTCCGGACCGTCAT
<i>Hvo</i> -DP2 FP Sph1	Cloning	GTTCTTT GCATGC GCGAGGAGGAAACCCGGTACTTCC
<i>Hvo</i> -DP2 RP Nhe1	Cloning	GTTCTTT GCTAGC CCTACATGAAGTCCGCGATGCCGACTGT
pSJ1/pSJ1B lacZα	Sequencing	CCCCAGGCTTTACACTTTATGCTTCC
<i>MjaMja</i> -Dp1 FP (H-A)	SDM	CAGGAAAC GCCG ATGCTGCTAGGCAAGCTATTA
<i>MjaMja</i> -DP1 RP (H-A)	SDM	CTAGCAGCATCG GCG TTTCTGGGGCAATGAAC
<i>S.YO3</i> FP exo ⁺	SDM	AATTATGGAGATGATATA GAC ATACCAAATCTTTTTG
<i>S.YO3</i> RP exo ⁺	SDM	TTGGTATGTCTATATCATCT CCATA ATTTGCTATG
<i>S.YO3</i> FP pol ⁺	SDM	TAATTCATATT GAC ACCGATTCTATCTGGGTTTAC
<i>S.YO3</i> RP pol ⁺	SDM	GAATCGGTGTCAATAT GAA TACTCTAAAGCCTTTTG
pET28a FP	PCR	GAGGCCCAAGGGGTTATG
pET28a RP	PCR	CAAGGAATGGTGCATGCAAGGAGATG

Table 2. 4 Oligodeoxynucleotide sequence of primers used in PCR amplification, DNA sequencing, and site-directed mutagenesis reactions. Restriction sites are highlighted in red, ligase independent cloning (LIC) 5' specific ends are highlighted in blue and mutated regions in site-directed mutagenesis (SDM) primers are highlighted green.

2.4 Agarose gel electrophoresis, DNA extraction, DNA purification and DNA quantification

2.4.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse DNA and RNA oligonucleotides. Agarose (0.8 g) was dissolved in 1 x TBE (80 ml) by heating (microwave) until dissolved to make 1 x agarose solution. Once cooled, the agarose was supplemented with 1 mg/ml ethidium bromide and the gel was poured. Oligonucleotide samples were supplemented with 1 x DNA loading dye (6.25 % glycerol, 6.25 % SDS, 0.1 % bromophenol blue and 0.1 % xylene cyanol) and loaded into the appropriate wells of the gel. 10 µl of GeneRuler 10 kb or 100 bp DNA ladder (Fermentas, Leicestershire, UK) was loaded in the outermost lane of the agarose gel and used as a visual molecular marker. Gels were covered in 1 x TBE and run for 1-2 hours (100 V, 100 mA, 10 W). The products were visualised using a UV transilluminator.

2.4.2 Extraction of DNA bands

Where DNA was required for further downstream applications, DNA bands were manually excised from the 1 % agarose gel using a surgical blade. The excised gel was then weighed and subjected to DNA gel extraction using a QIAquick gel extraction kit (Qiagen Ltd, Manchester, UK) according to the manufacturer's guidelines.

2.4.3 DNA purification

DNA samples that contained impurities such as salts and buffers were purified using a QIAquick PCR purification kit (Qiagen Ltd) according to the manufacturer's protocol.

2.4.4 DNA quantification

The concentration and purity of DNA and RNA samples were calculated based on the UV absorbance spectra using a NanopDrop 1000 Spectrophotometer. The A_{260} (absorbance at 260 nm) reading was used to estimate DNA concentration using the Beer Lambert Law. The extinction coefficient for

nucleic acids was assumed to be 50 ng/μl (1 cm lightpath length) for double stranded DNA and 33 ng/μl for single-stranded DNA. The $A_{260}:A_{230}$ ratio was used as an indication of purity with a ratio of 1.6 or above considered to be satisfactory for DNA and a ratio of 1.8-2.0 satisfactory for RNA.

2.5 Bacterial, *Haloferax volcanii* (Hvo) and *Methanococcus maripaludis* (Mma) growth media and agar

2.5.1 Bacterial growth media and agar

Luria-Bertani (LB) media (1 % Bacto yeast extract and 1 % NaCl) was used for the growth of bacterial cells. LB containing the appropriate antibiotic (50 mg/ml kanamycin and/or 100 mg/ml ampicillin and/or 10 mg/ml streptomycin) was used for the growth of transformed bacteria. 2 % Bacto agar was added to LB, prior to autoclaving, and poured to make bacterial agar plates (when required, antibiotic was added prior to pouring). Optimal growth on agar plates and LB media occurred at 37°C with shaking for LB.

2.5.2 *Haloferax volcanii* (Hvo) growth media and agar

Complete growth media for *Hvo* (*Hvo*-YPC) required large stocks of 30 % salt water (SW). 5 litres of 30 % SW (1200 g NaCl, 150 g $MgCl_2 \cdot 6H_2O$, 175 g $MgSO_4 \cdot 7H_2O$, 35 g KCl, 100 ml 1 M Tris-HCl [pH 7.5] made up to 5 litres with distilled H_2O) was autoclaved and stored in a dark cupboard. 333 ml of *Hvo*-YPC media consisting of 33 ml of 10 x YPC (8.5 g yeast extract (Becton-Dickinson Oxford Science, UK), 1.7 g peptone (Oxoid Limited, Hampshire, UK) 1.7 g casamino acids, distilled H_2O to a final volume of 167 ml and 3 ml 1 M KOH). 200 ml of 30 % SW and 100 ml distilled H_2O was autoclaved, and when cooled, supplemented with 0.5 M $CaCl_2$. *Hvo*-YPC media was placed in a cool dark cupboard for long term storage.

Hvo agar plates were made by mixing 5 g of agar, 100 ml distilled H_2O and 200 ml of 30 % SW and melting in a microwave (30 minutes). Once cooled, 33 ml of 10 x YPC media was added and the media was autoclaved. After cooling to ~57°C, the media was supplemented with 2 ml of 0.5 M $CaCl_2$. If required, 3.4

ml of thymidine (4 mg/ml in distilled H₂O) was added to the agar before pouring thick plates. Plates were incubated at room temperature for storage. Optimal growth on agar plates and *Hvo*-YPC media occurred at 45°C with shaking for *Hvo*-YPC.

2.5.3 *Methanococcus maripaludis* (*Mma*) growth media and agar

Mma growth media (*McCas*) consists of 250 ml distilled water, 250 ml of 50 % general salt solution (0.67 g KCl, 5.5 g MgCl₂·6H₂O, 6.9 g MgSO₄·7H₂O, 0.28 g CaCl₂·2H₂O, 1 g NH₄Cl dissolved in water to a final volume of 1 litre), 2.5 g NaHCO₃, 11 g NaCl, 0.7 g NaCH₃CO₂·3H₂O, 1 g casamino acids, 5 ml K₂HPO₄ solution (14 g/l), 2.5 ml FeSO₄ solution (0.19 g of FeSO₄·7H₂O per 100 ml of 10 mM HCl), 0.5 ml trace minerals (x 1000), (2.1 g Na₃Citrate·2H₂O [pH 6.5], 0.5 g MnSO₄·2H₂O, 0.1 g CoSO₄, 0.1 g ZnSO₄·7H₂O, 0.01 g CuSO₄·5H₂O, 0.01g AlK(SO₄)₂, 0.01 g H₃BO₄, 0.1 g Na₂MoO₄·2H₂O, 0.025 g NiCl₂·6H₂O, 0.2 g Na₂SeO₃, 0.01 g V(III)Cl, 0.0033 g Na₂WO₄·2H₂O, 100 ml water), 5 ml vitamin solution (x 100) (2 mg biotin, 2 mg folic acid, 10 mg pyridoxine HCl, 5 mg thiamine HCl, 5 mg riboflavin, 5 mg nicotinic acid, 5 mg DL-calcium pantothenate, 0.1 mg vitamin B12, 5 mg p-aminobenzoic acid, 5 mg lipoic acid) and 0.5 ml resazurin solution (1 g/L). *McCas* media was placed in a conical flask with stopper fixed, and heated until boiling under a stream of N₂/CO₂. Once boiled, the media was removed from the heat and supplemented with 0.35 g of cysteine. The stopper was returned to the conical flask and the media incubated under a stream of N₂/CO₂ for 10 minutes until the media changed from pink to clear and the precipitate had dissolved. The gassing cannula was removed and the stopper immediately fastened to prevent contamination with oxygen. The media was placed in an anaerobic chamber and 5 ml of media was aliquotted into glass tubes and fastened with stoppers. The media was gas exchanged with 80:20 H₂/CO₂ and pressurised to 30 psi. The media was removed from the anaerobic chamber and autoclaved. The media was left to cool to ~80°C then inverted several times, to reduce the amount of precipitate. *McCas* media was left to cool to room temperature prior to use. When required, media was supplemented with antibiotics (puromycin 2.5 mg/ml, neomycin 1 mg/ml (solid media) or 0.5 mg/ml (liquid media)).

Mma agar was initially made under aerobic conditions: 250 ml H₂O, 250 ml general salts solution, 1g NaHCO₃, 11 g NaCl, 5 ml K₂HPO₄, 2.5 ml FeSO₄ solution, 0.5 ml trace minerals, 5 ml vitamin solution (x 100), 0.5 ml resazurin solution, 0.7g NaOAc·3H₂O, 1 g casamino acids, 7.5 g Difco noble agar and 0.25 g DTT were mixed and immediately autoclaved. Once autoclaved, the agar was incubated under a stream of N₂/CO₂; allowed to cool to ~50°C then supplemented with 0.25 g cysteine. The agar was transferred to an anaerobic chamber, and when necessary the appropriate antibiotics were added, before pouring the plates. Once the plates had set, they were kept under strictly anaerobic conditions.

2.6 Competent cell preparation, vector transformations and vector preparation

2.6.1 Competent cell preparation

E.coli cells were streaked on an agar plate (containing the appropriate antibiotic) and incubated at 37°C for 15 hours. A single bacterial colony was used to inoculate 5 ml of LB which was incubated at 37°C for 15 hours in a shaker (150 rpm). 1 ml of the cell culture was added to 100 ml of LB and grown at 37°C in a shaker (150 rpm) until an O.D.₆₀₀ of 0.4-0.5 was reached. The cell culture was centrifuged for 10 minutes at 1000 g and 4°C. The supernatant was discarded and the pellet was re-suspended in 20 ml of ice cold 100 mM MgCl₂ and centrifuged at 4°C for a further 10 minutes at 1000 g. The supernatant was discarded and the remaining pellet was gently re-suspended in 4 ml of 100 mM cold, sterile CaCl₂ and incubated on ice for 2 hours. 1 ml of 100 % glycerol was added to the cells and gently mixed until homogeneous. The competent cells were aliquotted into pre-chilled microcentrifuge tubes and stored at -80°C.

2.6.2 Bacterial transformation

100 ng of vector DNA was mixed with 50 µl of the appropriate competent *E.coli* cells, gently swirled and placed on ice for 20 minutes. The cells were heat-shocked at 41°C for 35 seconds and incubated on ice for a further 15 minutes. 500 µl of LB was added to the cells prior to incubation at 37°C for a 1 hour

recovery period. After incubation, the mixture was re-suspended and 100 μ l was spread on a LB agar plate containing the appropriate antibiotic. Once dried, agar plates were incubated at 37°C for approximately 16 hours.

2.6.3 *Haloferax volcanii* (Hvo) transformation

10 ml of Hvo growth media was inoculated with 1-4 colonies of the desired Hvo strain and grown at 45°C with shaking overnight. When an O.D.₆₅₀ of 0.8 was reached, the cells were pelleted by centrifugation at 6,000 rpm for 8 minutes at 25°C. The pelleted cells were re-suspended in 2 ml of buffered spheroblasting solution (50 mM Tris-HCl [Ph 8.5], 1 M NaCl, 27 Mm KCl, 15 mM sucrose), transferred to a round bottomed, 2 ml micro-centrifuge tube and pelleted via centrifugation at 6,000 rpm at 25°C. The pellet was gently re-suspended in 600 μ l of buffered spheroplasting solution.

200 μ l of cells were transferred to a clean 2 ml round-bottomed tube and 20 μ l of 0.5 M EDTA [pH 8.0] was added and mixed by inverting the tube. The solution was left to incubate at room temperature for 10 minutes to form spheroplasts. 10 μ l of DNA to be transformed (~1-2 mg) was mixed with 15 μ l of unbuffered spheroplasting solution (1 M NaCl, 0.2 g KCl, 15 % sucrose, adjusted to pH [7.5] with NaOH) and 5 μ l of 0.5 M EDTA [pH 8.0]. The DNA was added to the spheroplasts in the same manner as EDTA and left to incubate at room temperature for 5 minutes. After 5 minutes add 250 μ l (equal volume) of 60% PEG 600 (480 μ l of PEG 600 and 320 μ l of unbuffered spheroplasting solution) was added to each transformation and left to incubate for 30 minutes. 1.5 ml of spheroplast dilution solution was then added, mixed and incubated at room temperature for 2 minutes. The cells were then pelleted via centrifugation at 6, 000 rpm for 8 minutes at 25°C, and the supernatant removed. 1 ml of regeneration solution (30 % SW, 1 x YPC, 15 % sucrose 3mM CaCl₂)(+ 60 μ g/ml thymidine, if required) was added and a wide-bore blue tip was used to suck up and transfer the whole pellet to a 4 ml sterile tube. The pellet was left undisturbed at 45°C for 1.5–2 hours to regenerate prior to re-suspension by gently tapping the side of the tube. The re-suspended pellet was incubated at 45°C for 3–4 hours, rotating. The cells were then transferred to a 2 ml round-bottomed tube, pelleted at 6,000 rpm for 8

minutes at 25°C and the supernatant removed. The pellet was gently re-suspended in 1 ml transformant dilution solution (18 % SW, 15 % sucrose, 3 mM CaCl₂) and 100 µl of cells plated onto Hv-YPC.

2.6.4 *Methanococcus maripaludis* (*Mma*) transformation

The following transformation protocol was performed under strictly anaerobic conditions: 100 µl of *Mma* (strain S0001) was added to 5 ml of *McCas*, supplemented with 100 µl of 2.5 % Na₂S. The media was pressurised to 40 psi with H₂/CO₂ and incubated at 37°C, with gentle shaking, for 48 hours. 1 ml of the incubated cell culture was added to 4 ml of *McCas*, supplemented with 100 µl of 2.5 % Na₂S, and incubated at 37°C with shaking. When an O.D.₆₀₀ of 0.7-1 was obtained (~1-2 hours) the cultures were pressurised to 30 psi with H₂/CO₂ and centrifuged (1500 g) at room temperature for 15 minutes until a pellet formed. The resulting supernatant was removed and 5 ml of transformation buffer (50 mM Tris-HCl [pH 7.5], 0.35 M sucrose, 0.38 M NaCl, 1 mM MgCl₂ and 0.00001 % rezasurin) was added and the pellet re-suspended. The cultures were re-pressurised to 30 psi with H₂/CO₂ and centrifuged at 1500 g for 15 minutes. The supernatant was, once again, removed and the remaining pellet re-suspended in 0.375 ml of transformation buffer. 5 mg (no more than 50 µl) of vector DNA was added to the cell culture and mixed by gently shaking the tube. 0.225 ml of 40 % polyethene glycol (PEG) solution was added to the transformation culture which was flushed with 100 % N₂ for 1 minute. The culture was pressurised to 30 psi with 100 % N₂, prior to incubation at 37°C for 1 hour (without shaking). A vacutainer needle was used to transfer 5 ml *McCas* media, supplemented with 100 µl of 2.5 % Na₂S, into the tube containing the transformed media. The long end of the vacutainer needle was inserted into the transformation tube until all the gas had escaped. The tube containing fresh *McCas* media was then inserted into the short end of the vacutainer needle and held vertically to allow gravity to maintain the flow of media into the transformation solution tube. The media was re-pressurised with H₂/CO₂ (30 psi) and spun at room temperature for 20 minutes at 1500 g. The supernatant was removed again using a vacutainer needle and 5 ml of *McCas* media (supplemented with 100 µl of 2.5 % Na₂S) was transferred to the

transformation tube and the pellet was re-suspended (by gentle tapping). The culture was flushed with H₂/CO₂ for 30 seconds, pressurised to 40 psi and incubated at 37°C with gentle shaking overnight.

After overnight incubation, the transformed culture was re-suspended and 100 µl was spread on *Mma* agar plates containing the appropriate antibiotics. Once dried, agar plates were incubated at 37°C in an anaerobic chamber pressurised to 30 psi with H₂/CO₂ for 3-5 days.

2.6.5 Bacterial and *Haloferax volcanii* (*Hvo*) vector preparation

Following the overnight incubation of transformant colonies on agar plates, individual colonies were picked and incubated in 5 ml of appropriate growth media (containing the required antibiotics) at 37°C (bacteria) or 45°C (*Hvo*) for a minimum of 15 hours (shaking 150 rpm). Plasmids were isolated from the resulting bacterial suspensions using a QIAprep Spin Miniprep Kit (Qiagen Ltd). Preparations were performed according to the manufacturer's protocol.

2.7 Restriction digest cloning

2.7.1 Vector and insert design

All DNA vectors and DNA inserts for cloning were designed using Clone Manager Professional Suite 8.0 (Scientific & Educational Software), NEB Cutter V2.0 (New England Biolabs) and Plasmapper Version 2.0 <http://wishart.biology.ualberta.ca/PlasMapper/jsp/librarySeq.jsp?id=Clontech15>

2.7.2 Restriction endonuclease digestion

Following vector preparation and DNA (insert) purification protocols, vectors and DNA inserts (reference table 1) for cloning were digested with restriction endonucleases (using the appropriate enzymes) (reference table 2.5). Enzymes were purchased from NEB and Thermo Scientific and reactions were performed according to manufacturer's protocol. Where possible, reactions were terminated by heat incubation at 65°C or as suggested by the manufacturer. DNA inserts were purified via QIAquick PCR purification (Qiagen Ltd), according to manufacturer's protocol, prior to ligation.

2.7.3 Vector dephosphorylation

Vector dephosphorylation was performed to prevent re-circularisation of vectors used in restriction digest cloning (Table 2.5) after endonuclease digestion. 1 x Antarctic phosphatase reaction buffer (NEB) and 1 µl of Antarctic phosphatase (NEB) (per 1-5 µg of DNA) was added to the digested vector and incubated at 37°C for 30 minutes. Vector dephosphorylation was stopped by incubation at 65°C for 10 minutes. The vector was purified using a QIAquick PCR purification kit (Qiagen), according to manufacturer's protocol, prior to ligation.

2.7.4 Ligation and harvesting

Enzyme digested DNA inserts (Table 2.5) were ligated to prepared vectors in 20 µl reactions: 10-100 ng DNA insert, 10-100 ng prepared vector, 1 µl of T4 ligation buffer (NEB) and 1 µl T4 Ligase (NEB) supplemented with distilled H₂O to a final volume of 20 µl. The reaction mixture was incubated at 4°C, 16°C or room temperature for a 4 or 24 hour incubation period. Ligation reactions were performed at varying DNA insert to vector ratios: 1:1, 1:3, 1:5 and 3:1. 5 µl of the ligated products were transformed into 50 µl of Top10 competent cells (Invitrogen) and 100 µl of the transformation solution was spread on LB plates containing the appropriate antibiotic. Plates were incubated at 37°C overnight and any colonies were subjected to vector preparation (page 55). Following vector preparation, 20 µl of DNA sample was sent to GATC biotech to be sequenced.

DNA insert	Vector	Cloning technique	Tag
<i>MjaMja</i> -DP2	pET28a	LIC	His
<i>MjaMja</i> DP1	pET28a pET22b	RDC RDC	- His
<i>Mth</i> -DP2	pET28a	LIC	His
<i>Mth</i> -DP1	pET28a pET22b	RDC RDC	- His
<i>MjaMja</i> -DP2	pAW42	RDC	His
<i>MjaMja</i> -DP1	pLW40	RDC	-
<i>R. marinus</i> DNA Pol II	pET28a	TOPO, RDC	His
<i>S. YO3</i> DNA Pol II	pET28a	TOPO, RDC	His
<i>Hvo</i> -DP2	pTA1392	TOPO, RDC	His
<i>Hvo</i> -DP1	pTA1392	TOPO, RDC	Strep

Table 2. 5 Summary of the vectors used to clones DNA inserts and the cloning technique that was utilised. “LIC” represents ligase independent cloning, “RDC” represents restriction digest cloning, “TOPO” represents Invitrogen TOPO T7 sequencing kit.

2.8 Construct preparation and gene insertion for ligase independent cloning (LIC)

2.8.1 Vector preparation

Two vectors were provided by Dr Fogg (University of York, England) for use in LIC: pET-YSBLIC and pET-YSBLIC3C (supplementary data). Both vectors were linearised in a restriction digestion reaction (50 µg vector DNA, 50 µl BseRI (NEB), 100 µl NEB buffer 2 (10 ×), and water to final volume of 1 ml), and incubated at 37°C for 2 hours. After incubation, the digested vector was supplemented with 1 x agarose loading dye (NEB) and analysed via agarose gel electrophoresis (page 55). The cut vector was excised from the gel using a surgical blade and the DNA was extracted using a QIAquick gel extraction kit (Qiagen Ltd). After gel extraction the salts and buffers were removed using a QIAquick PCR purification kit (Qiagen Ltd) according to manufacturer protocols.

2.8.2 LIC vector and DNA insert T4 polymerase reaction

Linearised vector was subjected to a T4 polymerase reaction containing 5 mM DTT, 2.5 mM dTTP, 1 x T4 reaction buffer, 20 U LIC qualified T4 DNA polymerase (Novagen/Merck) and 4 pmol of BseRI linearised vector. DNA inserts were also extended in a T4 polymerase reaction containing 2 pmol PCR product (DNA insert), 2 µl T4 polymerase reaction buffer (10 x), 2 µl dATP, 1 µl DTT, 0.4 µl LIC qualified T4 polymerase (Novagen/Merck), and H₂O to a final volume of 20µl. Both reaction mixtures were incubated at 22°C for 30 minutes before the reaction was stopped by incubation at 75°C for 20 minutes.

2.8.3 LIC annealing

2 µl of DNA insert LIC T4 polymerase reaction mixture was added to 1 µl of LIC T4 polymerase prepared vector (~15 ng/µl) and incubated for 10 minutes at room temperature. The mixture was supplemented with 1 µl EDTA (25 mM), mixed with a pipette tip, and incubated at room temperature for a further 10 minutes prior to transformation into *E.coli* Top10 cells (Invitrogen).

2.9 TOPO Cloning

An Invitrogen TOPO® TA Cloning® Kit for Sequencing, with pCR™4-TOPO® Vector was used to clone DNA inserts that were difficult to PCR amplify (

Table 2. 5). The use of the kit allowed rapid and successful cloning of small concentrations of PCR product into the TOPO vector, which could then be subjected to plasmid preparation and digestion with restriction endonucleases to produce large amounts of DNA insert required for restriction digest cloning or LIC into the desired vector. TOPO® TA Cloning® Kits for Sequencing were used according to manufacturer's guidelines.

2.10 Screening for successful cloning: restriction digest, analytical PCR, and DNA sequencing

2.10.1 Analytical restriction digests

Colonies were picked from agar plates and incubated in 5ml of LB (with the appropriate antibiotic), shaking overnight at 37°C. The cultures were subjected to vector preparation (QIAprep Spin Miniprep Kit) according to standard Qiagen protocol. Diagnostic restriction digests with the appropriate restriction enzymes (those that cut at either side of insert) (Table 2. 6) were performed according to standard manufacturer's protocol. The digested DNA was visualised using a 1 % agarose gel with a DNA ladder (Fermentas) loaded into the outermost lane.

2.10.2 Analytical PCR

Analytical PCRs were performed using colonies recovered from ligation transformations. Colonies were incubated in 5 ml LB (with appropriate antibiotic) at 37°C for 12 hours then subjected to plasmid preparation according to Qiagen protocol. The primers that were originally used to amplify the DNA insert (Table 2. 4) were used again, in a standard PCR. Successful amplification of the DNA insert, as indicated by size on an agarose electrophoresis gel, was indicative of successful cloning.

2.10.3 DNA sequencing

All sequencing reactions were performed by GATC Biotech AG (Konstanz, Germany) using 20 µl of 30-100 ng/µl DNA samples. Where appropriate, sequencing was initially performed using a primer complementary to the T7 promoter (TAATACGACTCACTATAGGG). Additional primers were designed to anneal to the template DNA at appropriate sites to allow further sequencing reactions to be performed (Table 2. 4). Sequencing results were analysed via NCBI Nucleotide BLAST(http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch).

2.11 Protein expression, harvesting and purification

2.11.1 Protein expression

A summary of expression vectors used in this thesis is given in Table 2. 6.

Gene	Vector	Insertion site	Tag	Selection marker
<i>Pfu</i> -DP2	pET22b	Nde1/Not1	No tag	Amp
<i>Pfu</i> -DP1	pET28a	Nde1/BamH1	His	Kan
<i>Pfu</i> -DP1 Exo-	pET28a	Nde1/BamH1	His	Kan
<i>MjaMja</i> -DP1	pET22a	Nde1/Sal1	His	Amp
	pET28b	Nde1/Sal1	No tag	Kan
<i>MjaMja</i> -DP2	pET-YSBLIC	LIC	His	Kan
<i>MjaMja</i> -DP1	pLW40	Nsi1/Bgl11	His	Amp, Neo
<i>MjaMja</i> -DP2	pAW42	Asc1Bgl11	His	Amp, Pur
<i>Mth</i> -DP1	pET22a	Nde1/Sal1	No tag	Amp
	pET28b	Nde1/Sal1	His	Kan
<i>Mth</i> -DP2	pET-YSBLIC	LIC	His	Kan
<i>Pfu</i> -Pol B	pET17b	Nde1/EcoRV	No tag	Amp
<i>Hvo</i> -DP2	SMH618	-	Strep	Thy
<i>Hvo</i> -DP1 & <i>Hvo</i> -DP2	SMH622	-	His, Strep	Thy
<i>Hvo</i> -DP2	pTA1392	Pci1/Nhe1	His	Amp
<i>Hvo</i> -DP1	pTA1392	Nhe1	Strep	Amp
<i>Hvo</i> -DP1 & <i>Hvo</i> -DP2	pTA1392	Pci1/Nhe1 (DP2) Nhe1 (DP1)	Strep His	Amp
<i>R.marinus</i> DNA Pol II	pET28b	Nhe1	His	Kan
<i>S.YO3</i> DNA Pol II	pET28b	Nhe1/Sal1	His	Kan

Table 2. 6 Summary of expression vectors used in this thesis. “Amp” represents ampicillin, “Kan” represents kanamycin, “Neo” represents neomycin, “Pur” represents puromycin and “Thy” represents thymidine.

The plasmids used for the expression of *Pfu*-Pol D were supplied by Professor Yoshi Ishino of Kyushu University, Japan. *Hvo* strains, SMH618 (*Hvo*-DP2) and SMH622 (*Hvo*-Pol D) were supplied by Dr Stuart MacNeil of the University of St Andrews. Plasmids containing *Hvo*-DP1 (pTA326) and *Hvo*-DP2 (pTA327) subunits were a gift from by Dr Thorsten Allers, University of Nottingham. These vectors were used as template DNA in PCR amplification of *Hvo*-DP1 and *Hvo*-DP2 subunits. Dr Allers also provided the modified pTA1392 overexpression vector and *Hvo* strain H1424.

2.11.2 Protein expression and harvesting in *E.coli*

Vectors were transformed into *E.coli* BL21 (DE3)-RIL codon+ cells (Invitrogen/Life Technologies) according to the standard bacterial transformation protocol (page 58). A single transformant colony was used to inoculate 5 ml of LB (with appropriate antibiotic) and incubated at 37°C for 8 hours shaking (150 rpm). 1 ml of this culture was used to inoculate 50 ml LB (with appropriate antibiotic) and incubated at 37°C with shaking (150 rpm) overnight. 15 ml of overnight culture was used to inoculate 500 ml of LB (with appropriate antibiotic) and incubated at 37°C with shaking (150 rpm) until an O.D.₆₀₀ of 0.6 was reached. The cells were then induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration 1 mM) and incubated at 37°C for a further 5-7 hours (shaking ~150 rpm). The culture was centrifuged at 3750 g for 10 minutes at 4°C and the supernatant discarded. The remaining pellets were re-suspended in 30 ml of re-suspension buffer (50 mM Tris-HCl [pH 7.5], 500 mM NaCl, 0.1 mM EDTA, 10 % glycerol, 20 mM imidazole, 1 x protease inhibitor tablet (EDTA free) (Roche Applied Sciences, Burgess Hill, UK) per 50 ml culture) and stored at -20°C overnight.

2.11.3 Protein expression in *Haloferax volcanii* (*Hvo*)

Transformation colonies were picked and incubated in 50 ml *Hvo*-YPC media at 45°C shaking at 150 rpm, for 36 hours. 1 L of *Hvo*-YPC media, was then inoculated with 30 ml of the grown culture and incubated by shaking (150 rpm) at 45°C until an O.D.₆₀₀ of 0.8 was reached (~ 26 hours). The cell culture was then pelleted via centrifugation at 3750 g for 10 minutes at 4°C and the supernatant discarded. The remaining pellets were re-suspended in 30 μl of *Hvo* re-suspension buffer (50 mM Tris-HCl [pH 8.0], 2 M KCl, 0.1 mM EDTA, 10 % glycerol, 20 mM Imidazole, 1 x protease inhibitor tablet (EDTA free) (Roche Applied sciences) per 50 ml culture) and stored at -20°C overnight.

2.11.4 Protein expression and harvesting in *Methanococcus maripaludis* (*Mma*)

Transformant colonies containing the vector with the desired DNA insert were grown in 4 x 5 ml of *McCas* media, supplemented with 100 µl of 2.5 % Na₂S and pressurised to 30 psi with H₂/CO₂ with shaking at 37°C overnight. When an O.D.₆₀₀ of 0.7 was obtained 20 ml of growth culture was used to inoculate 2 L of *McCas* fermenter media in an anaerobic fermenter. Growth media was filtered with H₂/CO₂ to promote growth and performed at room temperature. *Mma* cultures were grown until an O.D.₆₀₀ of ~2 was obtained. The culture was centrifuged at 3750 g for 10 minutes at 4°C and the supernatant discarded. The remaining pellets were re-suspended in 30 µl of re-suspension buffer (50 mM Tris-HCl [pH 8.0], 500 mM NaCl, 0.1mM EDTA, 10 % glycerol, 20 mM imidazole, 1 x protease inhibitor (Roche Applied Sciences, Burgess Hill, UK) per 50 ml culture) and stored at -20°C overnight.

2.11.5 Protein purification from *E.coli*, *Haloferax volcanii* (*Hvo*) and *Methanococcus maripaludis* (*Mma*)

Frozen cell pellets were thawed at 37°C (~10 minutes) and sonicated on ice for 10 minutes in 30 second pulses. DNase I (Roche Applied Sciences) was added to the cell suspension and incubated at 37°C for 20 minutes. After DNase I digestion, the cell suspension was incubated at 60°C (*MjaMja*-DP1 and *MjaMja*-DP2 subunits) or 75°C (*Pfu*-Pol B, *Pfu*-Pol D, *MjaMja*-Pol D, *R.marinus* DNA Pol II and *S.YO3* DNA Pol II) for 20 minutes. Insoluble, denatured protein and cellular debris were pelleted via centrifugation at 46000 g, at 4°C for 45 minutes. The resulting clarified lysate was filtered through a syringe-driven Millex 0.45 µm filter unit (Millipore) to remove any insoluble material and the clarified lysate was stored on ice.

The protocol used to purify the over-expressed protein from the (*Huber et al., 2002*) clarified lysate was determined by the presence of a “tag” (Table 2. 6). Target proteins that contained His-tags were incubated with 2 ml of NiNTA agarose (Qiagen Ltd), on ice, shaking for 1 hour. The lysate was then poured through a gravity flow column and the flow-through, reapplied to the column. The

resin was washed with 2 x 50 ml of the appropriate wash buffer prior to elution with 4 ml elution buffer (

Table 2. 7). Purification of Strep tagged proteins was performed using Strep-Tactin Superflow Plus (Qiagen Ltd) resin. Purification was performed using a gravity flow column in a similar manner to NiNTA agarose purification. Strep tagged proteins were washed and eluted in the appropriate buffers (

Table 2. 7). Proteins that contained both a His and Strep tag, were initially purified using NiNTA agarose and then further purified using Strep-Tactin Superflow plus (Qiagen Ltd). Proteins were visualised via SDS-PAGE.

Protein purified	Source organism	Purification method	Purification buffers
<i>MjaMja</i> -Pol D <i>MjaMja</i> -DP2 <i>MjaMja</i> -DP1 <i>Mth</i> -Pol D <i>Mth</i> -DP2 <i>Mth</i> -DP1 S.YO3 Pol II <i>R.marinus</i> Pol II	<i>E.coli</i>	Ni-NTA agarose	A)10 mM Tris-HCl [pH 7.5], 500 mM NaCl, 50 mM imidazole B)10 mM Tris-HCl [pH 7.5], 500 mM NaCl, 500 mM imidazole
<i>Hvo</i> -Pol D	<i>Hvo</i>	Strep resin	A) 50 mM NaH ₂ PO ₄ , 300 mM NaCl, adjusted to pH 8.0 using NaOH B)50 mM NaH ₂ PO ₄ , 300 mM NaCl, 2.5 mM desthiobiotin, adjusted to pH 8.0 using NaOH
<i>Hvo</i> -DP1 <i>Hvo</i> -Pol D	<i>Hvo</i>	Ni-NTA agarose	A)10 mM Tris-HCl [pH 6.8], 2 M KCl, 50 mM imidazole B)10 mM Tris-HCl [pH 6.8], 2 M KCl, 500 mM imidazole
<i>MjaMja</i> -DP2 <i>MjaMja</i> -Pol D	<i>Mma</i>	Ni-NTA agarose	A)10 mM Tris-HCl [pH 8.0], 500 mM NaCl, 50 mM imidazole B)10 mM Tris-HCl [pH 8.0], 500 mM NaCl, 500 mM imidazole
<i>Pfu</i> -Pol B	<i>E.coli</i>	Heparin column	A)10 mM Tris-HCl [pH 7.5], 200 mM NaCl B)10 mM Tris-HCl [pH 7.5], 1.5 M NaCl

Table 2. 7 Buffers used in the first column chromatography purification step of proteins over-expressed in *E.coli*, *Hvo* and *Mma*. Purification buffer: A) is the wash buffer B) is the elution buffer.

Pfu-Pol B which lacked purification tags was purified using a 1 x 5 ml HiTrap DEAE column and 1 x 5 ml HiTrap Heparin column on an AKTAprime plus chromatography purification system (GE Healthcare). Columns were equilibrated with wash buffer (

Table 2. 7) in a series with the DEAE before the Heparin column. The cell lysate was loaded onto the columns and washed with the correct buffer until the A_{280} reading was at approximately zero. As *Pfu*-Pol B does not bind to DEAE this column was removed to avoid contaminating the *Pfu*-Pol B elution process with contaminants that do bind DEAE. A 30 minutes linear gradient of 0-100 % elution buffer running at 1 ml/min was used to elute *Pfu*-Pol B from the Heparin column. 1 ml fractions were collected and 10 μ l of each fraction was visualised via SDS-PAGE.

If the desired level of purification was not obtained after the first round of column chromatography a second purification step was employed, gel filtration. Product from the first column was passed through a Superdex200 10/300 GL column (GE Healthcare) to separate the desired overexpressed proteins from native proteins, based on size. The protein was first buffer exchanged into gel filtration buffer and concentrated to 500 μ l. 50 ml of the gel filtration buffer, was used to equilibrate the column, before the protein was injected on to the column. The gel filtration buffer was washed over the column at 1 ml/minute and the eluted protein was collected in 1 ml fractions. 20 μ l of each fraction was analysed on a 12 % SDS-PAGE gel.

Due to the high levels of NaCl (2-4 M) required by *Hvo* proteins, they were not subjected to gel filtration.

2.12 Protein storage, analysis, identification and concentration determination

2.12.1 Protein storage

Proteins purified from *E.coli* were buffer exchanged into storage buffer (Table 2. 8), using a 30 kDa Amicon Ultra centrifugal filter (Millipore) and several rounds of centrifugation (1000 g) according to the manufacturer's protocol. Proteins purified from *Hvo* and *Mma* were dialysed against storage buffer (Table 2. 8) using dialysis tubing. Purified proteins were placed within dialysis tubing and sealed. The dialysis tubing was placed in storage buffer (Table 2.8) and stirred using a magnetic flea. The storage buffer was changed 4 x over a 24 hour period. The dialysis tubing was removed from the storage buffer and covered in PEG 20,000 (Sigma adrich) until the volume was reduced to ~500 μ l. Concentration of proteins from *Mma* was performed under strictly anaerobic conditions.

Host strain	Storage buffer
<i>E.coli</i>	20 mM Tris-HCl [pH 7.5], 200 mM NaCl, 1 mM DTT
<i>Hvo</i>	20 mM Tris-HCl [pH 6.8], 2 M KCl, 1 mM DTT
<i>Mma</i>	20 mM Tris-HCl [pH 8.0], 400 mM NaCl, 1 mM DTT

Table 2. 8 Buffers used for the long term storage of purified proteins from *E.coli*, *Hvo* and *Mma*.

Purified proteins stored in the appropriate storage buffer (Table 2. 8), were filter sterilised using a 0.22 μ m syringe driven filter (Millipore). Proteins that required short-term storage (maximum 5 days) were incubated on ice in a cold room,

4°C. Proteins that required long term storage were mixed with an equal volume of 100 % glycerol and stored at -80°C.

2.12.2 Protein analysis- Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed USING SDS-PAGE gels. SDS-PAGE gels were composed of a separating gel (7-15% Design A Gel 37.5:1 acrylamide: bisacrylamide (National Diagnostics), 375 mM Tris-HCl [pH 8.8], 0.1 % SDS, 0.05 % ammonium persulphate (APS) and 0.05 % tetramethylethylenediamine (TEMED)) and stacking gel (4 % acrylamide: bisacrylamide (37.5:1), 125 mM Tris-HCl (pH 6.8), 0.1 % SDS, 0.05 % APS and 0.05 % TEMED). Proteins were supplemented with 1 x SDS loading buffer (Tris-HCl [pH 6.8], 2 % SDS, glycerol, 0.1 % Coomassie blue, 0.1 % bromophenol blue) and heated above 80°C for approximately 10 minutes to denature the cells. Gels were covered in SDS running buffer (25 mM Tris-HCl [pH 7.0], 250 mM glycine and 0.05 % SDS) with 5 µl of Precision Plus Protein Dual Color standard (Bio-Rad) loaded in the outermost lane. Once fully resolved, gels were incubated in SDS-PAGE stain (10 % acetic acid, 10 % isopropanol, 0.25 % coomassie blue 79.75 %) on a platform shaker at room temperature for 30 minutes. The staining solution was discarded and the gels were incubated with SDS-PAGE destain (10 % acetic acid, 10 % isopropanol and 80 % distilled H₂O) on a platform shaker at room temperature until stained protein bands were visible.

2.12.3 Protein identification

Bands on the SDS-PAGE gels which corresponded to the expected protein size were excised using a surgical blade and identified by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry analysis performed by York University, Department of Biology. Results were confirmed using NCBI BLASTp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>).

2.12.4 Protein concentration determination

The concentrations of proteins extracted from *E.coli* and *Mma* were calculated using the appropriate Beer-Lambert equation:

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

Where C is the concentration of the protein (M), A_{280} is the absorbance measured at 280 nm using a NanopDrop™ spectrophotometer (Thermo Scientific), l is the light path length of the quartz cuvette (1) and ϵ is the extinction coefficient of the protein calculated using the protein amino acid sequence and ExPASy ProtParam (<http://www.expasy.ch/tools/protparam.html>).

The concentration of *Hvo* proteins was determined using Coomassie Plus (Bradford) Protein Assays (Thermo scientific) as described by the manufacturer.

2.12.5 Extinction coefficients of purified proteins

Table 2. 9 lists the extinction coefficients of purified proteins. Extinction coefficients were calculated from the protein's amino acid sequence using ExPASy ProtParam.

Protein	Extinction coefficient at 280 nm ($\text{mM}^{-1}\text{cm}^{-1}$)
<i>Pfu</i> -DP1	62.8
<i>Pfu</i> -DP2	156.4
<i>Pfu</i> -Pol B	129.2
<i>MjaMja</i> -DP1	48.5
<i>MjaMja</i> -DP2	120.3
<i>Mth</i> -DP1	54.3
<i>Mth</i> -DP2	123.0
<i>Hvo</i> -DP1	50.0
<i>Hvo</i> -DP2	171.8
<i>S. YO3</i> DNA Pol II	88.8
<i>R. marinus</i> DNAPol II	98.4
<i>E. coli</i> DNA Pol II	137.5

Table 2. 9 Extinction coefficients of proteins purified at 280 nm in $\text{mM}^{-1}\text{cm}^{-1}$. Values were calculated using ExPASy ProtParam (<http://web.expasy.org/protparam/>).

2.13 Oligodeoxynucleotide hybridisation, primer-template extension and exonuclease assays

2.13.1 Oligodeoxynucleotide hybridisation assays

Complementary primer-templates were annealed at a 1:1.5 ratio (200 nM: 300 nM). 1 x annealing buffer (10 mM Hepes [pH 7.5], 100 mM NaCl and 1 mM EDTA) was added to the DNA and the mixture was heated to 95°C, in a heat-block, for 10 minutes. Reaction mixtures were left in the heat-block to cool slowly to room temperature.

Analysis of primer-template annealing was conducted using 12 % non-denaturing polyacrylamide gels (12 % polyacrylamide (National Diagnostics), 1 x TBE, 10 µl TEMED and 200 µl of 10 % APS made up to 40 ml with H₂O). 40 nM primer-template DNA was mixed with 2 x loading buffer (20 mM Tris-HCl [pH 7.5], 20 mM NaCl, 40 % glycerol, 2 mM EDTA) at a 1:1 ratio. 20 µl of the sample was loaded into the native gel along with a control sample that contained 20 nM primers. Samples were run at 4 W per gel, for 3 hours and fluorescence analysis was performed using a Typhoon scanner and ImageQuant software (GE Healthcare).

2.13.2 Primer-template extension assays

Primer-template extension reactions were performed using the primer-template oligodeoxynucleotides listed in the appropriate results sections. The primer-template reactions contain 40 nM primer-template (unless otherwise stated), 400 µM of each dNTP, 1 x reaction buffer (Pol D: 10 mM Tris-HCl [pH 9.0], 50 mM KCl, 10 mM MgCl₂, H₂O, 10 mM DTT or Pol B: 20 mM Tris-HCl [pH 8.0], 10 mM KCl, 2 mM MgSO₄, 10 mM (NH₄)₂SO₄, 0.1 % Triton X-100, 0.1 mg/ml BSA) and enzyme. Where commercial enzymes were used, the supplier's reaction buffer was used. The reaction mixture was incubated at the appropriate temperature (see results section) for 10 minutes before the addition of the required polymerase (concentration stated in results section). Primer-template extensions assays were quenched at specific time points (as stated in the appropriate results sections) by the addition of the appropriate 2 x stop buffer (95 % formamide, 10 mM EDTA, 10 mM NaOH, 500 nM appropriate template

competitor sequences) mixed at a 1:1 ratio with the reaction mixture. Once quenched, the mixture was incubated for a further 10 minutes and stored on ice prior to analysis.

Denaturing acrylamide gel electrophoresis was used to analyse primer-template extension reactions. 15 μ l of sample was loaded into each well of a 17 % acrylamide gel (17% acrylamide (National Diagnostics), 8 M Urea, 1 x TBE, 10 μ l TEMED and 200 μ l APS (made up to 40ml with distilled H₂O)). Gel electrophoresis was performed for 4-6 hours at 4 W per gel. Results were visualised using a Typhoon scanner (GE Healthcare) detecting the relevant fluorophore and ImageQuant software (GE Healthcare).

2.13.3 Exonuclease reactions

Exonuclease reactions were performed with the primer-templates listed in the appropriate results section. Reactions were performed and visualised in the same manner as primer-template extension reaction with the only variation being the omission of dNTPs.

2.14 Determination of K_D - fluorescence anisotropy

Fluorescence anisotropy assays were performed using a SLIM-8199 (Aminco) fluorimeter and quartz cuvettes (Hellma) with a 1 ml reaction volume. An excitation wavelength of 535 nm was passed through a band pass filter (BG20) (Schott) to remove any light at 570 nm wavelength. Excitation slits were set at 8 mm. Emitted light was detected by a photomultiplier through an OG-570 longpass filter (Schott). Reactions were carried out at room temperature in the following anisotropy buffer: 20 mM Tris-HCl [pH 8.5], 20 mM KCl, 5 mM MgCl₂, and 1 mM EDTA.

1 nM of Hex labeled oligodeoxynucleotide was added to 1 ml of anisotropy buffer. The anisotropy of the free oligodeoxynucleotides was recorded and polymerase was added to the mixture (concentration stated in results section). The anisotropy was measured after each addition. This process was continued until no further increase in anisotropy was detected. GraFit (Version 3.09a, Erithacus Software, Staines, UK), was used to fit the data to a tight binding

equation: (this equation makes no simplifying assumptions based on $[enzyme] \gg K_D$):

$$A = A_{min} + \frac{(D + E + K_D) - \sqrt{(D + E + K_D)^2 - 4 * D * E}}{2 * D} * (A_{max} - A_{min})$$

A = Anisotropy (min = anisotropy with no enzyme bound, max = anisotropy with enzyme bound)

D = Total primer-template concentration

E = Total enzyme concentration

K_D = Dissociation constant

2.15 Protein thermostability assays

2.15.1 Differential scanning fluorimetry

A Rotor-Gene-6000 RT-PCR machine (Corbett Life Science, Crawley, UK) was used to detect changes in fluorescence at 555 nm. 1 x reaction buffer (as stated in results section) 5 x SYPRO Orange (Invitrogen) and 2.5 μ M protein (final volume of 100 μ l) were mixed and placed in the Rotor-Gene-6000. The samples were heated to every degree increment between 35-100°C for one minute. The excitation source was set to 470 nm and the emission detection was set at 555 nm. The Rotor-Gene-6000 detected the changes in fluorescence and plotted the results on a graph.

2.15.2 DESERVED analysis

For DESERVED protein analysis, DESERVED buffer (1 μ M 8-Anilino-1-naphthalenesulfonic acid (ANS), 40 mM Tris-HCl [pH 7.5], 400 mM NaCl, 2 mM DTT and distilled water) was mixed with 10 μ M protein at a 1:1 ratio to give a total volume of 250 μ l. The mixture was mixed well by pipetting and incubated at 80°C or 90°C on a heat block. 10 μ l aliquots were removed after 0, 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 seconds and placed on ice for 5 minutes. After incubation on ice, samples were left to equilibrate to room temperature prior to analysis on a NanoDrop fluorospectrometer (Thermo Scientific). The NanoDrop fluorospectrometer had an excitation wavelength using the UV LED (360 nm \pm 10) and was blanked with ANS buffer. Data

points were collected every nanometer increment from 395 to 751 nm. Samples were analysed in triplicate and multiple samples were analysed, non-consecutively, to minimise the effect of machine drift over the course of the experiment. Total fluorescence was calculated as the sum of fluorescence intensity between 395 and 751 nm for each sample.

The Barycentric mean wavelength (BCMw) was calculated using the area of ANS emission spectrum that is least susceptible to noise generated while using small volumes (475-543 nm). Data for the mean and the BCMw were calculated from 3 replicates and plotted using Microsoft Excel.

2.16 Plasmid based fidelity assay

Gapped pSJ3 plasmid was gifted by Brian Keith (Newcastle University) for use in fidelity reactions. Fidelity reactions consisted of 77 ng of gapped pSJ3 plasmid, 1 x Pol D reaction buffer (40 mM Tris-HCl [pH 9.0], 50 mM KCl, 10 mM MgCl₂, 10 mM DTT, H₂O), 1 mM dNTPs, 100 nM enzyme to a final volume of 20 µl. 4 x reaction mixtures, one control (no enzyme added) were incubated at 70°C for 1 hour. Following incubation an analytical EcoR1 digest was performed using 18 µl of each reaction mixture that contained enzyme. After digestion, the samples were run on a 1 % agarose gel, and the size was compared to the undigested control sample. A shift in size was indicative of successful filling of the gap by the enzyme. If successful, *E.coli* Top10 competent cells (Invitrogen) were transformed with 1 µl of undigested reaction mixture according to standard protocol. Transformed solution was diluted ¼ in L.B and 150 µl was spread on 6 x agar plates. Agar plates supplemented with 100 mg/ml ampicillin, 1 mM IPTG and 1 mM X-gal were incubated at 37°C overnight. Following incubation, a digital camera was used to photograph the agar plates and the total number of colonies on each plate was counted using ImageQuant software (GE Healthcare). The images were converted to monochromatic Tiff files using GIMP prior to analysis with ImageQuant. White colonies were counted manually. White colonies from each fidelity assay were picked and restrike onto agar plates (100 mg/ml ampicillin, 1 mM IPTG and 1 mM X-gal) and incubated at 37°C overnight. White colonies were grown in 5 ml of LB (Amp), subjected to plasmid preparation and sent to GATC for

sequencing. Fidelity rates were then calculated based on observations detailed in the results section. The background mutation rate was calculated by Brian Keith (Keith *et al.*, 2013).

2.17 DNA and protein ladders

The GeneRuler 1 kb or 100 bp DNA ladder (Fermentas) was loaded in the outermost lane of all agarose gels and used as an indicator of size for DNA samples. The Precision Plus Protein Dual Color standard ladder (Bio-Rad) was loaded in the outer most lanes of all SDS-PAGE gels as an indicator of the size of protein

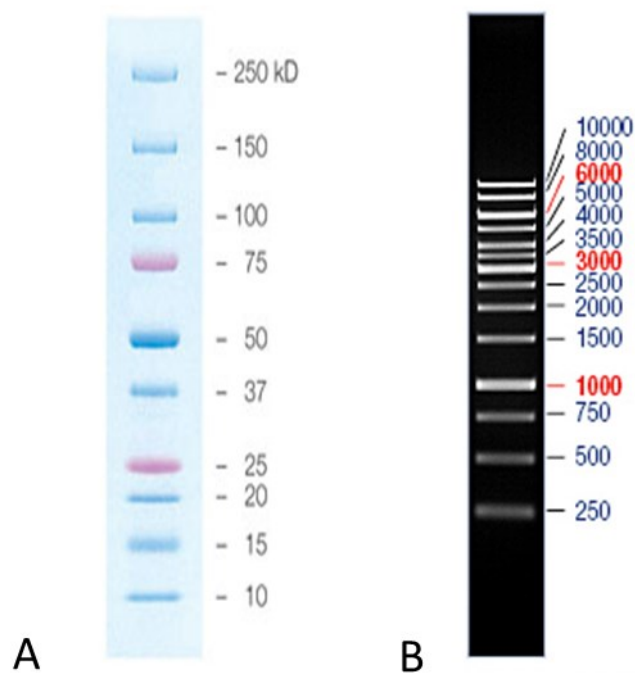


Figure 2.1 DNA and protein ladders. A) Precision Plus Protein Dual color standard ladder used in all SDS-PAGE gels. B) GeneRuler 1kb DNA ladder.

Inhibition of archaeal Pol D by uracil

3.1 Background

Current understanding of archaeal DNA replication has highlighted both Pol B and Pol D as the probable replicative polymerases in the archaea (Cann *et al.*, 1998; Cann and Ishino, 1999; Barry and Bell, 2006; Rouillon *et al.*, 2007; McCulloch and Kunkel, 2008), with targeted gene deletion experiments showing that both enzymes are essential for viability in halophilic euryarchaeon (Berquist *et al.*, 2007). Pol B has been characterised as a processive, high fidelity, replicative polymerase that possesses unique uracil “read ahead” recognition properties (Greagg *et al.*, 1999; Russell *et al.*, 2009). The uracil recognition mechanism of Pol B has been described in detail and crystallography structures which show the uracil binding pocket capturing a deaminated base are available (Firbank *et al.*, 2008).

Recent publications have shown that Pol D also possesses many properties associated with replicative polymerases. Pol D interacts with proliferating cell nuclear antigen (PCNA) at the replisome, has strand displacement and exonuclease activity, and is able to elongate both DNA and RNA (Isaac K. O. Cann, 1999; Henneke *et al.*, 2005; Rouillon *et al.*, 2007). These properties led to the hypothesis that Pol D acts after initiation of primase and at a later stage becomes responsible for lagging strand synthesis while Pol B continues the leading strand synthesis (Henneke *et al.*, 2005; Rouillon *et al.*, 2007; Castrec *et al.*, 2009). If this hypothesis is correct, then Pol D is responsible for the replication of 50 % of the genome within archaea (except Crenarchaeota).

However, recent experiments have shown that Pol B can be deleted from both *M.maripaludis* and *T.kodakarensis* while Pol D is found to be essential (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013). These findings have led to a second hypothesis regarding the role of Pol D in archaeal DNA replication. In this hypothesis, Pol D is the main replicative polymerase in archaea (except Crenarchaea) and is responsible for the majority or all of the DNA replication *in vivo*. Thus, it is conceivable that Pol D would possess uracil recognition properties similar to those identified in archaeal Pol B. At present little is known about Pol D's ability to recognize uracil, however, a brief report has indicated

that Pol D does not copy DNA strands containing uracil or incorporate dUTP into expanding DNA strands (Sawai *et al.*, 2007).

This chapter further investigates the uracil recognition properties of Pol D and proposes a hypothetical model for the molecular mechanism involved in uracil recognition.

3.2 Purification of *Pyrococcus furiosus* Pol D

Pol D consists of two subunits: DP1, the small subunit, possesses 3'-5' exonuclease activity and DP2, the large subunit, possesses polymerase activity. Vectors containing the DP2 and DP1 subunits of *Pyrococcus furiosus* (*Pfu*) Pol D were gifted by Professor Yoshi Ishino of Kyushu University, Japan. The DP1 and DP2 subunits were located on pET-28a (His-tagged) and pET-21a (non-His tagged) plasmids respectively.

The DP1 and DP2 subunits of *Pfu*-Pol D were co-expressed in BL21 (DE3) pLysS cells. The co-expressed *Pfu*-Pol D protein was initially purified using a heat step (80°C) to denature the native *E.coli* proteins. Further purification was performed using a His Trap column which bound the His-tagged DP1 subunit followed by purification using a gel filtration column to separate the proteins based on size. The use of three purification techniques resulted in relatively pure *Pfu*-Pol D protein with minimal contamination from endogenous *E.coli* proteins (

Figure 3. 1).

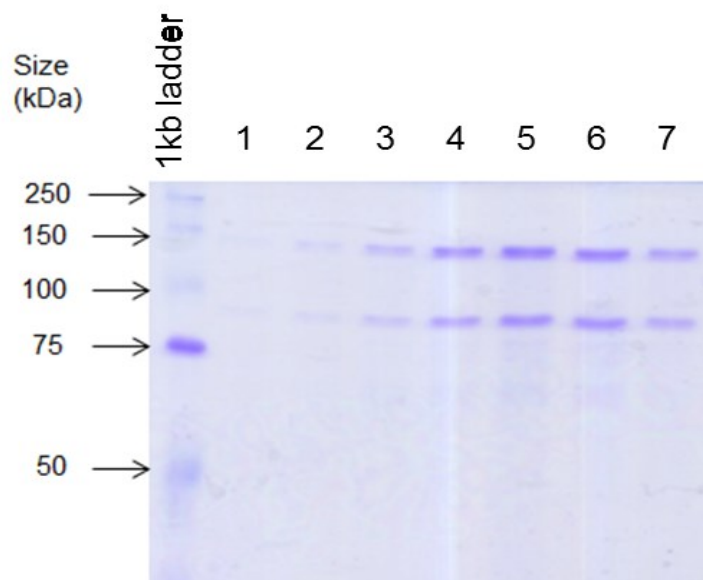


Figure 3. 1 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing purified *Pfu*-Pol D. Lanes 1-7 show the large, DP2 subunit (~148kDa), and small, DP1 subunit (~80kDa), of *Pfu*-Pol D after purification from *E.coli*

cells. Purification was performed on an Akta using a His Trap column and gel filtration column.

3.3 Design of uracil-containing oligodeoxynucleotides

To investigate if *Pfu*-Pol D is inhibited by uracil and how far ahead of the primer template junction the enzyme is able to recognize uracil, single stranded oligodeoxynucleotides were designed that contained uracil (thymine in controls) located at a defined position (Figure 3. 2). A complementary Cy5 labeled primer was annealed to the single stranded oligodeoxynucleotides to produce primer-template substrates (Figure 3. 2).

Primer-template

5' - **Cy5** - GCGTGATCTGATCAACGC - 3

3' CGCGACTAGACTAGTTGCGGCCTGCAAGTGCAATTCTGGCATGACTCGAATCTCGTAACG (U/T) TACTTAC

Primer-template T70/U70

5' - **Cy5** - GCGTGATCTGATCAACGC - 3

3 - CGCGACTAGACTAGTTGCGGCCTGCAAGTGCAATTCTGGCATGACTCGAATCTCGTAACGTTACTTAC
CATTGCAGTACGCAGTTAGA (U/T) CAATTAC - 5'

Primer-template T102/U102

5' - **Cy5** - GCGTGATCTGATCAACGC - 3

3' CGCGACTAGACTAGTTGCGGCCTGCAAGTGCAATTCTGGCATGACTCGAATCTCGTAACGTTACTTAC
CATTGCAGTACGCAGTTAGATCAATTACTATTGCACATCGCATATCGTCCAG (U/T) AAATCCG - 5'

Primer-template T34/U34

5' - **Cy5** - GCGTGATCTGATCAACGC - 3

3' CGCGACTAGACTAGTTGCGGCCTGCAAGTGCAATTCTGGCATGACTCGAATCTCGTAACGTTACTTAC
CATTGCAGTACGCAGTTAGATCAATTACTATTGCACATCGCATATCGTCCAGTAAATCCGGAATGAAG
TGACCGGAGTTCGACA (U/T) CTTAGTA - 5'

Figure 3. 2 Primer-templates containing template-strand uracil at a defined position. Uracil (thymine in controls), highlighted in red, is located 42, 70, 102 or 134 bases ahead of the primer template junction. The oligodeoxynucleotides are named based on the location of uracil ahead of the primer template junction.

3.4 Synthesis of uracil-containing, long oligodeoxynucleotides

Oligodeoxynucleotide synthesis is a routine technique; however it is limited by the lengths that can be easily prepared. Currently chemical synthesis using the phosphoramidite method is limited to about 120 bases (Brown and Brown, 1991; Hughes *et al.*, 2011). Thus, the template used to prepare T134/U134, 160 bases in length, was prepared in two halves and joined using DNA ligase and a “splint” oligodeoxynucleotide (Table 3.1) as illustrated in Figure 3. 3. Prior to the introduction of this technique, an alternative approach, based on PCR, illustrated in Figure 3. 4 was attempted (Keith *et al.*, 2013).

Oligodeoxy nucleotide	Sequence
A	5'ATGATTC <u>U</u> ACAGCTTGAGGCCAGTGAAGTAAGGCCTAAATGACCTGCTATACGCTACACGTTATCATTAACTAGATTGAC-OH-3'
B	5' <u>pGCATGACGTTACCATTCATTGCAATGCTCTAAGCTCAGTACGGTCTTAACGTGAACGTGAACGTCCGGCGTTGATCAGATCAGCG</u> C-3'
C	3'-GTAATTGATCTAACTGCGTACTGCAATGGTA-5'

Table 3. 1 DNA sequence of oligodeoxynucleotides used to synthesise single stranded T134/U134 templates. Underlined region represents DNA complementary to the “splint”, C, oligodeoxynucleotide. “p” represents a phosphate group and “OH” represents a hydroxide group. Uracil is highlighted in red.

In this approach PCR is used to amplify a long stretch of DNA. One of the PCR primers contains 3' phosphothioate residues at the 5' end; the second primer contains a 5' phosphate (Figure 3.4). A subsequent digestion of the PCR product with lambda exonuclease, which strongly degrades 5'-phosphate DNA but is unable to digest phosphorothioate capped DNA, gives a single stranded oligodeoxynucleotide (Figure 3.4). Unfortunately, this PCR based technique produced exceptionally low yields and so, after preliminary experiments, chemical synthesis of the oligodeoxynucleotide was pursued (Figure 3.3).

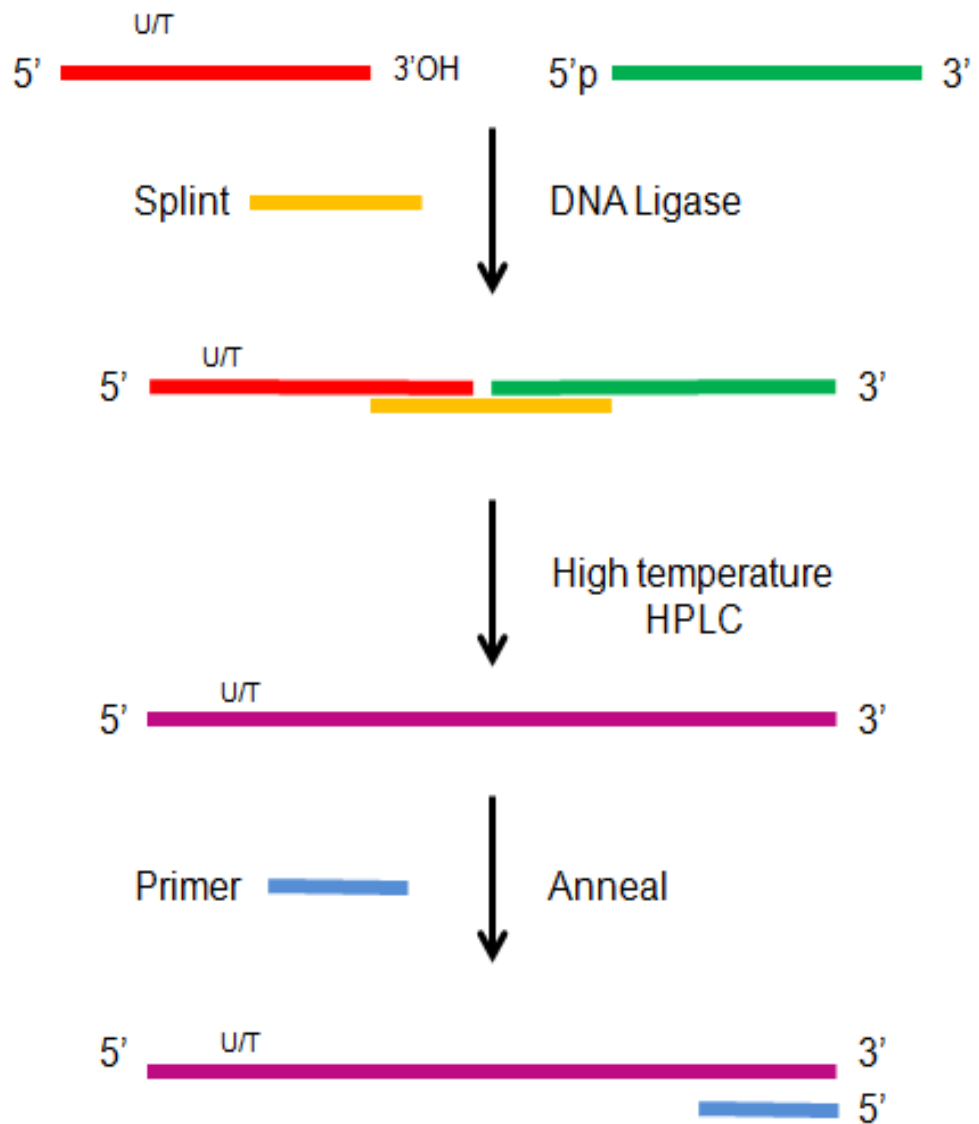


Figure 3.3 Synthesis of the T134/U134 single stranded long oligodeoxynucleotides with uracil (thymine in controls) located at a defined position. An 18mer, Cy5 labelled primer was annealed to create the desired primer-template with uracil located 134 bases ahead of the primer template junction.

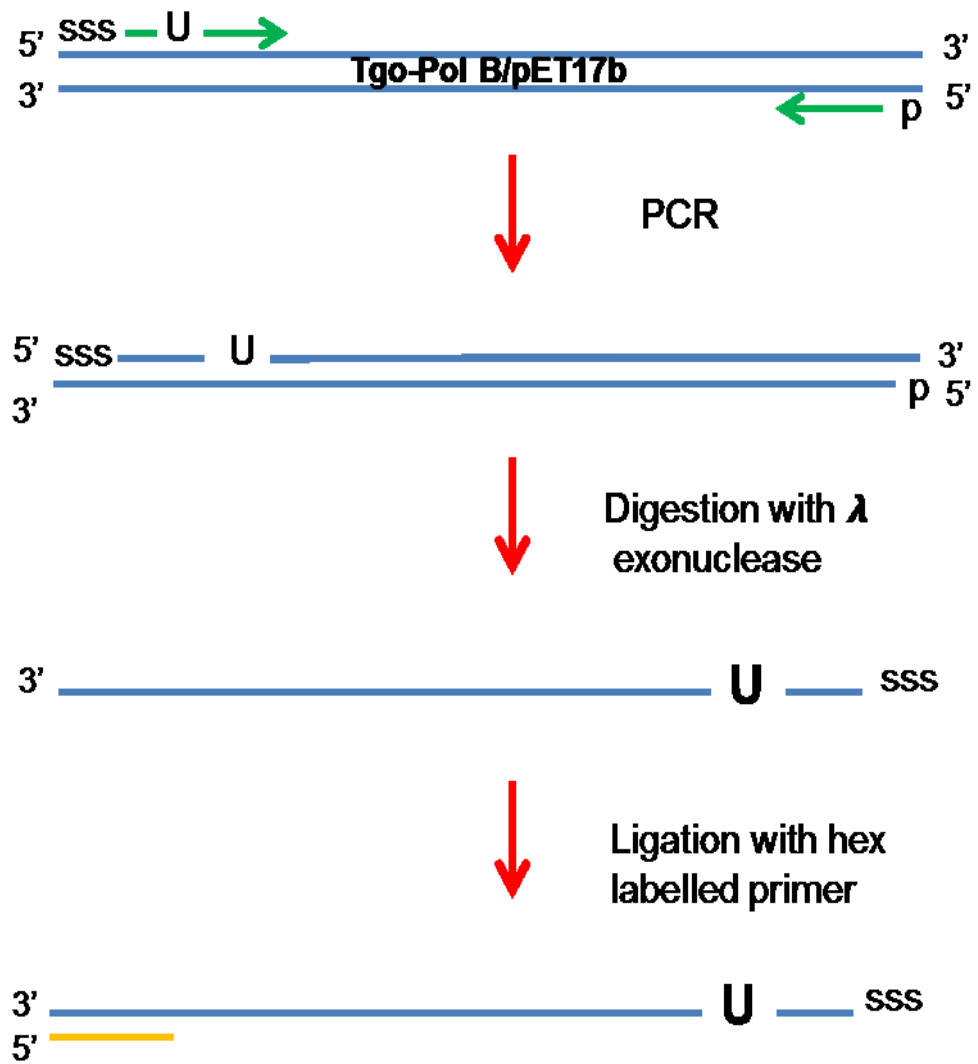


Figure 3. 4 PCR based method for synthesis of long oligodeoxynucleotides. “sss” represents three phosphothioates at the extreme 5’ end, “p” represents a 5’ terminal phosphate group, “U” represents uracil (thymine in controls).

3.5 Primer-template annealing assays

The long single stranded oligodeoxynucleotides (250 nM), were hybridised to the complementary Cy5 labelled primer (200 nM) and analysed via gel shift analysis to ensure that the substrates were fully annealed (Figure 3. 5). Full annealing was observed as shown in Figure 3. 5 for T70/U70, T102/U102 and T134/U134. Complete annealing was also seen for T42/U42 (result not shown).

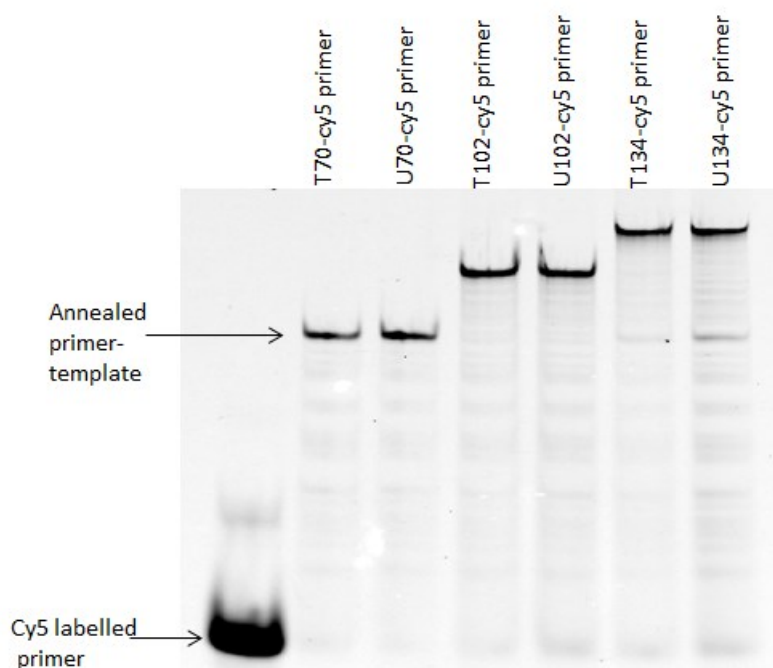


Figure 3. 5 Hybridisation of a Cy5 labelled primer to the T70/U70, T102/U102, and T134/U134 DNA templates. The Cy5 labelled primer was detected using a Typhoon Scanner (GE Healthcare) and visualised using ImageQuant software.

3.6 Extension of uracil containing primer-templates by *Pfu*-Pol D

Primer extension reactions were used to determine if the presence of uracil located 42, 70, 102 or 134 bases ahead of the primer template junction influenced the rate of DNA extension. Reactions were initiated by the addition of 80 nM of *Pfu*-Pol D.

All primer extension reactions were performed at 50°C with 40 nM of primer-template. T70/U70, T102/U102 and T134/U134 reactions were quenched after 0, 5, 15, 30 and 45 minutes and the T42/U42 reactions were quenched after 0, 2, 10, 20 and 30 minutes. Reactions were stopped in buffer containing 500 nM of competitor sequence (to sequester excess template DNA). The reaction mixtures were loaded into a 17 % acrylamide gel mix and analysed via PAGE (Figure 3.7, Figure 3.8, Figure 3. 9 & Figure 3. 10). The results were visualised using a Typhoon scanner (GE Healthcare) and the percentage of fully extended primer was calculated using ImageQuant software (Figure 3.6). The percentage of fully extended primer was calculated as the amount of fully extended product/total product x 100.

Analysis of the rate of primer extension highlighted a significant level of inhibition of polymerization when uracil was located 42, 70 and 102 bases ahead of the primer-template junction compared to thymine control reaction (Figure 3.7, Figure 3.8 and Figure 3.9). Inhibition was observed as a reduced rate of primer extension; however, fully extended primer was obtained in reactions containing U42, U70, U102 and U134 (Figure 3. 7- Figure 3. 10).

When uracil is located further ahead of the primer-template junction, 134 bases, an inhibition of polymerisation was no longer observed (Figure 3. 10). Experiments performed by Dr Tomas Richardson showed that *Pfu*-Pol D binds to uracil located in double stranded DNA. However, recognition of uracil in double stranded DNA only occurs when uracil is in close proximity to the primer-template junction (Richardson *et al.*, 2013). Thus, the lack of inhibition when uracil is located 134 bases ahead of the primer-template junction may be due to the long single stranded DNA folding to form stem-loop structures which locate the uracil within a double stranded region.

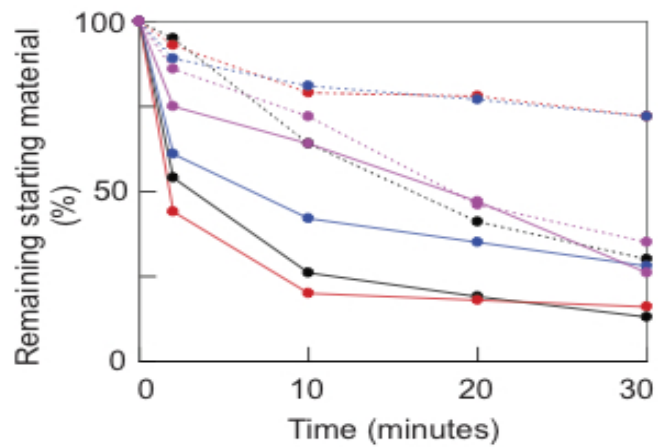


Figure 3. 6 Summary of data from Figure 3.7-Figure 3.10 showing remaining primer-template over time. Colour coding: black, +42; red, +70; blue, +102; magenta, +134. Solid lines, T; hatched lines, U.

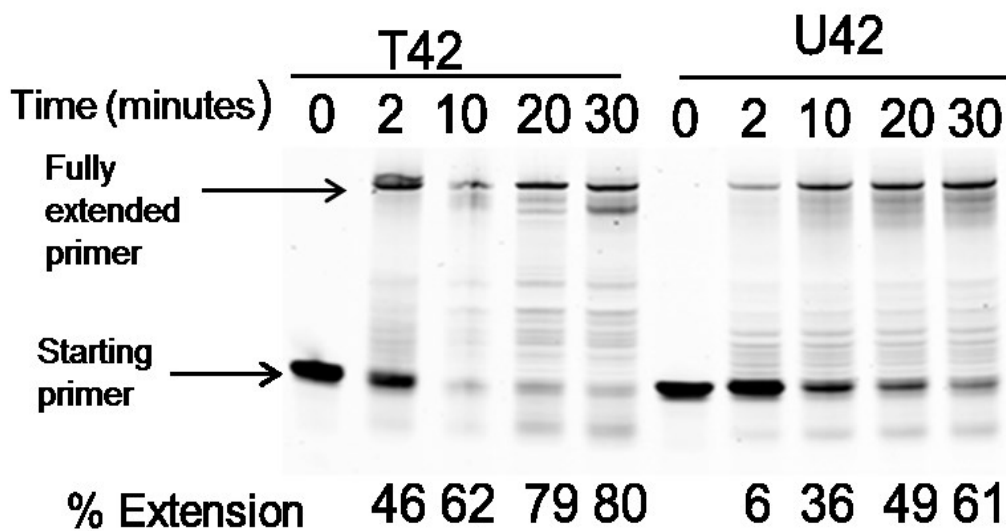


Figure 3. 7 Primer extension reaction with T42/U42. The percentage of starting primer extended was calculated with ImageQuant software.

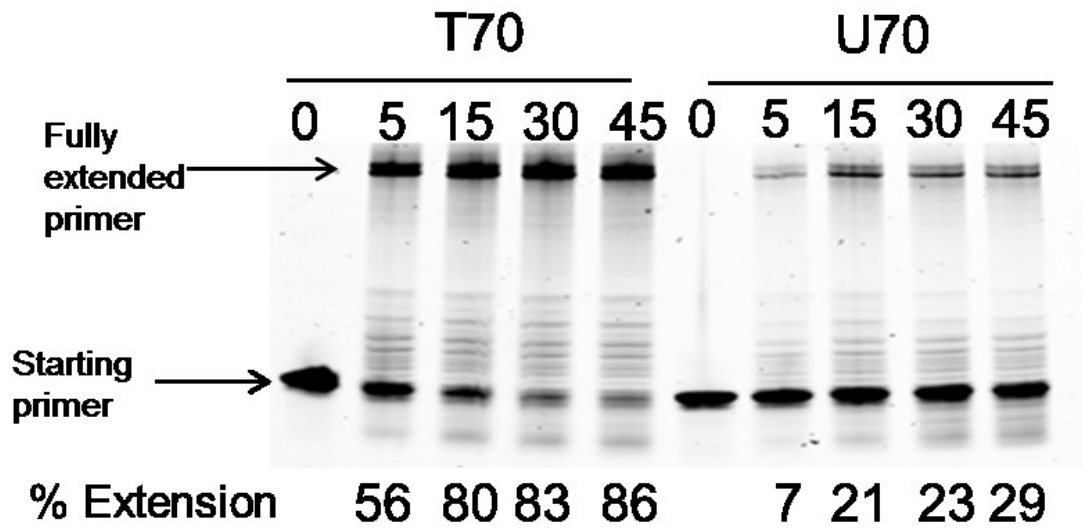


Figure 3. 8 Primer extension reaction with T70/U70. The percentage of starting primer extended was calculated with ImageQuant software.

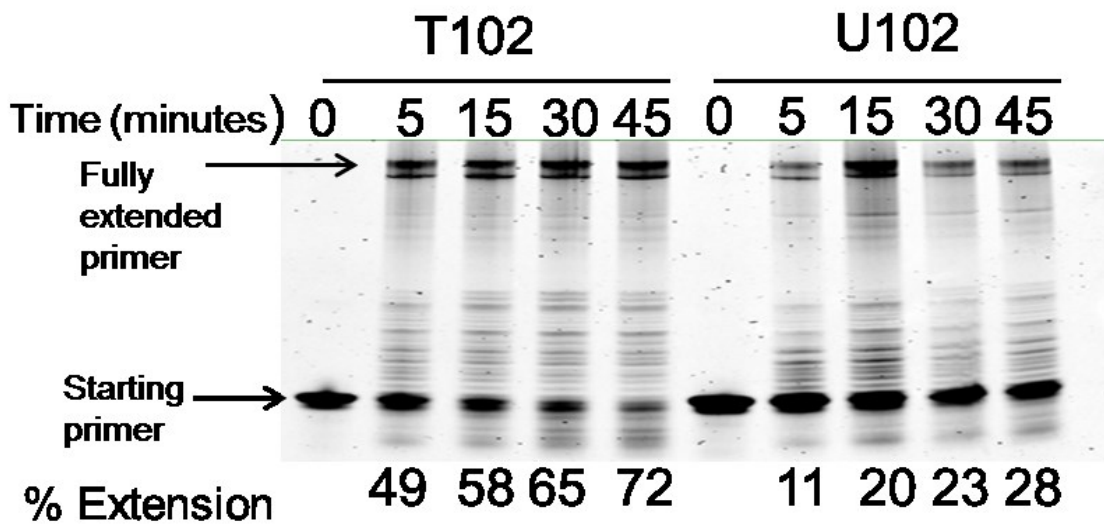


Figure 3. 9 Primer extension reaction with T102/U102. The percentage of starting primer extended was calculated with ImageQuant software.

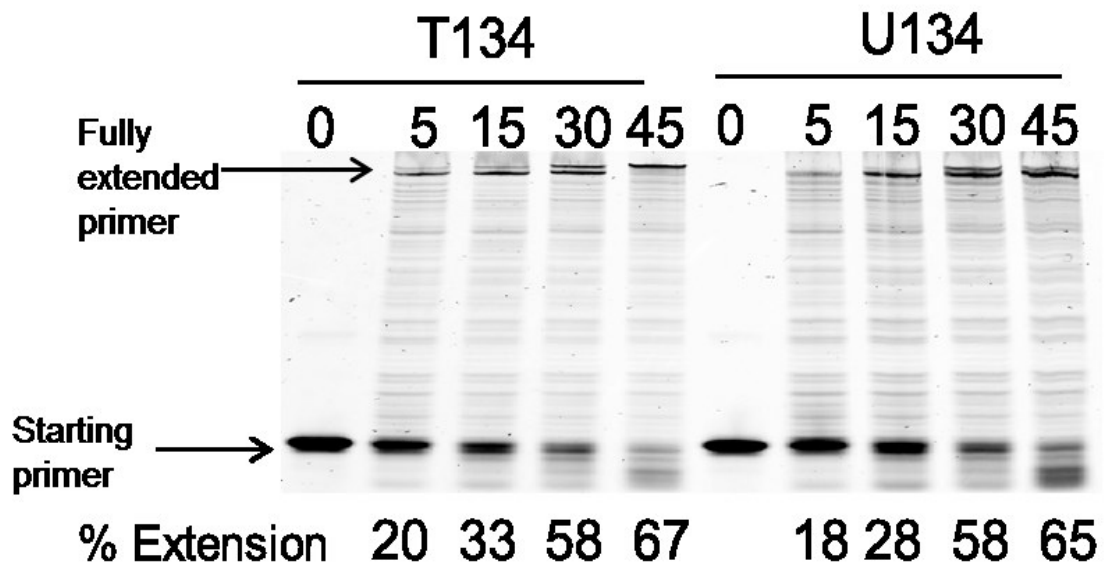


Figure 3. 10 Primer extension reaction with T134/U134. The percentage of starting primer extended was calculated with ImageQuant software.

3.7 Copying of replication forks by DNA Pol D: archaeal DNA replication *in vivo*

DNA replication is a complicated, multi-enzymatic process that is essential for life. Across the three domains of life, the general mechanisms and principals of DNA replication have a high degree of similarity and include recognition of defined origins, melting of double stranded DNA, RNA priming, recruitment of DNA polymerases and processivity factors, and replication fork formation (Barry and Bell, 2006; Hamdan and Richardson, 2009; O'Donnell *et al.*, 2013).

Despite these similarities, the protein machineries responsible for these processes are considerably different between the three domains of life e.g. replisome components, including polymerases and primases, in bacteria are unrelated or only distantly related to their counterparts in archaea and eukaryotes (Figure 1.6) (Edgell and Doolittle, 1997). To date, the majority of research on DNA replication has focussed on bacterial and eukaryotic systems. However, in recent years, research into archaeal DNA replication has expanded and revealed protein sub complexes predicted to constitute the replisome (

Figure 3. 11) (Li *et al.*, 2010). Despite increased understanding of the archaeal DNA replication process, the exact roles of the replicative polymerases, Pol B and Pol D are still unknown (Cubonova *et al.*, 2013; Richardson *et al.*, 2013). Thus, to further characterise the uracil recognition properties of Pol D, it was desirable to investigate how it identifies uracil present in the context of a replication fork. As studying DNA replication *in vitro* is exceptionally challenging, replication fork mimics were designed, which were amenable to experimentation. These forks were used to further investigate the polymerisation and uracil recognition properties of *Pfu*-Pol D.

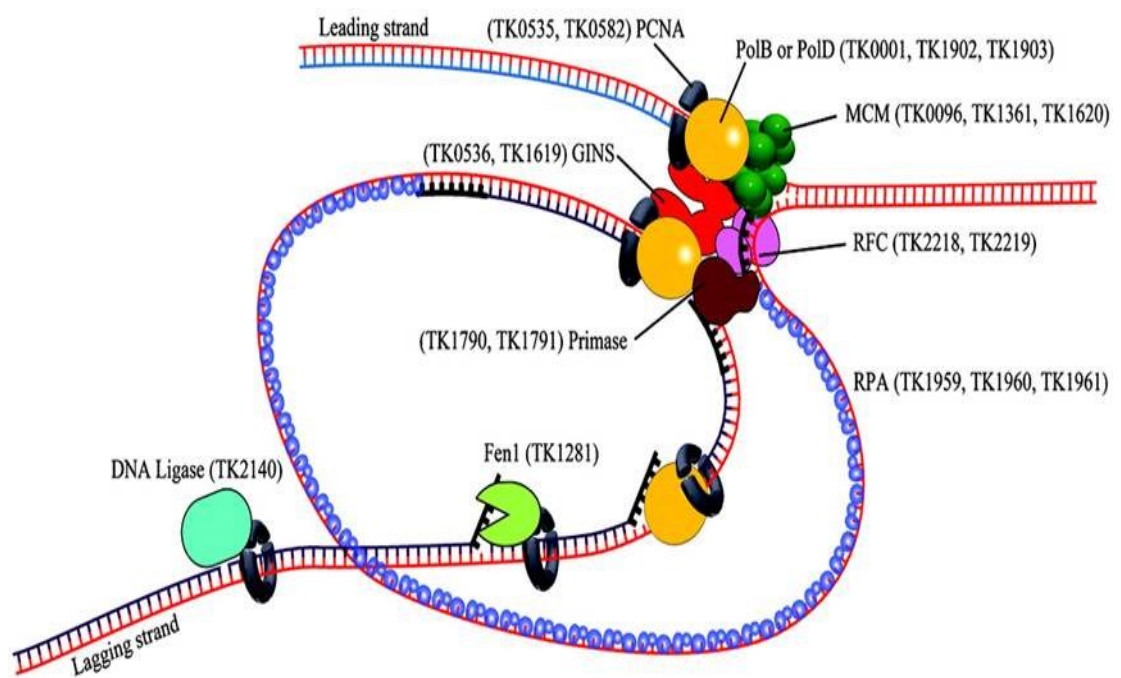


Figure 3. 11 Components of the archaeal replisome. The *T. kodakaraensis* numerical gene designations are listed adjacent to the protein sub complexes predicted to constitute the replisome. Image taken from Li *et al.*, 2010.

3.8 Replication fork mimic design

A replication fork mimic was designed, based around a single oligodeoxynucleotide, 106 bases in length (Figure 3.12). The 106-mer has a self-complementary double stranded region 16 bases long held together by a flexible linker composed of 4 thymidines. Two non-complementary single-stranded prongs, 35 bases in length, splay out to form leading and lagging strands (Figure 3. 12).

The design of the non-complementary leading and lagging strands facilitated the specific annealing of two different primers (Figure 3. 12). The primers were labelled with fluorophores (leading strand primer, Cy5; lagging strand primer, fluorescein), that had different spectral properties to allow each primer to be individually detected by scanning at different wavelengths (using a Typhoon scanner). The replication fork mimics were designed to contain uracil (thymine in controls) located 4 bases ahead of the primer template junction on both the leading and the lagging strands (Figure 3.12)

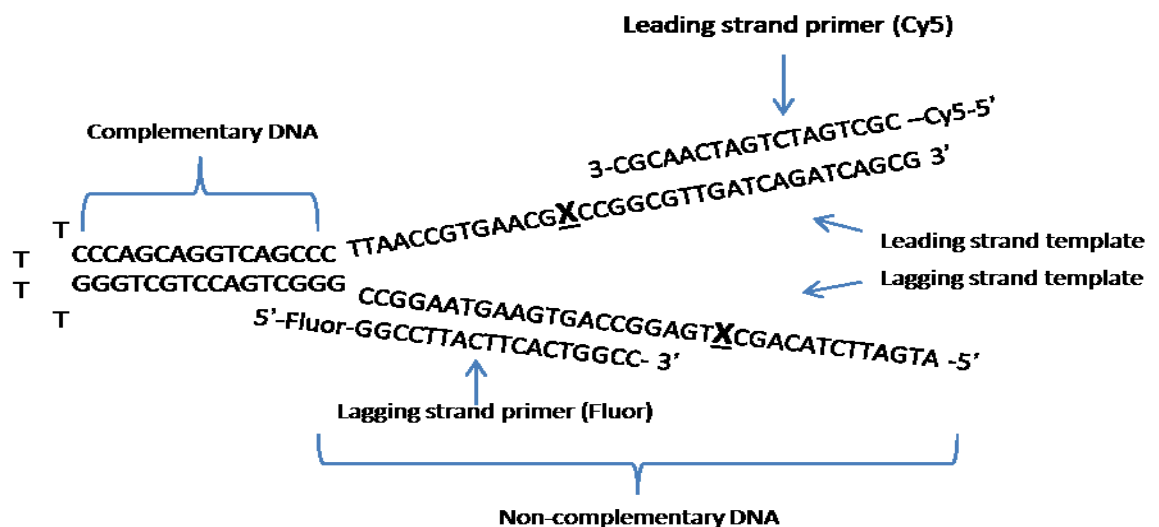


Figure 3. 12 106 base DNA fork mimics synthesised as a single oligodeoxynucleotide strand. There is a 16mer region of complementarity between the “leading” and “lagging” strands joined by a 4-thymidine linker. The remaining 35 nucleotides are non-complementary. Fluorescent labelled primers were designed that were complementary to the leading and lagging strands to allow primer-extension reactions to be performed. The 3’ end of both leading (Cy5) and lagging (Fluor) strand primers are positioned +4 bases from “X” (representing uracil or thymine in control).

3.9 Attempted “click” chemistry synthesis of replication fork mimics

As mentioned earlier, the length of oligodeoxynucleotides achievable by chemical synthesis is limited; the 106 base fork mimic being at the upper limits of accessibility. In an attempt to prepare longer forks, use was made of “click chemistry” (Brown and Brown, 1991; Kocalka *et al.*, 2008). The click chemistry technique required two chemically synthesised oligodeoxynucleotides, one containing a 3'-alkyne group, the other a 5'-azido group (Figure 3. 13). The two oligodeoxynucleotides were coupled in a ligation reaction (Figure 3. 13) to form the desired 124 base replication fork mimic.

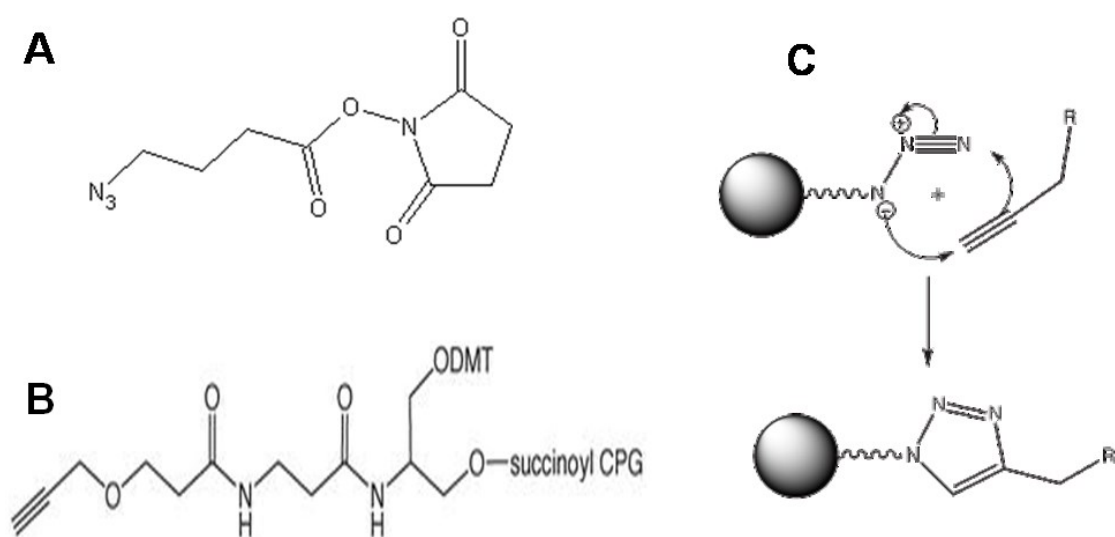


Figure 3. 13 Materials used for click chemistry synthesis of replication fork mimics. A) 5' end of top strand was modified with an amino group using an azide-containing 5'-amino modifier. B) The 3' end of the bottom strand was modified with alkyne using 3'-alkyne-modifier serinol CPG (Glen Research). C) Click chemistry mechanism, which covalently links two hybridised oligodeoxynucleotides which contain a 3'alkyne and a 5' azide. The reaction is driven by the proximity of the two reactive groups. Illustrations were drawn using ChemSketch.

Synthesis of the replication fork mimics via click chemistry did prove to be successful, however low yields of the final product were produced. Thus after preliminary experiments with the click chemistry replication fork mimics synthesis using the single long oligodeoxynucleotides was pursued, despite its length limitations (Figure 3.12)

3.10 Relication form mimics nomenclature

The replication fork mimics used in this study are shown in Figure 3.14. They are named according to the base at the +4 position ahead of the primer-template junction and the presence of fluorescent labelled primers.

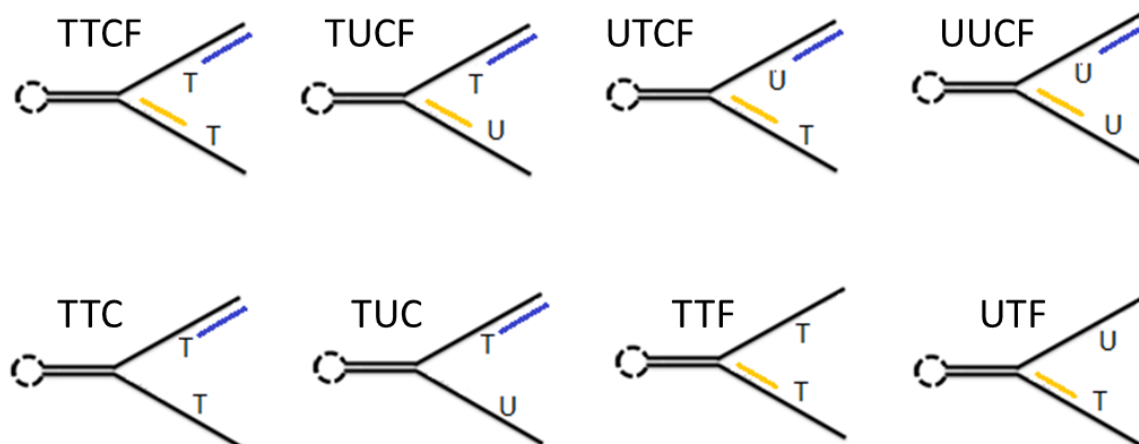


Figure 3. 14 Nomenclature of replication fork mimics. T (thymine) or U (uracil) indicates the base at the +4 position from the primer-template junction. C (Cyanine 5, abbreviated Cy5) represents the binding of the Cy5 labelled leading strand primer; F (Fluorescein, abbreviated Fluor) represents the binding of the lagging strand fluorescein primer.

3.11 Annealing of primers to replication fork mimic

Annealing of leading (Cy5) and lagging (Fluor) strand primers to the replication fork mimics was analysed using gel-shift analysis. The size of the annealed primer-template was compared to the primer (control) to detect if annealing was successful (Figure 3. 15). Detection of the two individual primers was achieved by scanning the gels at different wavelengths (~650/670 nM for Cy5 detection, 460 nM for Fluor detection) using a Typhoon scanner (Figure 3. 15). Full annealing was observed as shown in Figure 3. 15 for TTCF. Complete annealing was also seen for TUCF, UTCF and UUCF (not shown).

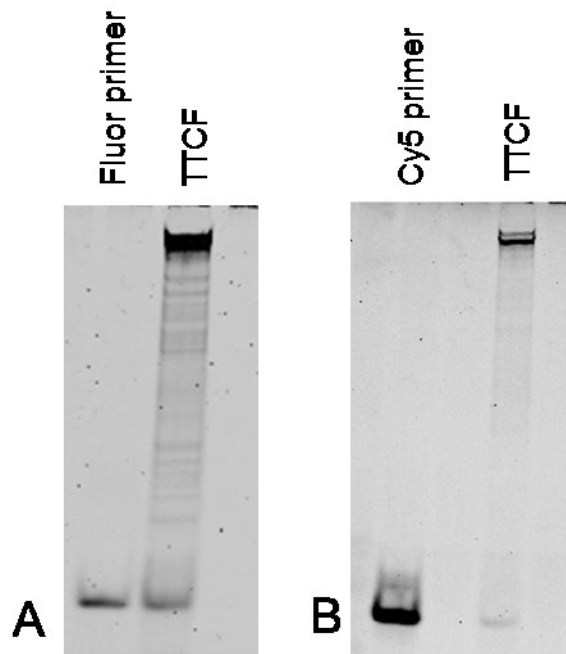


Figure 3. 15 TTCF replication fork mimic annealing assays visualised using a Typhoon scanner. A) Detection of Fluor B) Detection of Cy5.

3.12 Extension of replication fork mimics by *Pfu*-Pol D

After annealing with the desired primers, the replication fork mimics were used for primer extension reactions by *Pfu*-Pol D (80 nM) and performed at 60°C. Reactions were quenched in stop buffer containing 500 nM of competitor sequences that were complementary to the replication fork mimics' leading and lagging strands. Time points were taken at 0,1, 2, 5 and 10 minutes.

Primer extension reactions were initially performed on replication fork mimics that had both the leading (Cy5) and lagging (Fluor) strand primers annealed, TTCF, TUCF, UTCF and UUCF. The gels were scanned twice, using a Typhoon scanner, at wavelengths suitable for visualising Cy5 and Fluor and visualised with ImageQuant software (Figure 3.16-Figure 3. 19).

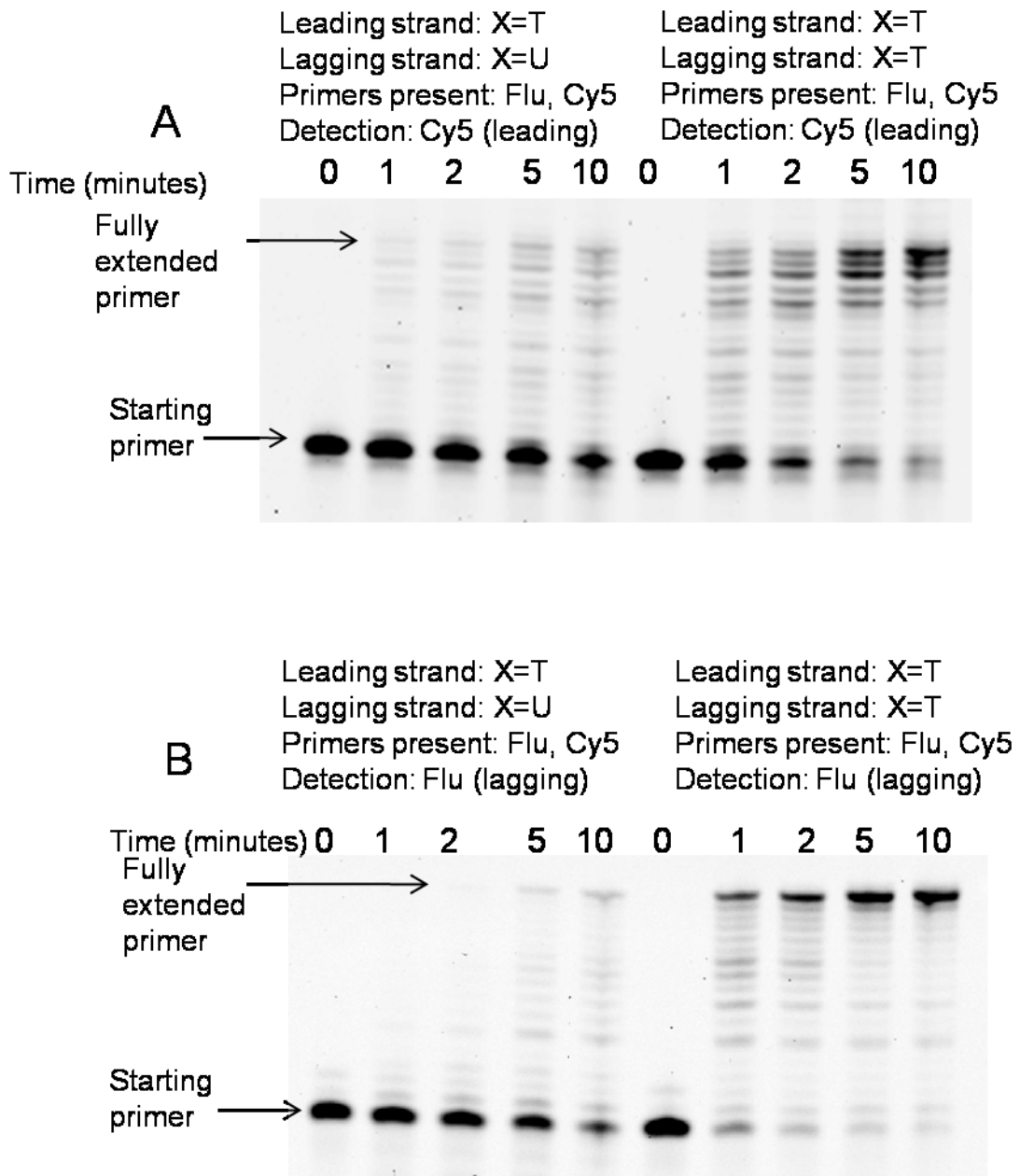


Figure 3.16 TTCF and TUCF replication fork mimic primer extension reactions.
 A) Detection of Cy5 (leading strand) labelled primer B) Detection of Fluor (lagging strand) labelled primer.

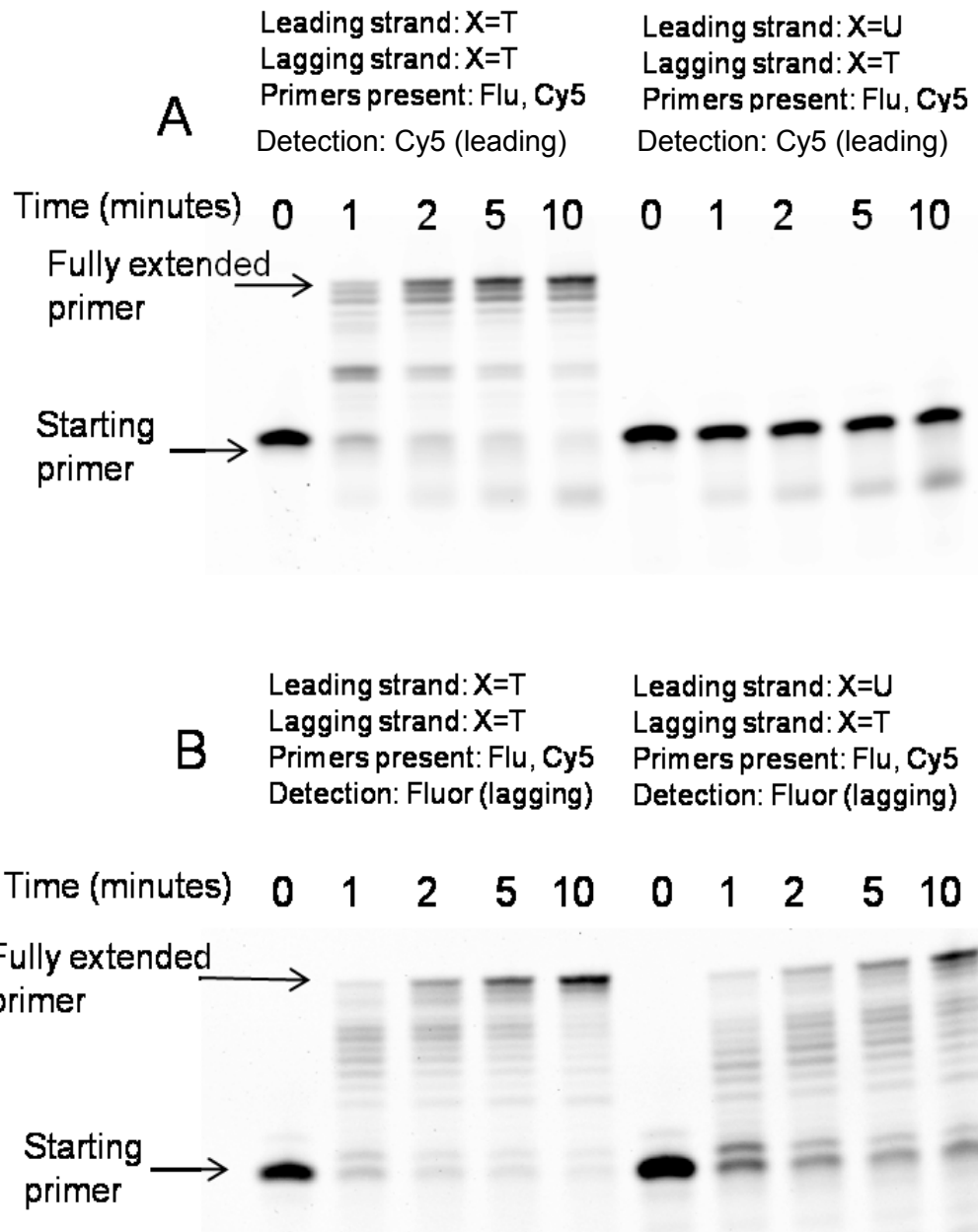


Figure 3.17 TTCF and UTCF replication fork mimic primer extension reactions. A) Detection of Cy5 (leading strand) labelled primer B) Detection of Fluor (lagging strand) labelled primer.

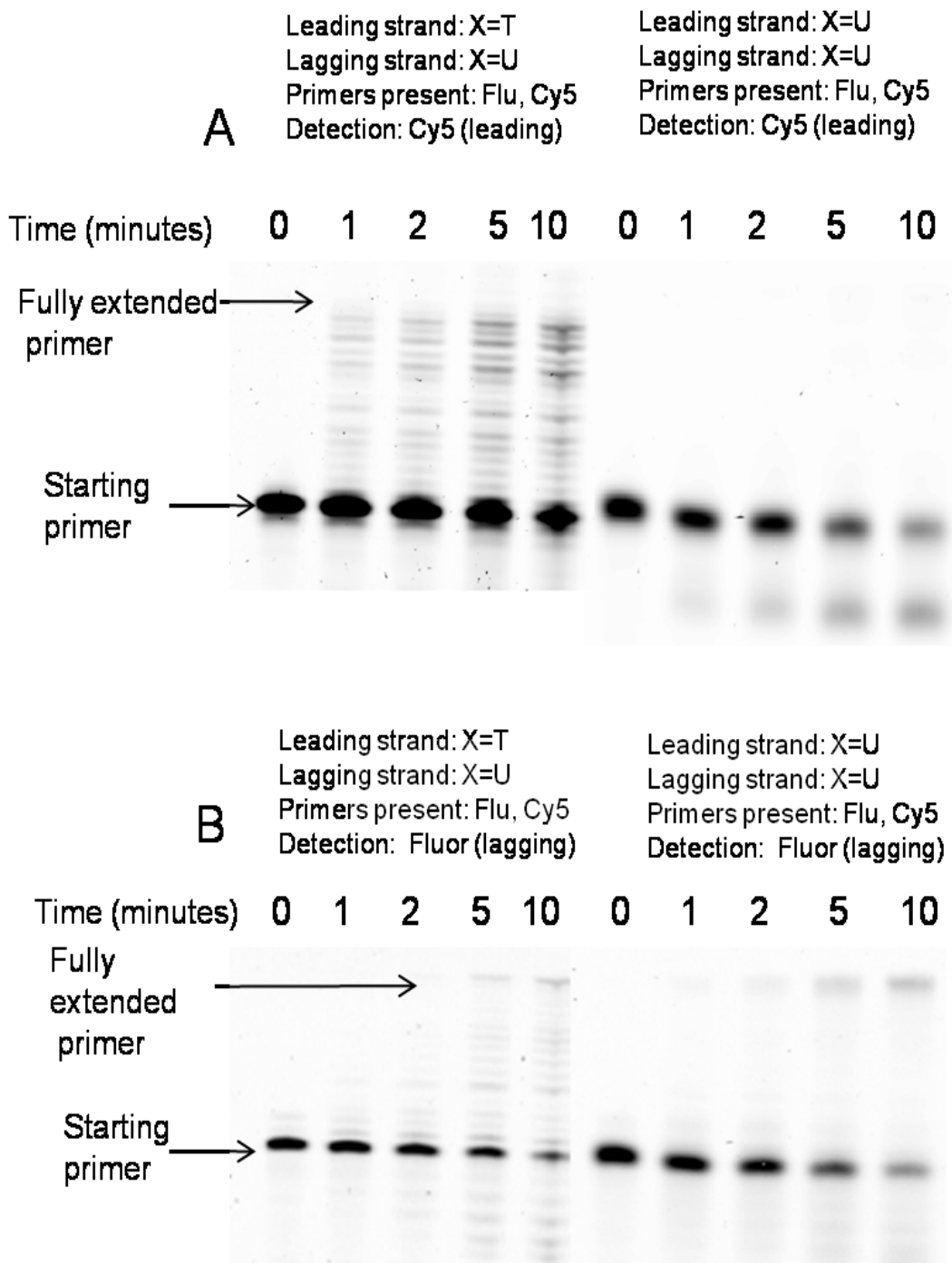


Figure 3.18 TUCF and UUCF replication fork mimic primer extension reactions.
 A) Detection of Cy5 (leading strand) labelled primer B) Detection of Fluor (lagging strand) labelled primer.

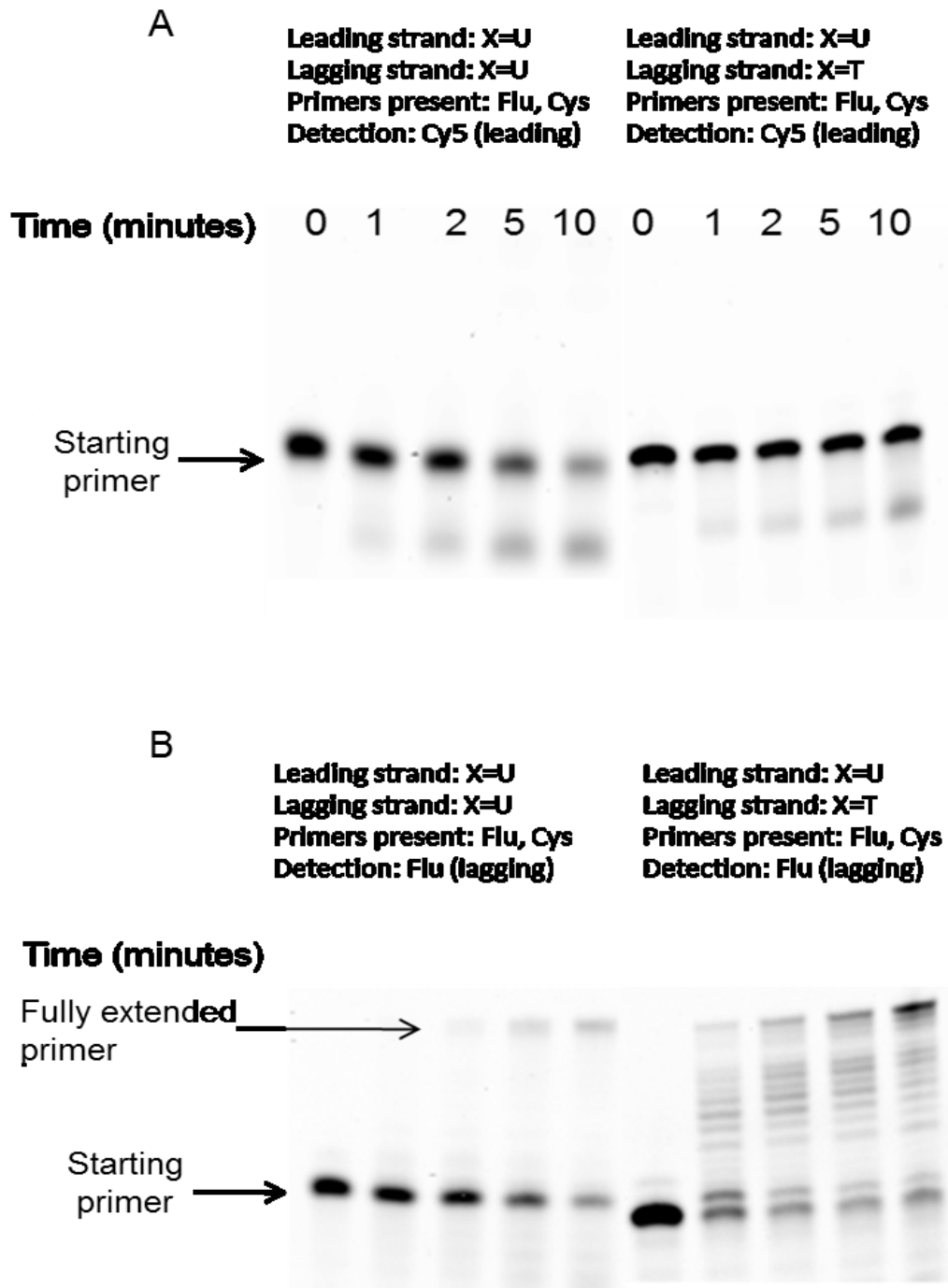


Figure 3. 19 UUCF and UTCF replication fork mimic primer extension reactions.
 A) Detection of Cy5 (leading strand) labelled primer B) Detection of Fluor (lagging strand) labelled primer.

Analysis of the extension assays shows the presence of uracil, located on the leading strand of the replication fork mimic, inhibits the rate of extension of the Cy5 primer (Figure 3.16 A, Figure 3.17 A, and Figure 3.18 A). Similarly, uracil located on the lagging strand, inhibits the rate of extension of the Fluor primer (Figure 3.16 B, B, Figure 3.18 B and B). Thus, the assays appear to show inhibition of polymerization by *Pfu*-Pol D when uracil is located in the leading or lagging strand template being copied. Figures 3.17-3.19 show primer degradation when uracil is located on the template strand being copied. This observation suggests that when uracil is detected, Pol D switches from a polymerisation mode to exonucleolysis. This observation is further supported by exonuclease assays that detect an increase in the rate of exonucleolysis in reactions initiated by Pol D when uracil is located within the template strand DNA (compared to thymine control DNA) (Figure 3.23).

Further extension assays were performed using replication fork mimics with only one primer annealed (Cy5 or Fluor) (Figure 3.20) to determine if the presence of uracil on a non-copied strand of the replication fork mimic affected the rate of extension on the adjacent copied strand. The use of ImageQuant software allowed the rate of primer extension in the TTC and TUC replication fork mimics to be quantified (Figure 3.21).

One minute after the initiation of polymerization, 53 % of starting primer in the TTC replication fork was extended compared to 44 % on the TUC fork (Figure 3.20 and Figure 3.21). This reduced rate of extension observed in the TUC fork is observed at every time point (Figure 3.21). These results suggest that the presence of uracil on the non-copied lagging strand of the replication fork, TUC, is affecting the rate of replication on the adjacent leading strand being copied by the Cy5 labeled primer. A similar result was observed when comparing the rate of primer extension in the TTF and UTF replication fork mimics (Figure 3.20 B & Figure 3.21). This reduced rate of lagging strand primer extension, observed when uracil is located on the opposite strand, UTF (Figure 3.21), suggests that Pol D possesses trans-inhibition properties. However, the rate of inhibition when uracil was located on the leading strand (Figure 3.20 A) appeared more pronounced than when it is located on the lagging strand.

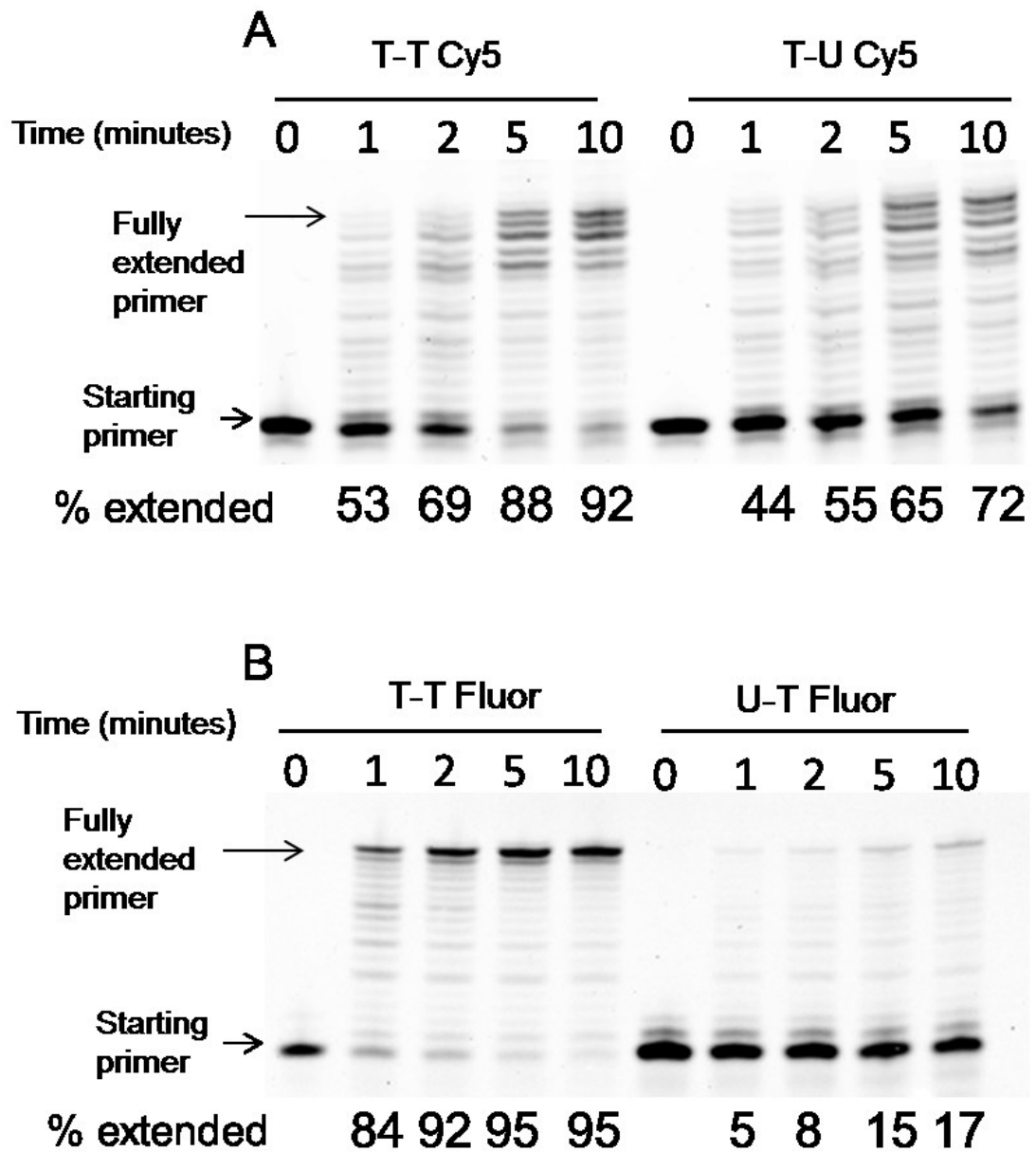


Figure 3.20 Replication fork primer-extension reactions. A) TTC and TUC replication fork mimics detecting Cy5. B) UTF and UUF replication fork mimics detecting fluorescein. “% extended” was calculated as the amount of fully extended product/total product x 100, using ImageQuant software.

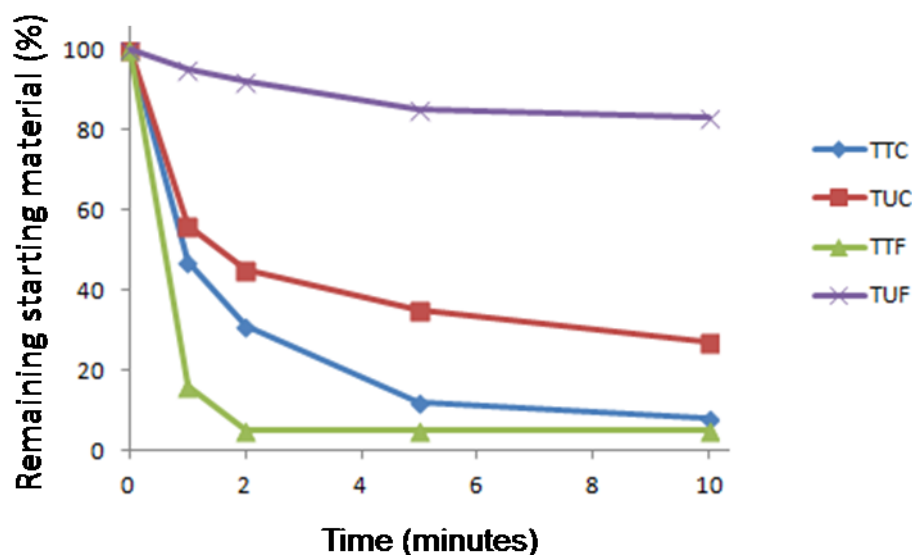


Figure 3. 21 Summary of data from Figure 3.20 showing remaining primer-template against time. Colour coding: blue, TTC; red, TUC; green, TTF and UTF, purple.

3.13 Rate of exonucleolysis in the presence of uracil

To determine if the presence of template strand uracil increased the rate of proofreading exonuclease activity of *Pfu*-Pol D, exonuclease reactions were performed using two annealed primer-templates. The primer-templates contained two A: T bases at the primer-template junction one had uracil at the +4 position, the second (control) had thymidine at +4 (Figure 3. 22). Reactions contained 20 nM primer-templates and were performed at 50°C, initiated with 140 nM *Pfu*-Pol D, an excess of protein over the concentration of primer-template.



Figure 3. 22 Cy5 labelled primer and complementary template used in single turnover assays. “X” represents uracil (thymidine in controls) located 4 bases ahead of the primer-template junction.

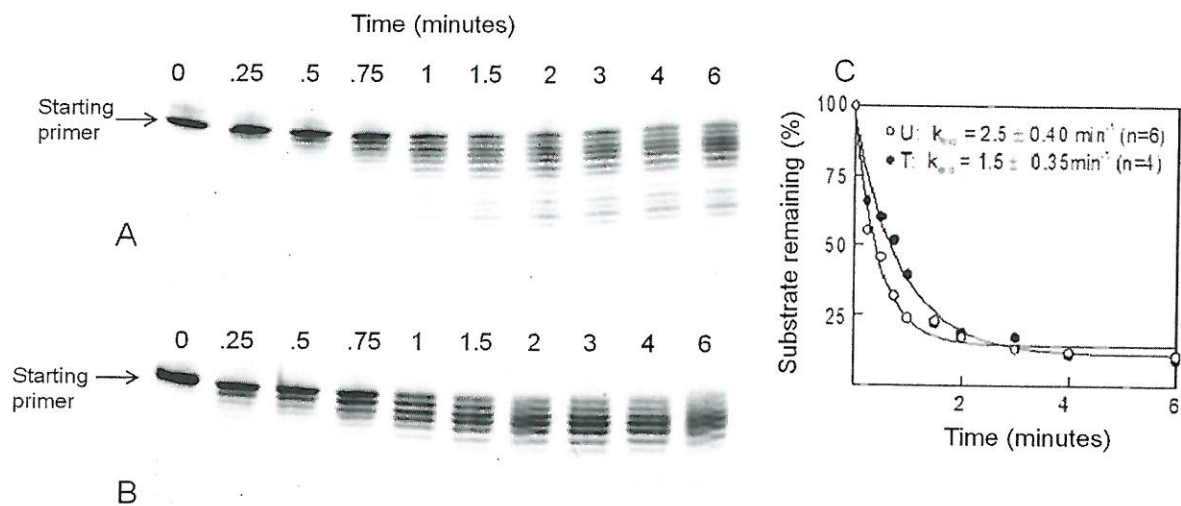


Figure 3. 1 Proof reading exonucleolysis of primer-templates containing uracil (thymine in control) initiated by *Pfu*-Pol D. A) Thymine (control) containing primer-templates B) uracil containing primer-templates C) the calculated k_{exo} values for reactions A) and B) shown on a graph. C) Represents the means (\pm standard deviation) for the number of experiments (n) performed. The full-length starting primers marked with an arrow in A) and B).

The degradation of the Cy5 labelled primer was observed using a Typhoon scanner and quantified using ImageQuant software. The amount of substrate remaining after a given time was fitted in a single exponential, using Grafit, to determine the rate constants of exonucleolysis. The k_{exo} values are shown on the graph and are the averages (\pm standard deviation) from four and six experiment repeats. The results show that the presence of uracil at +4 increased the exonuclease rate, by a factor of ~ 1.7 compared with the thymidine containing control (Figure 3. 1).

These results are consistent with findings by Dr Tomas Richardson, who performed the same reaction using a different primer template. Dr Richardson showed that primer-templates that contained two G: C base pairs at the primer-template junction and uracil 9 bases ahead of the primer-template junction resulted in an approximately three fold stimulation in the rate of 3'-5' exonucleolysis (Richardson *et al.*, 2013).

3.14 Discussion

The data presented in this chapter shows the presence of uracil, located up to 102 bases ahead of the primer-template junction in template strand DNA, reduces the rate of polymerisation of Pol D (Figure 3.7-Figure 3.10). Pol D continues to extend primers when uracil is present however a decrease in DNA synthesis is observed (Figure 3.6). Experimentation has also shown that Pol D is inhibited by uracil located on a non-copied strand of the replication fork (Figure 3.20) and that the presence of uracil in template strand DNA increases the rate of 3'-5' exonucleolysis (Figure 3. 23).

Further research into the uracil recognition properties of Pol D was conducted concurrently by Dr Tomas Richardson. Dr Richardson's research showed that Pol D binds to uracil containing DNA ~2 fold tighter than it binds to control DNA, increases 3'-5' proof reading exonuclease in the presence of uracil by a factor of three and is unable to detect uracil buried within double stranded DNA (Richardson *et al.*, 2013). These observations, combined with research detailed in this chapter have led to the proposal of a tentative model of the molecular mechanism that gives rise to uracil-dependent inhibition in Pol D (Figure 3. 24).

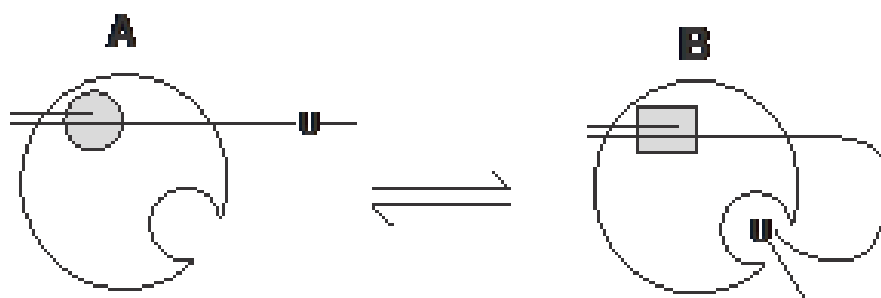


Figure 3. 24 Model for the interaction of *Pfu*-Pol D with uracil. Conformation A is found in the absence of uracil, with the primer-template binding active site (grey circle) having high polymerisation activity. Uracil, in single stranded templates, can bind at a remote site to give conformation B. Here the active site (grey square) shows a lowered extension rate. This form also has increased 3'-5' proof reading exonuclease activity.

Figure 3. 24 shows two possible polymerase-DNA confirmations that exist depending on whether uracil is present (up to 120 bases ahead of primer template junction) in the template strand. Confirmation A is found in the absence of uracil, and shows the Pol D in an active state which possesses strong polymerase activity and binds at the primer-template junction. Confirmation B occurs when uracil is located within the DNA template. Confirmation B shows the Pol D making an additional interaction with uracil thus changing the active site, resulting in a reduced rate of polymerisation and an increase in 3'-5' proofreading exonuclease activity. The recognition of uracil, regardless of its template strand position can be accounted for by looping out the flexible single stranded DNA, enabling the polymerase to interact with the deaminated base at a range of locations ahead of the primer-template junction.

Within the replisome, limited stretches of single stranded DNA are exposed at any one time. Therefore, the ability to recognise uracil up to 134 bases ahead of the primer-template junction would allow Pol D to recognise uracil immediately after unwinding of the double stranded DNA and slow the replication process, thereby, providing the maximum amount of time for repair mechanisms to replace the uracil. Thus, the read-ahead mechanism fits with the current knowledge of DNA replication *in vivo* as Pol D. An alternative hypothesis is that uracil inhibits the initiation of DNA replication by Pol D, rather than slowing replication on encountering uracil. This hypothesis is supported by the observation of fully extended product with no obvious intermediate products (Figures 3.16-3.19). However, this hypothesis is not favoured as *in vivo* replication is initiated by primase, and continued by Pol B and/or Pol D. Thus for this alternative hypothesis to hold true, *in vivo*, there would need to be a lag in the continued extension from the primer by Pol D when Pol D takes over from primase. As mentioned earlier, further research is required to understand the uracil recognition mechanism of archaeal Pol D – the proposed model is tentative (Figure 3.24).

The uracil recognition model is markedly different from the well-characterised uracil recognition observed in Pol B. Pol B stalls the replication fork when uracil

is located 4 bases ahead of the primer template junction (Greagg *et al.*, 1999; Firbank *et al.*, 2008b; Killelea *et al.*, 2010). The ability of both Pol B and Pol D to detect uracil supports the hypothesis that both enzymes play an important role in DNA replication as uracil detection is likely to be the first step of a novel DNA repair pathway that protects the archaea from the mutagenic consequences of uracil.

In DNA, the deamination of cytosine converts G: C base pairs to pro-mutagenic G: U mismatches, replication of which results in 50% of the progeny containing a G: C → A: T transition mutation (Figure 1.14). As archaea live in extreme environments and are therefore likely to be exposed to high rates of deamination, the ability to recognize and repair uracil would reduce the mutation rate and offer a selective advantage (Wardle *et al.*, 2008). Thus both the stalling observed in Pol B and the reduced rate of extension associates with Pol D in the presence of uracil are believed to reduce the error rate during DNA replication.

Initially, the reduced rate of polymerisation by Pol D in the presence of uracil appears to be less effective than the stalling of Pol B. However, it has been calculated that *P.abysssi* and *S.acidocaldarius*, initiate the synthesis of 2.2 and 1.6 Okazaki fragments per second, respectively (Matsunaga *et al.*, 2003). This rate is significantly higher than what has been observed in bacteria (0.4-0.8) or eukarya (~0.2). Pol D's ability to detect uracil distant from the primer-template junction, and slow the replication process, may be essential to give repair mechanisms time to remove uracil and replicate it with cytosine to ensure the integrity of the genome.

Further research is required to investigate the genetic properties of Pol D. At present, no high resolution crystal structure is available for Pol D, although information is available about the N terminal regions for both the small and large subunits (Yamasaki *et al.*, 2010; Matsui *et al.*, 2011). Unfortunately, sequence alignment does not identify any similar sequence to the uracil binding pocket characterised in Pol B (Firbank *et al.*, 2008b; Killelea *et al.*, 2010), although, the

lack of similarity is not surprising as the molecular mechanism of Pol B and Pol D appears to be different.

Recent research has shown that Pol D is essential for genome replication in *T.kodakarensis* and *M.maripaludis* while Pol B is not essential (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013). This discovery has led to the hypothesis that Pol D rather than Pol B is the main replicative polymerase in archaea. If this hypothesis is true, then it may be expected that Pol D would be more processive and possess stronger uracil recognition properties than Pol B. However, *in vitro* experimentation suggests that Pol B is the more processive polymerase that possesses strong uracil inhibition patterns.

At present, there is still much debate over the precise roles of archaeal Pol B and Pol D. However, the hypothesis that suggests that Pol D is the main replicative polymerase in archaea is supported by observation that eukaryotic Pol B enzymes possess Fe-S clusters *in vivo* that affect protein folding and thus function (Netz *et al.*, 2012). Pol D proteins have been found to contain the similar conserved cysteine domains that are responsible for Fe-S center in these Pol B enzymes (Table 5.1). Thus, it is suggested that Pol D is a metallo enzyme that possess an Fe-S cluster at the CTD of the DP2 subunit *in vivo*. It is believed that the Fe-S cluster is not formed when Pol D is overexpressed in *E.coli*, thus, biochemical experiments of Pol D are not giving a true representation of its function *in vivo*.

Chapter 4

Family D polymerases: characterisation of the individual subunits

4.1 Background

Archaeal polymerase D (Pol D), originally identified in 1997 from *Pyrococcus furiosus* (*Pfu*) (Uemori *et al.*, 1997a; Cann *et al.*, 1998), has been discovered in all Euryarchaeota, Thaumarchaeota, Korarchaeota, Aigarchaeota and Nanoarchaeota lineages (Brochier-Armanet *et al.*, 2011). Pol D is a heterodimeric, processive, replicative DNA polymerase that possesses proofreading, 3'-5' exonuclease activity and uracil recognition properties (Cann *et al.*, 1998; Jokela *et al.*, 2004a; Henneke *et al.*, 2005; Ishino and Ishino, 2012; Richardson *et al.*, 2013). Pol D is able to displace template strand DNA and interact with PCNA; facilitating the extension of long stretches of DNA during replication (Rouillon *et al.*, 2007; Tori *et al.*, 2007; Castrec *et al.*, 2009). These properties, combined with evidence that Pol D assembles into complexes with other replisome components (Zhuo Lia, 2012) and that Pol D is essential for genome replication (Cubonova *et al.*, 2013; Sarmiento *et al.*, 2013); support the hypothesis that Pol D is the main replicative polymerase in all archaea except Crenarchaea, from which the enzyme is conspicuously absent.

Archaeal Pol D is a dimeric protein made up of two subunits: a large, DP2 subunit and a small, DP1 subunit (Uemori *et al.*, 1997). Previous attempts have been made to overexpress and purify the individual subunits and characterise their biochemical properties (Uemori *et al.*, 1997; Ishino *et al.*, 1998; Jokela *et al.*, 2004). Characterisation of the large subunit has led to conflicting reports. One publication reports that DP2 expressed alone possesses DNA polymerase activity (~ 100 fold less than the Pol D enzyme) (Uemori *et al.*, 1997a) while another publication reports no polymerase activity when the large, DP2, subunit is expressed alone (Ishino *et al.*, 1998). Similarly, initial characterisation of the DP1 subunit reported that it did not possess any polymerase or exonuclease activity when purified alone (Uemori *et al.*, 1997a; Ishino *et al.*, 1998). However, a more recent, comprehensive investigation has identified the small, DP1 subunit as possessing manganese-dependent exonuclease activity (Jokela *et al.*, 2004). This report does not give any indication why DP1 requires manganese for exonuclease activity; however, it does show that activity can be abrogated via the mutation of a single amino acid (Jokela *et al.*, 2004).

This chapter further investigates the biochemical properties of the individual, DP2 and DP1, subunits of Pol D. Experiments were performed to determine if the individual subunits possessed polymerase or exonuclease activity or if either of the subunits was able to recognise uracil. These experiments were performed to clarify the conflicting reports and to determine where the uracil binding region of Pol D is located. Additionally, fidelity assays were performed to calculate the error rate of WT Pol D, Pol D exo^- and DP2 enzymes.

The family D polymerases used in these experiments were specifically chosen due to their potential application in future work. A hypothesis exists that states that Pol D *in vivo* possesses an Fe-S cluster that is lost when the enzyme is overexpressed and purified aerobically (Netz *et al.*, 2012). It is hypothesised that the Fe-S cluster facilitates the correct folding of the polymerase and interactions with other enzymes required for its activity. Thus, it is hypothesised that when Pol D is overexpressed in *E.coli* and purified aerobically, the protein is less processive and has a lower fidelity than the polymerase expressed *in vivo*. To test this hypothesis, attempts were made to overexpress Pol D within archaeal host species' and purify the protein with an intact Fe-S cluster (Chapter 5). However, prior to performing this work, archaeal Pol D as well as the individual DP1 and DP2 subunits were overexpressed in *E.coli* and the proteins purified aerobically.

To enable the biochemical properties of Pol D purified aerobically to be compared to Pol D purified anaerobically it was essential that the same Pol D enzymes that were overexpressed in *E.coli* were also suitable for overexpression within the natural host archaeal organism. Thus, Pol Ds from two closely related thermophilic archaeal species *Methanocaldococcus jannaschii* (*Mja*) and *Methanobacterium* (*Mth*) were selected for these experiments (Figure 4. 1). *Mja* and *Mth* are closely related members of the Type 2 Pol D family, and thus contain 6 pairs of highly conserved cysteine residues at the C terminus of their DP2 subunit (similar to those identified in eukaryotic Pol Bs) (Figure 5.1). They are also closely related to *Methanococcus maripaludis* (*Mma*), the anaerobic archaea intended to be used for the over expression of Pol D within an anaerobic archaeal host (Figure 4. 1)

(Chapter 5). Thus, it was hoped that the biochemical analysis of Pol D expressed and purified from *E.coli* could be compared to results of Pol D expressed and purified in an anaerobic *E.coli* host organism (Chapter 5).

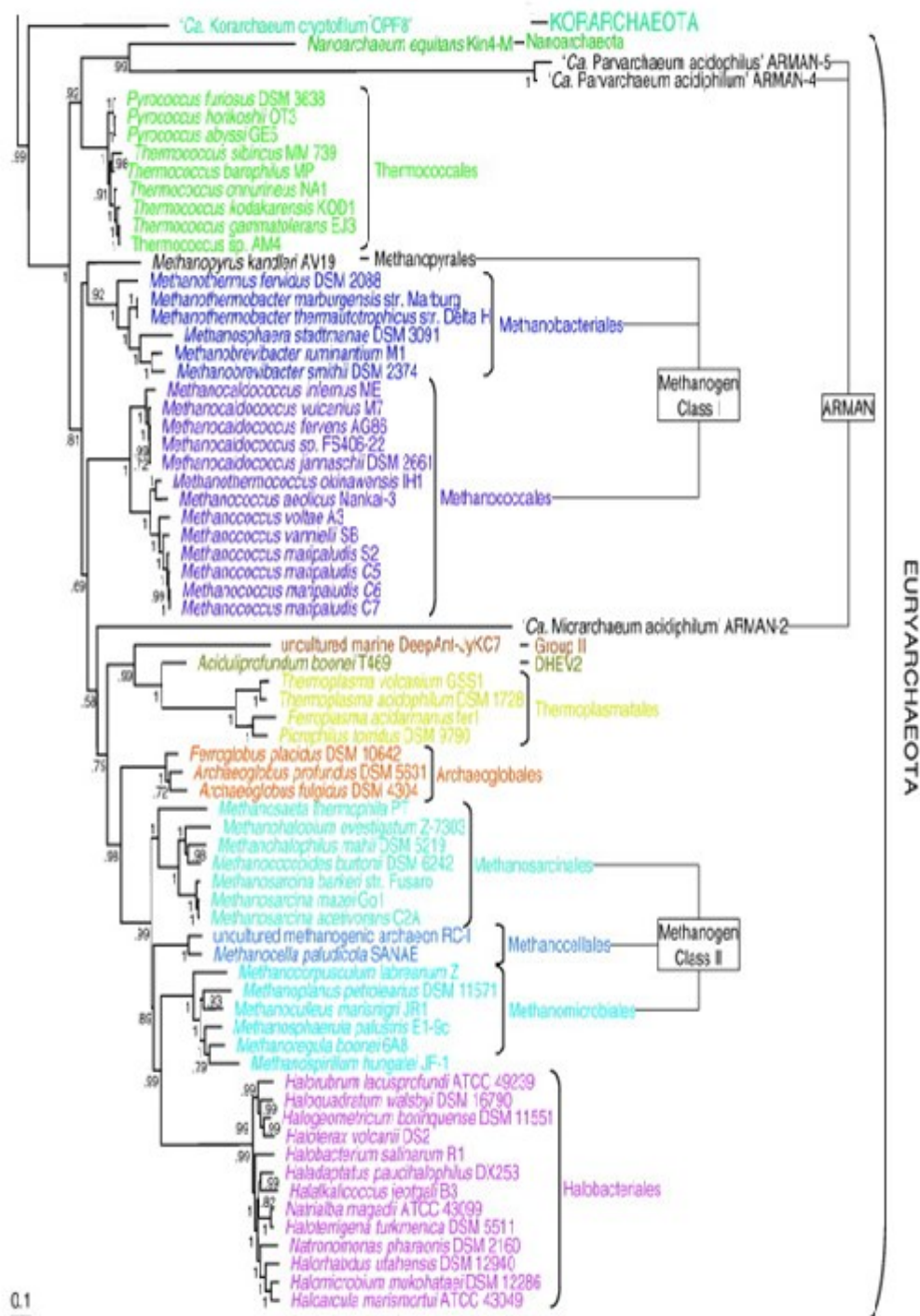


Figure 4. 1 Unrooted Bayesian tree of the archaeal Euryarchaeota based on a concatenation of 57 ribosomal proteins present in at least 89 of 99 genomes. Taken from (Brochier-Armanet *et al.*, 2011).

4.2 Cloning *Mja* and *Mth* DP2 and DP1 subunits into expression vectors

The large, DP2, and small, DP1, subunits of *MjaMja* and *Mth* Pol Ds were PCR amplified and cloned into expression vectors. The two DP2 subunits were cloned into pET28a vectors while the DP1 subunits were cloned into both pET28a and pET22b vectors (Table 4. 1). As pET28a provides resistance to kanamycin and pET22b to ampicillin; the use of the two vectors enabled the DP1 (pET22b) and DP2 (pET28a) subunits to be co-expressed and selected for using both antibiotics (Table 4. 2). The cloning technique also facilitated expression of the individual DP2 and DP1 subunits.

DP1 subunits were cloned via restriction digest cloning, however, the DP2 subunits proved difficult to clone using this technique and so a ligase independent cloning (LIC) method was used (Figure 4. 2) (Aslanidis and Dejong, 1990). LIC was performed using a modified pET28a vector, pET-YSBLIC, that was gifted by Dr Mark Fogg from *Mthe* University of York.

Gene	Vector	N-terminal tag	Selection
<i>MjaMja</i> DP2	pET28a	His	Kan
<i>MjaMja</i> DP1	pET28a	His	Kan
<i>MjaMja</i> DP1	pET22b	-	Amp
<i>Mth</i> DP2	pET28a	His	Kan
<i>Mth</i> DP1	pET28a	His	Kan
<i>Mth</i> DP1	pET22b	-	Amp

Table 4. 1 Summary of vectors used to clone the *MjaMja* and *Mth* DP2 and DP1 subunits. “Kan” represents kanamycin, “Amp” represents ampicillin.

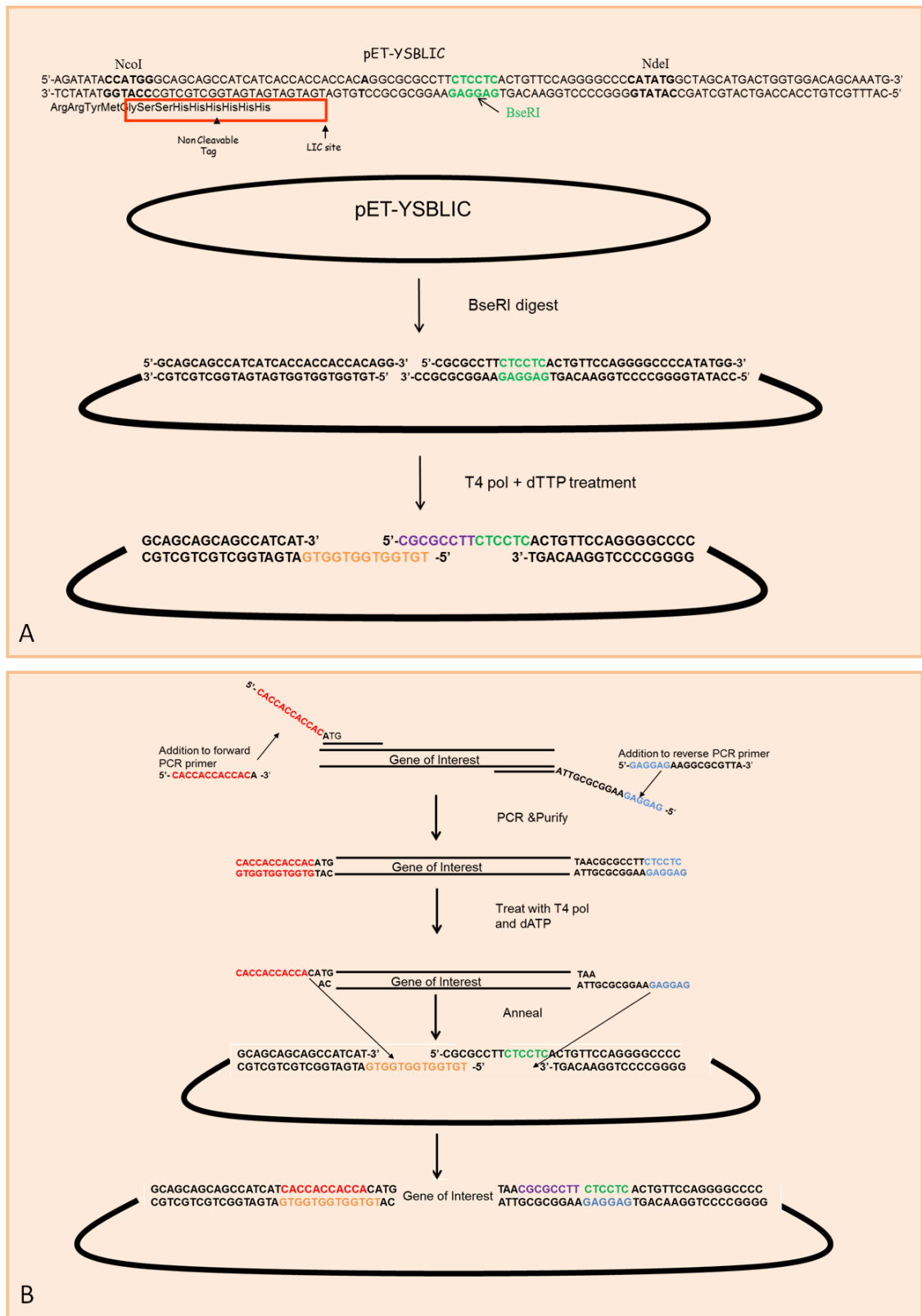


Figure 4. 2 Ligase independent cloning. A) Preparation of the pET-YSB LIC vector. B) Insertion of the target gene into the digested pET-YSB LIC vector.

4.3 Expression and purification of *Mja* and *Mth* Pol Ds as a holoenzyme and individual subunits

Mja-Pol D, *Mja*-DP2, *Mja*-DP1, *Mth*-Pol D, *Mth*-DP2, *Mth*-DP1, *Mja*-DP2/*Mth*-DP1 and *Mth*-DP2/*Mja*-DP1 proteins were overexpressed in *E.coli* BL21 (DE3) pLysS cells and purified (Table 4. 2). Hybrid proteins (combination of *Mja* and *Mth* subunits) were co-expressed and purified to investigate how the two subunits assemble *in vivo*. To facilitate purification, all overexpressed proteins contained a single N-terminal His tag (Table 4. 1). Thus pET28a vectors were used to overexpress DP2 and DP1 proteins while the Pol D proteins were obtained by co-expressing pET28a [DP2] and pET22b [DP1] (Table 4. 2).

Protein component	Vector	Purification
<i>Mja</i> -Pol D	pET22b [<i>Mja</i> -DP1] pET28a [<i>Mja</i> -DP2]	Successful
<i>Mja</i> -DP2	pET28a	Successful
<i>Mja</i> -DP1	pET28a & pET22b	Successful
<i>Mth</i> -Pol D	pET28a [<i>Mth</i> -DP2] pET22b [<i>Mth</i> -DP1]	Unsuccessful
<i>Mth</i> -DP2	pET28a	Unsuccessful
<i>Mth</i> -DP1	pET28a & pET22b	Successful
<i>Mja</i> -DP2/ <i>Mth</i> -DP1	pET28a [<i>Mja</i> -DP2] pET22b [<i>Mth</i> -DP1]	Successful
<i>Mth</i> -DP2/ <i>Mja</i> -DP1	pET28a [<i>Mth</i> DP2] pET22b [<i>Mja</i> -DP1]	Unsuccessful

Table 4. 2 Summary of vectors used to express Pol D and the proteins that were successfully purified.

E.coli BL21 (DE3) pLysS expression cells were transformed with the vectors containing the desired Pol D subunits and grown until an O.D.₆₀₀ of 0.7 was obtained. The strong T7 promoter was then induced with 1mM of IPTG and the cells were grown for a further 6 hours. The cells were harvested and the desired proteins were purified using a heat step and Ni-NTA agarose.

Lysate containing the overexpressed *Mja* and *Mth* proteins were heated to 70°C and 50°C, respectively, for 20 minutes. As thermopiles, the *Mja* and *Mth* proteins were able to withstand the high temperatures while many of the endogenous, mesophilic *E.coli* proteins were denatured. After heating, the expressed Pol D proteins were subjected to purification using Ni-NTA agarose. The protein lysates were passed over Ni-NTA agarose, bound to a gravity flow column, and subjected to several wash steps. The desired N-terminal His-tagged Pol D proteins were eluted from the Ni-NTA agarose using high concentrations of imidazole (500mM). Elutes were collected and visualised via SDS-PAGE (Figure 4. 3).

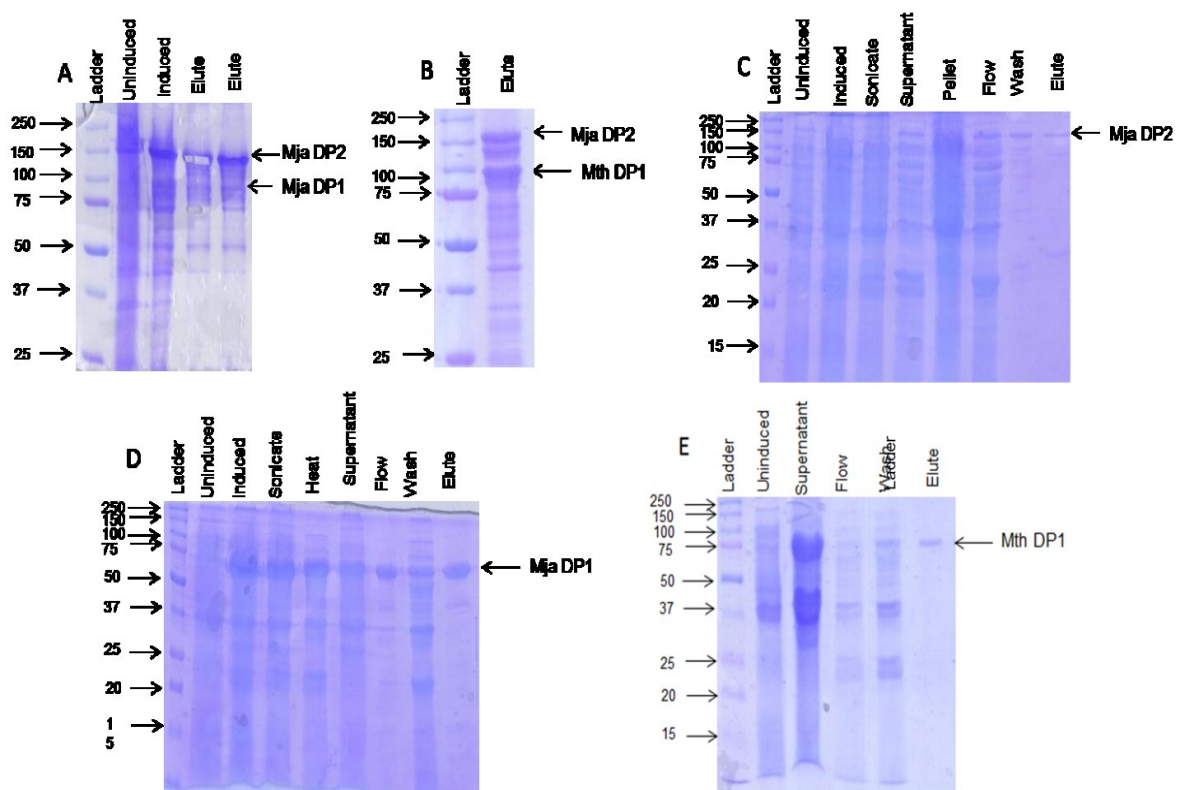


Figure 4. 3 Denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (12 %) showing the purified *Mja* and *Mth* proteins. A) *Mja*-Pol D (co-expressed) B) *Mja*-DP2/*Mth*-DP1 (co-expressed) C) *Mja*-DP2 D) *Mja*-DP1 E) *Mth*-DP1.

Figure 4. 3 shows the co-expressed *Mja*-Pol D and *Mja*-DP2/*Mth*-DP1 proteins were difficult to purify due to proteolytic degradation and contamination from native *E.coli* proteins. Similar difficulties have been reported by other groups attempting to purify Pol D (Gueguen *et al.*, 2001). As high levels of proteolytic degradation are associated with damaged or degraded proteins, it is hypothesised that the high levels of degradation are caused by the lack of Fe-S cluster in the Pol D proteins due to the aerobic purification process.

Unfortunately, *Mth*-DP2 and *Mth*-Pol D were expressed at exceptionally low levels in BL21 (DE3) pLysS and the proteins appeared to be insoluble. Thus, attempts to purify these proteins were unsuccessful (results not shown). Purification of the individual *Mja*-DP2, *Mja*-DP1 and *Mth*-DP1 subunits was less problematic, however, low yields of purified protein were obtained (~0.25 mg of protein was purified from 8 litres of cell culture). Purification of the individual subunits produced enzymes that were more pure than the Pol D enzyme (Figure 4. 3).

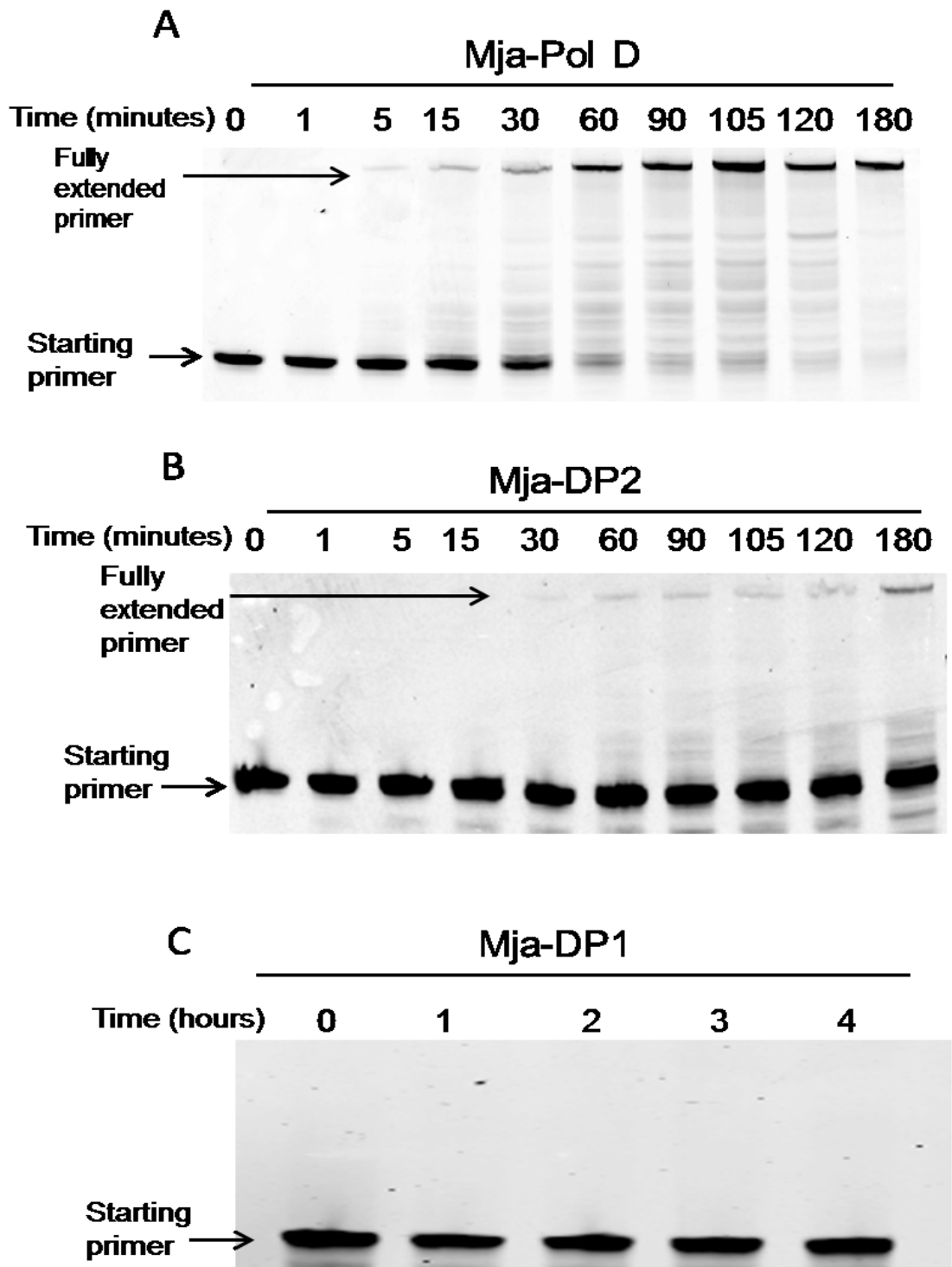


Figure 4.5 Primer extension reactions initiated by A) 150 nM *Mja*-Pol D B) 150 nM *Mja*-DP2 C) 150 nM *Mja*-DP1.

4.5 Extension of uracil containing DNA primer-templates by *Mja*-Pol D and *Mja*-DP2

The ability to recognise template strand uracil was originally believed to be unique to archaeal Pol B, which has a well characterised uracil binding pocket (Firbank et al., 2008; Killelea et al., 2010). However, recent research has shown that *Pfu*-Pol D is also able to recognise uracil (Sawai et al., 2007; Richardson et al., 2013). Thus, primer-template extension reactions were performed to determine if *Mja*-Pol D or *Mja*-DP2 recognise uracil in a similar manner as has been observed with *Pfu*-Pol D.

Primer extension reactions contained, 20U-Cy5 (20T-Cy5 in control reactions), primer-templates (Figure 4. 6) and were performed and analysed exactly as stated in section 4.4.

Figure 4. 6 shows the rate of extension in reactions initiated with *Mja*-Pol D is greatly reduced when uracil (rather than thymine) is located in the template strand being copied. Fully extended primer is detected after 15 minutes in the control (thymine) reaction (Figure 4. 6), whereas only a small amount of fully extended primer is detected in the uracil containing template after 90 minutes (Figure 4. 6). A similar pattern of inhibition was observed in the reactions initiated with *Mja*-DP2 (Figure 4. 6). Thus uracil located in template strand DNA is inhibiting the rate of primer extension by *Mja*-Pol D and *Mja*-DP2 in a similar manner as described for *Pfu*-Pol D (Richardson et al., 2013). The inhibition of *Mja*-DP2 by uracil located in the template strand indicates that the uracil recognition site of *Mja*-Pol D is located within the large, DP2, subunit.

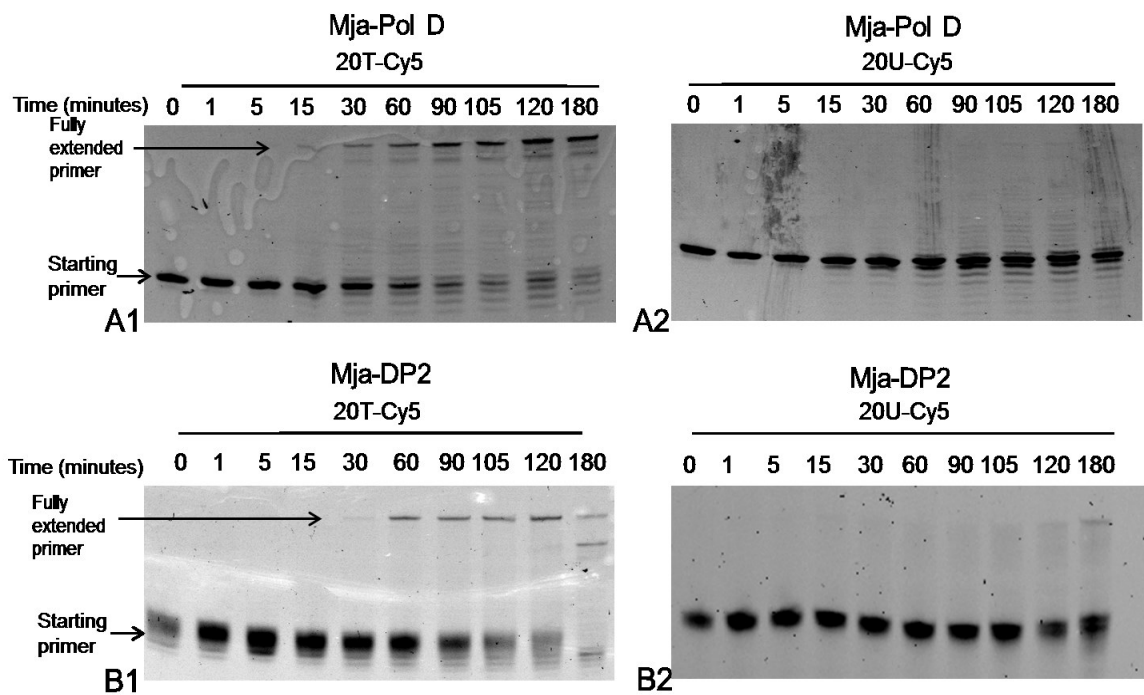


Figure 4. 6 Primer extension reactions initiated by A) 150 nM *Mja*-Pol D B) 150 nM *Mja*-DP2.

4.6 Extension of DNA primer-templates by *Mja*-DP2/*Mth*-DP1

Primer-extension reactions were performed to determine if *Mja*-DP2/*Mth*-DP1 possessed polymerase activity similar to that observed in *Mja*-Pol D (Figure 4.7). Reaction were performed and analysed as described in section 4.4. Figure 4.7 shows that *Mja*-DP2/*Mth*-DP1 possesses polymerase activity and is inhibited by the presence of template strand uracil. *Mja*-DP2/*Mth*-DP1 fully extended primer in the control (thymine) reaction within 30 minutes, whereas it required 60 minutes to fully extend the primer in the uracil containing template (Figure 4. 7). Comparisons of Figure 4. 7 with Figure 4. 6, shows that *Mja*-DP2/*Mth*-DP1 has a similar rate of polymerisation as *Mja*-Pol D and that *Mja*-DP2/*Mth*-DP1 extends primer-template DNA quicker than *Mja*-DP2. This suggests that *Mja*-DP2 and *Mth*-DP1 assemble *in vivo* to produce an active Pol D hybrid. As *Mja* and *Mth* are closely related species, with highly conserved Pol D genes (Figure 4. 1), it is likely that *Mja*-DP2 and *Mth*-DP1 form interactions similar to those formed in the wild type *Mja*-Pol D protein.

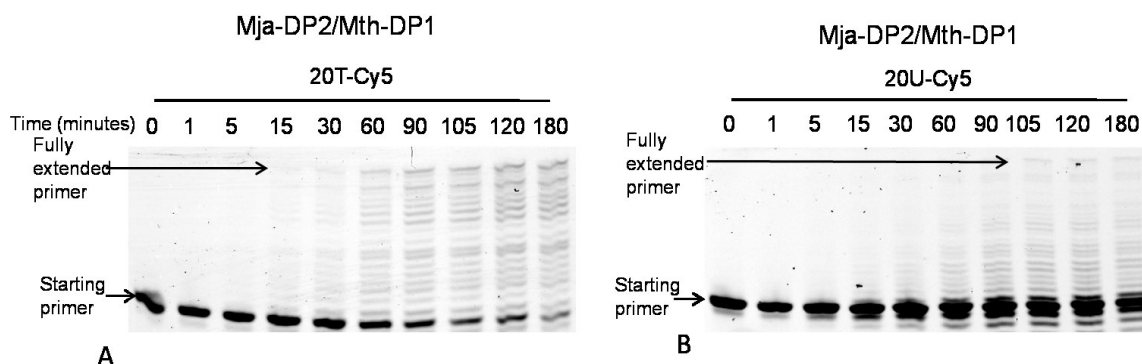


Figure 4.7 Primer extension reactions initiated with 150 nM *Mja-DP2/Mth-DP1*.

4.7 3'-5' exonucleolysis of template DNA by *Mja*-Pol D, *Mja-DP2/Mth-DP1*, *Mja-DP1* and *Mth-DP1*

Family D polymerases possess 3'-5' proofreading exonuclease activity (Cann *et al.*, 1998). It was originally believed that both DP1 and DP2 subunits were required for exonucleolysis (Uemori *et al.*, 1997a; Castrec *et al.*, 2010), however, *Mja-DP1* has recently been described as possessing Mn²⁺ dependent exonuclease activity (Jokela *et al.*, 2004a; Shen *et al.*, 2004). To confirm this report, exonuclease reactions were performed using *Mja*-Pol D, *Mja-DP2/Mth-DP1*, *Mja-DP1* and *Mth-DP1* (Figure 4.9).

Exonuclease reactions contained 20 nM Hex ssT (Figure 4.8), reaction buffer (Table 4.3) and 80 nM polymerase. Reactions were performed at 65°C or 50°C (reactions containing *Mth* DP1) with time-points taken after 1, 2, 5, 10, 15, 20, 25 and 30 minutes. Reactions were quenched in stop buffer, containing 500 nM of the appropriate competitor sequence, and visualised and analysed as described earlier (Section 4.4).

Hex ssT 5' -Hex-TTTCTGGTTCCAGCTGGACCATTTCGCCTATAGGACCTATT-3'

Figure 4.8 Hex labelled oligodeoxynucleotide used in exonuclease reactions.

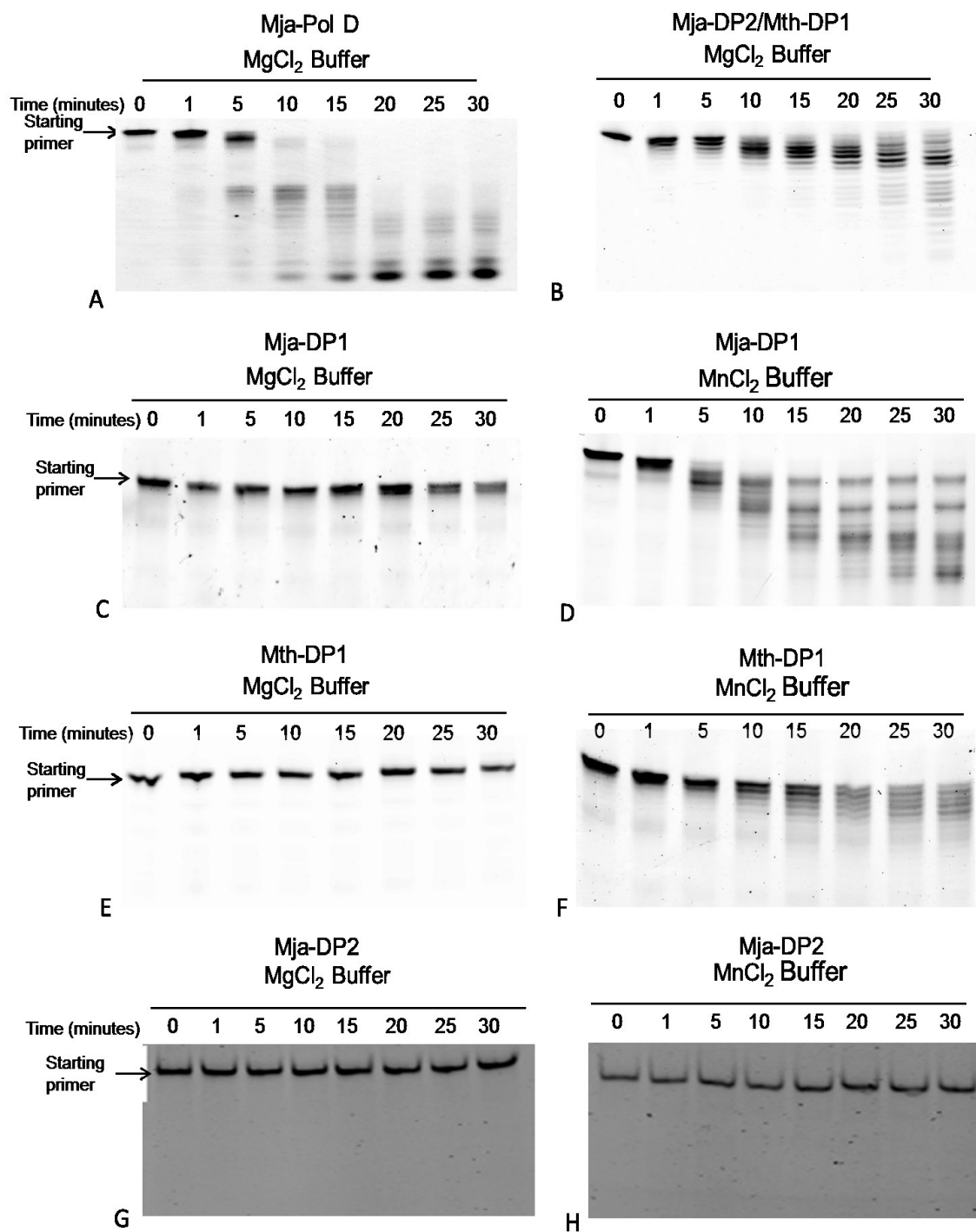


Figure 4. 9 Exonuclease reactions A) Reaction containing $MgCl_2$ and *Mja*-Pol D B) Reaction containing $MgCl_2$ and *Mja*-DP2/*Mth*-DP1 C) Reaction containing $MgCl_2$ and *Mja*-DP1 D) Reaction containing $MnCl_2$ and *Mja*-DP1 E) Reaction containing $MgCl_2$ and *Mth*-DP1 F) Reaction containing $MnCl_2$ and *Mth*-DP1. G) Reaction containing $MgCl_2$ and *Mja*-DP2. H) Reaction containing $MnCl_2$ and *Mja*-DP2. Buffer components are detailed in Table 4.3.

Initial reactions were performed in a reaction buffer, Pol D Mg^{2+} , which contained $MgCl_2$ (Table 4.3). Figure 4.9 A & B shows that both *Mja*-Pol D and *Mja*-DP2/*Mth*-DP1 possess exonuclease activity in these reactions. However, *Mja*-DP1, *Mth*-DP1 and *Mja*-DP2 did not possess exonuclease activity in these reactions (Figure 4.9 C, E and G). Thus, the experiments were repeated using an alternative reaction buffer, Pol D Mn^{2+} , which contained $MnCl_2$ (Table 4.3).

Buffer Name	Buffer components
Pol D Mg^{2+}	10 mM Tris-HCl [pH 8.0], 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT
Pol D Mn^{2+}	10 mM Tris-HCl [pH 8.0], 50 mM KCl, 5 mM $MnCl_2$, 10 mM DTT

Table 4.3 Reaction buffers used in exonuclease reactions.

Figure 4.9 D & F show the *Mja*-DP1 and *Mth*-DP1 proteins possess exonuclease activity in reactions containing $MnCl_2$. *Mja*-DP1 degraded all starting DNA within 10 minutes whereas *Mth*-DP1 was less active and required 10 minutes for a small amount of exonucleolysis to occur. These reactions confirm that the DP1 subunits from both *Mja* and *Mth* possess Mn^{2+} dependent exonuclease activity (Jokela *et al.*, 2004a). Experiments containing *Mja*-DP2 show that the enzyme did not possess any exonuclease activity in either the Mg^{2+} or the Mn^{2+} reaction buffer (Figure 4.9). This finding suggests the *Mja*-DP1 subunit is responsible for exonuclease activity within Pol D and supports a previous hypothesis based on gene deletion experiments, yeast two-hybrid assays and surface plasmon assays (Shen *et al.*, 2001; Tang *et al.*, 2004)

4.8 Assembling Pol D by mixing *Mja*-DP2 and *Mja*-DP1 subunits *in vitro*

Purified *Mja*-DP2 and *Mja*-DP1 were mixed *in vitro* and used to initiate primer-extension reactions. These reactions were performed to determine if the two subunits were able to interact and acquire the processivity observed in reactions initiated by *Mja*-Pol D produced by co-expression *in vivo* (Figure 4.5 & Figure 4.6). *Mja*-DP2 and *Mja*-DP1 were mixed at various ratios and a range of incubation times and temperatures (as indicated on Figure 4.). After mixing primer-extension reactions were performed as described in section 4.4.

Figure 4.10 shows comparisons of the rate of primer extension in reactions initiated by *Mja*-Pol D (co-expressed) and *Mja*-DP2/*Mja*-DP1 (mixed on bench) shows an accelerated rate of extension when the proteins are co-expressed *in vivo* (Figure 4.10). No noticeable difference in primer-extension rate was observed when the DP1 and DP2 subunits were mixed on the bench compared to the DP2 subunit expressed alone (Figure 4.5 & Figure 4.6). The inability to obtain the processivity of Pol D (co-expressed) by mixing the DP2 and DP1 subunits *in vitro* is further evidence that the DP2 and DP1 subunits form interactions *in vivo*.

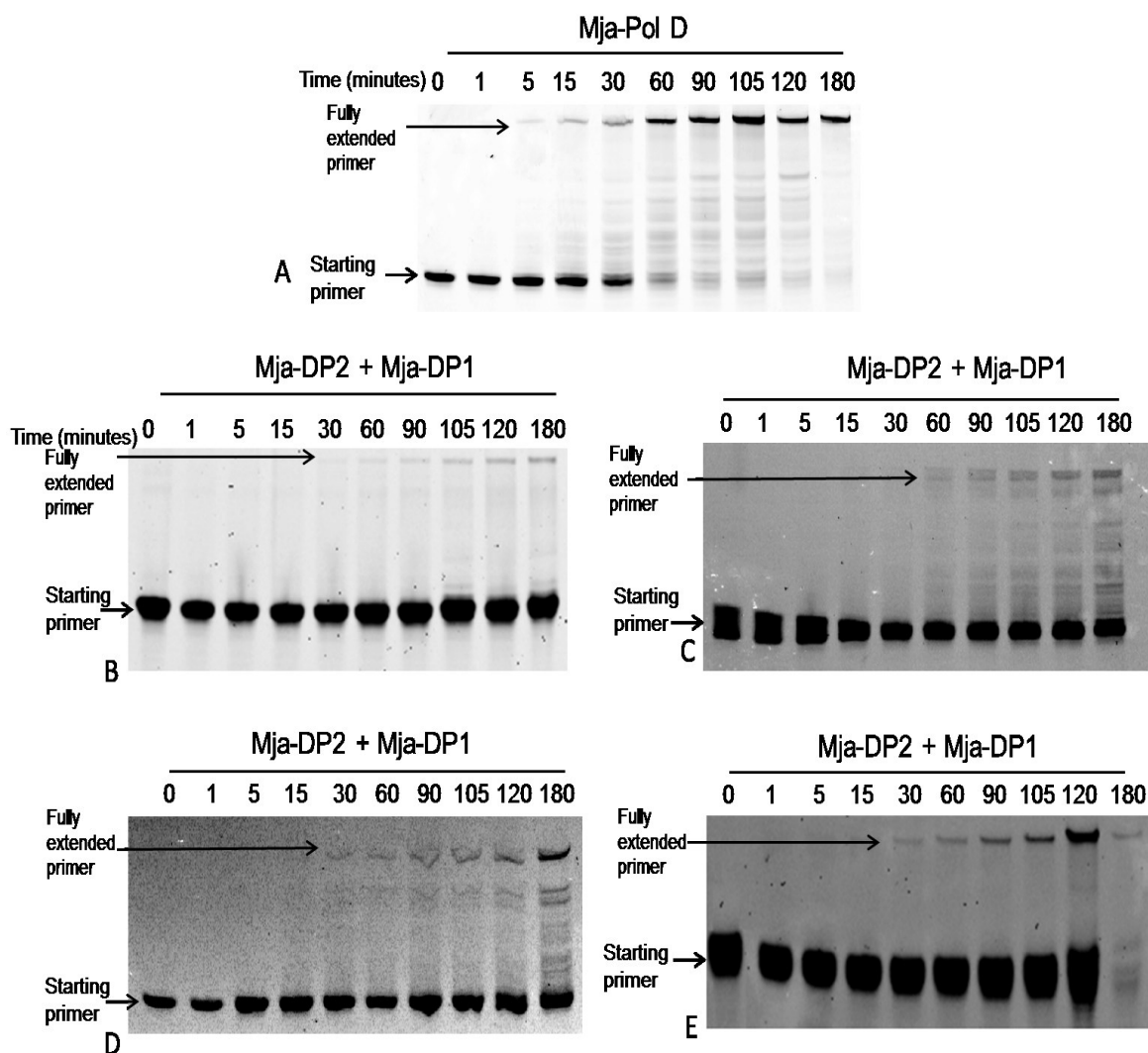


Figure 4.10 Primer extension reactions. Reactions were initiated by: A) 150 nM *Mja*-Pol D produced by co-expression in vivo B) 150 nM *Mja*-DP2 and 150 nM *Mja*-DP1 mixed at room temperature and incubated for 20 minutes C) 150 nM *Mja*-DP2 and 300 nM *Mja*-DP1 mixed at 4°C and incubated for 1 hour D) 150 nM *Mja*-DP2 and 300 nM *Mja*-DP1 mixed at 60°C for 1 hour E) 150 nM *Mja*-DP2 and 150 nM *Mja*-DP1 mixed at 60°C for 16 hours.

4.9 Binding of Pol D and its individual subunits to DNA

The interaction of Pol D with DNA has been monitored using fluorescence anisotropy (Figure 4.11). This method has been widely applied by other groups to study protein-DNA interactions (LiCata and Wowor, 2008).

Fluorescence anisotropy uses the excitation of polarised light to measure the rotational diffusion of fluorescently labelled DNA via polarised detection (Heyduk and Lee, 1990; LiCata and Wowor, 2008). Fluorescently labelled DNA molecules are illuminated with polarised light. If the fluorophore has absorption transition dipoles in the same plane as the polarised light then the fluorophore is excited. Polarised light is then emitted, in a plane defined by the fluorophore's emission transition dipoles.

If the fluorescently labelled DNA is bound to a protein, the rate at which the DNA tumbles is significantly decreased. This reduced rate of tumbling affects the polarisation of the emitted light (LiCata and Wowor, 2008). The larger the protein that is bound, the slower the fluorescence DNA tumbles. Therefore by calculating the anisotropy of the free DNA molecule, partially bound molecule and fully bound molecule, measured by titrating the DNA molecule with protein the dissociation constant (K_D) for the protein-DNA interaction can be calculated.

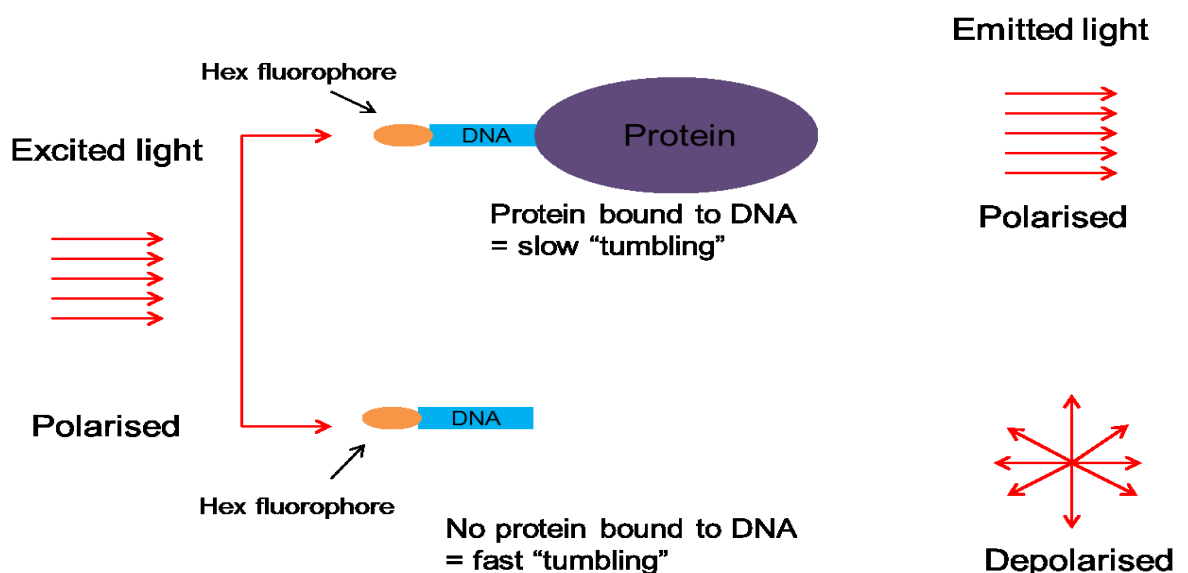


Figure 4. 11 Illustration of fluorescence anisotropy.

4.10 Interaction of Pol D with DNA

The binding affinity of the *Mja* and *Mth* proteins for uracil (thymine in control) containing single-stranded DNA (Table 4.4) was measured using fluorescence anisotropy. Increasing amounts of the polymerase variants were added to a reaction mixture containing the hex-labelled oligodeoxynucleotides and the anisotropy value was recorded (Figure 4.12 and 4.13). Polymerase was added until the anisotropy value reached a plateau. Each set of data points was repeated 3 times and the dissociation constant (K_D) value of the proteins calculated using the following equation (Reid *et al.*, 2001):

$$A = A_{min} + \frac{(A_{max} - A_{min}) \left(\frac{D}{D + E + K_D} \right)}{1 + \frac{D}{D + E + K_D}}$$

A = Anisotropy (min: anisotropy with no protein bound) (max; anisotropy with protein bound)

D = Total primer template concentration

E = Total protein concentration

KD = Dissociation constant

Oligodeoxy nucleotide	Sequence (5'-3')
Hex ss T/U	Hex-TTTCTGGT T /UCCAGCTGGACCATTTCGCCTATAGGACCTATT

Table4. 4 Oligodeoxynucleotide used in fluorescence anisotropy. Oligodeoxynucleotides were named according to the base highlighted in red.

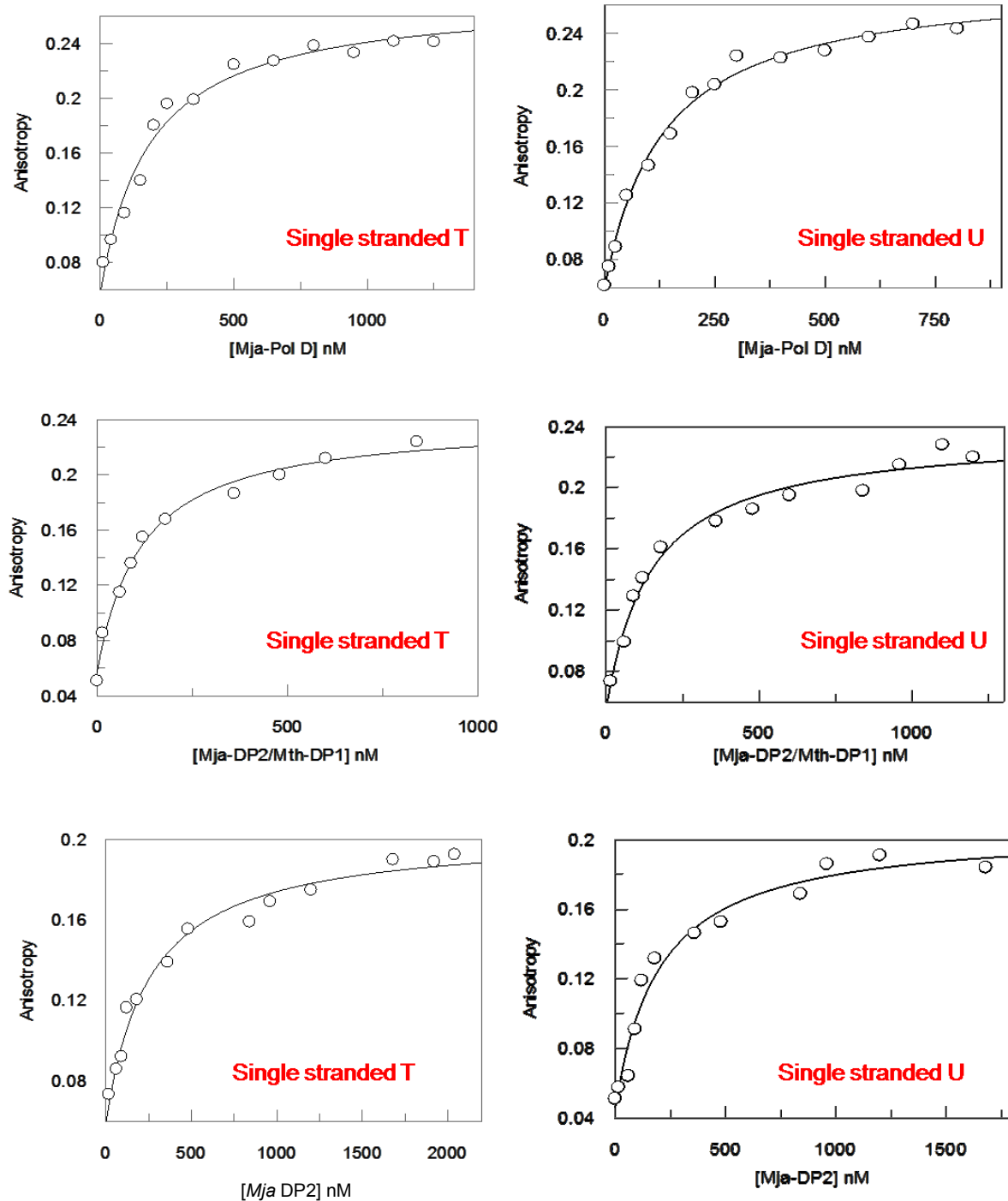


Figure 4. 12 Binding curves of Pol D proteins with the DNA substrates listed in Table 4.4 as determined via fluorescence anisotropy. Graphs were compiled using GraFit software.

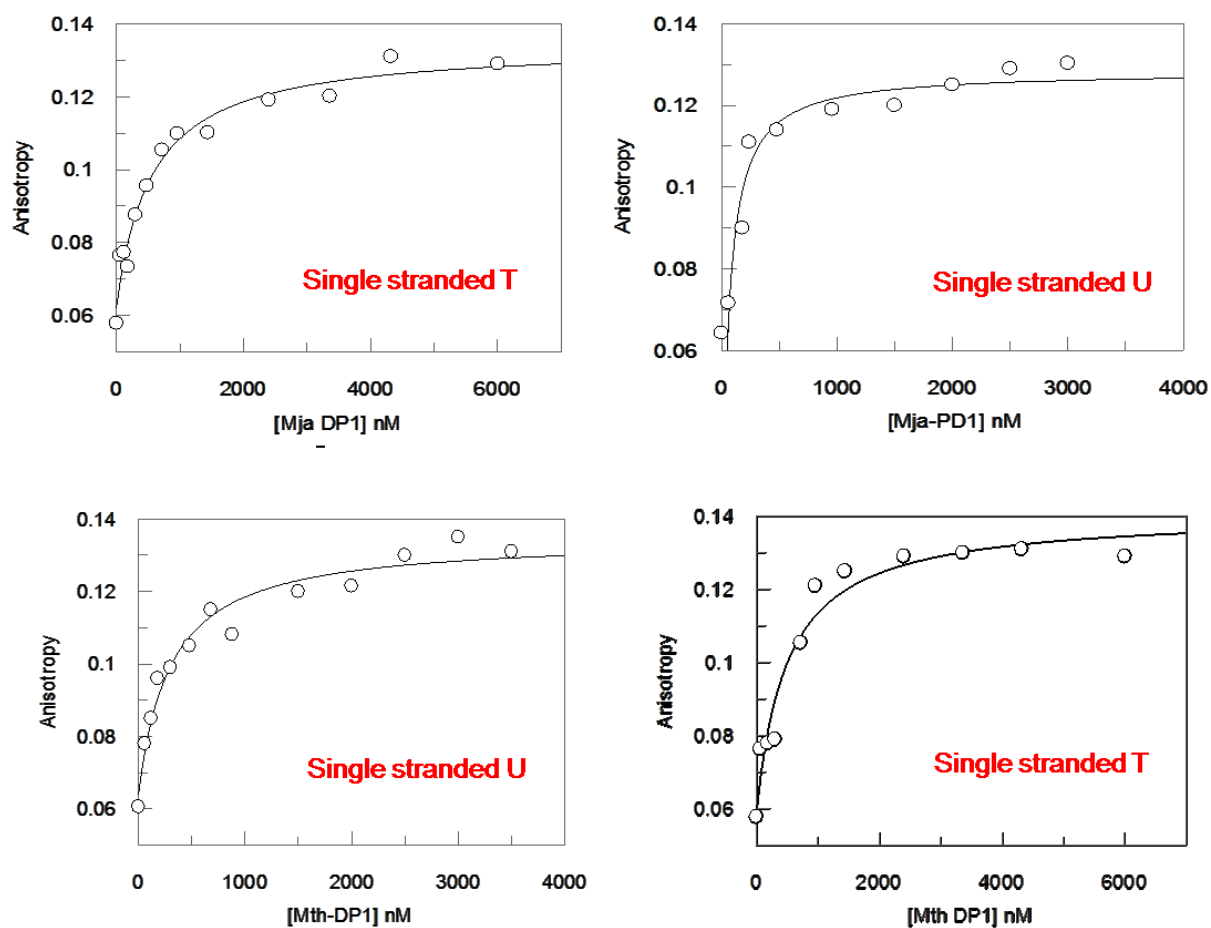


Figure 4.13 Binding curves of Pol D proteins with the DNA substrates listed in Table 4.4 as determined via fluorescence anisotropy. Graphs were compiled using GraFit software.

Oligodeoxy-nucleotide	Protein				
	<i>Mja</i> -Pol D	<i>Mja</i> -DP2/ <i>Mth</i> -DP1	<i>Mja</i> -DP2	<i>Mja</i> -DP1	<i>Mth</i> -DP1
Hex ssT	224±14	219±17	303±45	557±68	414±46
Hex ssU	39±6	42±7	246±24	127±13	156±12

Table 4.5 The K_D values for the binding of protein (mean \pm standard deviation from at least three determinations) to oligodeoxynucleotides containing uracil (thymine in controls). The fluorophore, hexachlorofluorescein, was used to determine the K_D value by fluorescence anisotropy titration.

Fluorescence anisotropy titrations revealed that *Mja*-Pol D and *Mja*-DP2/*Mth*-DP1 binds ~ 5 x more tightly to uracil containing single stranded DNA than thymine containing DNA (Table 4.5). These results follow similar patterns that have previously been described for *Pfu*-Pol D, however, the binding constants of *Mja*-Pol D are significantly higher than those observed in *Pfu*-Pol D (Richardson *et al.*, 2013)

Table 4.5 shows that *Mja*-DP2 did not show strong preference for uracil containing DNA compared to the thymine control, only a very slight increase in binding was observed in the uracil containing DNA, ~1.2 x. This result is much lower than expected as primer-template extension assays showed that *Mja*-DP2 was able to recognise uracil (Figure 4.6). Thus it was expected that the protein would show significant preference for uracil containing DNA.

Surprisingly, both *Mja*-DP1 and *Mth*-DP1 proteins showed a preference and bound more tightly to the uracil containing DNA (~3-5 x) than the control DNA. This finding suggests that the uracil binding pocket is in the small subunit. However as primer extension assays show that DP2 was inhibited by uracil it is likely that the uracil binding pocket is located in the large, DP2, subunit. Thus, these findings are unexpected and may suggest that both the large and small subunits of archaeal Pol D play a role in uracil recognition.

4.11 Preparation of *Mja*-Pol D lacking 3'-5' proofreading exonuclease activity

Site-directed mutagenesis (Weiner *et al.*, 1994) was used to eliminate the 3'-5' exonuclease activity of *Mja*-Pol D. An H421N mutation was introduced into the small, DP1, subunit (Jokela *et al.*, 2004a; Shen *et al.*, 2004) and exonuclease assays were performed to confirm the absence of hydrolytic degradation of DNA substrates (Figure 4.14). Reactions were performed as described in section 4.8. Figure 4.14 reveals that the mutated *Mja*-Pol D protein does not possess exonuclease activity.

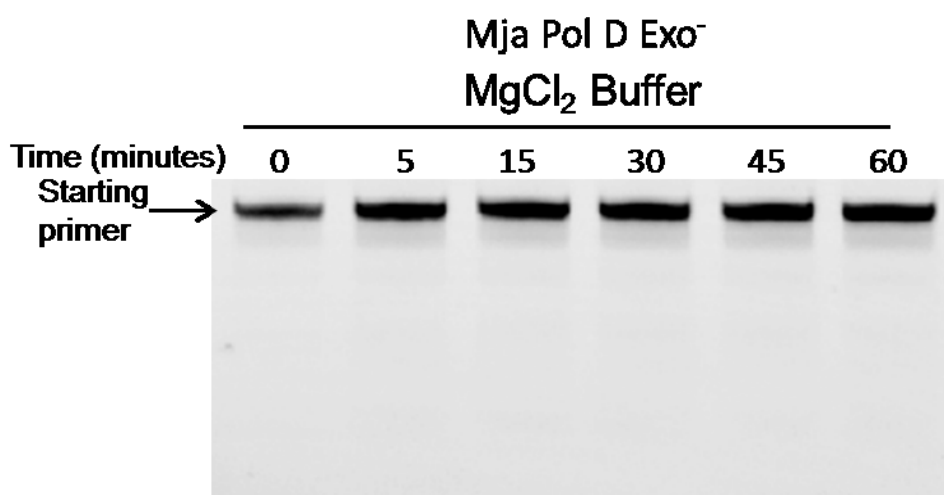


Figure 4. 14 Exonuclease reaction initiated by 80 nM *Mja*-Pol D Exo⁻.

4.12 Fidelity of Pol D

The fidelity of *Mja*-Pol D, *Mja*-Pol D Exo⁻, *Mja*-DP2, *Pfu*-Pol D and *Pfu*-Pol D Exo⁻ was calculated using a plasmid based assay (Table 4.6) (Keith *et al.*, 2013). The assay utilised pSJ3, a modified pUC18 vector that contains a segment encoding lacZ α flanked by two single-stranded nicking endonuclease sites (Figure 4.15) (Keith *et al.*, 2013). The gapped plasmid was then used to calculate the fidelity of the polymerases.

The gapped, pSJ3 plasmid used in this assay was provided by Brian Keith, who had previously calculated the background mutation rates (transformed gapped vector) (Keith *et al.*, 2013). The fidelity assay performed by incubating 1 nM of polymerase with 77 ng of pSJ3 and 10 mM of dNTPs at 70°C for 90 minutes. This allowed the polymerase to copy the lacZ α gene and fill the gapped plasmid *in vitro*. As successful filling of the pSJ3 vector creates an EcoRI site within the plasmid, following incubation with the polymerase a sample (Figure 4.15) of the filled plasmid was subjected to incubation with EcoRI (according to manufacturer protocol) and analysed via agarose gel electrophoresis (Figure 4.16).

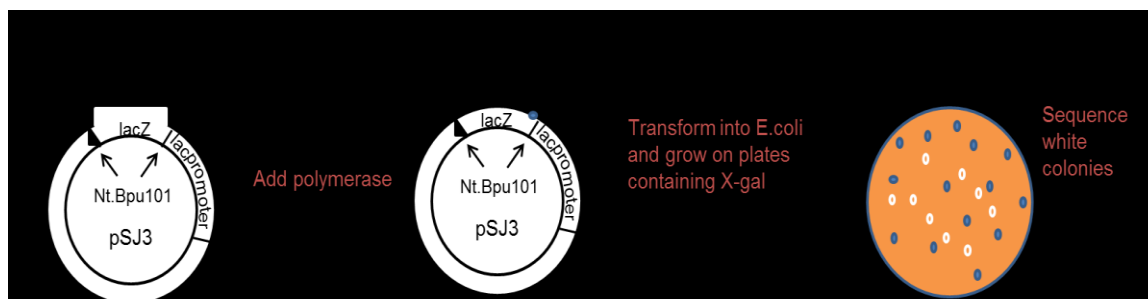


Figure 4.15 Gapped plasmid used in the fidelity assay and principle of the assay. “dcs” represents downstream cutting site and “ucs” represents upstream cutting sites.

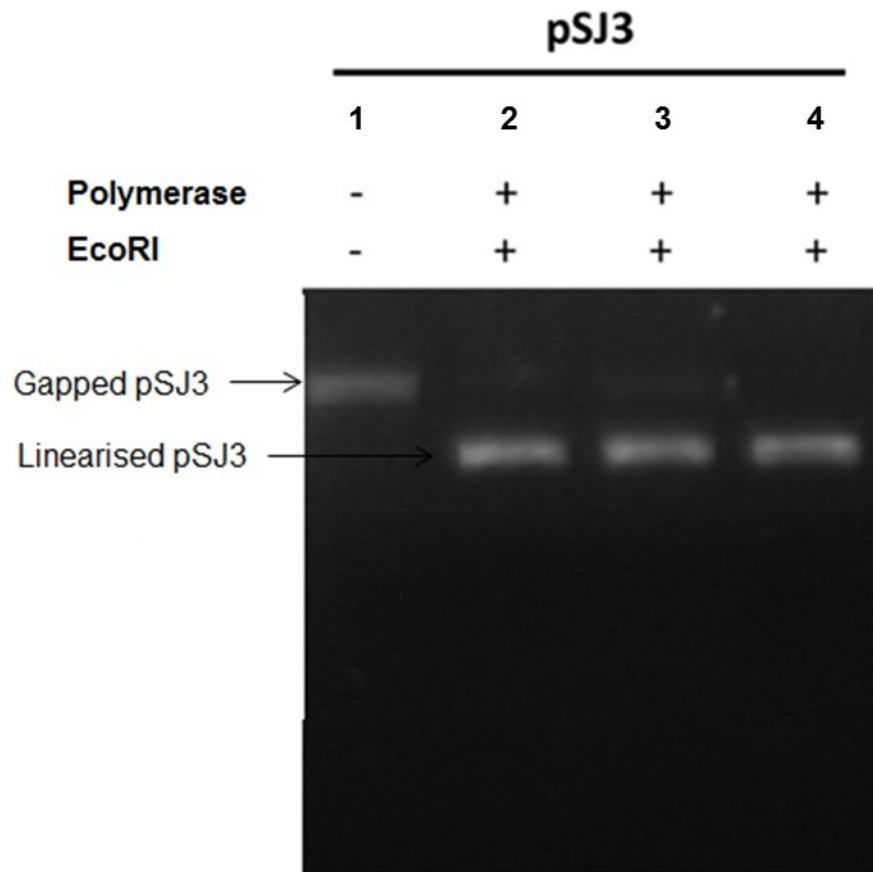


Figure 4.16 Agarose gel (1 %) electrophoresis showing filling of gapped pSJ3 by polymerases. Lane 1, starting gapped pSJ3 (Enzyme⁻, EcoRI⁻), lanes 2-4, filled pSJ3 plasmid linearised via incubation with EcoRI (Enzyme⁺, EcoRI⁺).

Plasmids that had been successfully filled were identified via gel electrophoresis following digestion with EcoRI. Control gapped pSJ3 plasmid was run on the same 1 % agarose gel as the filled plasmids (via addition of polymerase) that had been linearised via EcoRI digestion. The EcoRI digested, linearised, plasmids were identified as they travelled further through the gel than the nicked, gapped pSJ3 control vector. After identification of successfully filled plasmids, *E.coli* Top10 cells were transformed with 1 µl of the plasmid and blue/white screening, using X-gal, was performed to calculate the number of mutants (white colonies). The ratios of white/blue plaques were used as a reflection of fidelity, the higher the proportion of white plaques, the less accurate the polymerase.

Fidelity results were determined from 5 independent observations carried out in triplicate (Table 4.6).

Protein	Total colonies ^a	Mutant (white) colonies ^a	Corrected mutation frequency ^b	Error rate ^c
<i>Pfu</i> -Pol D	36690	93	0.2529	1.7×10^{-5}
<i>Pfu</i> -Pol D <i>exo</i> ⁻	70903	301	0.425	2.9×10^{-5}
<i>Mja</i> -Pol D	37748	109	0.288	2.0×10^{-5}
<i>Mja</i> -Pol D <i>exo</i> ⁻	45895	298	0.648	4.4×10^{-5}
<i>Mja</i> -DP2	39852	376	0.992	6.4×10^{-5}

Table 4. 6 Error rate of DNA polymerases determined using pSJ3 in the plasmid based fidelity assay.

a Sum of three independent experiments each consisting of 5 repeats.

b The corrected mutation frequency is the observed mutation rate (white colonies)/(total colonies) X 100. with the background mutation frequency of pSJ3 (3.1×10^{-5}) subtracted.

c The error rate was determined using the equation in the text which makes explicit the reasons for determining the detectable sites (D) and expression frequency (P) (0.44) was used.

The error rate was calculated from the mutation frequency using the following equation (Keith *et al.*, 2013)

$$ER = \frac{Ni/Fx MF}{D x P}$$

Ni=number of a particular type of mutation

N*=total number of mutations

MF=observed mutation frequency-background mutation frequency

D= number of detectable sites for a particular mutation

P= probability of expressing the mutant *lacZα* gene (expression frequency)

***Ni can only be used if the type of mutation has been identified, whereas in the absence of sequence data Ni/N=1 and can only be used to determine total mutations.**

The corrected mutation frequency is calculated as the observed mutation frequency (number of white colonies/number of blue colonies x 100) with the background mutation frequency of the pSJ3 plasmid subtracted (3.1×10^{-5}). The background mutation frequency was calculated by Brian Keith (Keith *et al.*, 2013). The corrected mutation frequency states the frequency of errors incorporated throughout the entire fidelity experiment. However, the error rate, calculated using the above formula, calculates the number of mistakes the polymerase makes per nucleotide incorporated. This figure facilitates the comparison of results from other assays and thus is the most widely used observation.

Analysis of table 4.6 reveals that the error rates of *Pfu*-Pol D, 1.71×10^{-5} , and *Mja*-Pol D, 1.96×10^{-5} are similar. Both of the wild type Pol D enzymes showed a reduction in error rate, of about 2 fold, when the 3'-5' proofreading exonuclease activity was eliminated (Table 4.6). These findings are consistent with observations of error rates of wild type and *exo⁻* variants of other well characterised polymerases, including Klenow, Vent and *Pfu*-Pol B polymerases (Byrappa *et al.*, 1995; Keith *et al.*, 2013), where the elimination of exonuclease activity reduces fidelity, 2-40 folds, depending on the reaction conditions used.

The error rate of *Mja*-DP2, 6.44×10^{-5} , was slightly higher than *Mja*-Pol D *exo⁻*, 4.42×10^{-5} . As both of these enzymes lack 3'-5' exonuclease activity, the lowered fidelity of *Mja*-DP2 is likely to be due to the protein not forming an optimally folded structure, thus, reducing the intrinsic polymerase accuracy. This indicates that the DP1 subunit contributes to fidelity even when 3'-5' exonuclease activity is inactivated. As Pol D is a dimeric protein, it is likely that interactions between the DP2 and DP1 subunits affect the final tertiary structure. Thus, when DP1 is absent, the structure of the DP2 subunit will be significantly different from the Pol D heterodimer, which may affect the function of the polymerase and therefore reduce the fidelity.

The results of the plasmid-based fidelity assays suggest that Pol D is a high fidelity enzyme (Table 4.6), in keeping with the hypothesis that it plays an important role in DNA replication. The fidelity rates of the Pol D enzymes were slightly lower than the fidelity of well-characterised, replicative, family B

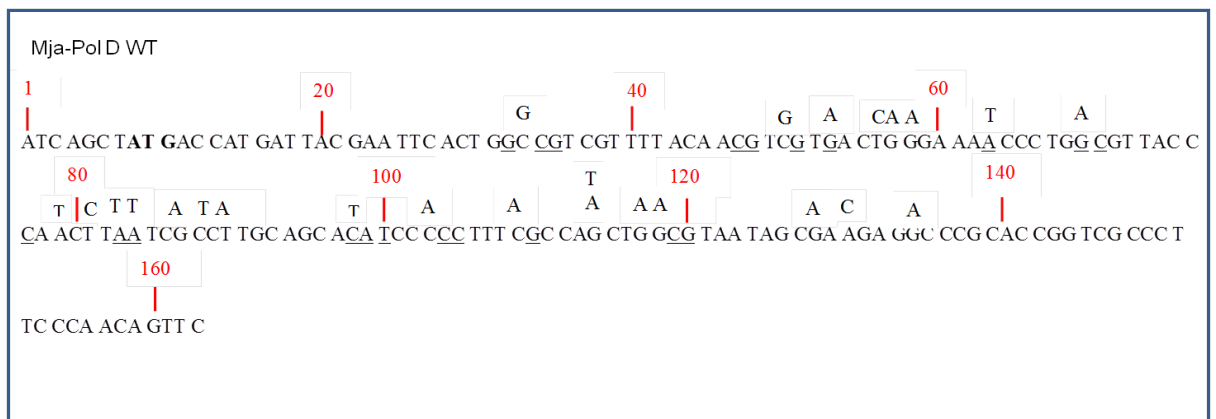
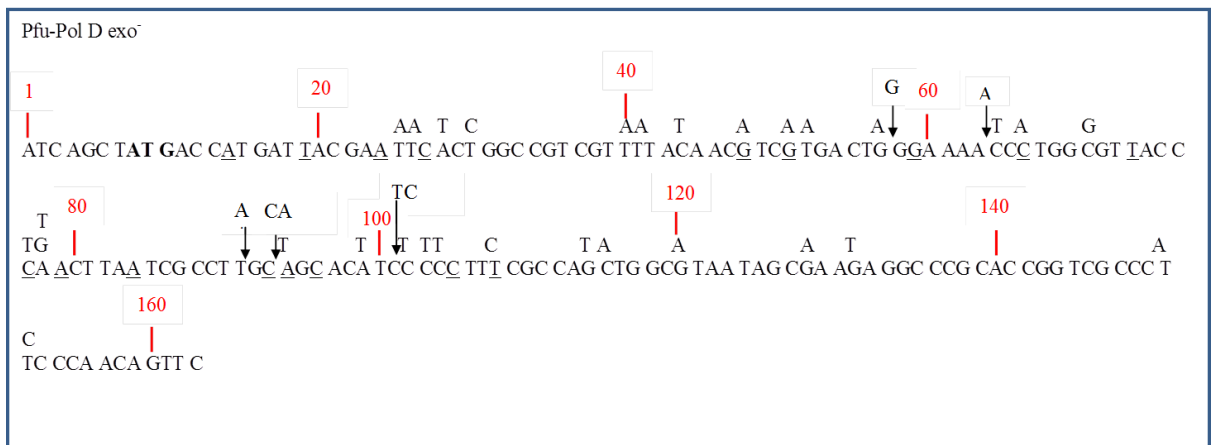
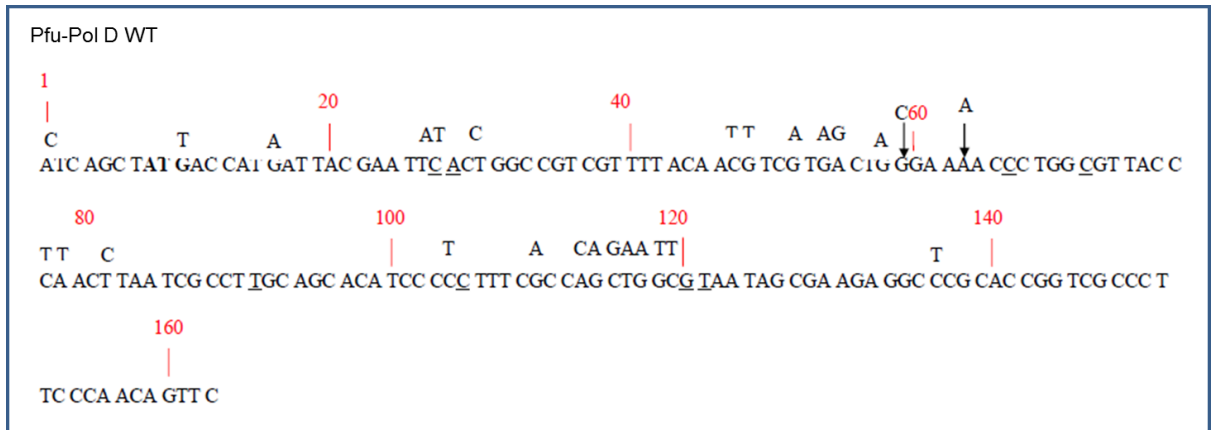
polymerases (Table 6.2). However, as the fidelity assays were performed *in vitro*, they may not be a true representation of the accuracy of the enzyme *in vivo*. In fact, it is probably that the fidelity of Pol D *in vivo* is higher than calculated in Table 4.6 as the *in vivo* environment is likely to provide the optimal conditions for the enzyme. Additionally, other proteins may interact with Pol D *in vivo* and thereby improve fidelity. Therefore although this assay highlights the accuracy of Pol D and supports the hypothesis that Pol D is a replicative polymerase— further *in vivo* testing is required to accurately calculate Pol D fidelity rates.

4.12.1 Sequencing of mutant colonies obtained during the plasmid based fidelity assay

Characterisation, by DNA sequencing, of 50 mutant colonies obtained during each plasmid based fidelity assay was performed to determine if there were any mutation hotspots (Figure 4.17) or if one specific mutation type was more common than others (Figure 4.18). Figure 4.17 shows the 163 base pair length of the lacZ α gene that was analysed during the fidelity assay and details the location of mutations detected via sequencing. This 163 base pair region contains 329 possible alterations in lacZ α that can result in an inactive gene and a white phenotype. Figure 4.17 reveals that mutations were most common within the first 120 bases of the lacZ α gene with very few mutations detected within the last 43 bases. This observation is correlated with the number of detectable sites as mutations towards the end of a gene have less chance of inactivating the lacZ α gene. This observation was consistent for all 5 polymerases tested. Analysis did not detect any mutation hotspots or sequences that were more prone to mutation.

Figure 4.18 was compiled to determine if the polymerases had a preference for a specific type of mutation e.g. substitution, insertion or deletion. Analysis of figure 4.18 reveals that for all 5 polymerases analysed, single base substitutions were more common than deletions and deletions were more common than insertions. The most common mutation type in all 5 polymerases was G-A/C-T substitution. Figure 4.17 also reveals that more than one mutation was detected in many of the mutant colonies. The wild-type Pol D polymerases had

the fewest colonies with multiple mutations while the Exo⁻ variants and *Mja* DP2 had slightly more (Figure 4.18).



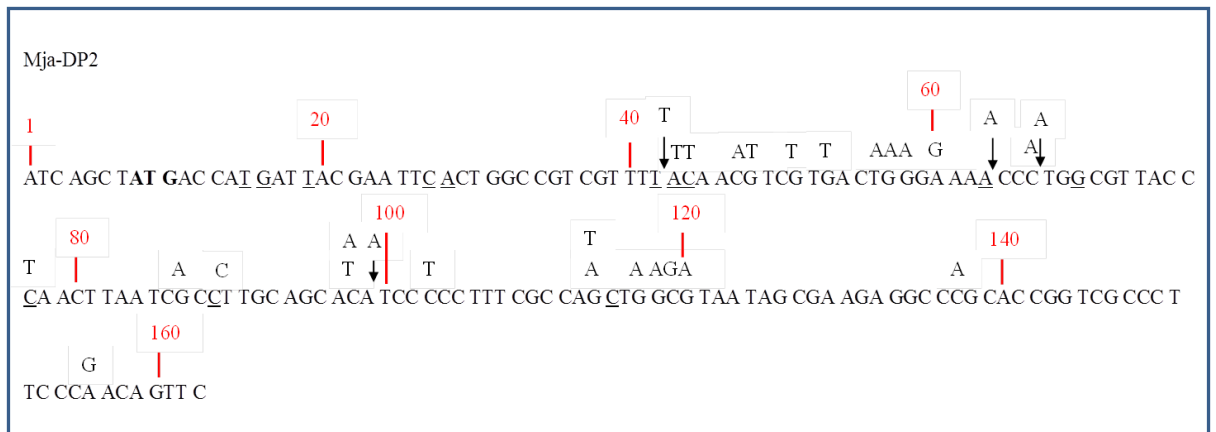
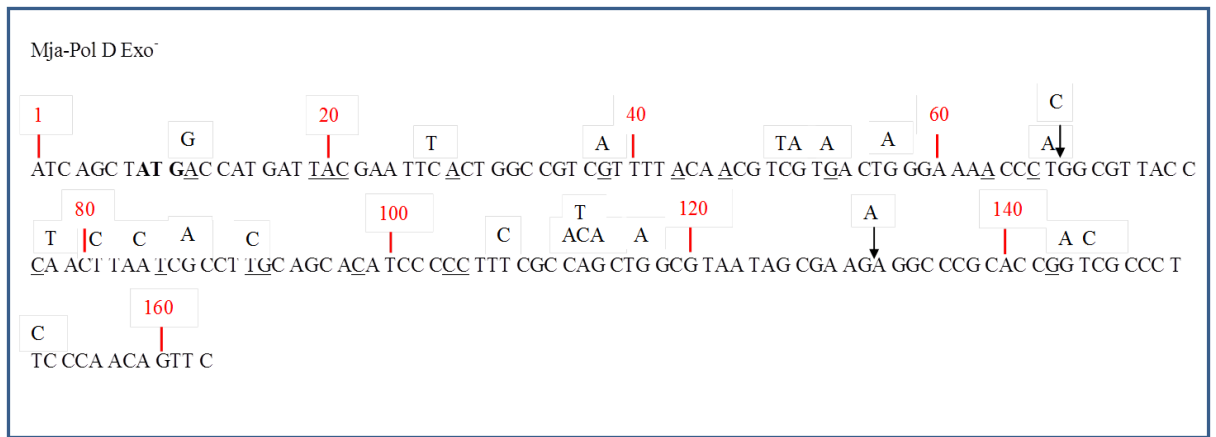


Figure 4. 10 Mutations detected, via sequencing, in 50 white colonies from the DNA polymerases during the plasmid-based fidelity assay. Underlined regions are single base deletions; letters above the sequence indicate single base substitutions and arrows represent the insertion site of a single base.

A		
Mutation Type	Number of mutations	Mutation rate (%)
A → T / T → A	2	3.8
A → G / T → C	3	5.7
A → C / T → G	4	7.5
G → A / C → T	32	60.3
G → C / C → G	2	3.8
G → T / C → A	2	3.8
Insertion	3	5.7
Deletion	5	9.4
Total number	53	100

B		
Mutation Type	Number of mutations	Mutation rate (%)
A → T / T → A	9	14.3
A → G / T → C	4	6.3
G → A / C → T	14	22.2
G → C / C → G	1	1.6
G → T / C → A	7	11.1
Insertion	6	9.5
Deletion	22	34.9
Total number	63	100

C		
Mutation Type	Number of mutations	Mutation rate (%)
A → T / T → A	3	5.7
A → G / T → C	2	3.8
G → A / C → T	27	50.9
G → C / C → G	2	3.8
G → T / C → A	1	1.9
Insertion	2	3.8
Deletion	16	30.2
Total number	53	100

D		
Mutation Type	Number of mutations	Mutation rate (%)
A → T / T → A	8	15.5
A → G / T → C	7	12.7
A → C / T → G	1	1.8
G → A / C → T	16	29.1
G → C / C → G	0	0
G → T / C → A	2	3.6
Insertion	2	3.6
Deletion	20	36.4
Total number	55	100

E		
Mutation Type	Number of mutations	Mutation rate (%)
A → T / T → A	4	6.2
A → G / T → C	3	4.6
G → A / C → T	20	30.8
G → C / C → G	1	1.5
G → T / C → A	8	12.3
Insertion	6	9.2
Deletion	23	35.4
Total number	65	100

Figure 4. 11 Analysis of mutation type obtained during the plasmid-based fidelity assay. Tables A-E list the mutations detected in fidelity assays performed on 5 different polymerases: A) *Pfu*-Pol D. B) *Pfu*-Pol D Exo⁻ C) *Mja*-Pol D WT D) *Mja*-Pol D Exo⁻. E) *Mja*-DP2. Mutation type was identified via sequencing performed by GATC.

4.13 Discussion

The data presented in this chapter provides further characterisation of family D polymerases overexpressed in *E.coli* and purified aerobically. Experimentation shows that *Mja*-Pol D possesses polymerase activity and uracil recognition properties similar to those observed in *Pfu*-Pol D (Richardson *et al.*, 2013). Although this result was unsurprising, it served to confirm previous reports of uracil recognition by *Pfu*-Pol D and suggests that this property may be universal for family D polymerases. Unfortunately, *Mth*-Pol D could not be purified from *E.coli* and thus it was not possible to test this polymerase for uracil recognition properties.

Mja-DP2 was found to polymerise DNA but at a slower rate than *Mja*-Pol D (Figure 4.6). This experiment helped to clarify conflicting reports regarding whether the DP2 subunit of family D polymerases possess polymerase activity in the absence of the small subunit (Figure 4.5 and Figure 4.6). *Mja*-DP2 was found to possess weak polymerase activity and did not possess any exonuclease activity (Figure 4.9). Further experimentation showed that *Mja*-DP2 was inhibited by the presence of uracil in template strand DNA. This is the first direct evidence that the DP2 subunit has uracil recognition properties. Unfortunately, as the DP2 subunit of Pol D does not share any significant sequence homology with other DNA polymerases (Cann & Ishino, 1999) and crystal structures are not available, the uracil recognition mechanism remains unknown.

Experiments also confirmed that the *Mja*-DP1 and *Mth*-DP1 required Mn ions to facilitate exonuclease activity. This manganese dependent exonuclease activity, previously described by Jokela *et al*, indicates that the DP1 subunit is less efficient at exonucleolysis than the *Mja*-Pol D enzyme. To explain why DP1 requires Mn ions to exhibit exonuclease activity, we looked at the well characterised reactions performed by restriction endonucleases (Imhof *et al.*, 2009). Restriction endonucleases digest DNA at specific “recognition sites”. For digestion to occur, the correct protein-DNA contacts must form to obtain enough binding energy to achieve the transition state and therefore catalysis (Imhof *et al.*, 2009). For many restriction endonucleases, Mg ions are

essential (binds scissile phosphate and a number of water molecules) for this reaction to occur. However, other metal ions, including Mn, are able to bind to the restriction endonucleases as necessary for the reaction to take place (Imhof *et al.*, 2009).

When Mn ions are present in restriction endonucleases reactions there is an increased level of “Star” activity. Star activity, cleavage at a site that differs by at least one base, is indicative of the restriction endonuclease losing specificity and therefore being less efficient. Mn ions are able to form DNA-protein interactions similar to those formed by Mg ions; however, Mn ions require less DNA-protein interactions to obtain the energy for the reaction to occur. Thus, proteins that do not make the correct DNA-protein interactions in the presence of Mg ions may be able to achieve the transition state in the presence of Mn ions. This indicates that the DP1 proteins have weak activity when expressed alone (requiring Mn to complete the reaction), and much stronger activity is observed when the DP2 subunit is present (Pol-D).

This observation indicates that interactions between the DP2 and DP1 subunits improve the rate of exonucleolysis. This observation is further supported by experimentation that shows that the purified DP2 and DP1 proteins mixed *in vitro* did not obtain the processivity observed in the *Mja*-Pol D (co-expressed *in vivo*) protein (Figure 4.11) and that *Mja*-DP2 was significantly less processive than *Mja*-Pol D (Figure 4.6). As a dimeric protein, it is likely that the large, DP2, and small, DP1 subunits form interactions (e.g. Van der Waal’s) *in vivo* that affect the tertiary structure and folding of the protein.

At present the interactions that occur are unknown, however surface plasmon resonance (SPR) assays have shown that the DP1 subunit interacts with the N-terminus of the large, DP2, subunit and that intra-subunit interactions form between the N-terminus and the C-terminus of the DP2 subunit (Tang *et al.*, 2004). Attempts to obtain a crystal structure of a family D polymerase, using X-ray crystallography, to identify the structure and the interactions between the DP2 and DP1 subunits have been unsuccessful (Yamasaki *et al.*, 2010; Matsui *et al.*, 2011). It is hypothesised that crystallography studies have failed as the Pol D proteins studied have been purified aerobically. It is believed that these

proteins have lost their intrinsic Fe-S cluster, which affects the folding of the protein and prevents crystals forming. Thus, to determine the structure of Pol D, it is believed that the protein needs to be overexpressed in an archaeal host species and purified anaerobically.

As DNA polymerases are central to DNA replication and repair processes, low error rates are essential to prevent DNA mutagenesis. Therefore, if Pol D is the main replicative polymerase in archaea, it is expected to have a low error rate. However, comparisons of *Mja*-Pol D (1.96×10^{-5}) and *Pfu*-Pol D (1.71×10^{-5}) error rates with other replicative polymerases including: *Pfu* Pol B (3.5×10^{-6}) (Keith *et al.*, 2013) and *Saccharomyces cerevisiae* Pol ϵ (5.0×10^{-8}) (McCulloch and Kunkel, 2008) indicate that *Mja*-Pol D and *Pfu*-Pol D have relatively low fidelity rates. However, fidelity is affected by many factors including: pH, temperature, concentration of dNTPs, protein conformation, accessory proteins and exonuclease activity (Cline *et al.*, 1996). Thus the results obtained from the fidelity assay may not be giving a true representation of the accuracy of Pol D *in vivo*.

Due to the dimeric structure of Pol D and the hypothesis that Pol D purified from *E.coli* lacks Fe-S clusters, which may affect protein folding and structure, it is possible that Pol D, *in vivo*, has a higher fidelity than these results suggest. Therefore, although the results provided in this chapter allow for comparisons of processivity between enzymes, they may not provide a true representation of *in vivo* function. Thus Pol D, *in vivo*, may be more processive and accurate than currently reported and may be the main replicative polymerase in archaea. To test this hypothesis, attempts were made to overexpress Pol D within archaeal host species and purify the protein with an intact Fe-S cluster (Chapter 5).

It should be noted that the proteins used in this chapter contained high levels of contamination from endogenous *E.coli* proteins (Figure 4.5), thus determination of protein concentration was tricky. Therefore, it is important to focus on the general trends observed in the results presented in this chapter e.g. *Mja*-Pol D is more processive than *Mja*-DP2 rather than attempt to extract precise measurements of extension rates. It is hypothesised that overexpression of Pol D within an archaeal host organism will lead to improved purification of the

protein as the presence of an intact Fe-S will reduce proteolysis and improve the thermostability of the protein and thus aid purification.

Chapter 5

Overexpressing family D polymerases in archaea

5.1 Background

As mentioned previously, in chapter 4, it has been hypothesised that family D polymerases are metallo-enzymes that possess an Fe-S cluster at the C-terminus domain (CTD) of the DP2 subunit (Tahirov *et al.*, 2009; Netz *et al.*, 2012) This hypothesis is based on the profile of highly conserved cysteine residues, within the DP2 CTD, that is similar to that which has previously been observed in eukaryotic family B polymerases (Figure 5.1 and Table 5.1) (Netz *et al.*, 2012)

Research has shown that all four yeast family B polymerases possess an Fe-S cluster *in vivo* that is labile under aerobic conditions and thus does not form when the Pol Bs are overexpressed in *E.coli* (Netz *et al.*, 2012). It is believed that the Fe-S cluster in eukaryotic family B polymerases facilitates the correct folding of the protein and interactions with other subunits required to assemble the polymerase holoenzyme (Netz *et al.*, 2012). Thus, due to the similarity in the profile of conserved cysteine residues (Table 5.1) it is hypothesised those family D polymerases may contain an Fe-S cluster *in vivo* (at the CTD of DP2 subunit) that performs a similar role.

Figure 5.1 identifies an unusually large number of highly conserved cysteine residues, almost all of which are situated in pairs (represented by yellow Cs). This profile of cysteine residues is similar to that observed in yeast family B polymerases and is often associated with metal-binding sites including iron sulphur clusters (Table 5.1) (Otvos and Armitage, 1980; Passerini *et al.*, 2006). Eukaryotic family-B polymerases possess two conserved cysteine-rich metal-binding motifs, referred to as CysA and CysB, at their CTD (Table 5.1). CysA binds Zn^{2+} and is required for PCNA-mediated Pol δ processivity while CysB is an 4Fe-4S centre (Netz *et al.*, 2012).

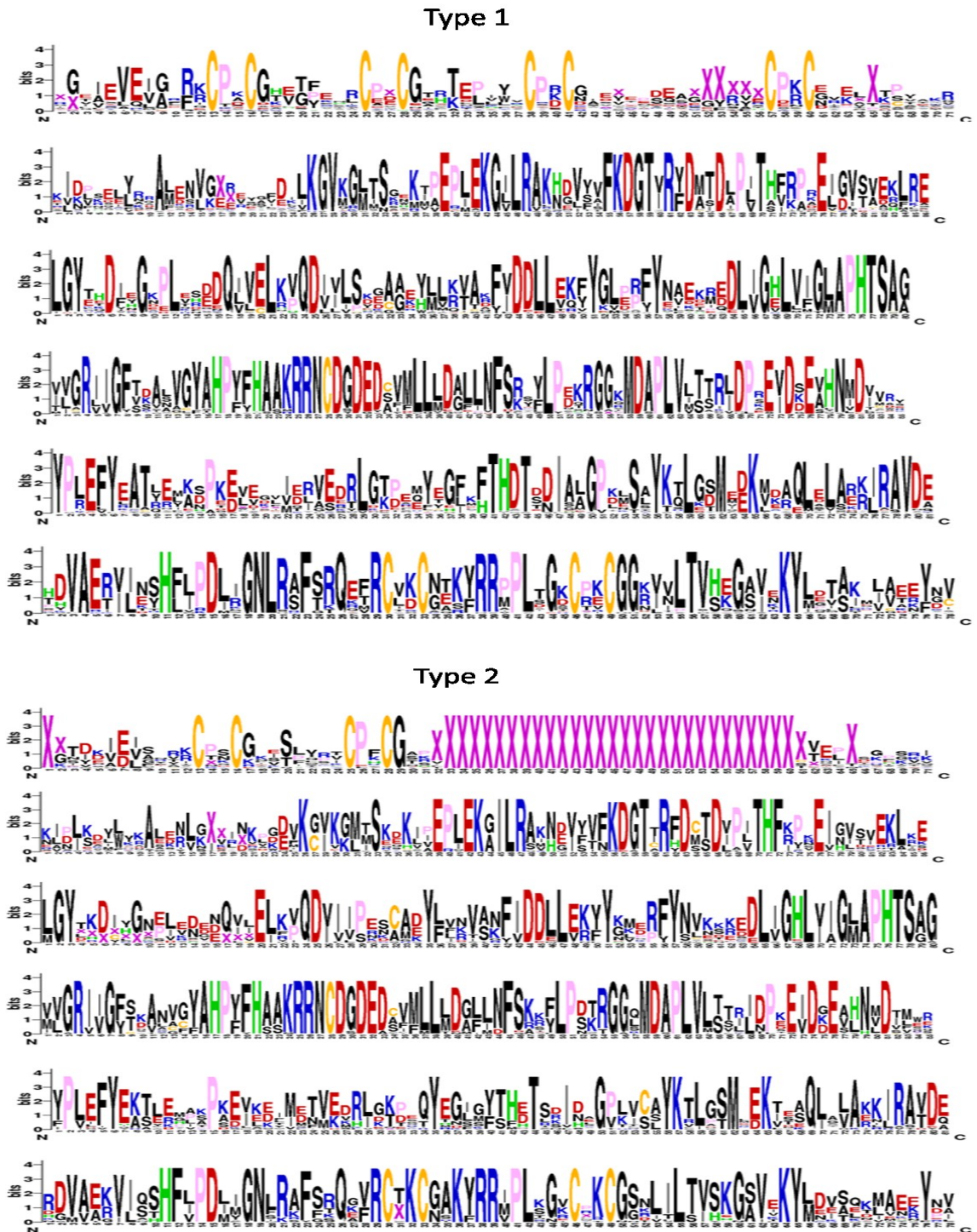


Figure 5.1 Amino acid sequence conservation of the large subunit C-terminus of family D DNA polymerases. The degree of conservation of each residue is indicated by the size of each letter. The two types are almost identical except for a region containing 30 amino acids, which is absent from type 2 sequence and is indicated by magenta X's on line 1. Family D polymerases from *Pyrococcus* and *Thermococcus* genera are invariably type 1; while family D polymerases from *Mthe* other genera can be classified as type 2 (Image produced using Weblogo {<http://weblogo.berkeley.edu/>}).

Polymerase	Cys A (Zn)	Cys B (FeS)
Eukaryotic Pol-B	C(N) ₂ C(N) _{~10} C(N) ₂ C	C(N) _{~2} C(N) _{~10} C(N) _{~2} C
Pol-D Type 1)	C(N) ₂ C(N) _{~8} C(N) ₂ C(N) _{~9} C(N) ₂ C(N) _{~15} C(N) ₂ C	C(N) ₂ C(N) _{~12} C(N) ₂ C
Pol-D(Type 2)	C(N) ₂ C(N) _{~8} C(N) ₂ C	C(N) ₂ C(N) _{~12} C(N) ₂ C

Table 5.1 All eukaryotic family-B polymerases contain two cysteine motifs, designated CysA (binds Zn²⁺) and CysB (4Fe-4S centre). Two similar motifs can be seen with family-D enzymes (on the large DP2 subunit), which appear to come in two flavours. The *Mja* type is very similar to eukaryotes, while the *Pfu* type has one of the motifs expanded. At present it is not possible to say which Pol-D motif corresponds to Cys A and Cys B.

Within eukaryotic family B polymerases the 4Fe-4S centre (CysB) plays an essential role in stabilizing the CTD. CysB facilitates complex formation and maintenance of the catalytic polymerase subunit with its respective accessory proteins (Netz *et al.*, 2012). Thus it can be hypothesised that Pol D, *in vivo*, possesses an Fe-S centre that also facilitates stabilization of the CTD and assists in protein folding and binding. It is hypothesised, that correct protein folding, in the presence of the Fe-S cluster confers increased processivity and a higher fidelity rate than has been observed in studies involving family D polymerases that have been purified aerobically from *E.coli* and lack an Fe-S cluster (Chapter 4).

Although iron and sulphide can spontaneously assemble into an Fe-S cluster under laboratory conditions, genetic studies have shown that the assembly of Fe-S clusters *in vivo* is mediated by a consortium of highly conserved proteins (Zheng *et al* 1998; Takahashi & Tokumoto 2002). Across the three domains of life, three different Fe-S pathways have been identified: the Isc (iron sulphur cluster) system, the Suf (sulphur formation) system, and the Nif (nitrogen fixation) system (Johnson *et al.*, 2005). In all three pathways, cysteine desulfurase is required to free sulphur atoms from cysteine for use in cluster assembly (Zheng *et al.*, 1993).

The free sulphur is then donated to a second protein that acts as a “scaffold” for nascent Fe-S cluster assembly. The process of iron sequestration for Fe-S formation is not well understood, however, it is believed to be donated by a chaperone protein (such as CyaY) (Castro *et al.*, 2008). The formation of Fe-S clusters *in vivo* is a complex multi-step process that requires many accessory proteins. The final structure of Fe-S clusters is dependent on the availability of sulphur, iron, and accessory proteins (Figure 5.2) (Raulfs *et al.*, 2008). The most common Fe-S clusters are the rhombic cluster [2Fe-2S] and the cubane [4Fe-4S] that normally attach to their protein via cysteine thiolate ligands (Figure 5.2) (Raulfs *et al.*, 2008). The profile of cysteine residues in CTD of the DP2 subunit (Figure 5.1) and observations in eukaryotic Pol B indicates that Pol D forms a 4Fe-4S cubane cluster *in vivo*.

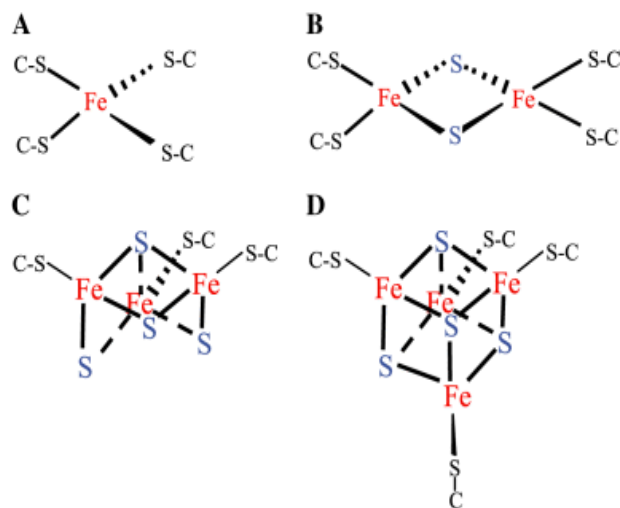


Figure 5.2 Structures of common iron-sulphur clusters. A) Rubredoxin-style iron centre B) [2Fe-2S] cluster C) [3Fe-4S] cluster D) [4Fe-4S] cluster. CS represents coordinating cysteine residues from the polypeptide backbone. Taken from Imlay, J.A., 2006.

Despite being involved in many essential biochemical processes including electron transfer, substrate binding and activation, DNA replication and repair, regulation of gene expression, and tRNA modification (Johnson *et al.*, 2005; Ayala-Castro *et al.*, 2008), Fe-S clusters are relatively unstable and sensitive to

oxidation (Figure 5.3) (Imlay, J.A., 2006). When exposed to oxygen cubane Fe-S clusters are oxidised converting the $[4\text{Fe-4S}]^{2+}$ form to an unstable $[4\text{Fe-4S}]^{3+}$ state which releases iron (Figure 5.3). Further degradation of the $[4\text{Fe-4S}]^{3+}$ cluster has been observed *in vitro* and *in vivo*, however the rate of degradation and the final products obtained are not well established (Imlay, J.A., 2006).

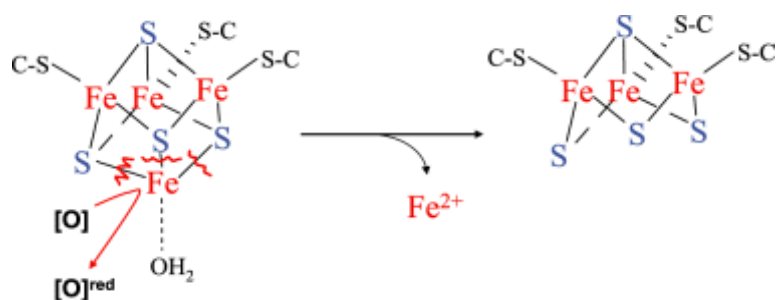


Figure 5.3 Fe-S cluster destruction by oxidation. Univalent oxidants, including superoxide, hydrogen peroxide, molecular oxygen and peroxyxynitrite [collectively denoted (O)], convert the exposed $[4\text{Fe-4S}]^{2+}$ cluster to the unstable +3 oxidation state. The cluster then spontaneously decomposes to the inactive $[3\text{Fe-4S}]^{1+}$ form. Taken from Imlay, J.A., 2006.

The lability of iron sulphur centres under aerobic conditions makes it tricky to purify metallo-enzymes with the clusters intact. Thus, when eukaryotic family B polymerases are expressed in *E.coli* and purified under aerobic conditions the Fe-S centre is destroyed and CysB aberrantly binds Zn^{2+} (Netz *et al.*, 2012). Therefore, to detect the Fe-S cluster, eukaryotic Pol Bs were overexpressed in yeast and purified anaerobically.

Thus, if family D polymerases do contain Fe-S clusters *in vivo*, it is likely that they are degraded (or do not form at all) when the proteins are overexpressed in *E.coli* and purified aerobically (UV/visible absorbance spectroscopy gives no indication of the presence of a $[4\text{Fe-4S}]$ cluster). Thus, to investigate if family D polymerases possess Fe-S clusters *in vivo*, and what role the centre plays it is necessary to overexpress the family D polymerase in an archaeal host organism and purify the desired protein anaerobically. Archaea host species

should have the correct systems for instigating Fe-S centres into archaeal proteins. Also many anaerobes limit degradation. Once purified, it would be possible to test the protein sample for metal ions via atomic absorption spectroscopy (AAS) or inductively coupled plasma mass spectrometry (ICPMS) (Beauchemin, 2010)

5.2 *Methanococcus maripaludis* as a model organism

Methanococcus maripaludis (*Mma*) is a methanogenic mesophile, originally isolated from salt marsh sediment in North Carolina (Jones *et al.*, 1983), that has been developed into a model organism used in archaeal genetic and biochemical studies (Leigh *et al.*, 2011). *Mma* grows on a defined liquid media with a doubling time of two hours and on solid media with a growth time of 5 days, under optimal growth conditions (38°C, pH 6.8-7.2) (Jones *et al.*, 1983). As a strictly anaerobic methanogen, *Mma* requires either formate or a mixture of H₂ (reducing agent) and CO₂ (carbon source) as growth substrates (Jones *et al.*, 1983). Thus specialist equipment is required for growth within a laboratory environment.

The genetic tools available for *Mma* include shuttle vectors (Gardner & Whitman, 1999), reporter genes, a PEG based transformation protocol (Tumbula *et al.*, 1994), and a markerless mutagenesis protocol that allows the deletion of non-essential genomic regions and the introduction of mutations or foreign genes (Moore & Leigh, 2005). *Mma* is the best anaerobic archaeal model organism currently available.

Mma was chosen as the host organism to overexpress *Mja*-Pol D as *Mma* and *Mja* are closely related members of the *Methanococcus* phyla (Figure 4.2), and thus contain similar DNA and protein sequences. As *Mma* is a methanogenic anaerobe, it was hypothesised that overexpressed *Mja*-Pol D purified under strict anaerobic conditions would possess an intact Fe-S cluster at the CTD of the DP2 subunit. As a closely related archaea species it was predicted that *Mma* would possess the necessary enzymes required for Fe-S insertion and thus facilitate Fe-S formation. It was anticipated that the biochemical properties of *Mja*-Pol D overexpressed and purified in *Mma* (containing Fe-S cluster) could

be compared to previous experiments performed with *Mja*-Pol D overexpressed and purified in *E.coli* (no Fe-S cluster) (Chapter 4).

5.3 Overexpressing recombinant proteins in the Euryarchaea

The heterologous expression of proteins in euryarchaea is in its infancy. At present, there is a limited availability of suitable plasmids that facilitate overexpression within euryarchaeal species. Due to the lack of knowledge regarding suitable strong promoters and gene control elements there are no controllable promoters currently available. The lack of suitable plasmids with strong, inducible promoters limits the overexpression of proteins and thus further work is required to improve the selection of plasmids currently available. In particular, systems for expressing proteins composed of two subunits are required to allow multi-subunit proteins to be co-expressed and their properties investigated.

5.4 Cloning of *Mja*-DP2 subunit into pAW42, a plasmid suitable for overexpression in *M.maripaludis*

Mja-DP2 was cloned into a *Mma* shuttle plasmid, pAW42 (Figure 5.4), using the restriction digest cloning technique. It would have been preferential to overexpress both DP2 and DP1 subunits within the same vector; however, there are no vectors currently available for expression within *Mma* that can facilitate more than one DNA insert. Thus initial attempts were made to overexpress *Mja*-DP2, as previous experiments had shown that *Mja*-DP2 expressed without the DP1 subunit possessed polymerase activity. Cloning was designed to insert *Mja*-DP2 between *AscI* and *BglII*, downstream of the *Phmva* promoter (the *Methanococcus voltae* histone promoter) (Figure 5.4). This design ensured that *Mja*-DP2 contained an N-terminal (His)₆ tag to facilitate purification of the expressed protein.

E.coli Top10 cells were transformed with the pAW42 vector containing the ligated *Mja*-DP2 gene. The presence of an ampicillin resistance marker and Ori within the pAW42 vector facilitated its passage through *E.coli* and allowed successful transformants to be selected for. The use of *E.coli* cells allowed successful clones to be identified (confirmed by sequencing) prior to

transformation of *Mma* cells. *Mma* strain S0001 was then transformed with the pAW42 [*Mja*-DP2] vector and selected for on agar containing puromycin. Transformation of *Mma* strain S0001 was performed under strictly anaerobic conditions.

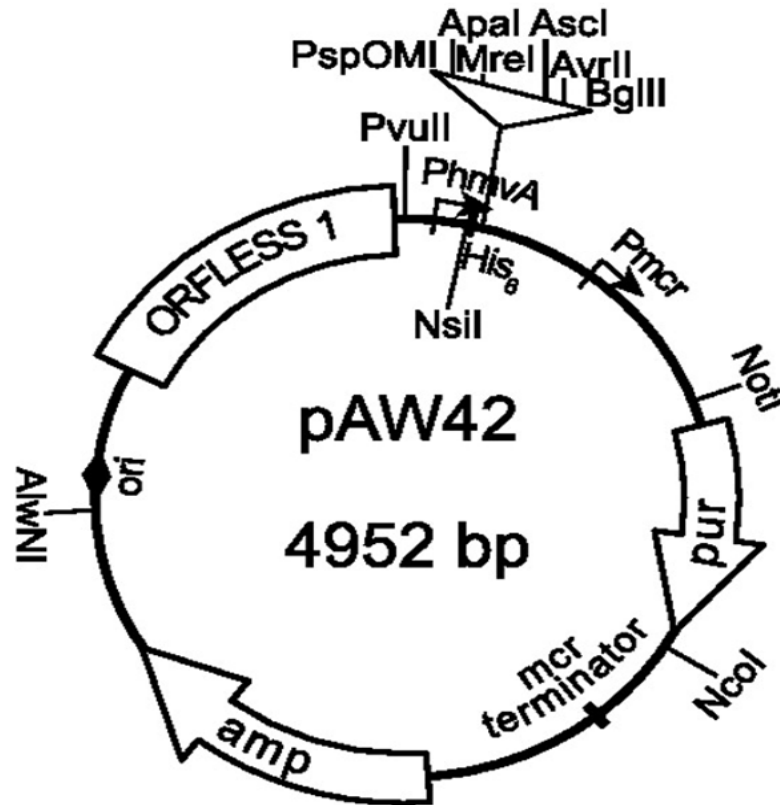


Figure 5.4 pAW42 shuttle plasmid for protein expression in *Mma*. Ampicillin marker provides ampicillin resistance for selection in *E.coli* and puromycin marker provides resistance for selection in *Mma*. Taken from Walters *et al.*, 2011.

5.5 Expression and purification of *Mja*-DP2 from *Mma* strain S0001

Following transformation of *Mma* strain S0001 with pAW42 [*Mja*-DP2], utilising puromycin for selection, individual colonies were picked. 1 litre of *Mma* growth media was inoculated with one transformant colony and grown in a fermenter under anaerobic conditions until an O.D.₆₀₀ of 2.0 was obtained. The *Mma* cells were then pelleted via centrifugation and re-suspended in an appropriate buffer. The cells were sonicated for 5 minutes, incubated with DNase at 37°C for 20 minutes and heated to 60°C for 30 minutes prior to a final round of

centrifugation. It was hypothesised that the thermostable *Mja*-DP2 protein would be able to tolerate the high temperatures while the mesophilic *Mma* proteins would be denatured, and thus improve the purification of *Mja*-DP2. Expression and purification of *Mja*-DP2 was performed under strict anaerobic conditions. The supernatant was analysed via SDS PAGE (Figure 5.5).

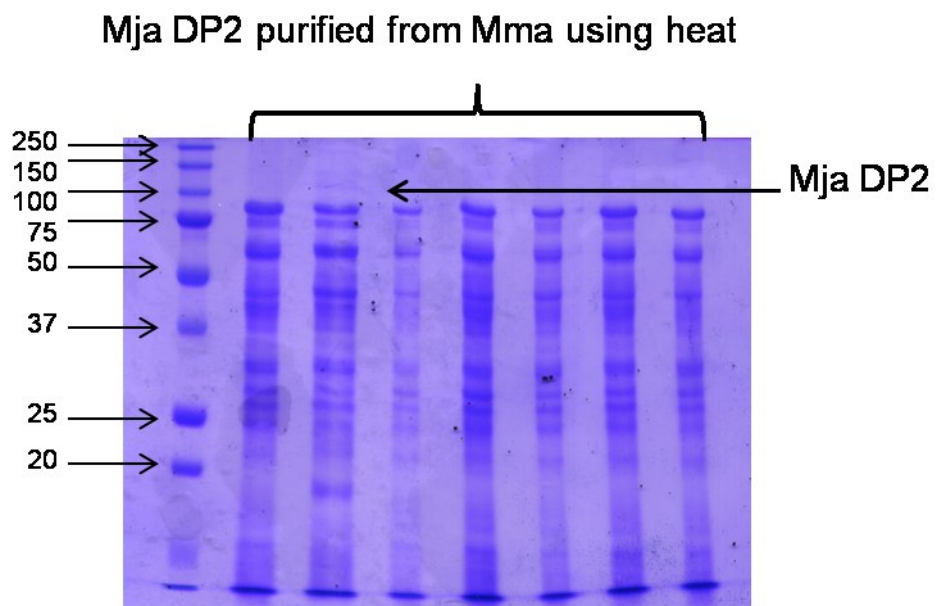


Figure 5.5 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing heated and centrifuged *Mma* cell extract. Each lane contains repeats of the same *Mma* extract.

Analysis of the SDS-PAGE revealed an exceptionally faint protein band of ~ 100 kDa. This protein band was excised and identified as *Mja*-DP2 via mass spectrometry (Figure. 5.6). Unfortunately the *Mja*-DP2 protein appeared to have been expressed in very low levels (faint band) and the purification protocol resulted in high levels of contamination by native *Mma* proteins (Figure 5.5).

Mascot Search Results

{*MATRIX*}
 {*SCIENCE*}

User : tf2012
 Email :
 Search title : Ultraflex ProteinID
 MS data file : 13229323905421392.mgf
 Database : NCBIInr 20101130 (12348165 sequences; 4221604734 residues)
 Timestamp : 2 Nov 2012 at 15:24:04 GMT
 Enzyme : Trypsin
 Fixed modifications : [Carbamidomethyl \(C\)](#)
 Variable modifications : [Oxidation \(M\)](#)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 100 ppm
 Fragment Mass Tolerance: ± 0.5 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 10
 Protein hits : [gi|15669826](#) DNA polymerase II large subunit [Methanocaldococcus jannaschii DSM 2661]
 [gi|45358548](#) acetyl-CoA decarboxylase/synthase complex subunit alpha [Methanococcus maripaludis S2]

Select Summary Report

Format As: Select Summary (protein hits)

Significance threshold p< Max. number of hits

Standard scoring MudPIT scoring Ions score or expect cut-off Show sub-sets

Show pop-ups Suppress pop-ups Require bold red

Re-Search All queries Unassigned Below homology threshold Below identity threshold

1. [gi|15669826](#) Mass: 130460 Score: 546 Matches: 5(5) Sequences: 5(5)
 DNA polymerase II large subunit [Methanocaldococcus jannaschii DSM 2661]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3	1482.8234	1481.8161	1481.7847	21.2	0	57	0.027	1	U	R.ILELIGCCHLVR.N
4	1545.7905	1544.7833	1544.7524	20.0	0	92	1.1e-05	1	U	K.ANVGYAHFYFHAAK.R
6	1747.9652	1746.9579	1746.9265	18.0	0	136	3.4e-10	1	U	R.SAGGTAQALSVLVGFVR.K
8	1912.9772	1911.9699	1911.9366	17.4	0	127	2.5e-09	1	U	K.NPDGTEYLAIYYAGPIR.S
10	2152.1038	2151.0965	2151.0444	24.2	0	134	4e-10	1	U	R.NIPIEITGEATDDVEVSGHR.D

Proteins matching the same set of peptides:

[gi|289192128](#) Mass: 187589 Score: 546 Matches: 5(5) Sequences: 5(5)
 DNA polymerase II, large subunit DP2 [Methanocaldococcus sp. FS406-22]

2. [gi|45358548](#) Mass: 87223 Score: 122 Matches: 2(2) Sequences: 2(2)
 acetyl-CoA decarboxylase/synthase complex subunit alpha [Methanococcus maripaludis S2]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
5	1694.9045	1693.8973	1693.8716	15.2	0	61	0.012	1	U	K.YFGELPPDLELFVR.N
9	1965.0426	1964.0353	1964.0003	17.8	0	62	0.0083	1		K.KPLSGNFEEVADYILNR.V

Figure 5.6 Mass spectrometry results identifying the large protein band highlighted on figure 5.5 as *Mja*-DP2. Mass spectrometry was performed by York University who provided the above report.

5.6 Thermostability of *Mja*-DP2

In an attempt to improve the purification, of *Mja*-DP2 from *Mma*, and reduce the amount of endogenous *Mma* proteins, the purification process was repeated with higher temperatures. Protein samples were heated for 30 minutes at 60°C, 70°C, 80°C and 90°C (Figure 5.7).

Unfortunately, despite a visible reduction of endogenous *Mma* proteins, the *Mja*-DP2 protein appeared to be degraded by the high temperatures (Figure). SDS-PAGE analysis detected low amounts of *Mja*-DP2 protein in samples that had been heated at 60°C and 70°C, however, no *Mja*-DP2 protein was detected in samples that had been heated to 80°C or 90°C. This finding confirms previous reports that *Mja*-DP2 is less stable when at high temperatures than the *Mja*-Pol D itself (Jokela *et al.*,2004).

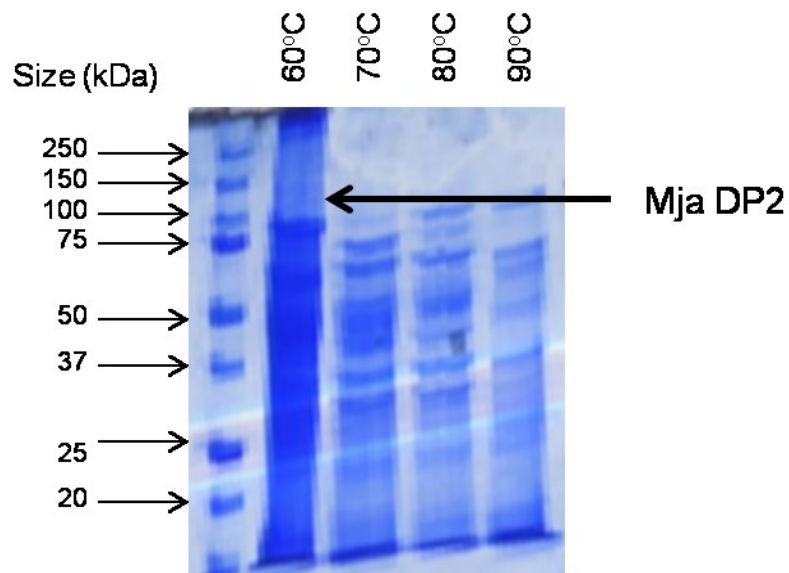


Figure 5.7 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing elute samples containing *Mja*-DP2 purified from *Mma*. Samples were heated to 60°C, 70°C, 80°C and 90°C.

5.7 Purification of *Mja*-DP2 using the N-terminal His tag

Further attempts to purify *Mja*-DP2 expressed in *Mma* were made utilising the N-terminal His tag. Purification was performed as stated in section 5.5 prior to the protein sample being mixed with Ni-NTA agarose and passed over a gravity flow column. The Ni-NTA agarose was washed 3 times then subjected to high concentrations of imidazole to elute the *Mja*-DP2 protein. Eluate samples were collected and analysed via SDS-PAGE. Purification was performed under strictly anaerobic conditions. Unfortunately, SDS-PAGE analysis did not identify any protein bands corresponding to the size of *Mja*-DP2 (~ 100 kDa) (result not shown).

It is likely that *Mja*-DP2 was not identified due to the low level of protein expression obtained under the control of the non-inducible *Phmva* promoter in pAW42. Purification using Ni-NTA agarose would have diluted the protein sample further than the amounts obtained from crude extracts and shown in Figures 5.5 & Figure 5.7. Thus, it is unsurprising that *Mja*-DP2 protein was not detectable following Ni-NTA agarose chromatography.

5.8 Purification of *Mja*-DP2 via gel filtration

As *Mja*-DP2 is significantly larger than almost all the endogenous *Mma* proteins (Figure 5.5), it was anticipated that it may be possible to purify the protein based on size. Thus, a final attempt was made to purify *Mja*-DP2 from *Mma* using a gel filtration column. The protein purification process was performed as stated in section 5.5, followed by application of the supernatant to a gel filtration column.

Analysis of protein fractions collected from the gel filtration column (Figure 5.8) shows a reduction in endogenous *Mma* proteins compared to purification using heat alone (Figure 5.7). However, despite a slight improvement in the purification process, a significant level of endogenous *Mma* proteins remained, rendering the purification process unsuitable (Figure 5.8). Gel filtration was performed under aerobic conditions. Thus, it was primarily used to determine if the technique would be successful. If it had been, plans would have been made to purify the protein using gel filtration anaerobically.

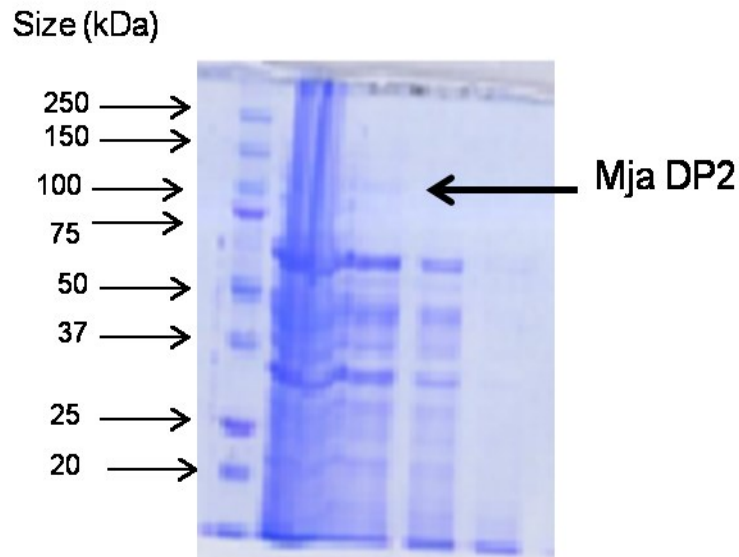


Figure 5.8 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %). Final samples after initial purification using a heat step (60°C) followed by gel filtration.

5.9 Cloning of *Mja*-DP1 into pLW40

To facilitate co-expression of *Mja*-DP2 and *Mja*-DP1 in *Mma*, *Mja*-DP1 was cloned into the pLW40neo expression plasmid via restriction digest cloning (Figure 5.9). pLW40neo possesses an ampicillin marker and Ori to facilitate its passage and selection through *E.coli*. Thus, *E.coli* was transformed with ligation mixture containing pLW40neo and *Mja*-DP1 and successful transformants were selected for using ampicillin. *Mja*-DP1 was inserted between the *Apal* and *BglIII* sites and did not possess a His tag at either the N or C terminal (Figure 5.9). Transformed plasmids were extracted from transformant colonies and sequenced to confirm successful insertion of the *Mja*-DP1 gene.

Mma S0001 was co-transformed with pLW40neo [*Mja*-DP1] and pAW42 [*Mja*-DP2] and successful transformants selected for by growth on agar containing both puromycin and neomycin. Several attempts were made to co-transform pAW42 [*Mja*-DP2] and pLW40neo [*Mja*-DP1] using various concentrations and ratios of the plasmids as literature searches did not reveal any publications describing transforming *Mma* with two plasmids. The successful transformation

mixture contained 10 mg of pLW40neo [*Mja*-DP1] and 5 mg of pAW42 [*Mja*-DP2] and produced a few transformant colonies (~ 20) that were used to overexpress *Mja*-Pol D.

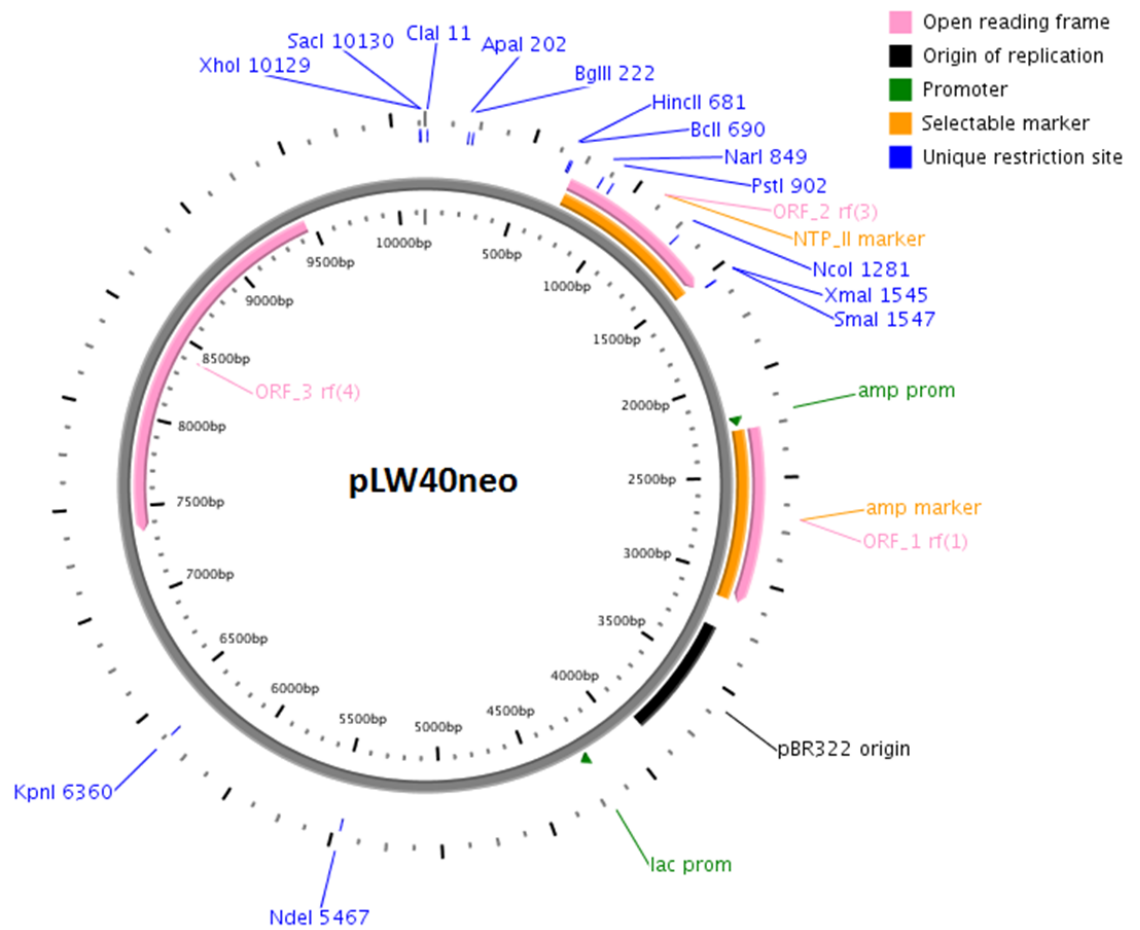


Figure 5.9 pLW40neo shuttle plasmid for protein expression in *Mma*. Ampicillin marker provides ampicillin resistance for selection in *E.coli* and neomycin marker (NTP) provides resistance for selection in *Mma*. Image created using PlasMapper version 2.0.

5.10 Expression and purification of *Mja*-Pol D from *Mma* strain S0001

A single transformant *Mma*, strain S0001, colony containing pAW42 [*Mja*-DP2] and pLW40neo [*Mja*-DP1] was used to inoculate 1 litre of *Mma* growth media. The colony was grown in a fermenter under strict anaerobic conditions until an O.D.₆₀₀ of 2.0 was obtained. The *Mma* cells were harvested and the *Mja*-Pol D protein was purified using a heat-step and Ni-NTA agarose as described for *Mja*-DP2 in section 5.5. Protein lysate was analysed using SDS-PAGE (Figure 5.10).

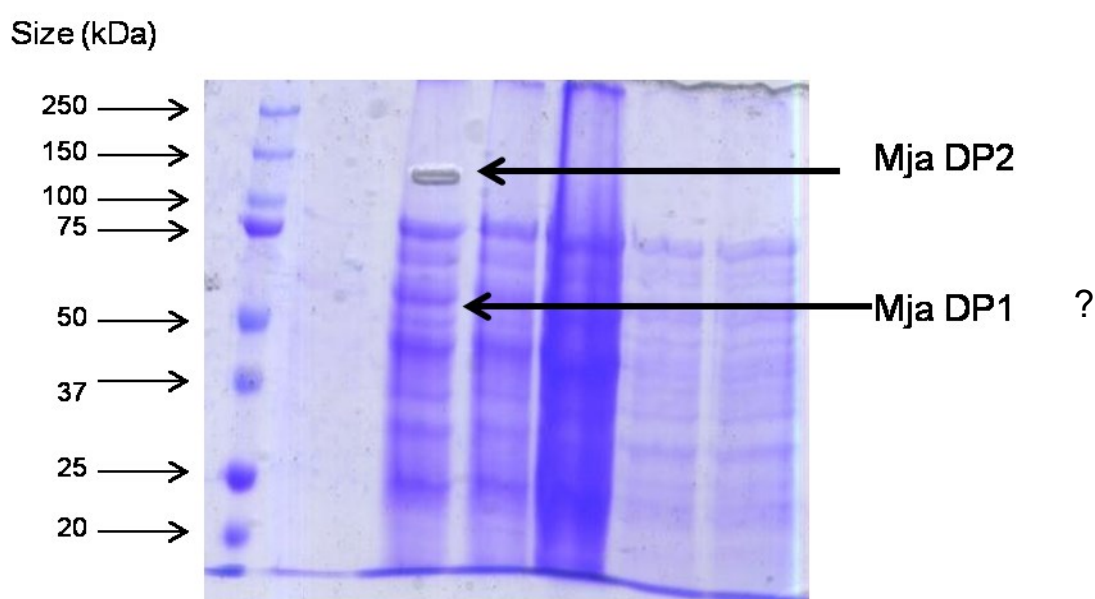


Figure 5.10 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing elute samples containing *Mja*-DP2 and *Mja*-DP1 purified from *Mma* using heat (60°C). Each lane contains repeats of the same *Mma* extract.

Analysis of Figure 5.10 identified a band that corresponded to the *Mja*-DP2 subunit (~ 100 kDa). Unfortunately, due to the large amount of endogenous *Mma* proteins, it was not possible to visually identify the *Mja*-DP1 subunit. Thus, 3 protein bands that corresponded to the size of *Mja*-DP1 (~ 48.5 kDa) were excised and analysed via mass spectrometry. Unfortunately all protein bands analysed were identified as endogenous *Mma* proteins (appendix).

Thus, despite the transformant *Mma* strain conferring puromycin and neomycin resistance, it was not possible to conclusively confirm the presence of *Mja*-DP1.

Figure 5.10 revealed that the expression of *Mja*-DP2 under the control of the *Phmva* promoter was exceptionally low, with only a very faint protein band identified. For this reason, further attempts were not made to purify *Mja*-Pol D using the N-terminal His tag on the *Mja*-DP2 subunit. Due to the limitations of growing *Mma* (limited to growing 1 litre cultures in the anaerobic fermenter available to us at York University), it was not possible to try and “scale-up” the process to obtain more protein for purification. Thus, experiments with *Mma* were abandoned and attempts were made to overexpress another family D polymerase within a natural host species, *Haloferax volcanii*.

5.11 *Haloferax volcanii* as a model organism

Haloferax volcanii (*Hvo*) is an obligate halophile, isolated from the Dead Sea that can be grown on a simple defined media without the need for specialist equipment or extensive training (Elazari-Volcani, 1943). Unlike *Mma*, *Hvo* is an aerobe and thus does not require anaerobic growth conditions making it easier to work with. It is an obligate mesophile and has a generation time of 3 hours when grown in liquid media and 5 days when grown on agar plates, at 45°C (Hartman *et al* 2010). *Hvo* has a stable genome and efficient natural homologous recombination properties (Leigh *et al.*, 2011) making it a desirable model organism.

As such, many genetic tools have been developed for *Hvo* and shuttle vectors, inducible promoters, protein purification protocols, transformation protocols, a gene knock-out strategy, and genetic and physical maps are all available (Bitan-Banin *et al* 2003; Hartman *et al* 2010; Allers *et al.*, 2004; Leigh *et al* 2011). Thus, *Hvo* was chosen as a model organism to overexpress and purify a family D polymerase. It was anticipated, that despite being aerobic, the high salt concentration required for *Hvo* growth would reduce the solubility of O₂ in water and thus prevent oxygenisation and thus the Fe-S cluster would remain intact.

5.12 Overexpression of *Haloferax volcanii* Pol D in *Haloferax volcanii*

Hvo strains, SMH610 (*Hvo* DP1) and SMH618 (*Hvo* Pol D), were supplied by Dr Stuart MacNeill from the University of St Andrews (Table 5.2). The strains contain chromosomally located Pol D genes that are wild type except for the presence of purification tags at their termini. Both strains were constructed by inserting the tagged Pol D genes onto an integrative shuttle vector, pTA131, and using them to replace the wild type Pol D in strain H53 via the “pop-in/pop-out” technique (Figure 5.11) (Bitan-Banin *et al.*, 2003; Allers *et al.*, 2004). *Hvo* strain H53 has two non-essential genes deleted: *pyrE2* and *trpA* (Allers *et al.*, 2010). *PyrE* encodes an enzyme involved in uracil synthesis while *TrpA* encodes an enzyme involved in tryptophan synthesis (Allers *et al.*, 2004). Thus, *Hvo* strain H53 is dependent on the presence of uracil and tryptophan in the media for growth (Allers *et al.*, 2010).

The pTA131 integrative shuttle vector has a multiple cloning site (MCS) located within the *lacZ* gene that is under the control of a *lac* promoter. pTA131 contains a copy of the *pyrE2* gene and thus restores the uracil synthesising properties of *Hvo* strain H53 when the vector is integrated. Thus when transformed cells are grown on media without uracil and can only proliferate if the plasmid integrates into the chromosome, thus providing an efficient selection method (Figure 5.11).

Strain	Gene used to select for gene integration	Replaced genes	Tag
SMH610	<i>PyrE2</i>	<i>Hvo</i> DP1	His (C terminal)
SMH622	<i>PyrE2</i>	<i>Hvo</i> DP1 <i>Hvo</i> DP2	His (C terminal) Strep (N terminal)

Table 5.2 *Haloferax* strains and their associated properties. Strains were constructed by Dr Søren Overballe-Petersen from the University of Copenhagen. Both strains require tryptophan for growth.

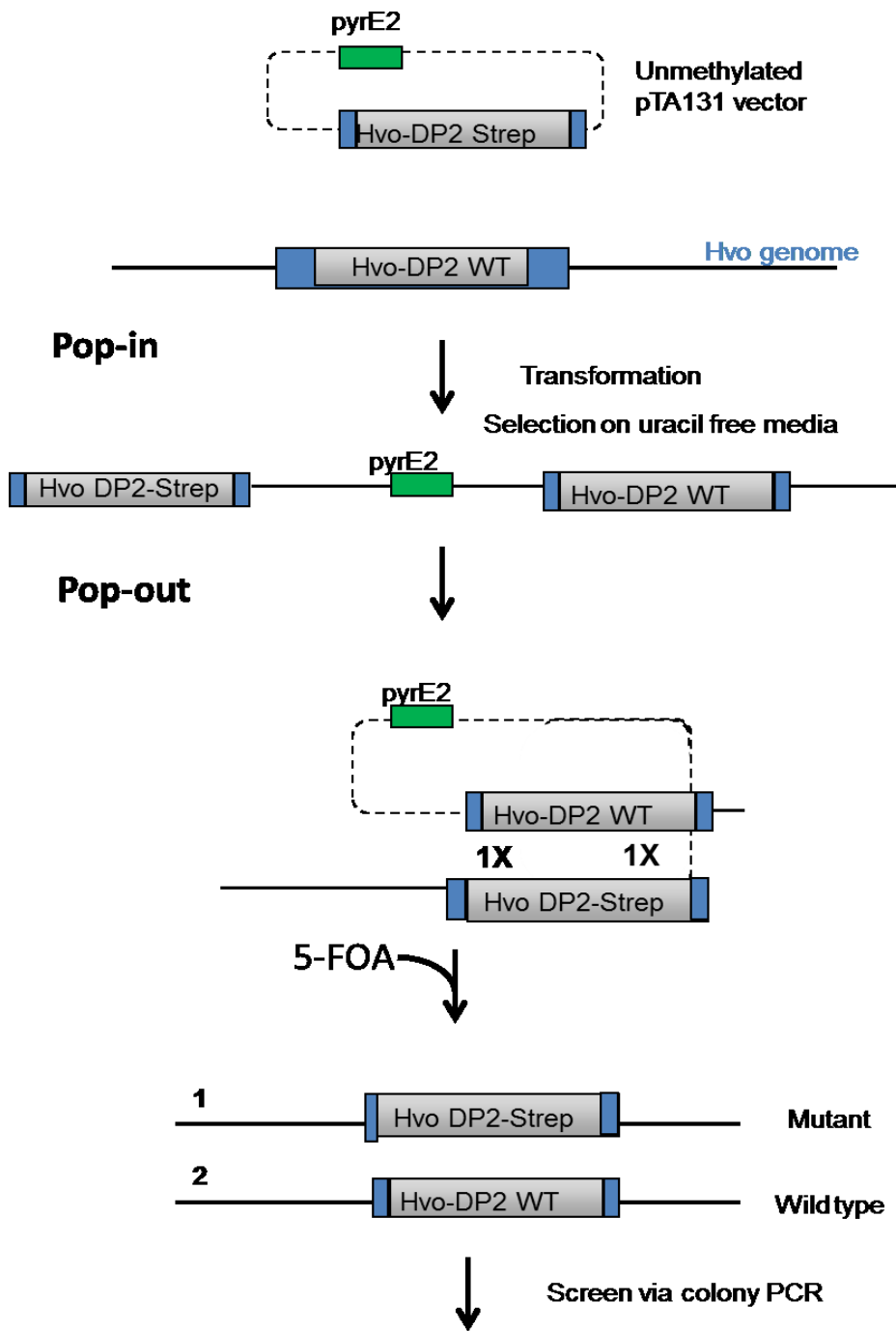


Figure 5.11 The pop-in/pop-out cloning system. The plasmid recombines into the chromosome by a crossover in the homologous sequence. The PyrE2 gene enables the cell to propagate without uracil. Transfer to media with 5-FOA and uracil selects for loss of the plasmid. Depending on the position of the second crossover, either “wild type” or mutant cells will arise. The blue areas represent homologous sequences.

5.12.1 Purification of *Haloferax volcanii* (*Hvo*) Pol D

SMH610 and SMH622 were grown in *Hvo* growth media supplemented with tryptophan (both strains have the *trpA* gene deletion and thus require supplementation with tryptophan for growth) until an O.D.₆₀₀ of 2.0 was obtained. The cells were then pelleted via centrifugation and subjected to sonication prior to centrifugation to remove the cellular debris. Attempts were made to extract the desired, tagged, proteins from the supernatant using gravity flow columns containing the appropriate resin. SMH610 supernatant was passed over Ni-NTA agarose resin and washed 3 times. The resin was washed with a buffer containing a high concentration of imidazole to elute the His-tagged, *Hvo*-DP1, protein. SMH622 was initially passed over Strep-Tactin resin that was also washed 3 times. The Strep tagged, *Hvo*-Pol D, protein was then eluted from the resin using a low concentration of d-desthiobiotin. The eluate was then passed over Ni-NTA agarose in the same manner as described for SMH610. The collected eluate samples were analysed via SDS-PAGE and the observed protein bands (Figure 5.12) were identified via mass spectrometry (appendix).

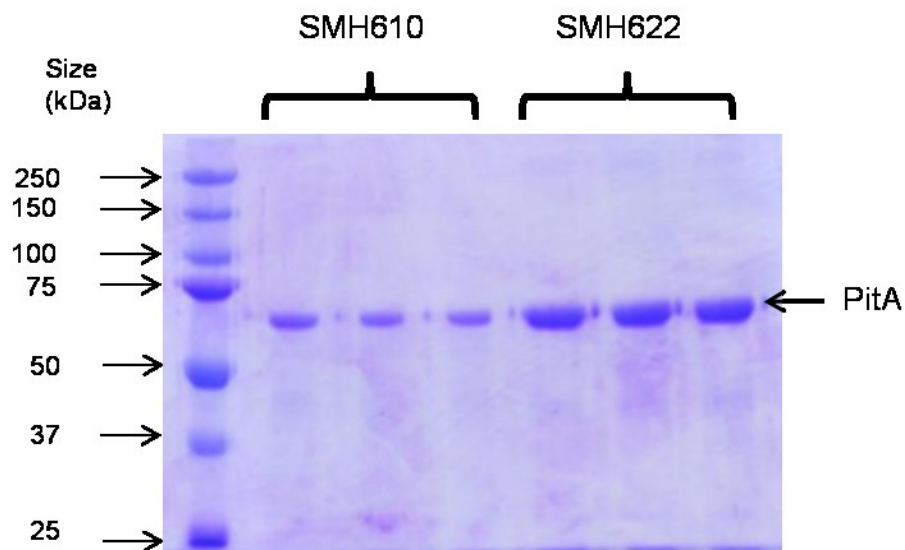


Figure 5.12 Denaturing (sodium dodecyl sulphate) polyacrylamide gel (12 %) showing samples purified from SMH610 and SMH622. Mass spectrometry results identified the single observed band, ~65 kDa, as *Hvo*-PitA. Lanes 1-3 contain repeats of the same *Hvo* extract and lanes 4-6 contain repeats of the same *Hvo* extract.

Analysis of the SDS-PAGE revealed a single band of ~ 65 kDa in both SMH610 and SMH622 protein samples (Figure 5.12). As *Hvo*-DP1 has a molecular weight of 57.8 kDa it was initially hypothesised that the band observed on the SDS-PAGE was *Hvo*-DP1. However, mass spectrometry results identified the protein as *Hvo*-PitA an essential, histidine-rich *Hvo* protein (Appendix).

Unfortunately, mass spectrometry did not detect any *Hvo*-DP1 protein. Additionally, SDS-PAGE analysis of SMH622 did not identify protein that corresponded to the size of *Hvo*-DP2 (~ 134.0 kDa). Thus the attempts to purify *Hvo*-DP1 and *Hvo*-Pol D from SMH610 and SMH622 were unsuccessful due to the low expression of the proteins (not detected on the SDS-PAGE) and contamination by the endogenous, histidine-rich *Hvo*-PitA protein. Although *Hvo*-Pol D contains purification tags it is in the correct chromosome position under control of its natural endogenous promoter, thus the inability to purify the protein suggests that the protein is expressed at low levels within the cell.

5.13 Cloning of *Haloferax volcanii* Pol D into pTA1392

Further attempts were made to overexpress *Hvo*-Pol D in *Haloferax volcanii* strain H1424 using the expression vector, pTA1392 (Figure 5.13). pTA1392 has a strong tryptophan-inducible promoter derived from the *Haloferax tnaA* tryptophanase promoter which allows high levels of protein expression (Allers *et al.*, 2010). pTA1392 allows a His and StrepII tag to be added to the proteins to assist in purification, depending on the restriction sites used for gene insertion. An *E.coli* origin and ampicillin resistance gene enables plasmid shuttling through *E.coli* while a copy of the *PyrE2* gene facilitates growth in the absence of uracil within Δ *PyrE2*-*Hvo* strains and can allow for selection of plasmid uptake.

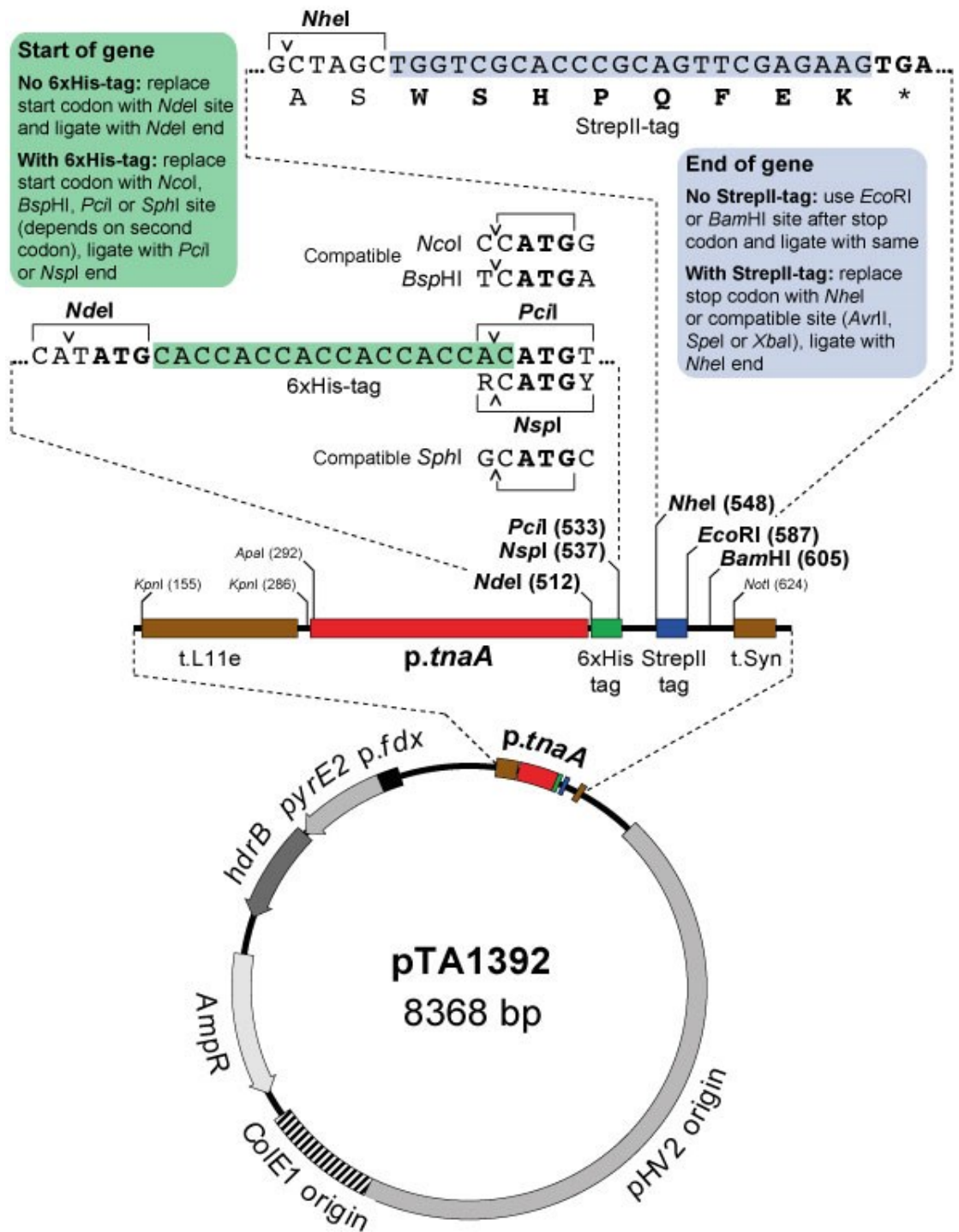


Figure 5.13 pTA1392 vector. Strong promoter, antibiotic resistance marker, uracil selection, thymidine selection, Figure produced by Kayleigh Wardle, Nottingham University.

Experiments were designed to insert *Hvo*-DP2 between *Pcil* and *NheI* and to insert *Hvo*-DP1, unidirectionally, at the *NheI* site of pTA1392 (Figure 5.13). The cloning was designed to ensure the heterodimeric *Hvo*-Pol D protein would possess a C-terminal Strep tag and an N-terminal His tag. Due to the difficulties of amplifying large amounts of GC-rich *Haloferax* DNA via PCR, the *Hvo*-DP2 and *Hvo*-DP1 subunits were PCR amplified and cloned into Topo sequencing vectors. This allowed large amounts of *Hvo*-DP2 and *Hvo*-DP1 to be obtained via plasmid isolation and restriction digest reactions to facilitate cloning into the pTA1392 vector

Attempts were made to clone *Hvo*-DP2 and *Hvo*-DP1 into pTA1392 and transform *E.coli* Top 10 cells. Initial attempts were made to transform Top10 *E.coli* cells with the ligated product. Unfortunately cloning *Hvo*-DP2 and *Hvo*-DP1 into pTA1392 using *E.coli* cells was unsuccessful. Restriction endonuclease reactions were performed to determine if the cloning had been successful, however, these reactions revealed DNA fragments of inexplicable sizes. Thus, attempts were made to get the recombinant DNA sequenced; however these reactions were also unsuccessful. Despite no obvious explanation regarding why the ligation was unsuccessful, correspondence with other laboratories revealed similar observations of “scrambled DNA” when pTA1392 *E.coli* was transformed (Figure 5.14).

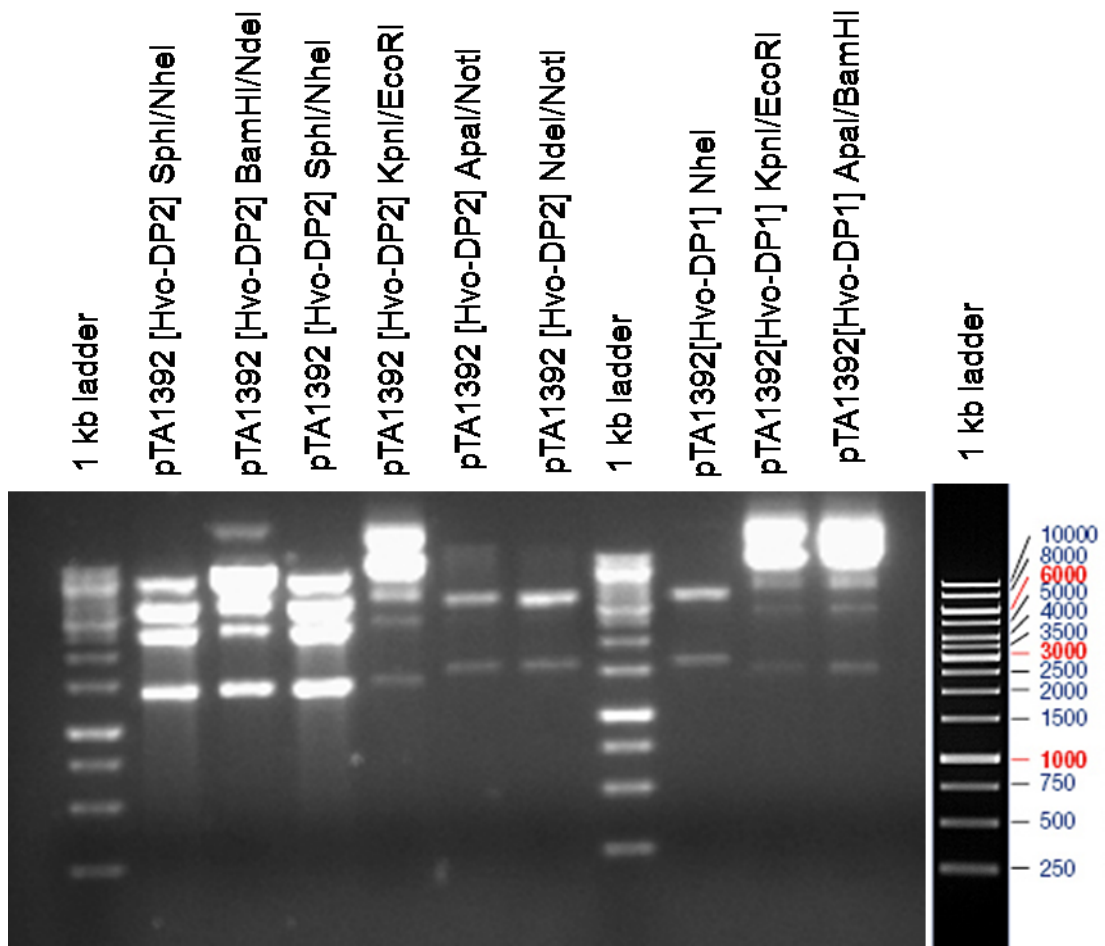


Figure 5.14 Agarose gel (1 %) electrophoresis showing restriction digested plasmid following transformation of Top10 cells with ligated pTA1392[*Hvo*-DP2] and pTA1392 [*Hvo*-DP1].

Thus, attempts were made to directly transform the *Haloferax volcanii* strain, H1412, with the ligation mixture. H1412 is a modified strain in which the endogenous PitA gene has been deleted and replaced with an ortholog from *Natronomonas pharaonis*. This PitA gene contains fewer histidine residues than the endogenous *Hvo*-PitA thus reduces the high levels of contamination of His tagged recombinant proteins purified previously observed (Figure 5.12) (Allers *et al.*, 2010). H1412 also has the *mrr* restriction endonuclease gene deleted, allowing direct transformation without the need to passage DNA through an *E.coli* *dam*⁻ strain.

Unfortunately attempts to transform H1412 with pTA1392 [*Hvo*-DP2] ligation product also proved unsuccessful. The transformation protocol did yield some

colonies, however sequencing results revealed that they did not contain the desired gene inserts. Due to the design of the cloning, it was necessary to insert the large, DP2, subunit prior to cloning the small, DP1, subunit in order to successfully clone both subunits into the vector. The DP2 subunit of archaeal Pol D is notoriously difficult to clone; for reasons currently unknown.

5.14 Discussion

This chapter details attempts to overexpress family D polymerases within a natural archaeal host organism and purify the polymerase with an intact Fe-S cluster. Unfortunately, attempts to overexpress *Mja*-Pol D in *Mma* and purify the protein anaerobically proved unsuccessful due to a low level of protein expression and contamination by endogenous *Mma* proteins. Regrettably, the availability of *Mma* shuttle vectors is limited and only two promoters are currently available for selection: Phmva and pMcr (used for antibiotic marker). Thus, it was not possible to clone the genes into another vector with a stronger, inducible promoter in order to improve protein expression. Additionally the requirement for anaerobic growth conditions made it tricky to up-scale cell harvesting to obtain more protein. Therefore attempts to overexpress *Mja*-Pol D in *Mma* were halted, and attention was paid to overexpress *Hvo*-Pol D in *Hvo*.

Attempts to overexpress *Hvo*-Pol D in *Hvo* also proved unsuccessful due to the low levels of protein expression and contamination by *Hvo*-PitA. Initial strains contained *Hvo*-Pol D and *Hvo*-DP1 cloned onto the chromosome and therefore was under the control of the natural promoter and was not inducible. As Pol D is expressed at low levels within the cell, this technique was not suitable for overexpression.

Thus, attempts were made to clone *Hvo*-Pol D into an improved shuttle vector, pTA1392, which contains a strong, inducible promoter. It was hoped that cloned genes could be expressed in H1424 which has the endogenous PitA gene deleted reducing contamination and facilitating purification of the protein. Unfortunately, attempts to clone the *Hvo*-DP2 and *Hvo*-DP1 subunits into pTA1392 proved unsuccessful. Nevertheless, this proposed strategy has the potential to be successful. Thus, future work should continue to attempt to

clone the *Hvo*-Pol D genes into pTA1392, perhaps using different restriction endonuclease sites.

In order to fully characterise family D polymerases and gain understanding of their role *in vivo*, it is critical to carry out studies with enzymes with the correct metallo-status. Thus, attempts to overexpress and purify Pol D with an intact Fe-S cluster are on-going. Dr Javier Abellón Ruiz (Newcastle University) has explored the use of yeast as an expression host for *Pfu*-Pol D using vectors supplied by Peter Burgers (Washington University, St. Louis). Although this system has been successful in expressing Fe-S containing yeast DNA polymerases, so far it has been unsuccessful for the overexpression of family D polymerases.

Dr Javier Abellón Ruiz has also attempted to overexpress family D polymerases in *E.coli* strains purported to enhance the Fe-S cluster synthesis. Unfortunately the use of these *E.coli* strains has made expression considerably worse. Thus, Dr Ruiz is currently focussing on improving the expression system available for *Mma*. Attempts are being made to replace the *Phmva* promoter with a stronger, inducible promoter.

It is hoped that such improvements will facilitate overexpression of family D polymerases with an intact Fe-S cluster. It is anticipated that such a polymerase will assist in obtaining a crystal structure for Pol D, by stabilising the correct protein fold, as it is hypothesised that the lability and complexity of Fe-S clusters have inhibited previous attempts. It is expected that such structures will aid in the identification of a uracil binding pocket within Pol D.

The availability of a family D polymerase with an intact Fe-S cluster will facilitate characterisation of the novel polymerase. Biochemical methods used to determine how the polymerase interacts with DNA and, in particular, the damaged base uracil and genetic approaches, aimed at probing the function of the polymerase and its key metal ion cofactors in archaeal cells are expected to provide evidence that replication is performed by Pol-D in most archaea, inferring the use of different replicative polymerases in the three domains of life. Such a conclusion raises wider issues concerning the origin(s), evolution, and

benefits of different DNA replication enzymes and replicative strategies, widely debated and poorly understood concepts critical for understanding how life arose.

**Characterisation of bacterial DNA Polymerase II
from *Rhodothermus marinus* and
Sulfurihydrogenibium sp. YO3AOP1**

6.1 Background

DNA polymerases constitute a large class of enzyme that is subdivided into distinct families based on sequence homology (A, B, C, D, X, Y, and RT) (Ito and Braithwaite, 1991; Bebenek and Kunkel, 2004; Joyce and Benkovic, 2004). Many DNA polymerases are utilised in molecular biology research and are used for many routine applications including: PCR, reverse transcriptase PCR, DNA sequencing and molecular cloning (Hamilton *et al.*, 2001). However, the application of DNA polymerases extends beyond such routine uses and many polymerases with unusual or unique properties, including strand displacement and genetic recombination, have been commercialised to initiate highly specific reactions (Hamilton *et al.*, 2001).

Despite the large selection of DNA polymerases that are currently available the characterisation of new polymerases, as well as the genetic modification of existing enzymes, is vital to identify polymerases with novel properties (Henry and Romesberg, 2005; Kranaster and Marx, 2010). The identification of new, unique polymerases has the potential to enable the development of new molecular biology techniques. This chapter describes the characterisation of two bacterial, family B, DNA polymerases and investigates their potential application in biotechnology.

Family B DNA polymerases are normally highly processive, replicative polymerases with a high fidelity rate and 3'-5' exonuclease activity. However, a unique family B polymerase, DNA Pol II, that has been well characterised in *E.coli*, is known to play a role in translesion synthesis (TLS) (Paz-Elizur *et al.*, 1996; Becherel and Fuchs, 2001; Wang and Yang, 2009), a feature normally associated with family Y polymerases (Ohmori *et al.*, 2001; Yang and Woodgate, 2007). *E.coli* DNA Pol II is able to extend primers past mutagenic DNA, and is involved in nucleotide excision repair (Berardini *et al.*, 1999) and replication restart following UV exposure (Rangarajan *et al.*, 1999). However, unlike the family Y polymerases, *E.coli* DNA Pol II maintains 3'-5' exonuclease activity and the efficient DNA polymerase activity normally associated with family B polymerases.

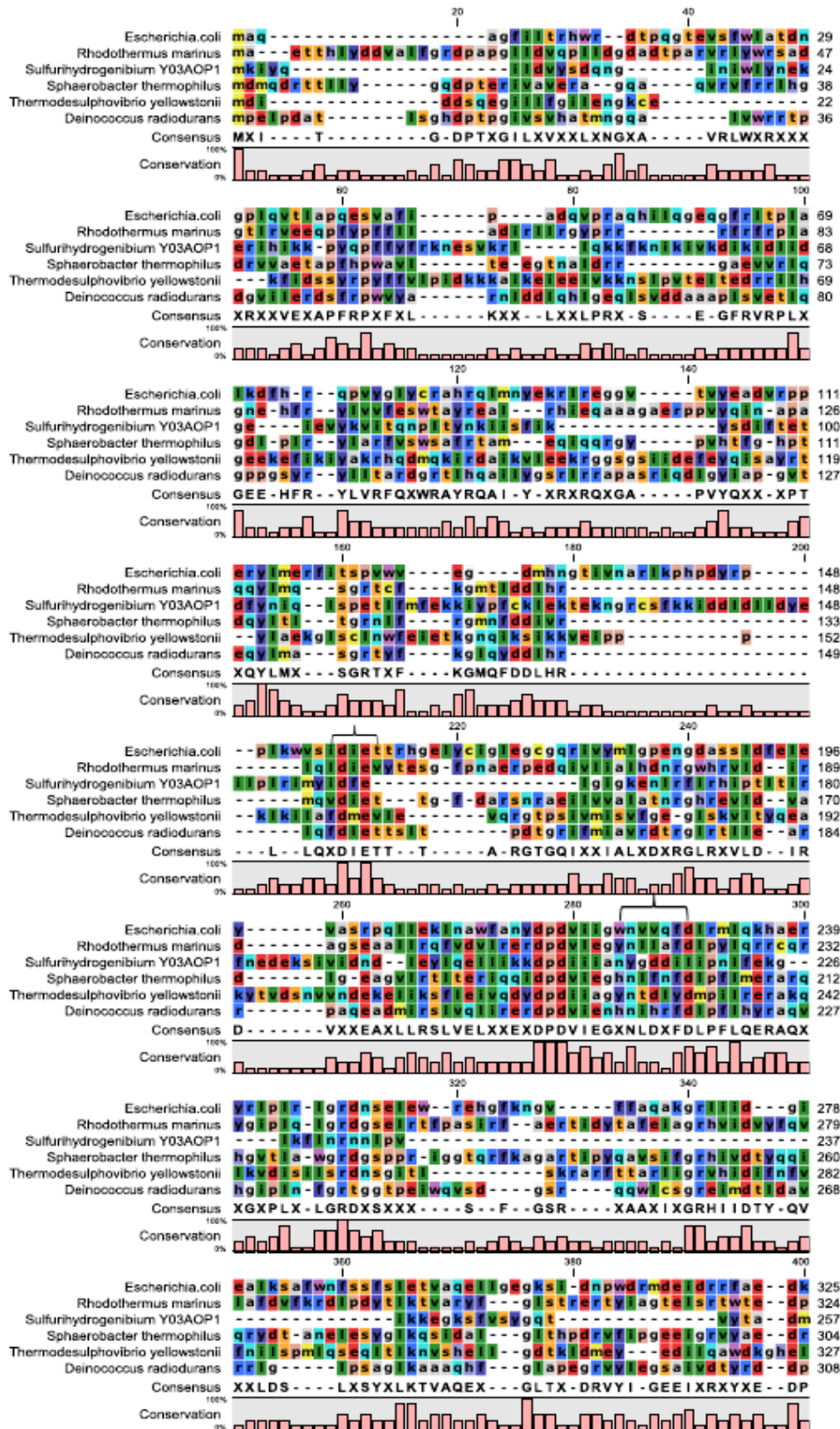
The unique TLS properties found in *E.coli* DNA Pol II combined with its high fidelity rate (Banach-Orlowska *et al.*, 2005), 3'-5' exonuclease activity and strong polymerase activity, may have potential applications in biotechnology. However, as *E.coli* is an obligate mesophile, purified *E.coli* DNA Pol II is sensitive to high temperatures and has a reduced rate of activity when heated (activity is reduced by more than 50 % after heating to 45°C for 10 minutes) (Kornberg and Gefter, 1971). Therefore *E.coli* DNA Pol II is expected to be unsuitable in PCR-based applications. Thus, database searches were performed to identify bacterial DNA Pol II's from thermostable organisms with a high degree of homology to *E.coli* DNA Pol II. These searches identified two DNA Pol II enzymes from the thermostable organisms: *Rhodothermus marinus* (*R.marinus*) and *Sulfurihydrogenibium* sp. *YO3AOP1* (*S.YO3*) (maximum growth temperatures 80°C and 78°C, respectively) (Andresson and Fridonsson, 1994; Silva *et al.*, 1999; Nakagawa *et al.*, 2005; Bjornsdottir *et al.*, 2006). These enzymes were cloned, overexpressed and purified and their properties investigated to determine if they had characteristics desirable for biotechnology applications.

6.2 Sequence of *R.marinus* and *S.YO3* DNA Pol II

Thermostable bacterial DNA Pol II sequences were found in initial BLAST searches using *E.coli* DNA Pol II as a template. Some of the polymerases found are shown in Figure 6.1. Unfortunately, DNA Pol II homologues appear to be rare in nature (at least using BLAST searches) and they were not identified in many of the highly thermostable species often used as sources of thermostable enzymes e.g. *thermus*, *thermotoga* and many of the *aquifex* species. Further attempts to identify DNA Pol II sequences from thermostable species was performed using the thermophilic DNA Pol II sequences from *R.marinus* and *S.YO3* to initiate BLAST searches, however, these searches failed to reveal homologues in very thermophilic bacteria.

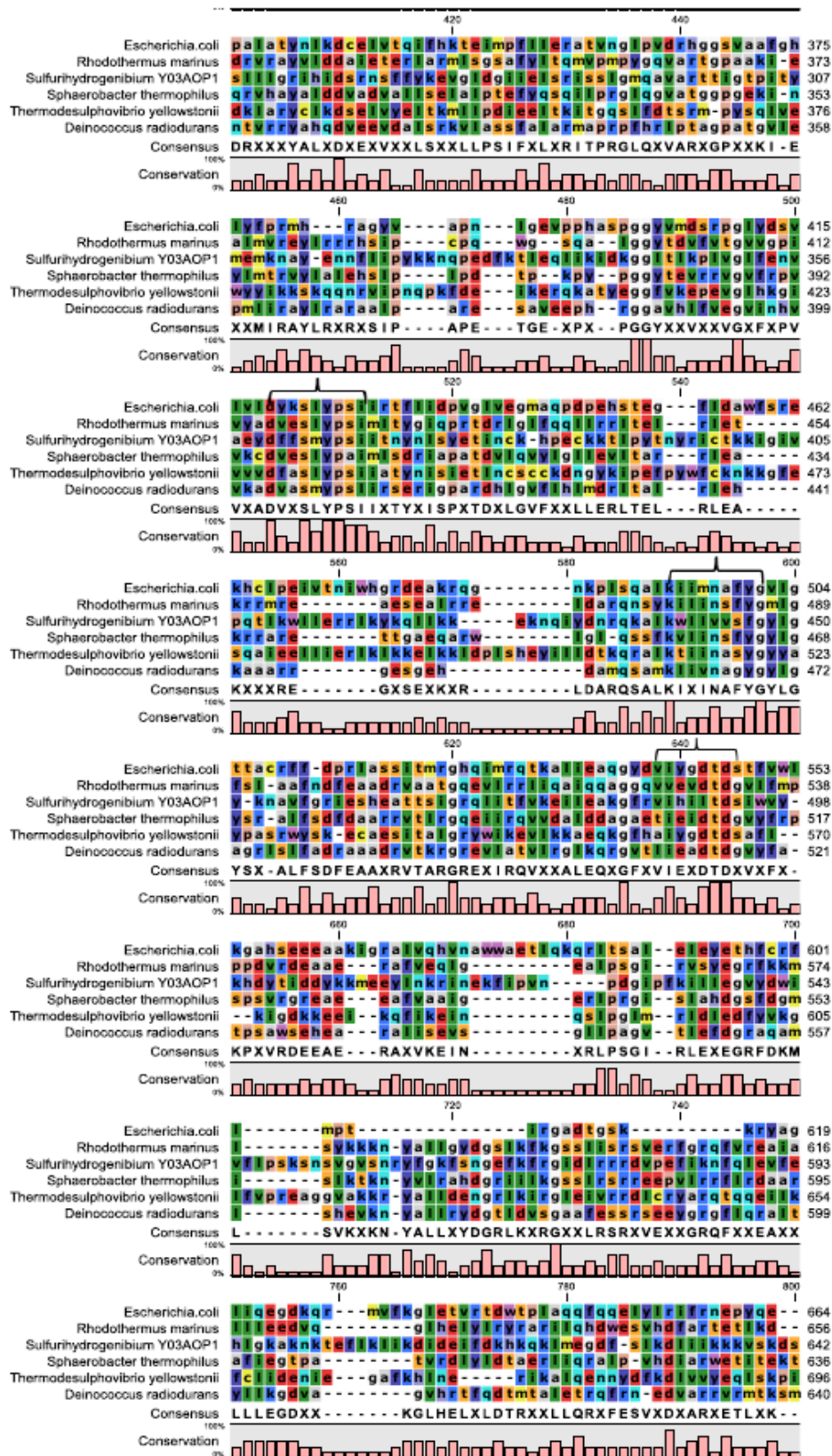
Sequence alignments of the thermostable DNA Pol II proteins revealed some conserved regions as indicated on Figure 6.1. Analysis of the sequences identifies the highly conserved aspartic acid and glutamic acid residues within the exonuclease and polymerase regions (Figure 6.1). These aspartic acid

residues are known to play an important role in the binding of metal ions, essential for processive polymerase activity and proofreading exonuclease activity (Steitz.,1999). However, *S.YO3* appears to be missing an aspartic acid residue in both a polymerase and exonuclease domain that has been highly conserved in the other five polymerases (Figure 6.1).



Exonuclease I

Exonuclease II



Polymerase active site (region II)

dNTP recognition

Polymerase active site (region I)

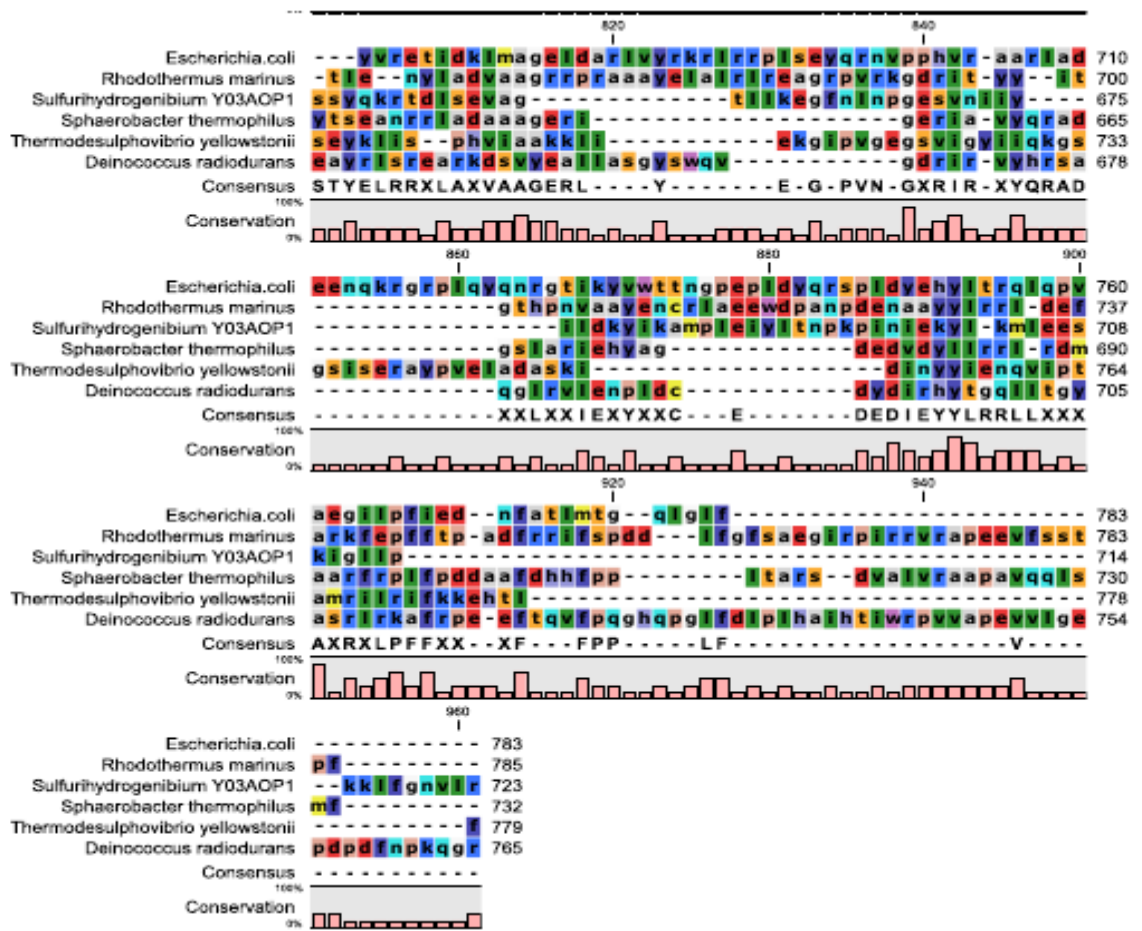


Figure 6.1 Alignment of DNA Pol II sequences identified in BLAST searches using *E.coli* DNA Pol II. The conserved polymerase, exonuclease and dNTP recognition domains are highlighted. Image created using CLc sequence viewer, version 6.

6.3 Cloning of *R.marinus* and *S.YO3* Pol II

R.marinus and *S.YO3* DNA Pol II gene sequences were amplified with appropriate primers using the PCR and inserted into pET28a vectors using specific restriction endonuclease sites (Table 2.4). The insertion site was chosen to ensure a His-tag, (His)₆, was located at the N-terminus of the Pol II protein sequences. Due to the high GC content of the *R.marinus* genome (~ 64 mol %)(Alfredsson *et al.*, 1988) the initial PCR amplification of the Pol II gene proved to be challenging, thus “touchdown” PCR was used (Don *et al.*, 1991). This PCR technique utilises initial high annealing temperatures, to improve the specificity of the primers, and temperatures are reduced by 0.5°C in each subsequent amplification round. The initial high annealing temperatures ensure that the target fragment out-competes other non-specific sequences to which the primers may initially bind at lower temperatures (Figure 6.2).

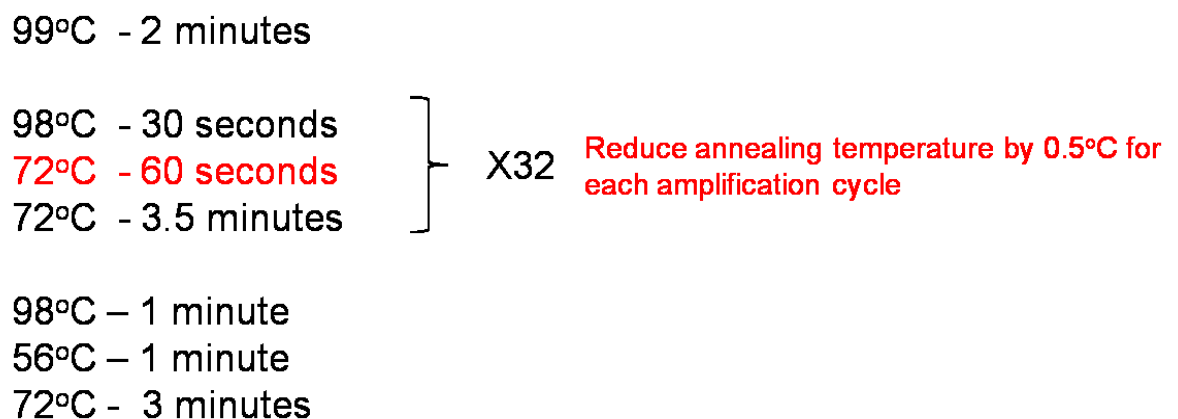


Figure 6.2 Thermocycler conditions used in touchdown PCR. All reaction mixtures involved in touchdown PCR contained 5 % dimethylsulphoxide (DMSO).

6.4 *R.marinus* and *S.YO3* Pol II protein purification

Plasmids containing the His-tagged *R.marinus* and *S.YO3* Pol II genes were used to transform *E.coli* BL21 (DE3) pLysS and the expressed proteins were purified using a heat step (70°C) followed by Ni-NTA agarose or a His Trap column (Figure 6. & Figure 6.3). The His-tagged proteins were eluted from the Ni-NTA agarose using increasing concentrations of imidazole and the eluted proteins were analysed via SDS-PAGE (Figure 6.3 & Figure 6.4). The protein purification protocols were efficient, with very few endogenous *E.coli* proteins remaining in the final purified samples (Figure 6. & Figure 6.4). Protein bands of ~ 70 kDa were excised from the SDS gels and their identity was confirmed using mass spectrometry (Figure 6.5 & Figure 6.6).

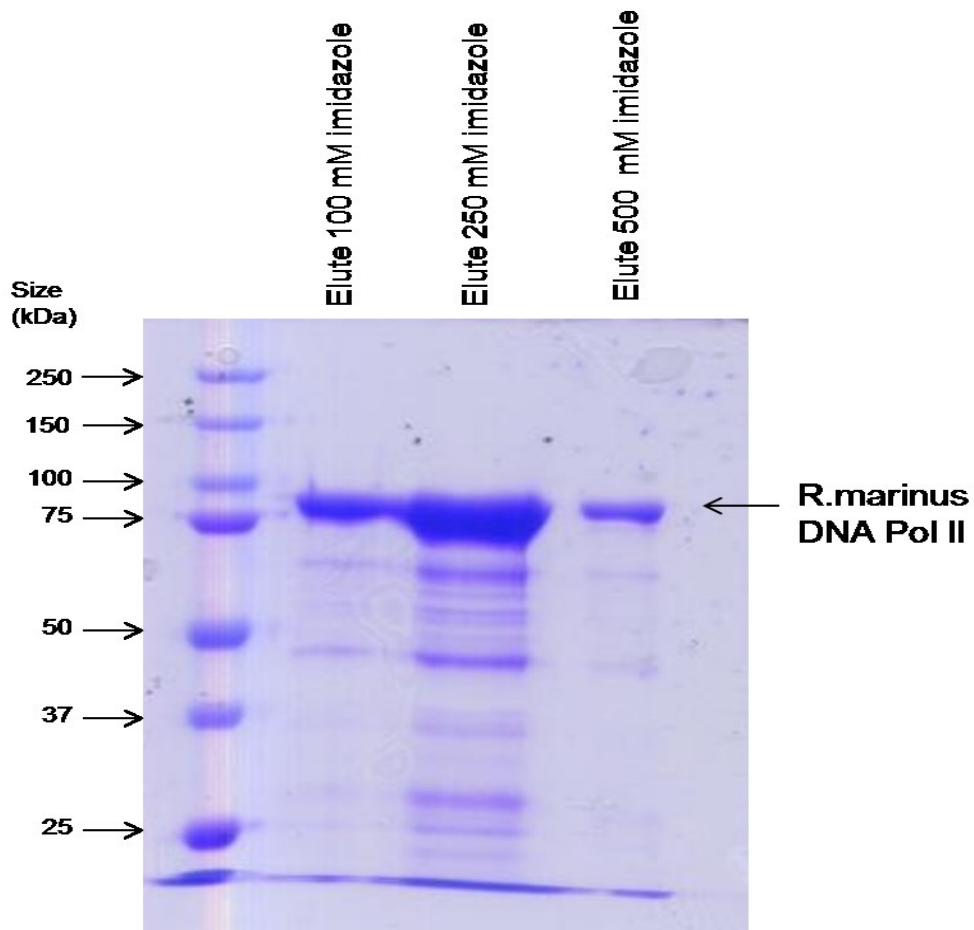


Figure 6. 3 Denaturing sodium dodecyl sulphate polyacrylamide gel (10 %) showing purified *R.marinus* DNA Pol II. The pol II protein is visualised as a single band of ~75kDa.

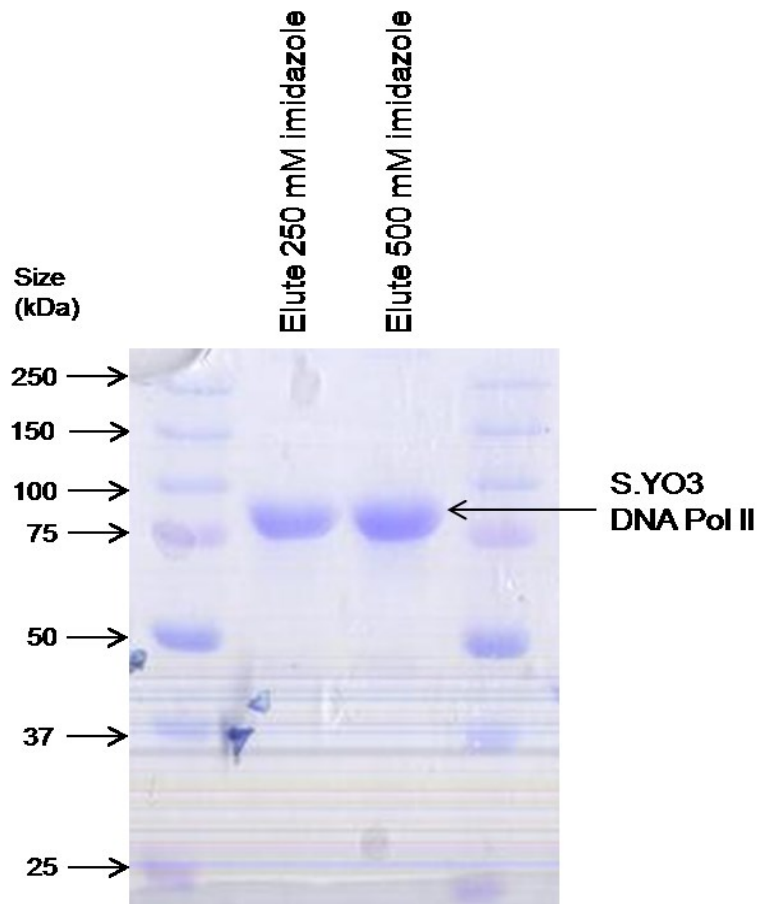


Figure 6. 4 Denaturing sodium dodecyl sulphate polyacrylamide gel (10 %) showing the purified *S.YO3* DNA Pol II protein. The pol II protein is visualised as a single band of ~70kDa.

E.coli DNA Pol II and *E.coli* Pol II exo^- proteins were gifted by Wei Yang, Bethesda. The purified *E.coli* DNA Pol II proteins provided were used in all the experiments shown in this chapter.



Mascot Search Results

User : tf2013
 Email :
 Search title : Ultraflex_ProteinID
 MS data file : 13229323905421996.mgf
 Database : NCBI nr 20101130 (12348165 sequences; 4221604734 residues)
 Taxonomy : Bacteria (Eubacteria) (7020473 sequences)
 Timestamp : 22 Jan 2013 at 11:57:44 GMT
 Enzyme : Trypsin
 Fixed modifications : [Carbamidomethyl \(C\)](#)
 Variable modifications : [Oxidation \(M\)](#)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 100 ppm
 Fragment Mass Tolerance : ± 0.5 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 10
 Protein hits : [gi|268316969](#) DNA polymerase B region [Rhodothermus marinus DSM 4252]

Select Summary Report

Format As [Help](#)
 Significance threshold p < Max. number of hits
 Standard scoring MudPIT scoring Ions score or expect cut-off Show sub-sets
 Show pop-ups Suppress pop-ups Require bold red
 Re-Search All queries Unassigned Below homology threshold Below identity threshold

1. [gi|268316969](#) Mass: 90533 Score: 1042 Matches: 10(10) Sequences: 10(10)
 DNA polymerase B region [Rhodothermus marinus DSM 4252]

Query	Observed	Mr (expt)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1	1087.6731	1086.6658	1086.6550	10.0	0	72	0.00042	1	U	R.LGLFQQLLR.R
2	1356.7002	1355.6929	1355.6721	15.3	0	100	9.5e-07	1	U	R.TIDYTAFEIAGR.H
3	1501.7665	1500.7592	1500.7402	12.7	1	77	0.00018	1	U	R.KFEFFFTPADFR.R
4	1533.7928	1532.7856	1532.7664	12.5	0	94	3.5e-06	1	U	R.YLVVFESWTAYR.E
5	1586.8748	1585.8675	1585.8464	13.3	0	120	8e-09	1	U	R.AFVEQLGEALPSGIR.V
6	1752.8512	1751.8439	1751.8380	3.40	0	107	1.5e-07	1	U	R.ILQHDWESVHDFAR.T
7	1984.0205	1983.0132	1983.0142	-0.48	0	125	1.9e-09	1	U	R.VEEQPFYPPFLLADIR.L
8	2137.0955	2136.0882	2136.1004	-5.71	0	96	1.5e-06	1	U	R.IFSPDDLFGFSAEGIRPIR.R
9	2236.9929	2235.9856	2236.0072	-9.65	0	99	6.1e-07	1	U	R.LAEEDWPANPDENAAYLR.R
10	2324.2239	2323.2166	2323.2423	-11.07	0	155	1.9e-12	1	U	R.EAIALLLLEEDVQGLHELYLR.Y

Figure 6.5 Mass spectrometry results identifying the large protein band highlighted on Figure 6.3 as *R.marinus* DNA Pol II. Mass spectrometry was performed by York University who provided the above report.



Mascot Search Results

User : tf2012
 Email :
 Search title : Ultraflex_ProteinID
 MS data file : 13229323905421547.mgf
 Database : NCBI nr 20101130 (12348165 sequences; 4221604734 residues)
 Timestamp : 30 Nov 2012 at 14:49:17 GMT
 Enzyme : Trypsin
 Fixed modifications : [Carbamidomethyl \(C\)](#)
 Variable modifications : [Oxidation \(M\)](#)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 100 ppm
 Fragment Mass Tolerance : ± 0.5 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 10
 Protein hits : [gi|188997124](#) DNA polymerase B region [Sulfurihydrogenibium sp. YO3AOP1]
 [gi|1628447](#) tipA [synthetic construct]

Select Summary Report

Format As	Select Summary (protein hits) ▾	Help
Significance threshold p<	0.05	Max. number of hits AUTO
Standard scoring	<input checked="" type="radio"/> MudPIT scoring <input type="radio"/>	Ions score or expect cut-off 0.05
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups <input type="radio"/>	Show sub-sets 0
		Require bold red <input checked="" type="checkbox"/>

Re-Search All queries Unassigned Below homology threshold Below identity threshold

1. [gi|188997124](#) Mass: 85366 Score: 639 Matches: 7(7) Sequences: 7(7)
 DNA polymerase B region [Sulfurihydrogenibium sp. YO3AOP1]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2	950.5814	949.5741	949.5709	3.43	0	56	0.029	1	U	R.HIPTLTIR.F
4	1300.7438	1299.7365	1299.7034	25.5	0	104	6.4e-07	1	U	K.EVGLDGIIELSR.I
5	1343.7683	1342.7610	1342.7285	24.2	0	77	0.0003	1	U	K.FIPVNPDGIFPK.I
6	1460.7917	1459.7845	1459.7460	26.4	0	87	3.5e-05	1	U	K.NFQLEVFELGK.A
7	1586.9229	1585.9156	1585.8868	18.1	0	96	3.3e-06	1	U	R.VIHILTDSIWVYK.H
9	1816.0054	1814.9981	1814.9666	17.4	0	127	2.4e-09	1	U	K.IDDLDLLDYEILPLR.I
10	1944.0864	1943.0792	1943.0615	9.08	1	94	4e-06	1	U	K.KIDLDLLDYEILPLR.I

Figure 6.6 Mass spectrometry results identifying the large protein band highlighted on Figure 6.4 as *S.YO3* DNA Pol II. Mass spectrometry was performed by York University who provided the above report.

6.5 Extension of DNA primer-templates by DNA Pol II enzymes

The ability of bacterial DNA Pol II to extend primer-templates was determined with the enzymes from *E.coli*, *R.marinus* and *S.YO3*. Two fluorescent labelled primer-templates were used in the experiment: 45T-Fluor and 45U-Fluor (Figure 6.7). 45U-Fluor contained uracil and was used to determine if the presence of uracil inhibited polymerisation.

```
45T-Fluor 5'-Flu-ACCGCGGGATATCGGCCCTTT-3'  
          3'TGGGCGCCCTATAGCCGGGAAATCCGTTCGTCCGAACAGAGGTAT-5'  
45U-Fluor 5'-Flu-ACCGCGGGATATCGGCCCTTT  
          3'UGGGCGCCCUAUAGCCGGGAAAUCCGUUCGUCCGAACAGAGGUAAU-5'
```

Figure 6. 7 Oligodeoxynucleotides used in primer-extension reactions. In 45U-Fluor all thymine residues are substituted with uracil.

Primer-template extensions were performed using 40 nM primer-template in the presence of Pol D reaction buffer (40 mM Tris-HCl [pH 9.0], 50 mM KCl, 10 mM MgCl₂, 10 mM DTT), 10mM dNTPs at 50°C and initiated with 200 nM of polymerase. Reactions were quenched at the timepoints indicated in Figure 6.8 and analysed using a 17 % acrylamide gel, visualised using a Typhoon scanner and ImageQuant software (Figure 6.8).

The extensions observed with *E.coli* and *R.marinus* Pol II were largely complete after 5 seconds whereas the *S.YO3* Pol II required 60 minutes to fully extend the primer (Figure 6.8). This observation suggests that *S.YO3* DNA Pol II is a less active polymerase than *E.coli* or *R.marinus* DNA Pol II. This may result from lower intrinsic activity i.e. the rate at which each dNTP is inserted or from decreased processivity. For all three enzymes, the rate of extension in the uracil containing templates (45U-Fluor) is similar to that observed in the control (45T-Fluor) templates, indicating that uracil does not inhibit the rate of polymerisation (Figure 6.8).

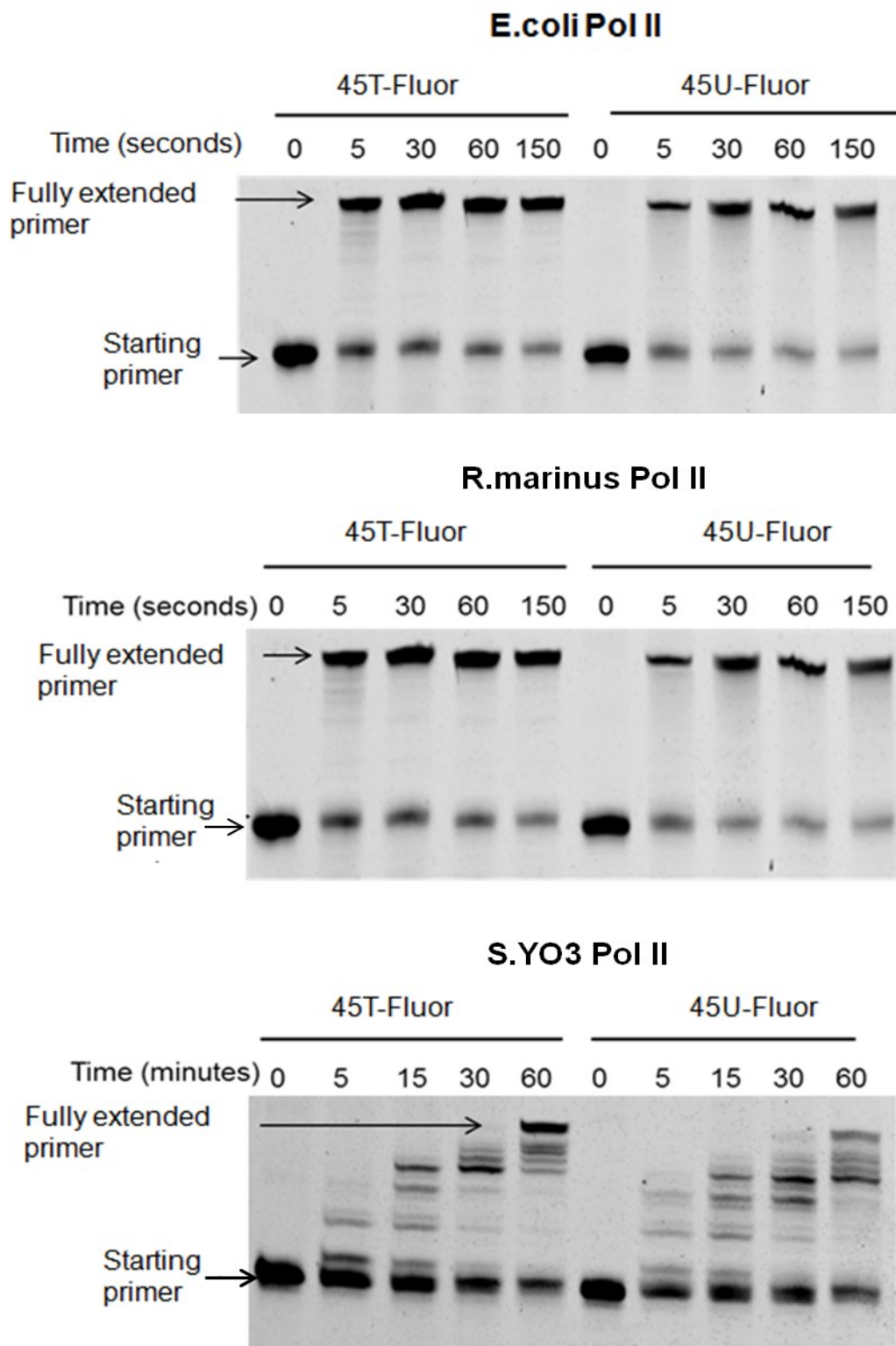


Figure 6.8 Primer-template extensions with bacterial DNA Pol II observed.

6.6 DNA Pol II 3'-5' exonuclease assays

Assays were performed to determine if the DNA Pol II's possessed the proofreading, 3'-5' exonuclease activity, normally associated with family B polymerases. Reactions were performed using both single stranded DNA and complementary primer-template DNA (Figure 6.9). Reactions were initiated with 200 nM of *E.coli* DNA Pol II (Figure 6.10) 200 nM *R.marinus* DNA Pol II (Figure 6.11), 400 nM *S.YO3* Pol II (Figure 6.) or 200 nM *Pfu*-Pol B (control). All reactions were performed at 50°C, for the times indicated on figures 6.10-6.12 and analysed using a 17 % denaturing polyacrylamide gel visualised using a Typhoon scanner and ImageQuant software.

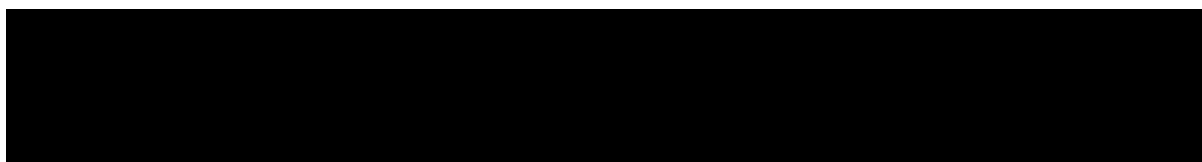


Figure 6. 9 Oligodeoxynucleotides used in 3'-5' exonuclease reactions. “ss” represents single stranded and “Fluor” represents fluorescein.

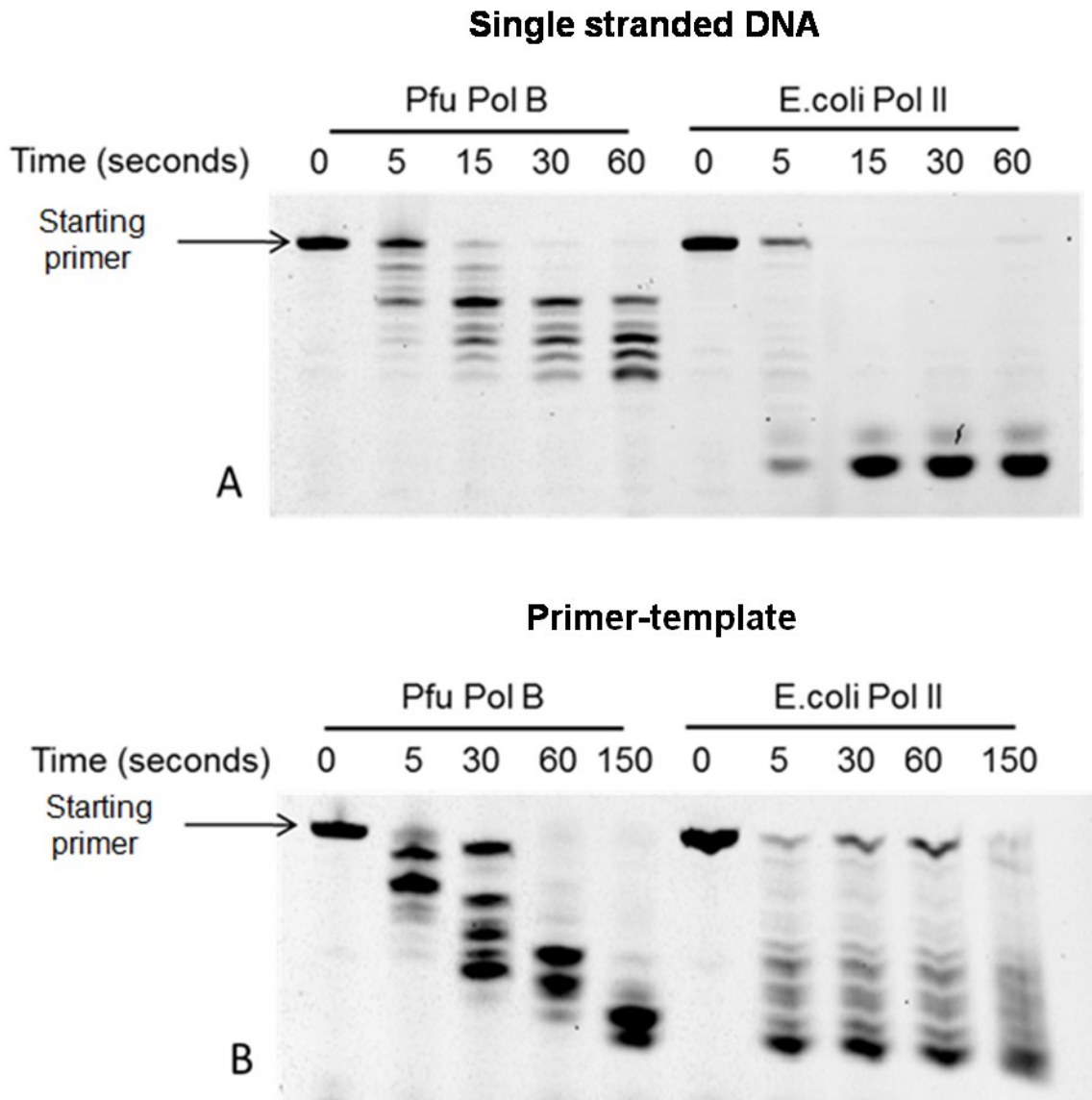


Figure 6.10 3'-5' exonuclease reactions initiated with 200 nM *Pfu*-Pol B (control) or 200nM *E.coli* DNA Pol II. A) Single stranded DNA template B) Primer-template DNA.

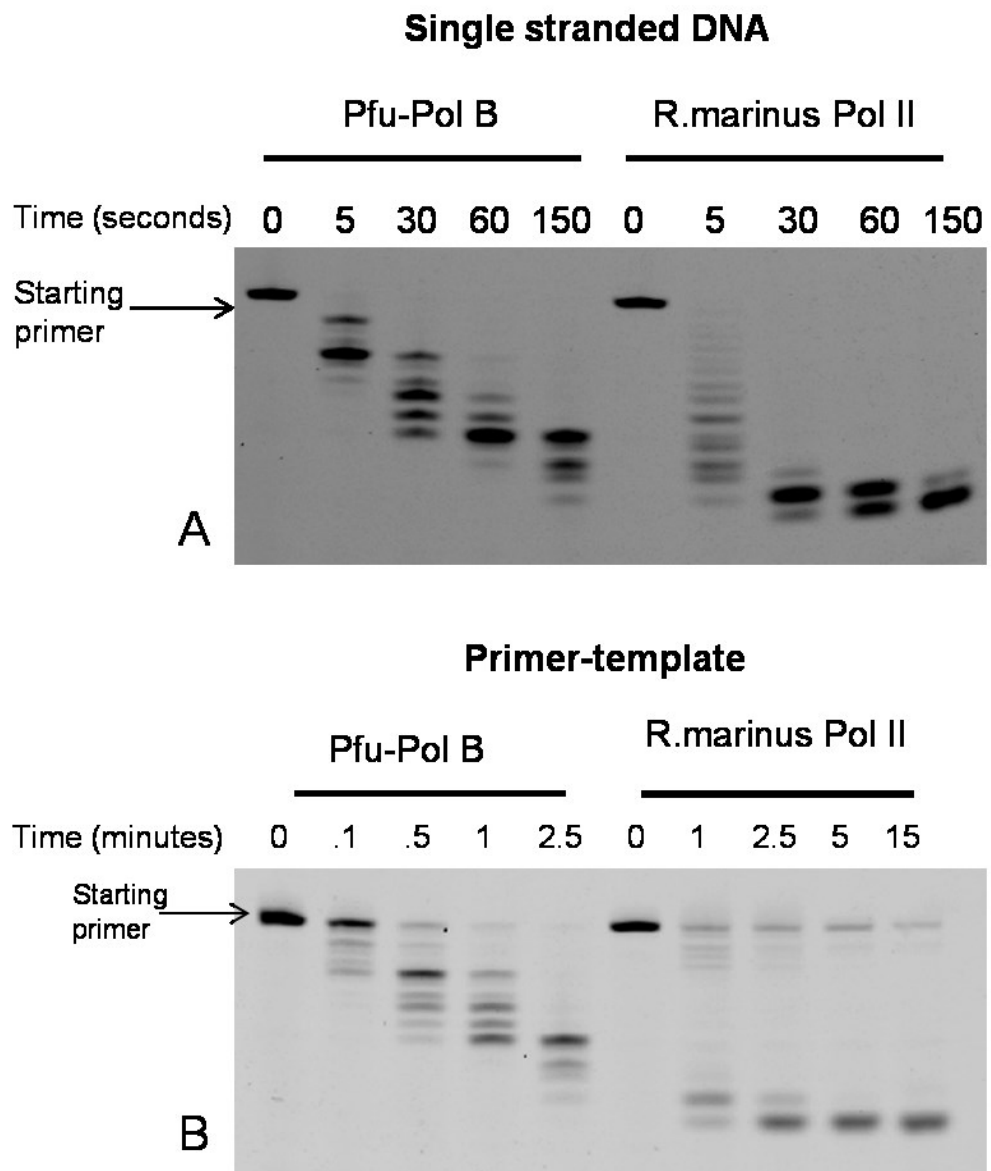


Figure 6.11 3'-5' exonuclease reactions initiated with 200 nM *Pfu*-Pol B (control) and 200nM *R.marinus* DNA Pol II. A) Single stranded DNA template B) Primer-template DNA.

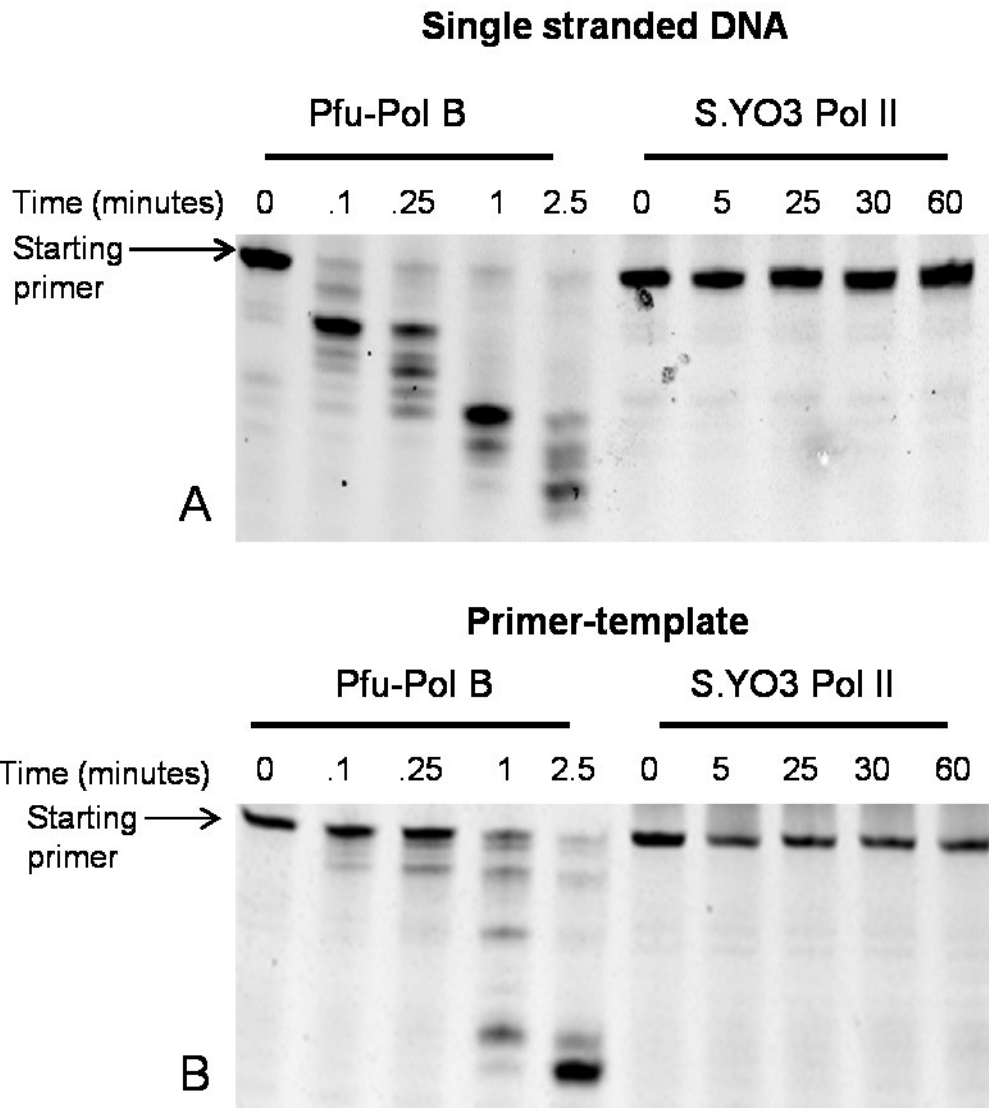


Figure 6.12 3'-5' exonuclease reactions initiated with 200 nM *Pfu*-Pol B (control) and 400 nM *S.YO3* DNA Pol II. A) Single stranded DNA B) primer-template DNA.

E.coli and *R.marinus* DNA Pol II enzymes degraded the single stranded DNA and the primer-template DNA (Figure 6.12). For both enzymes, the rate of exonucleolysis was faster with single stranded DNA compared to primer-template material (Figure 6.12). *R.marinus* Pol II degraded the single stranded DNA so that no starting material remained, after 5 seconds (Figure 6.12). However, with primer-template, some starting material still remained after 15 minutes (Figure 6.12). A similar observation was made with *E.coli* Pol II; complete exonucleolysis of single stranded DNA within 15 seconds, starting

primer still visible in the reaction containing primer-template after 60 seconds (Figure 6.10).

As the primer-template sequence used is fully base-paired it is a poor substrate for exonucleolysis. Therefore the enzymes would be expected to degrade the single stranded substrate more efficiently (Beese and Steitz, 1991). The presence of exonuclease activity within *E.coli* and *R.marinus* DNA Pol II enzymes was expected due to the highly conserved sequence within the exonuclease domain (Figure 6.1) which is known to be responsible for exonuclease activity (Beese and Steitz, 1991). As stated in the introduction, *S.YO3* does not possess these highly conserved residues (Figure 6.1) therefore, it was not surprising that *S.YO3* did not exonuclease either the single-stranded or primer-template DNA (Figure 6.12)

Reactions carried out with *S.YO3* DNA Pol II were performed over a longer time course (60 minutes) and contained a higher concentration of enzyme than the reactions performed with the *E.coli* and *R.marinus* DNA Pol IIs. However, no exonuclease activity was detected in any reaction initiated by *S.YO3* Pol II (Figure 6.12). This observation suggests that *S.YO3* Pol II does not possess any pronounced 3'-5' exonuclease activity.

6.7 RNA polymerase activity of DNA Pol IIs

Primer-extension reactions were performed to investigate if *E.coli*, *R.marinus* or *S.YO3* Pol II enzymes possessed RNA polymerase activity, an activity that has potential use in preparing RNA sequences. Reactions contained 20 nM of DNA primer-template (DNA 45-Fluor) (Figure 6.13), 10 mM NTP's and 200 nM enzyme. Reactions were quenched after 0, 15, 30, 60 and 90 minutes. A positive control was performed that contained dNTPs rather than NTPs. Reactions were performed at 50°C and analysed as discussed above (Section 6.5).

DNA 45-Fluor 5'-Flu-ACCCGCGGGATATCGGCCCTTT -3'
3'-TGGGCGCCCTATAGCCGGGAAATCCGTTTCGTCGGAACAGAGGTAT-5'

Figure 6. 13 Fluorescent labelled DNA primer-template used in RNA polymerase assays.

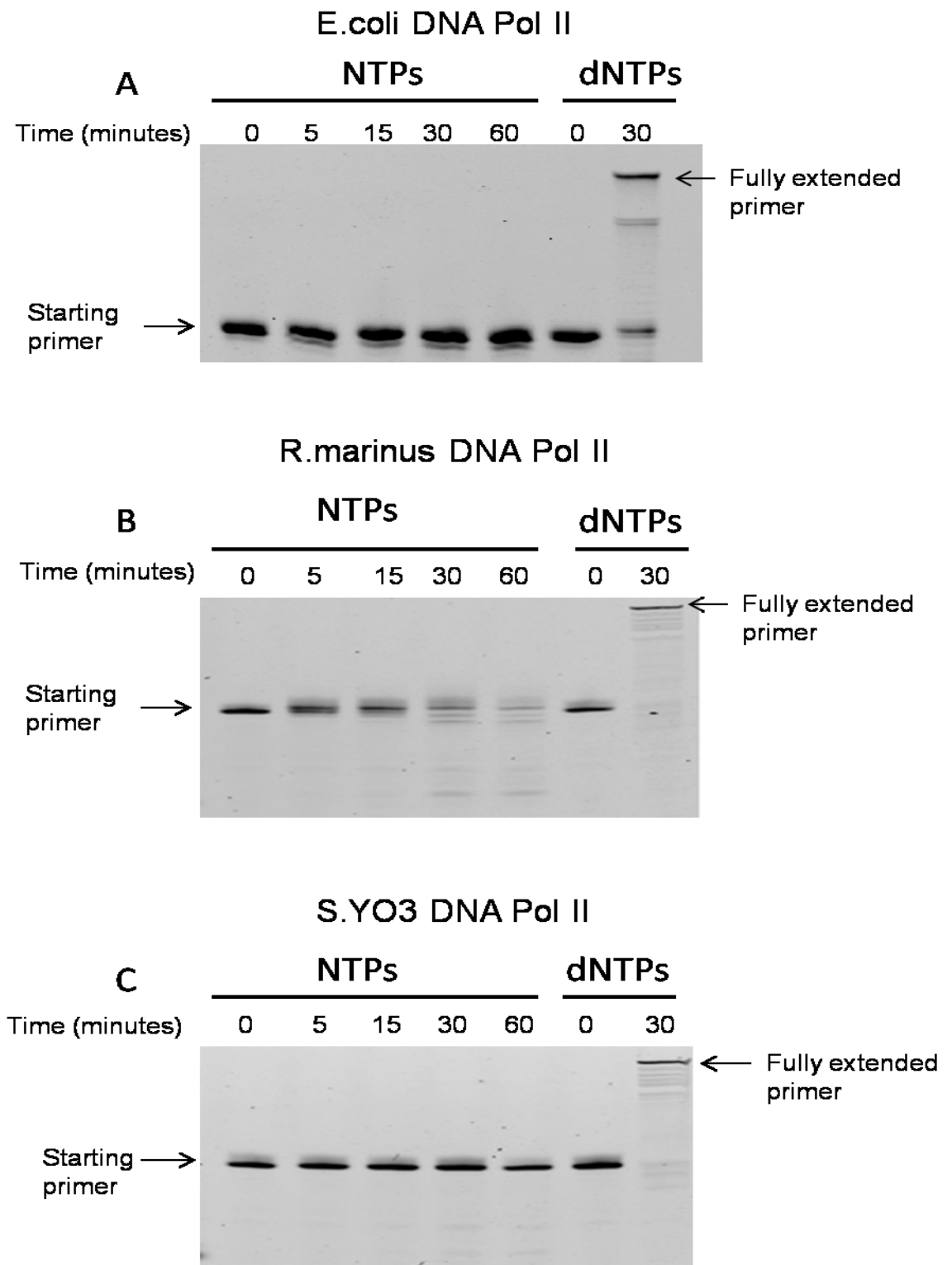


Figure 6.14 RNA polymerase reactions. Reactions were initiated with 200 nM of A) *E.coli* Pol II B) *R.marinus* Pol II and C) *S.YO3* Pol II enzymes. Either NTPs (RNA polymerase) or dNTPs (DNA polymerase, control) were used in the reactions.

E.coli, *R.marinus* and *S.YO3* DNA Pol II enzymes were not able to incorporate NTP's opposite the DNA template during the 60 minute reaction (Figure 6.14). However, the control reactions containing dNTPs all showed full extension of the primer within 30 minutes, indicating that the enzymes were active under the reaction conditions tested. Exonucleotic degradation of the primer was seen with the *E.coli* and *R.marinus* enzymes; however no degradation was seen in the reaction containing *S.YO3* DNA Pol II (Figure 6.14). Thus the DNA Pol II enzymes did not possess RNA polymerase activity under the conditions tested.

6.8 Reverse transcriptase activity of DNA Pol II's

Reverse transcription (RT) is the process in which complementary DNA is generated from an RNA template (Gilboa *et al.*, 1979). Many polymerases with RT activity have been commercialised for use in reverse transcriptase (RT) PCR. However, these polymerases are often limited by their sensitivity to heat or low fidelity rates (Eckert and T A Kunkel, 1991). RT activity is normally associated with retroviruses, however, some eukaryotic polymerases such as telomerase and retrotransposons possess RT activity (Lingner *et al.*, 1997; Finnegan, 2012). Thus, reverse transcriptase activity assays were performed to determine if the DNA Pol II enzymes possessed this activity.

Primer-template extension reactions were performed using an RNA template (DNA in control) hybridised to a DNA primer (Figure 6.15). Reactions were performed at 50°C and contained 40 nM of RNA-U or DNA-T (control) primer-template (Figure 6.15), 10 mM dNTPs and 200 nM Pol II enzymes or 1 µl of commercial reverse transcriptase (control). Analysis was performed using a 17 % acrylamide gel as described above (Section 6.5) (Figure 6.16).

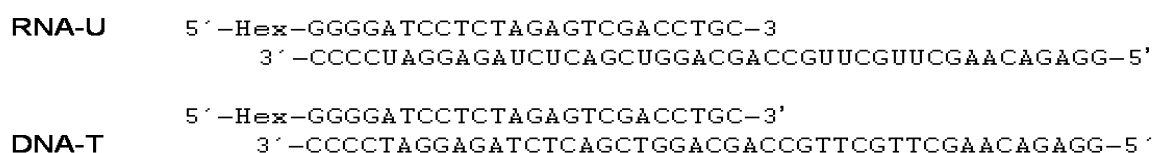


Figure 6.15 Oligonucleotide sequences used in reverse transcriptase activity assay.

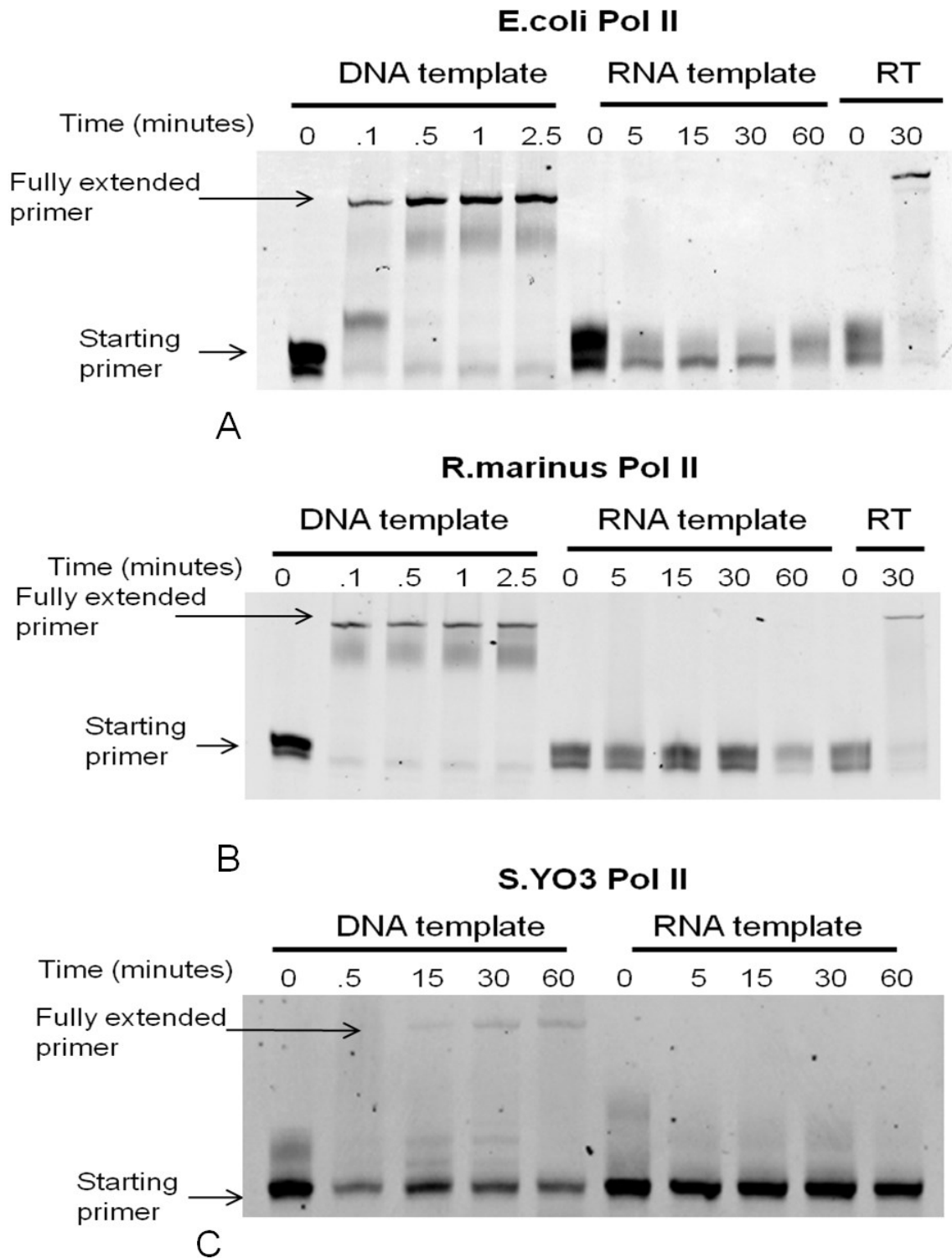


Figure 6.16 Reverse transcriptase assays initiated with **A) 200 nM *E.coli* Pol II** **B) 200 nM *R.marinus* Pol II** or **C) 400 nM *S.YO3* Pol II**. Control assays (marked RT on the gel above) were initiated with 1 μ l of reverse transcriptase and contain DNA/RNA primer-template (Figure 6.15).

Figure shows the DNA Pol II enzymes were unable to incorporate dNTPs opposite the template RNA. The control reactions initiated by *E.coli* and *R.marinus* DNA Pol II, containing the DNA template, show full extension of the primer within 5 seconds whereas the reaction initiated by *S.YO3* required 60 minutes to generate a small amount of fully extended product (similar to observations in Figure 6.8). The second control reaction performed contained the RNA template and commercially available reverse transcriptase. This reaction showed full extension of the DNA primer, thus confirms that the RNA was of a good standard. The DNA Pol II's lack of reverse transcriptase activity was not surprising as it is a feature rarely found in eukaryotic polymerases.

6.9 DNA Pol II exonucleolysis of DNA/RNA primer-templates

The lack of reverse transcriptase activity observed in Figure 6.16 may be due to the exonucleotic degradation of the primer. To test this, exonuclease reactions were performed using the RNA-U template (Figure 6.15). Reactions were initiated with 200 nM of *E.coli* DNA Pol II, 200 nM *R.marinus* DNA Pol II, 400 nM *S.YO3* Pol II or 200 nM *Pfu*-Pol B (control) (Figure 6.17). All reactions were performed at 50°C, for the times indicated on Figure 6.17 and analysed using a 17 % denaturing polyacrylamide gel visualised using a Typhoon scanner and ImageQuant software.

Figure 6.17 shows that both *E.coli* and *R.marinus* DNA Pol II enzymes are able to degrade the RNA-U template. However, no degradation is observed in the reaction initiated with *S.YO3* DNA Pol II. This is further confirmation that the *S.YO3* Pol II enzyme does not possess exonuclease activity. The exonucleolysis of the template RNA may be responsible for the lack of RNA polymerase activity in *E.coli* and *R.marinus* DNA Pol IIs.

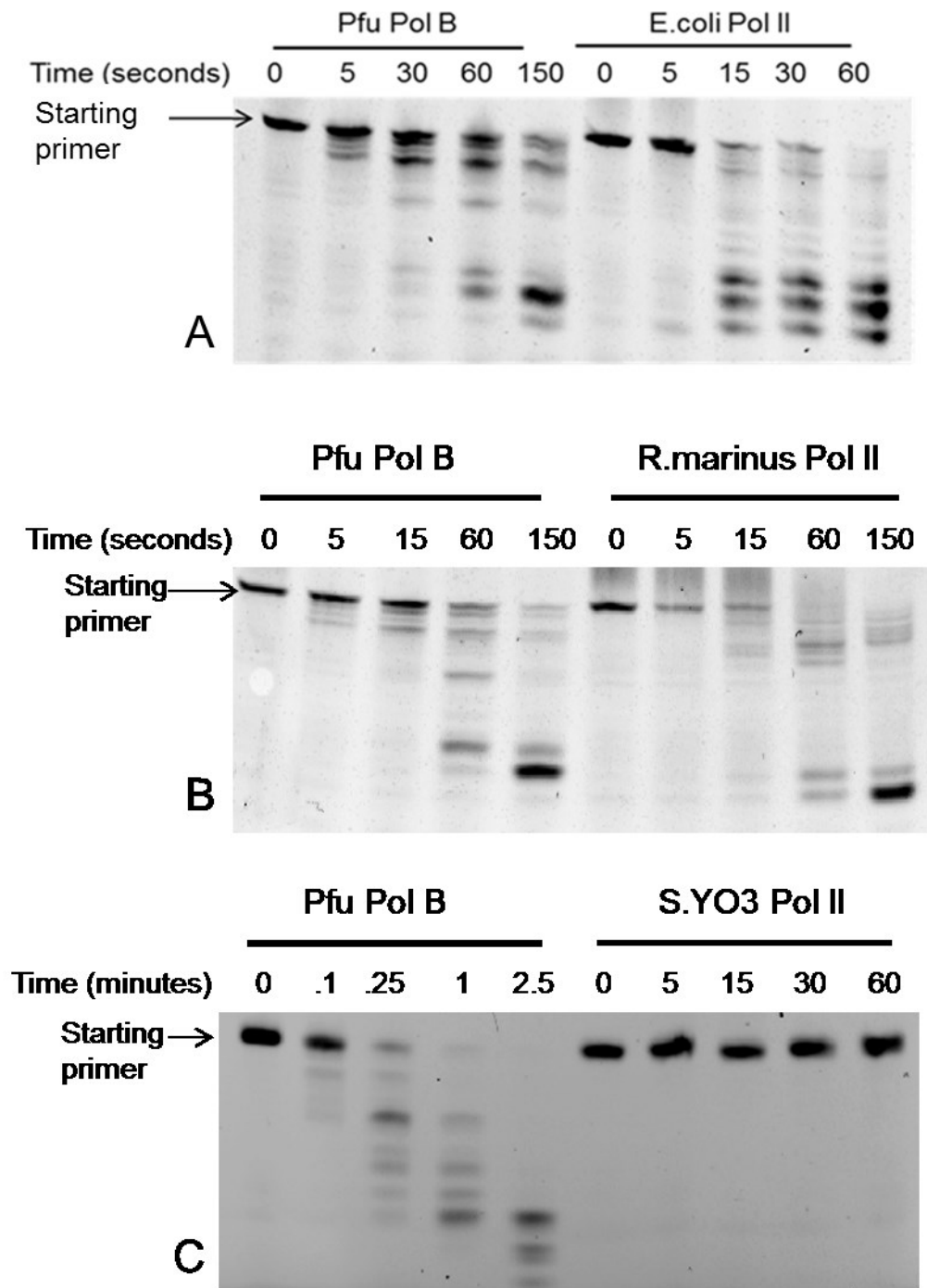


Figure 6.17 3'-5' exonuclease reactions containing RNA primer-template. Reactions initiated with A) 200 nM *E.coli* DNA Pol II B) 200 nM *R.marinus* DNA Pol II C) 400 nM *S.YO3* DNA Pol II.

6.10 Thermostability of *R.marinus* and *S.YO3* DNA Pol II

Proteins extracted from *R.marinus*, a member of the bacteroidetes, have been found to be highly thermostable (highest growth temperature ~ 85°C) (Andresson and Fridjonsson, 1992; Bjornsdottir *et al.*, 2006), whereas very little literature is available about the thermostability of proteins purified from *S.YO3*, a member of the *Aquificales* (growth temperatures up to 78°C) (Nakagawa *et al.*, 2005). To measure the thermostability of *R.marinus* and *S.YO3* DNA Pol II, two different techniques were employed; In-depth Simple Rapid Small Volume detection (DESERVED) analysis (Chalton and Lakey, 2010) and differential scanning fluorimetry (DSF) (Niesen *et al.*, 2007).

Both techniques monitor the change of tertiary structure that occurs when proteins unfold and, therefore, lose activity. Unfolding may be reversible i.e. on cooling the proteins re-fold and recover activity (Creighton, 1980). Recovery of activity is unusual and difficult, particularly when unfolding is thermally induced (Hilser *et al.*, 1993). Most proteins irreversibly denature on unfolding, forming insoluble aggregates rendering the protein inactive (Tanford, 1970). The use of two different techniques allowed for comparison of results and an improved consensus of the temperature at which unfolding occurs.

6.10.1 DESERVED analysis

DESERVED analysis detects small changes in the tertiary structure of proteins using a fluorophore which binds to hydrophobic regions (Chalton and Lakey, 2010). When proteins unfold, breakdown of secondary and tertiary structure results in clusters of, normally buried, hydrophobic amino acids being exposed (Chalton and Lakey, 2010). DESERVED analysis takes advantage of the increase in hydrophobicity by using the fluorophore, 1-anilino 8-naphthalene sulfonic acid (ANS), which produces an increased fluorescence signal when bound to hydrophobic amino acids (Gabellieri and Strambini, 2006). The increases in fluorescent signal can be measured and thus provides a qualitative measure of protein unfolding and consequently determination of protein thermostability.

To measure heat induced denaturation, proteins (10 μ M) were mixed with 1mM ANS and incubated at 80°C or 90°C. After 0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200 and 220 seconds, 20 μ l of reaction mixture was removed and placed on ice for 5 minutes. The samples were then left to equilibrate to room temperature and the level of ANS fluorescence was measured using a fluorescence-nanodrop set to detect light emitted at 395-755 nm.

The ANS fluorescence measurements made it possible to distinguish differences in the time taken for the unfolding and eventual denaturation of *R.marinus* and *S.YO3* DNA Pol II proteins (Figure 6.18). The total sum of fluorescence emitted between 395 and 751 nm was calculated for each time point and plotted on a line graph (Figure 6.18). Measurements were also performed, under identical conditions, using *Pfu*-Pol B, a well characterised thermostable protein (Killelea and Connolly, 2011).

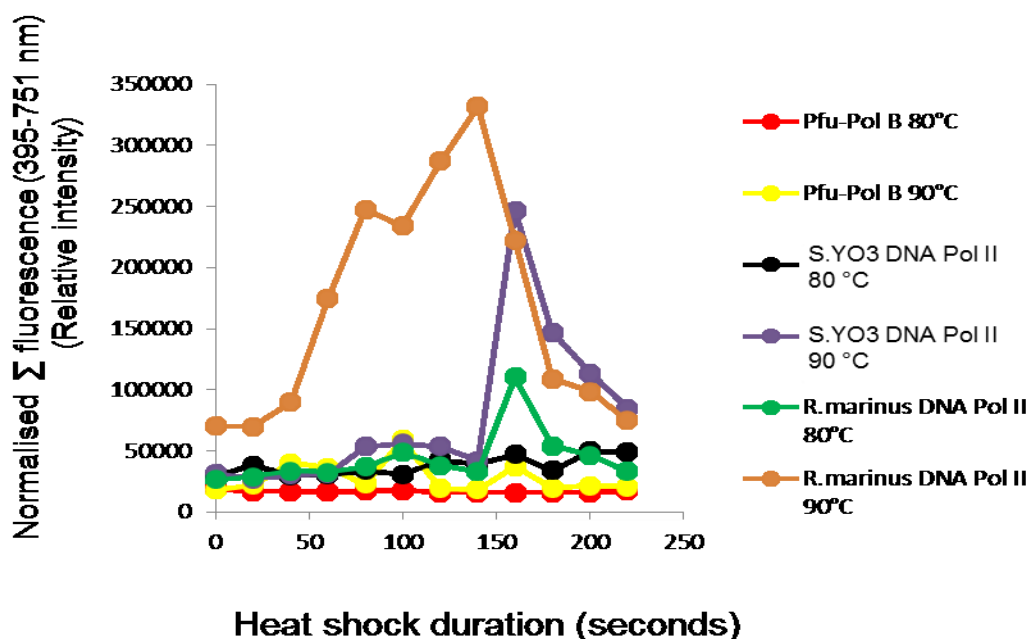


Figure 6.18 Total fluorescence emitted between 395 and 751 nm plotted against heating time. Reactions were performed at both 80°C and 90°C for *R.marinus* DNA Pol II, *S.YO3* DNA Pol II and *Pfu*-Pol B.

Figure 6. shows that the *S.YO3* DNA Pol II protein remains in its folded state when incubated at 80°C for 220. This is determined as the level of fluorescence for *S.YO3* remains low over the entire 220 second timecourse similar to *Pfu*-Pol B. The lack of fluorescence indicates that the protein has not unfolded, exposing hydrophobic regions, and thus has not bound ANS. *R.marinus* shows a slight increase in fluorescence after incubation at 80°C for 160 seconds. The increase in fluorescence indicates unfolding of the protein, although the increase does not appear to be significant (Figure 6.18). However, when heated at 90°C, both proteins appeared to unfold as indicated by an increase in total fluorescence (Figure 6.18).

R.marinus DNA Pol II begins to unfold after the 20 second time point whereas *S.YO3* DNA Pol II starts to unfold after 140 seconds. The unfolding of the proteins results in an increased exposure of hydrophobic groups which bind ANS and thus increase the fluorescence emitted (Chalton and Lakey, 2010). Following the initial increase in fluorescence, further heating resulted in a drop in the fluorescence signal of the ANS dye. It is likely that aggregation and precipitation of the unfolded state (i.e. irreversible denaturation) was the cause of the fluorescence decrease. *Pfu*-Pol B, is highly thermostable (Killelea and Connolly, 2011), and was stable at both 80°C and 90°C as indicated by no increase in fluorescence emission (Figure 6.18).

The use of the total fluorescence signal of the ANS spectrum is an established technique for interpreting DESERVED analysis data. However, areas of the ANS spectrum are more susceptible to noise when measured at low volumes (Figure 6.19) (Chalton and Lakey, 2010). Therefore, it is advantageous to corroborate total fluorescence data using a measurement known as the Barycentric wavelength mean (BCWM). The value of the BCWM is calculated using the following equation (Chalton and Lakey, 2010).

$$\lambda_m = \frac{\sum F(\lambda)x(\lambda)}{\sum F(\lambda)}$$

Where λ_m is the Barycentric wavelength mean and $F(\lambda)$ is the fluorescence intensity at wavelength λ .

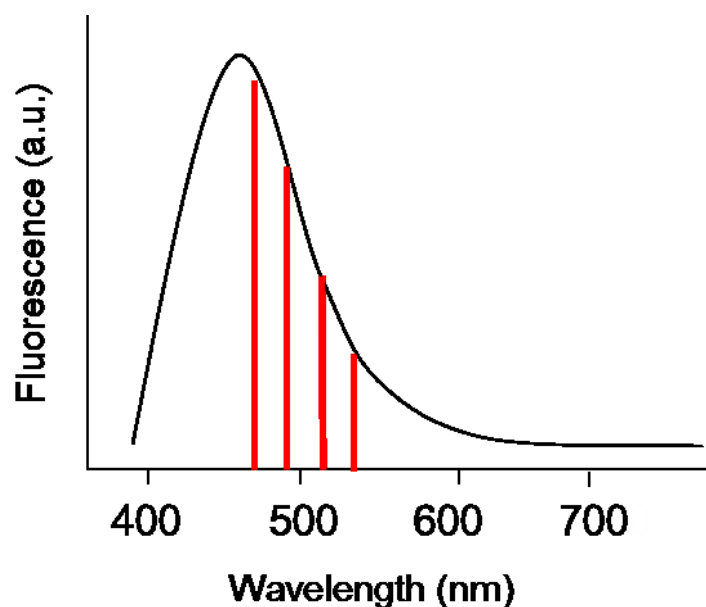


Figure 6. 19 A graphical representation of the ANS spectra between 395 and 751 nm. The red area is the section that is least susceptible to noise at low volumes (475-540 nm) according to the manufacturer (Sigma Aldrich UK).

It can be seen that the *R.marinus* DNA Poll II and *S.YO3* experienced a decrease in the BCM over the course of time (Figure 6.20), which suggests unfolding of the polymerase and exposure of hydrophobic regions. There were variations in the starting time of the decrease in BCM, mirroring the data plotted for the mean total fluorescence (Figures 6.18 & 6.20). The similarity between the total fluorescence data and the BCM data indicates that the total fluorescence data was not influenced by noise in the ANS spectrum (Figure 6.19).

Again, as was seen with the total fluorescence analysis of the data, the polymerase underwent an aggregation event. An increase in the value of the BCM indicated a decrease in the fluorescence signal emitted by ANS, mirroring the analysis of the total fluorescence signal (Figure 6.18).

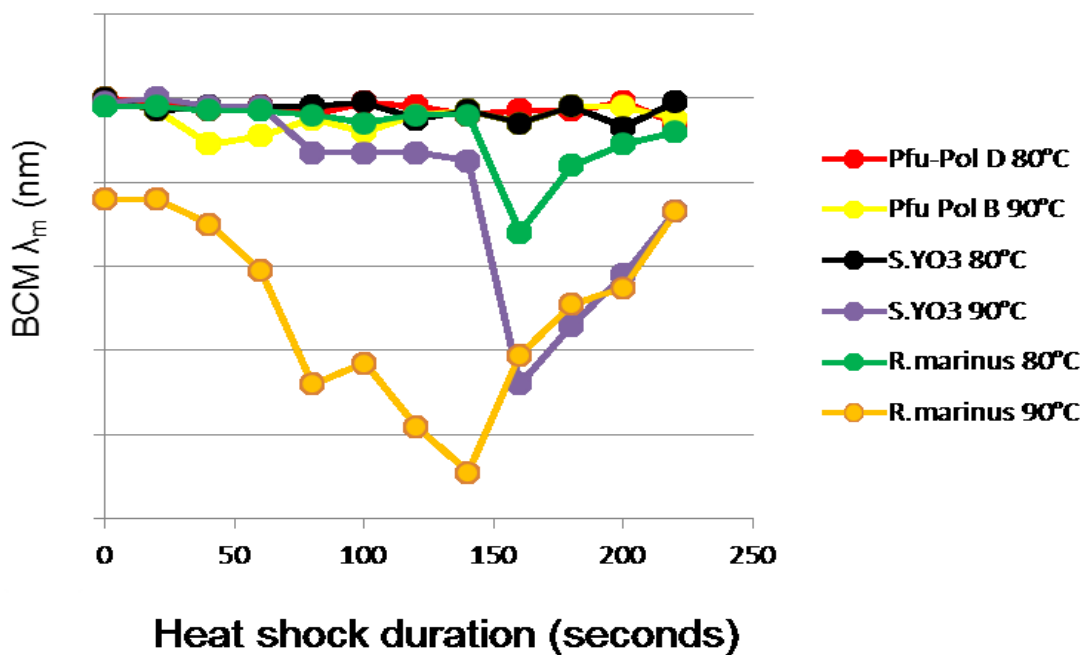


Figure 6.20 The barycentric wavelength mean as established through DESERVED analysis. Plotted is the average BCM (calculated from 3 repeats) for *R.marinus* DNA Pol II, *S.YO3* DNA Pol II and *Pfu*-Pol B.

6.10.2 Differential scanning fluorimetry (DSF)

To unequivocally confirm the thermostability of the proteins an additional thermal denaturation technique, differential scanning fluorimetry (DSF) was performed (Niesen *et al.*, 2007). DSF allows analysis of protein unfolding as a function of temperature in real time. A dye, SYPRO orange, known to fluoresce when bound to exposed hydrophobic regions of proteins, was incubated with protein samples. Samples were heated at a rate of 1°C per minute, between 25°C-100°C, and the level of fluorescence measured, in real time, using a Rotor-Gene-6000 thermocycler (Figure 6.21).

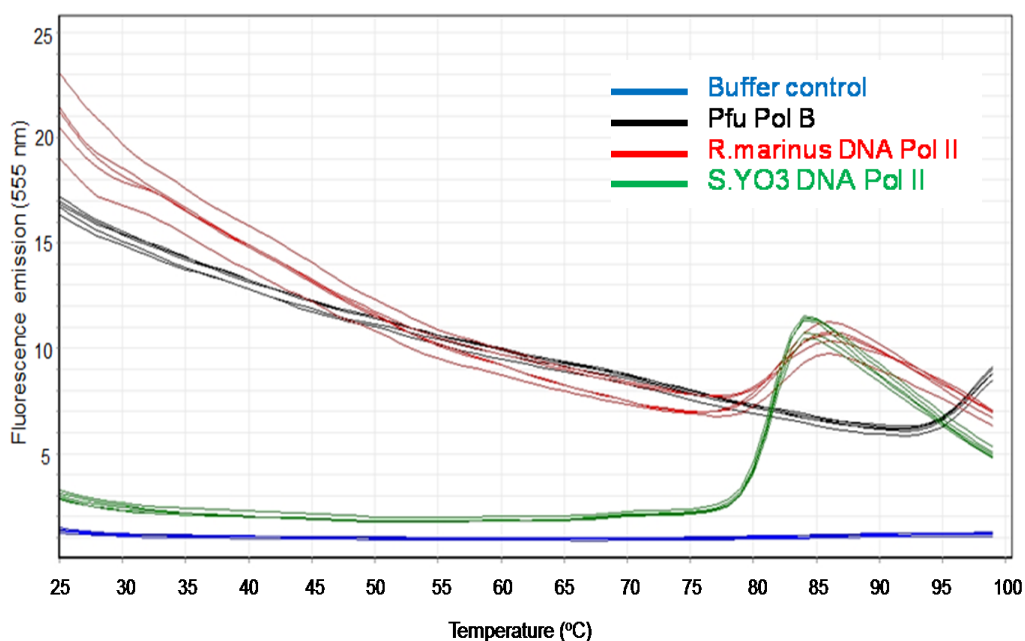


Figure 6.21 Differential scanning fluorimetry of DNA Pol II.

The buffer control (no protein) fluorescence remains constant for the duration of the experiments as there is no protein present and therefore no change in hydrophobicity. *Pfu*-Pol B shows a slight increase in fluorescence at ~96°C. *Pfu*-Pol B is a well characterised thermostable protein and does not unfold at temperatures near 100°C (Killelea and Connolly, 2011). *R.marinus* and *S.YO3* DNA Pol IIs show an increase of fluorescence emission at ~80°C. The increase in fluorescence peaks at ~84°C for *R.marinus* and ~86°C for *S.YO3*. The subsequent decrease in fluorescence may represent aggregation and precipitation reducing the fluorescence as seen in the DESERVED analysis.

The data shown in Figure 6.21 is problematic as the baselines are not constant for all enzymes tested. *S.YO3* DNA Pol II has a flat baseline, whereas *Pfu*-Pol B and *R.marinus* Pol II's baselines both slowly decrease from a high starting level, prior to the unfolding event. To make the results more clear a differential plot of the DSF data is presented (Figure 6.22).

The differential plot makes the data clearer and allows the thermal transition midpoint (T_m) for protein unfolding to be calculated. The T_m values were calculated from 4 individual repeats of the DSF experiment. The results show

the *S.YO3* and *R.marinus* DNA Pol II proteins have a melting temperature of 82.3°C and 83.5°C, respectively, whereas the melting temperature of *Pfu*-Pol B is significantly higher. The melting temperature of *R.marinus* Pol II is consistent with the melting temperature of other purified *R.marinus* proteins (Bjornsdottir *et al.*, 2006).

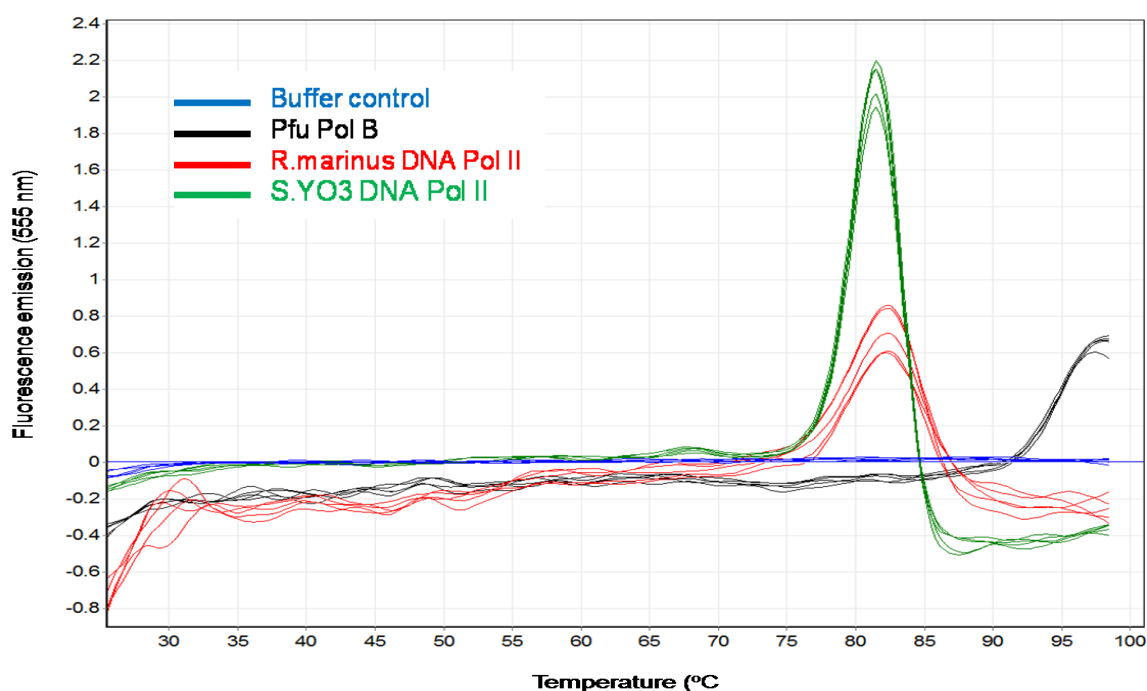


Figure 6.22 Differential plot of DSF data shown in Figure 6.21.

6.11 Attempted PCR with *R.marinus* and *S.YO3* DNA Pol II

The polymerase chain reaction has revolutionised molecular biology and enables specific regions of DNA to be amplified in a quick and accurate manner (Bartlett & Stirling., 2003). PCR mixtures with primers designed to amplify a ~200 bp region of the pET28a vector (Table 2.4), were mixed with varying concentrations of *R.marinus* and *S.YO3* DNA Pol II (20 nM-1 μ M) (Figure 6.23) and placed in a thermocycler under standard reaction conditions. The PCR was repeated at three different primer-annealing temperatures, 56°C, 57°C and 58°C to try and improve the specificity of the reaction (Figure 6.23). Reactions were performed in GC and HF reaction buffers (purchased from New England

Biolabs). Controls containing Phusion polymerase and the manufacturer's GC and HF buffers, were performed simultaneously (Figure 6.23).

Under the conditions tested no amplification of the target 200 bp DNA region was detected in reactions initiated by *R.marinus* or *S.YO3* DNA Pol II (Figure 6.23). At each annealing temperature tested, 56°C, 57°C and 58°C, the commercially available Phusion polymerase successfully amplified the target DNA (Figure 6.23). This suggests that the primers used were suitable for PCR amplification.

To further investigate the bacterial polymerases' ability to amplify DNA in the PCR, the experiment was repeated using different reaction buffers (Table 6. 1) and a range of annealing temperatures (55°C-60°C) (Figure 6.24). Buffer solution, salt concentration, pH and temperature all affect the activity of polymerases thus a wide range of reaction conditions were used to try and optimise the PCR (Figure 6.24).

Neither *R.marinus* nor *S.YO3* DNA Pol II enzymes were able to amplify the ~200 bp regions of the pET28a vector under any of the reaction conditions tested (Figure 6.23 & 6.24). It is likely that both proteins are denatured on exposure to the high temperatures (~95°C) in the initial rounds of PCR.

Buffer name	Buffer composition
GC Buffer	20 mM Bicine/KOH [pH 8.6], 2 mM MgSO ₄ , 100 mM KCl , 150 mM PPS 1-(3-Sulfopropyl) pyridinium hydroxid inner salt 98 % , 0.25 % (w/v) Nonidet P40
HF Buffer	Provided by manufacturer
Buffer 1	30 mM Tricine-NaOH pH [7.5], 3 mM MgSO ₄ , 40 mM KCl, 10 ug/ml BSA
Buffer 2	30 mM Tricine-NaOH pH [8.5], 3 mM MgSO ₄ , 100 mM KCl, 10 ug/ml BSA
Buffer 3	30 mM Bicine-NaOH pH [7.5], 3 mM MgSO ₄ , 40 mM KCl, 10 ug/ml BSA
Buffer 4	30 mM Bicine-NaOH pH [8.5], 3 mM MgSO ₄ , 100 mM KCl, 10 ug/ml BSA

Table 6. 1 Buffers used in PCRs by Phusion, *R.marinus* DNA Pol II and *S.YO3* DNA Pol II. GC and HF buffers were supplied with the Phusion enzyme (New England Biolabs).

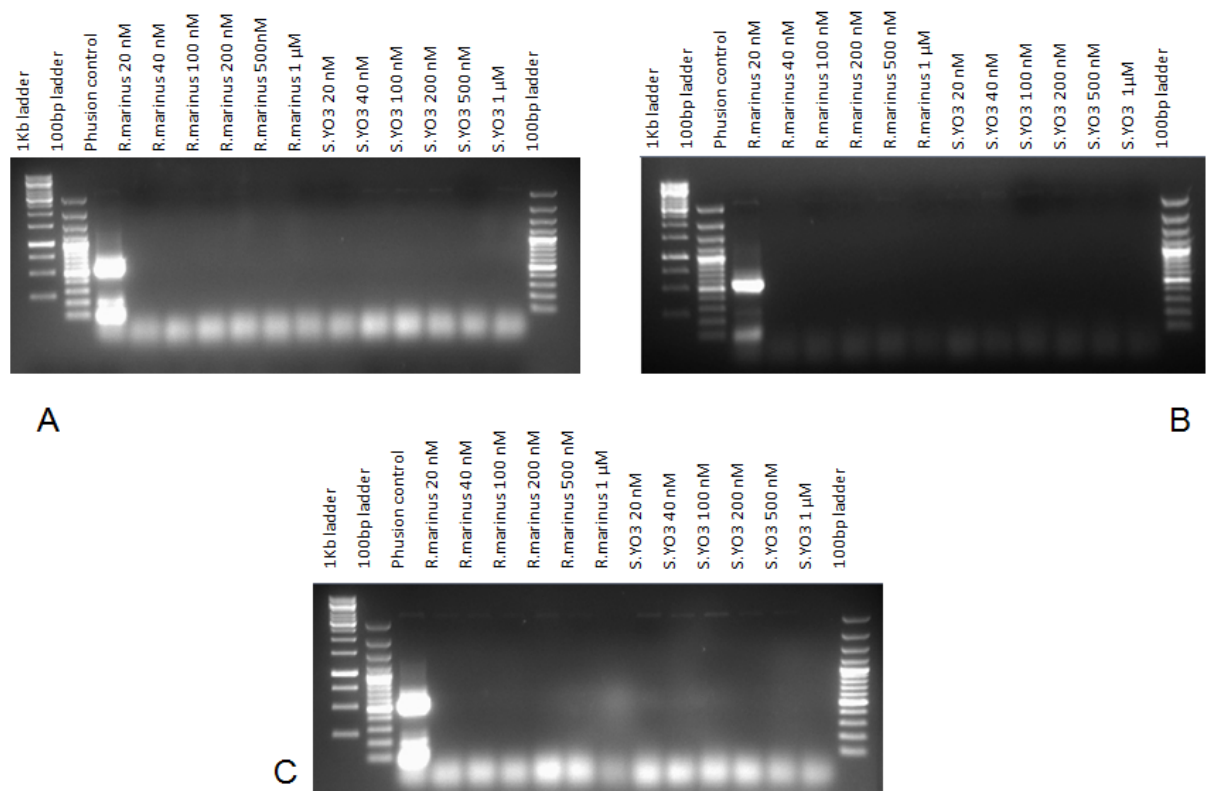


Figure 6.23 Products from PCRs, targeting a ~ 200 bp region of pET28a, visualised on a 1 % agarose gel. Various concentrations of enzyme were used in the PCRs (stated on gels) and GC buffer was used in the reaction. The experiment was repeated using three different annealing temperatures: A) 56°C B) 57°C C) 58°C.

6.11.1 Real time PCR

Real time PCR enables the build up of amplified DNA products to be monitored in real time and has been used for quantification of small amounts of nucleic acid (Dorak, 2007). RT-PCR reaction mixtures are normally similar to those used in standard PCR but additionally contain a fluorescent indicator, capable of monitoring the increasing levels of DNA as the PCR progresses. The dye, SYBR-Green, which binds strongly to double-stranded DNA with a profound increase in fluorescence intensity, is commonly used. (Dorak, 2007) Thus, in a RT-PCR experiment a small amount of starting DNA is amplified in the presence of SYBR-Green and the increase in fluorescence monitors the reaction in real time and also allows quantification of the initial DNA amount (Dorak, 2007).

RT-PCR was employed as an additional test to determine if *R.marinus* or *S.YO3* DNA Pol II enzymes were able to amplify DNA. A short, 250 base region of the *Saccharomyces cerevisiae* Pol 2 gene was targeted for amplification. A control reaction was performed targeting the same DNA sequence using *Pfu*-Pol B. Varying concentrations of enzyme were used in the reactions: 20 nM, 40 nM, 400 nM and 1 μ M (Figure 6.25).

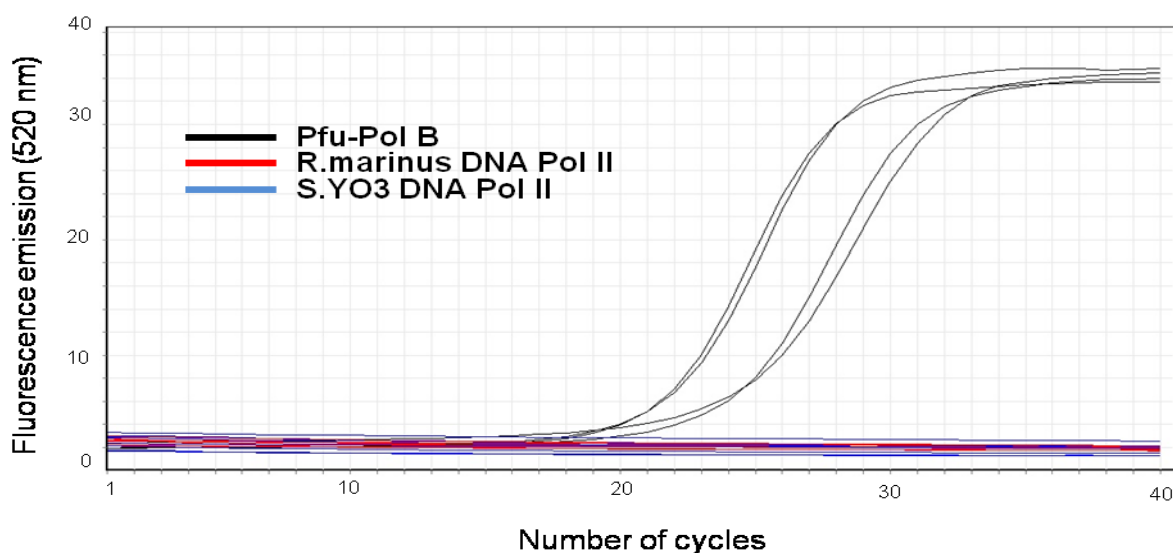


Figure 6.25 RT-PCR targeting a 250 bp region of the *S. cerevisiae* Pol 2 gene. Reactions initiated with varying concentrations of *Pfu*-Pol B, *R.marinus* DNA Pol II or *S.YO3* DNA Pol II. 20 nM, 40 nM, 400 nM and 1 μ M of each enzyme were used. Amplification was detected in all reactions initiated by *Pfu*-Pol B.

As double stranded DNA builds up during the real time PCR, there should be an increase in fluorescence. The RT-PCR data show that *R.marinus* and *S.YO3* Pol II did not amplify the target DNA under the conditions tested (Figure 6.25). The positive control, *Pfu*-Pol B, shows successful amplification of the target DNA at all four concentrations of enzyme used (Figure 6.25).

Further analysis using the differential plot makes the data clearer and allows the thermal transition midpoint (T_m) of the product produced via PCR amplification by *Pfu*-Pol B to be calculated. Figure 6.26 shows a single peak, indicating that the amplification product was likely to be the desired product rather than primer-dimers or a by-product. Additionally the T_m of the product was calculated as 86.8°C (calculated as an average of 4 repeats) (Figure 6.26), which is similar to the T_m for this product has been previously measured (Killelea and Connolly, 2011), indicating that *Pfu*-Pol B was amplifying the target region of DNA.

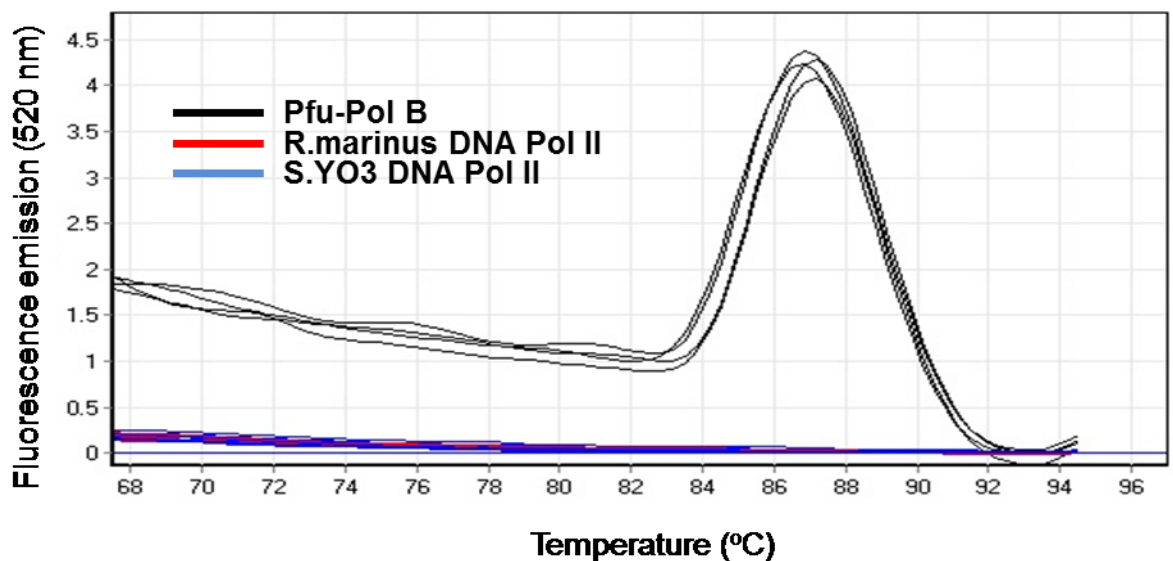


Figure 6. 26 Analysis of the melting temperatures of the DNA products produced during RT-PCR. The DNA produced by *Pfu*-Pol B has a melting temperature of $\sim 86.8^\circ\text{C}$.

6.12 Fidelity of *E.coli*, *R.marinus* and *S.YO3* DNA Pol IIs

Polymerases with a low error rate are desired in many application that require accurate replication of the target DNA e.g. PCR and sequencing. In order to calculate the fidelity rate of the purified Pol II proteins, a plasmid-based fidelity assay was employed, as discussed previously in chapter 4. The gapped pSJ3 vector was kindly gifted by Brian Keith (Newcastle University) who had already calculated the background mutation frequency of the vector (3.075×10^{-5}) (Keith *et al.*, 2013).

The fidelity assay consisted of three independent experiments each of which was repeated five times. The results found are shown in Table 6. 2

Protein	Number of colonies counted ^a	Number of white colonies ^a	Number of white colonies sequenced	Observed mutation rate (%) ^b	Corrected mutation frequency ^c	Error rate ^d
<i>E.coli</i> Pol II	59,235	28	10	.047	4.42×10^{-4}	3.03×10^{-6}
<i>R.marinus</i> Pol II	33,843	14	10	.041	3.83×10^{-4}	2.62×10^{-6}
<i>S.YO3</i> Pol II	38,086	15	10	.039	3.63×10^{-4}	2.49×10^{-6}
<i>E.coli</i> Pol II <i>exo</i> ⁻	41,903	128	10	.305	3.02×10^{-3}	2.07×10^{-5}
<i>Taq</i> ^e	20,756	34	-	.05	1.6×10^{-3}	1.1×10^{-5}
<i>Pfu</i> -Pol B ^e	20, 116	11	-	.016	5.2×10^{-4}	3.5×10^{-6}

Table 6. 2 Error rates of DNA polymerases determined using the gapped, pSJ3. a Sum of three independent experiments each consisting of 5 repeats. b Observed mutation rate is defined as (white colonies)/(total colonies) x 100. c The mutation frequency here have the background mutation frequency of pSJ3 subtracted. d Error rates were calculated using the formula given in the text (Chapter 4). e enzyme's fidelity rates calculated by Brian Keith (Keith *et al.*, 2013)

E.coli, *R.marinus* and *S.YO3* DNA Pol II's were found to be high fidelity polymerases with low error rates similar to those calculated for other family B polymerases including *Pfu*-Pol B (Kunkel and Alexander, 1986; Keith *et al.*, 2013). *E.coli* DNA Pol II exo^- had a considerably higher error rate than the WT *E.coli* DNA Pol II, with almost 10 x more errors detected (Table 6.2). This increased error rate is expected as 3'-5' exonuclease activity proof-reads the DNA during replication, excises mis-incorporated bases and therefore improve the fidelity of the enzymes. However, the high fidelity observed in *S.YO3* DNA Pol II was unexpected as *S.YO3* did not appear to possess 3'-5' exonuclease activity (Figure 6.) and thus, was expected to have a lower fidelity rate than *R.marinus* and *E.coli* DNA Pol II. Many factors contribute to a polymerases fidelity rate including pH, temperature and Mg concentration; therefore it is possible that the conditions used were favourable for *S.YO3*. An additional hypothesis is that, despite the lack of 3'-5' exonuclease activity, *S.YO3* has a high fidelity due to the evolution of other properties such as an improved rate of correct nucleotide incorporation.

6.13 Discussion

The data presented in this chapter characterises two previously undescribed bacterial family B polymerases, *R.marinus* DNA Pol II and *S.YO3* DNA Pol II. Primer-template extensions revealed that both enzymes possess DNA polymerase activity which is not inhibited by the presence of template strand uracil. Comparisons of the conserved polymerase domain show that both *R.marinus* (D416 and D532) and *S.YO3* (D360 and D492) possess two essential aspartic acid residues responsible for polymerase activity (Figure 6.1). These aspartic acid residues bind two metal ions (Mg^{2+}), during polymerisation, allowing DNA replication to occur (Bernad *et al.*, 1990).

The phosphoryl transfer reaction of all polymerases is catalysed by a two metal ion mechanism which requires these two aspartic acid residues (sometimes glutamic acid) for binding (Steitz, 1991). For polymerases to have high processivity, conserved regions of importance have been identified (Figure 6.1) (Bernad *et al.*, 1990). In particular an aspartic acid residue located within the polymerase activity active site II region (Figure 6.1) is known to be important.

S.YO3 contains a single amino acid change within this region (residue 542 for *E.coli*, Figure 6.1). This single change could be responsible for the reduced rate of polymerisation, observed in *S.YO3* DNA Pol II, compared to *E.coli* and *R.marinus* DNA Pol IIs. Additionally the *S.YO3* DNA Pol II lacks an aspartic acid residue that is normally associated with polymerase DNA Pol IIs are known to be responsible for translesion synthesis, and previous studies have shown they are able to bypass damaged DNA bases either directly or by “template-skipping” (Wang and Yang, 2009). Previous work has focussed on abasic (AP) lesions, C and AAF adducts and no previous reference to uracil recognition could be found. However, the ability to bypass uracil in template DNA was expected as, to date, only archaeal DNA replicase family B and D have been found to respond to, and be inhibited by, uracil (Richardson *et al.*, 2013).

Replicative, family A, B and C, polymerases normally possess intrinsic 3'-5' exonuclease activity (Bebenek and Kunkel, 2004; Reha-Krantz, 2010). This activity is involved in removing mis-incorporated bases from extending primers, serving as a proofreading function to improve the enzymes fidelity. Reactions showed that *E.coli* and *R.marinus* DNA Pol II's possessed 3'-5' exonuclease activity (Figure 6. and Figure 6.). Both enzymes were able to degrade single stranded DNA, primer-template DNA and primer-template DNA-RNA.

However, no exonuclease activity was detected in the *S.YO3* DNA Pol II. The mechanism of 3'-5' exonuclease activity has been well characterised and involves a similar 2 metal ion mechanism as seen in DNA polymerase activity (Bernad *et al.*, 1989; Beese and Steitz, 1991). As such, three conserved aspartic acid residues are required within the exonuclease domain to bind two metal ions. Analysis of the amino acid alignments of the DNA Pol IIs identifies a leucine at position 216 within the *S.YO3* DNA Pol II's exonuclease domain II (residue 290 within *E.coli*) (Figure 6.1). The amino acid alignment sequences show the other 5 DNA Pol II sequences have an aspartic acid residue located at this position (Figure 6.1). Therefore, as aspartic acid is known to be essential for 3'-5' exonuclease activity, it is likely that this single amino acid change results in the lack of exonuclease activity observed in *S.YO3* DNA Pol II.

Despite the absence of exonuclease activity, *S.YO3* DNA Pol II was found to possess a similar fidelity rate to *E.coli* and *R.marinus* DNA Pol II (Table 6.2). As 3'-5' exonuclease activity provides proofreading activity and thus improves fidelity, this observation was unexpected. The fidelity rate of the *E.coli* exo⁻ DNA Pol II was substantially lower than the WT *E.coli* DNA Pol II (Table 6. 2). This result was expected as the lack of proofreading activity would prevent the excision of incorrectly incorporated dNTPs.

Thus, the *S.YO3* DNA Pol II must possess other properties which ensure a high fidelity rate. The fidelity of polymerases is affected by many factors including the intrinsic accuracy of nucleotide selection (Joyce and Benkovic, 2004). Therefore, it is likely that despite the lack of 3'-5' exonuclease activity, *S.YO3* DNA Pol II has evolved to develop a high fidelity rate similar to those found in other family B polymerases. Unlike family B polymerases, most TLS polymerases (family Y) have a low fidelity rate when copying normal DNA (10^{-2} to 10^{-4}) (Bebenek and Kunkel, 2004). Thus, as DNA Pol II's play a role in TLS, it may be expected that they would have much lower fidelity rates than those observed (Table 6. 2). However, comparisons of crystal structures of *E.coli* DNA Pol II copying normal and damaged DNA shows that a change in polymerase conformation occurs during the replication of normal DNA, leading to the high fidelity rate observed when DNA Pol II's replicate undamaged DNA (Wang and Yang, 2009).

The two thermostability assays performed, DESERVED analysis and DSF, yielded similar results and showed that the melting temperatures of *R.marinus* and *S.YO3* DNA Pol II enzymes were 82.3°C and 83.5°C, respectively (Figure 6.22). Due to the known growth temperature of both species this result was as expected, although slightly disappointing as other proteins extracted from *R.marinus* have been identified with melting temperatures up to 100°C (Halldorsdottir *et al.*, 1998). *Pfu*-Pol B has much higher T_m value (Figure 6.22), which is why it is routinely used in PCR (Takagi *et al.*, 1997)

Despite the lower melting temperature of the bacterial Pol II compared to *Pfu*-Pol B, attempts at amplifying a 200 bp region of DNA via PCR were made. However, amplification of the target DNA was not detected under a wide variety

of conditions tested (Figure 6.23 and Figure 6.24). RT-PCR was performed as an additional test to determine if the enzymes were active in the PCR. Unfortunately, no DNA amplification was detected (Figure 6.25 & Figure 6.26).

Bacterial DNA Pol IIs have high polymerase activity, high fidelity and the ability to bypass damaged bases/regions of DNA (TLS). These enzymes, in contrast to archaeal DNA Pol Bs, are also not inhibited by template strand uracil and so not subject to “uracil poisoning”. DNA Pol II enzymes are interesting as they are the only known polymerases with translesion synthesis properties that also possess 3'-5' exonuclease activity (Yang and Woodgate, 2007). Thus bacterial Pol IIs have potential applications in the PCR especially with ancient or damaged DNA where the template quality would be expected to be poor.

Unfortunately the two DNA Pol IIs tested, *R.marinus* and *S.YO3*, were unable to carry out PCR presumably due to insufficient thermostability. The DNA Pol II enzymes investigated here are the most thermophilic identified (as examples were not found in highly thermophilic bacteria such as *Thermus* and *aquaferex* (Figure 6.1).

DNA Pol II's are rare in nature with few homologous enzymes found in database searches (Figure 6.1). The evolution of DNA Pol II presents an enigma as cells possess family Y Pols for TLS and family A, B and C polymerases for DNA replication and repair. Why Pol II's exist in some organisms and not in others is currently unknown.

As DNA polymerases have evolved to fulfil specific roles within the cell, often requiring specific substrates, naturally occurring polymerases often have limited applications in biotechnology (Hamilton, 2001). The creation of polymerases with unnatural abilities often develops enzymes with greater utility than those extracted directly from nature. As such many techniques have been described that can improve specific properties of DNA polymerases, making them more suitable to applications in biotechnology (Hamilton, 2001).

6.14 Future Work

The *S.YO3* DNA Pol II lacked 3'-5' exonuclease activity and was found to have a slow rate of polymerisation compared to *R.marinus* and *E.coli* DNA Pol II. Further research was performed by an undergraduate student at Newcastle University, James Dunce, under my supervision. James attempted to introduce exonuclease activity into *S.YO3* DNA Pol II (*S.YO3* DNA Pol II *exo*⁺) and improve the polymerase activity (*S.YO3* DNA Pol II *pol*⁺) using site directed mutagenesis. Two mutations were introduced into the *S.YO3* DNA Pol II gene using site-directed mutagenesis and exonuclease and polymerase reactions were repeated (same reaction conditions as used in figures 6.8 and 6.12). Figure 6.27 shows that the introduction of an aspartic acid residue at position 216 within the exonuclease domain did not improve exonuclease activity. Likewise, Figure 6.27 shows that the introduction of aspartic acid within the polymerase domain at position 480 of figure 6.1 (residue 542 for *E.coli*, Figure 6.1) stimulated polymerase activity, but only slightly (Figure 6.27). This result shows that the lack of exonuclease activity and the weak polymerase activity of *S.YO3* is not caused by a single amino acid change as previously hypothesised.

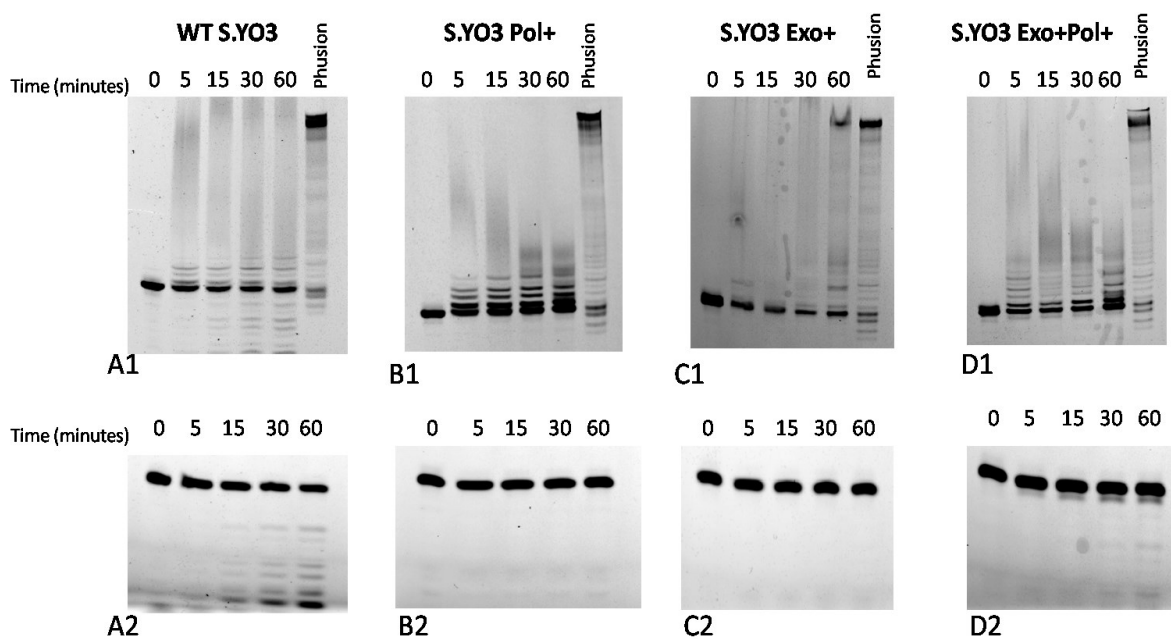


Figure 6.27 Primer extension and exonuclease reactions performed using modified *S.YO3* DNA Pol II.

The most important improvement of the polymerases would be an increase in thermostability. Several techniques could be employed to improve the thermostability of the enzymes, including random mutagenesis, site directed mutagenesis and potentially the best approach, compartmentalised self replication (CSR) (Ghadessy and Holliger, 2007) (Figure 6.28).

During CSR, *E.coli* cells overexpressing a polymerase library are combined with an aqueous solution containing specific primers for polymerase gene amplification, dNTPs and Mg^{2+} . The suspension is stirred into a mineral oil to generate an emulsion composed of segregated compartments that usually contain a single cell. The emulsion is subjected to PCR and the polymerase molecules present in each compartment replicate their own gene. Only thermostable mutants would be able to survive the high temperatures and replicate their own gene. After polymerase mediated primer extension achieved by thermal cycling, the aqueous phase is collected and the PCR products are appropriately processed and sub-cloned further rounds of selection.

This would be the quickest and most efficient technique which would result in the fewest number of “mutant screens” being performed. If a thermostable protein is obtained, it should be sequenced and tested for polymerase activity, exonuclease activity, thermostability and ability to initiate the PCR.

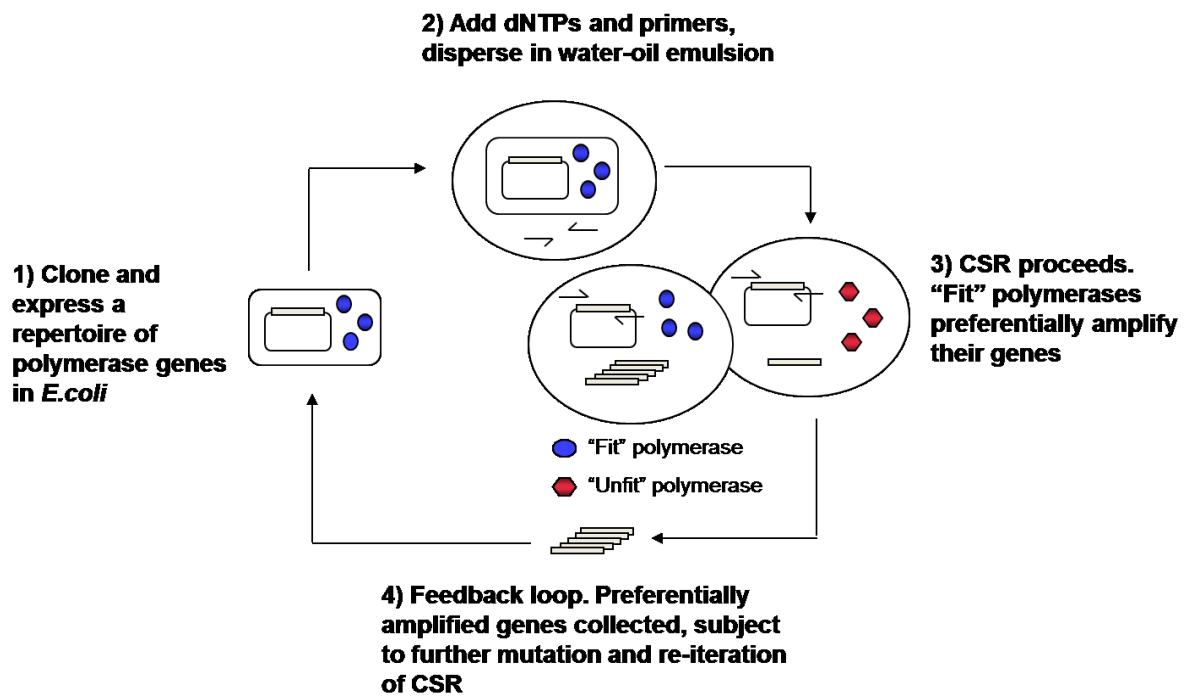


Figure 6. 28 Compartmentalised self-replication.

Conclusion

The data presented in this thesis provides compelling evidence that the occurrence of uracil in DNA templates diminishes extension by euryarchaeal family D DNA polymerases. Unlike observations with Pol B, where polymerisation is stalled in the presence of uracil, Pol D is able to continue DNA replication, but at a greatly reduced rate. Reactions performed using replication fork mimics showed that uracil located on an adjacent non-copied strand of DNA is also able to inhibit polymerisation by Pol D. This finding, combined with the observation that uracil resulted in a slight increase in binding affinity and 3'-5' exonuclease rates, led to the proposal of a model of uracil recognition for archaeal Pol D (Figure 3.24).

Characterisation of the large (DP2) and small (DP1) subunits of Pol D confirmed that the DP2 subunit possesses polymerase activity, while the DP1 subunit possesses 3'-5' exonuclease activity. The activities of the individual DP1 and DP2 subunits was significantly reduced compared to the activities observed in the Pol D holoenzyme. Experiments were performed to determine if full activity of the Pol D holoenzyme could be reconstituted by mixing the two subunits *in vitro*. Despite mixing the subunits using a number of conditions, the high activities characteristic of the Pol D holoenzyme were not obtained.

The C-terminal domain of the DP2 subunit of archaeal Pol D has strong homology to eukaryotic family B polymerases, especially regarding the position and spacing of cysteine residues. With the eukaryotic enzymes these cysteines are responsible for binding Zn^{2+} and forming an Fe-S cluster, both of which play a role in the folding of the polymerase and interaction with other proteins. Thus archaeal Pol D may also be a Zn^{2+} /Fe-S cluster protein. Fe-S clusters are assembled *in vivo* and are labile under aerobic conditions; therefore, should Pol D possess an Fe-S cluster it may not form correctly in *E.coli* or degrade during purification under aerobic conditions. Therefore, attempts were made to express Pol D in archaeal hosts in order to identify whether an Fe-S was present. Unfortunately attempts to overexpress Pol D in archaea were unsuccessful due to low levels of protein expression and contamination by endogenous proteins.

Fidelity assays were performed to calculate the previously undetermined error rate of Pol D. These assays revealed that Pol D is an accurate enzyme, however the fidelity rate was lower than expected if Pol D is, in fact, the main replicative polymerase in archaea. Fidelity assays were also performed to determine the error rate of the large, DP2, subunit and Pol D *exo*⁻ variants. These assays revealed that the isolated DP2 subunit had a higher error rate than even the Pol D *exo*⁻ enzyme. This finding suggests that the DP1 subunit increases fidelity both by supplying proof reading exonuclease activity and by facilitating folding of DP2 to increase the intrinsic accuracy of the polymerase reaction itself.

The research conducted during this thesis has contributed to the existing evidence that Pol D is a replicative polymerase that may be responsible for DNA replication in archaea. It is anticipated that purification of Pol D with an intact Fe-S will enable crystal structures to be obtained and may lead to identification of a uracil binding pocket. Work with eukaryotic Pol B has shown that the correct metallo-enzyme status greatly influences the folding of the protein and, thus, is critical for correct functioning. Therefore, it is hypothesised that Pol D with the correct metallo-enzyme status, and Pol D *in vivo*, will demonstrate higher fidelity than has been observed with currently purified samples. Further genetic and biochemical analysis of Pol D with the correct metallo-status may help close the debate regarding whether Pol B or Pol D is the main replicative polymerase in archaea. The role of Pol D within archaea raises wider issues regarding the origin and evolution of species and is critical for understanding how life arose

Research was performed to characterise two previously undescribed, thermostable bacterial DNA Pol II enzymes. This work was performed in an attempt to identify a polymerase that would possess properties with suitable application in biotechnology. *R. marinus* DNA Pol II was identified as possessing polymerase activity, strong exonuclease activity and thermostability of ~ 85°C. Unfortunately attempts to use the enzyme to initiate PCR were unsuccessful and it did not possess reverse transcriptase activity. Similarly, *S. YO3* DNA Pol II was found to possess polymerase activity, albeit much weaker than that of

R.marinus and did not possess exonuclease activity. *S.YO3* DNA Pol II was also unable to amplify DNA during attempts at PCR. Although this research did not identify enzymes that were immediately suitably for PCR, the properties of *R.marinus* DNA Pol II, particularly its ability to read through damaged bases, make it an ideal target for adaption for use in biotechnology. It is recommended that the CSR technique is used to improve the thermostability of the enzyme. It is likely that with increased thermostability combined with strong polymerase activity, high fidelity rate, and the intrinsic TLS properties would make *R.marinus* DNA Pol II an ideal polymerase for initiation of the PCR.

References

- Acharya, N., Johnson, R.E., Prakash, S. and Prakash, L. (2006) 'Complex formation with Rev1 enhances the proficiency of *Saccharomyces cerevisiae* DNA polymerase for mismatch extension and for extension opposite from DNA lesions', *Molecular and Cellular Biology*, 26(24), pp. 9555-9563.
- Alfredsson, G.A., Kristjansson, J.K., Hjorleifsdottir, S. and Stetter, K.O. (1988) 'Rhodothermus-Marinus, Gen-Nov, Sp-Nov, a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in Iceland', *Journal of General Microbiology*, 134, pp. 299-306.
- Andresson, O.S. and Fridjonsson, O.H. (1994) 'The Sequence of the Single 16s Ribosomal-Rna Gene of the Thermophilic Eubacterium-Rhodothermus Marinus Reveals a Distant Relationship to the Group Containing Flexibacter, Bacteroides, and Cytophaga Species', *Journal of Bacteriology*, 176(20), pp. 6165-6169.
- Arezi, B. and Kuchta, R.D. (2000) 'Eukaryotic DNA primase', *Trends in Biochemical Sciences*, 25(11), pp. 572-576.
- Aslanidis, C. and Dejong, P.J. (1990) 'Ligation-Independent Cloning of Pcr Products (Lic-Pcr)', *Nucleic Acids Research*, 18(20), pp. 6069-6074.
- Avery, O.T., Macleod, C.M. and McCarty, M. (1995) 'Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types - Induction of Transformation by a Deoxyribonucleic-Acid Fraction Isolated from Pneumococcus Type-iii (Reprinted from Journal of Experimental Medicine, Vol 79, Pg 137-158, 1944)', *Molecular Medicine*, 1(4), pp. 344-365.
- Banach-Orlowska, M., Fijalkowska, I.J., Schaaper, R.M. and Jonczyk, P. (2005) 'DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in *Escherichia coli*', *Molecular Microbiology*, 58(1), pp. 61-70.
- Barnes, D.E. and Lindahl, T. (2004) 'Repair and genetic consequences of endogenous DNA base damage in mammalian cells', *Annu Rev Genet*, 38, pp. 445-76.
- Barns, S.M., Delwiche, C.F., Palmer, J.D. and Pace, N.R. (1996) 'Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences', *Proc Natl Acad Sci U S A*, 93(17), pp. 9188-93.
- Barry, E.R. and Bell, S.D. (2006) 'DNA replication in the archaea', *Microbiol Mol Biol Rev*, 70(4), pp. 876-87.
- Bartlett, J.M. and Stirling, D. (2003) 'A short history of the polymerase chain reaction', *Methods Mol Biol*, 226, pp. 3-6.
- Beauchemin, D. (2010) 'Inductively coupled plasma mass spectrometry', *Anal Chem*, 82(12), pp. 4786-810.
- Bebenek, K. and Kunkel, T.A. (2004) 'Functions of DNA polymerases', *DNA Repair and Replication*, 69, pp. 137-165.
- Becherel, O.J. and Fuchs, R.P. (2001) 'Mechanism of DNA polymerase II-mediated frameshift mutagenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8566-71.
- Beese, L.S. and Steitz, T.A. (1991) 'Structural Basis for the 3'-5' Exonuclease Activity of *Escherichia-Coli* DNA-Polymerase-I - a 2 Metal-Ion Mechanism', *Embo Journal*, 10(1), pp. 25-33.

- Bell, S.P. (2002) 'The origin recognition complex: from simple origins to complex functions', *Genes & Development*, 16(6), pp. 659-672.
- Berardini, M., Foster, P.L. and Loechler, E.L. (1999) 'DNA polymerase II (polB) is involved in a new DNA repair pathway for DNA interstrand cross-links in *Escherichia coli*', *Journal of Bacteriology*, 181(9), pp. 2878-2882.
- Berg JM, T.J., Stryer L. 5th Edition (2002) 'Biochemistry'. 12/07/2011. New York: W H Freeman. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK22374/>.
- Bernad, A., Blanco, L., Lazaro, J.M., Martin, G. and Salas, M. (1989) 'A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases', *Cell*, 59(1), pp. 219-28.
- Bernad, A., Blanco, L. and Salas, M. (1990) 'Site-Directed Mutagenesis of the Ycdtds Amino-Acid Motif of the Phi-29 DNA-Polymerase', *Gene*, 94(1), pp. 45-51.
- Berquist, B.R., DasSarma, P. and DasSarma, S. (2007) 'Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1', *BMC Genet*, 8, p. 31.
- Bienko, M., Green, C.M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A.R., Hofmann, K. and Dikic, I. (2005) 'Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis', *Science*, 310(5755), pp. 1821-1824.
- Biles, B.D. and Connolly, B.A. (2004) 'Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR', *Nucleic Acids Research*, 32(22).
- Bjornsdottir, S.H., Blondal, T., Hreggvidsson, G.O., Eggertsson, G., Petursdottir, S., Hjorleifsdottir, S., Thorbjarnardottir, S.H. and Kristjansson, J.K. (2006) '*Rhodothermus marinus*: physiology and molecular biology', *Extremophiles*, 10(1), pp. 1-16.
- Bochkareva, E., Korolev, S. and Bochkarev, A. (2000) 'The role for zinc in replication protein A', *Journal of Biological Chemistry*, 275(35), pp. 27332-27338.
- Bock, J.B., Matern, H.T., Peden, A.A. and Scheller, R.H. (2001) 'A genomic perspective on membrane compartment organization', *Nature*, 409(6822), pp. 839-41.
- Braithwaite, D.K. and Ito, J. (1993) 'Compilation, alignment, and phylogenetic relationships of DNA polymerases', *Nucleic Acids Research*, 21(4), pp. 787-802.
- Brill, S.J. and Stillman, B. (1991) 'Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase', *Genes Dev*, 5(9), pp. 1589-600.
- Brochier-Armanet, C., Bousseau, B., Gribaldo, S. and Forterre, P. (2008) 'Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota', *Nature Reviews Microbiology*, 6(3), pp. 245-252.
- Brochier-Armanet, C., Forterre, P. and Gribaldo, S. (2011) 'Phylogeny and evolution of the Archaea: one hundred genomes later', *Current Opinion in Microbiology*, 14(3), pp. 274-281.
- Brown, T. and Brown, D.J.S. (1991) *Oligonucleotides and Analogues*.
- Bruce Alberts, Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter., P. (2002) *Molecular Biology of the Cell, 4th edition*. New York: Garland Science.
- Byrappa, S., Gavin, D.K. and Gupta, K.C. (1995) 'A highly efficient procedure for site-specific mutagenesis of full-length plasmids using Vent DNA polymerase', *Genome research*, 5(4), pp. 404-7.

Cann, I.K., Ishino, S., Nomura, N., Sako, Y. and Ishino, Y. (1999) 'Two family B DNA polymerases from *Aeropyrum pernix*, an aerobic hyperthermophilic crenarchaeote', *J Bacteriol*, 181(19), pp. 5984-92.

Cann, I.K. and Ishino, Y. (1999) 'Archaeal DNA replication: identifying the pieces to solve a puzzle', *Genetics*, 152(4), pp. 1249-67.

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', *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), pp. 14250-14255.

Capson, T.L., Peliska, J.A., Kaboord, B.F., Frey, M.W., Lively, C., Dahlberg, M. and Benkovic, S.J. (1992) 'Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4', *Biochemistry*, 31(45), pp. 10984-94.

Carlini, L., Curth, U., Kindler, B., Urbanke, C. and Porter, R.D. (1998) 'Identification of amino acids stabilizing the tetramerization of the single stranded DNA binding protein from *Escherichia coli*', *FEBS Letters*, 430(3), pp. 197-200.

Carpentieri, F., De Felice, M., De Falco, M., Rossi, M. and Pisani, F.M. (2002) 'Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from *Mthe* crenarchaeon *Sulfolobus solfataricus*', *Journal of Biological Chemistry*, 277(14), pp. 12118-12127.

Castrec, B., Laurent, S., Henneke, G., Flament, D. and Raffin, J.P. (2010) 'The glycine-rich motif of *Pyrococcus abyssi* DNA polymerase D is critical for protein stability', *J Mol Biol*, 396(4), pp. 840-8.

Castrec, B., Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J.P. (2009) 'Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B', *J Mol Biol*, 394(2), pp. 209-18.

Chalton, D.A. and Lakey, J.H. (2010) 'Simple Detection of Protein Soft Structure Changes', *Analytical Chemistry*, 82(7), pp. 3073-3076.

Chong, J.P.J., Hayashi, M.K., Simon, M.N., Xu, R.M. and Stillman, B. (2000) 'A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase', *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), pp. 1530-1535.

Cline, J., Braman, J.C. and Hogrefe, H.H. (1996) 'PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases', *Nucleic Acids Res*, 24(18), pp. 3546-51.

Collins, K.L., Russo, A.A.R., Tseng, B.Y. and Kelly, T.J. (1993) 'The Role of the 70-Kda Subunit of Human DNA Polymerase-Alpha in DNA-Replication', *Embo Journal*, 12(12), pp. 4555-4566.

Corn, J.E. and Berger, J.M. (2006) 'Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions', *Nucleic Acids Research*, 34(15), pp. 4082-4088.

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(10), pp. 2322-8.

Curth, U., Genschel, J., Urbanke, C. and Greipel, J. (1996) 'In vitro and in vivo function of the C-terminus of *Escherichia coli* single-stranded DNA binding protein', *Nucleic Acids Research*, 24(14), pp. 2706-2711.

Curth, U., Urbanke, C., Greipel, J., Gerberding, H., Tiranti, V. and Zeviani, M. (1994) 'Single-Stranded-DNA-Binding Proteins from Human Mitochondria and Escherichia-Coli Have Analogous Physicochemical Properties', *European Journal of Biochemistry*, 221(1), pp. 435-443.

Dabrowski, S., Olszewski, M., Piatek, R., Brillowska-Dabrowska, A., Konopa, G. and Kur, J. (2002) 'Identification and characterization of single-stranded-DNA-binding proteins from *Mthermus thermophilus* and *Thermus aquaticus* - new arrangement of binding domains', *Microbiology-Sgm*, 148, pp. 3307-3315.

Dahm, R. (2008) 'Discovering DNA: Friedrich Miescher and the early years of nucleic acid research', *Human Genetics*, 122(6), pp. 565-581.

Dalhus, B., Arvai, A.S., Rosnes, I., Olsen, O.E., Backe, P.H., Alseth, I., Gao, H., Cao, Tainer, J.A and Bjoras, M (2009) "Structures of Endonuclease V with DNA Reveal Initiation of Deaminated Adenine Repair", *Nat Struct Mol Biol*.

Dickey, T.H., Altschuler, S.E. and Wuttke, D.S. (2013) 'Single-Stranded DNA-Binding Proteins: Multiple Domains for Multiple Functions', *Structure*, 21(7), pp. 1074-1084.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) "'Touchdown' PCR to circumvent spurious priming during gene amplification", *Nucleic Acids Res*, 19(14), p. 4008.

Drake, J.W. (1991) 'A constant rate of spontaneous mutation in DNA-based microbes', *Proc. Natl. Acad. Sci. USA*, 88, pp. 7160-7164.

Echols, H. and Goodman, M.F. (1991) 'Fidelity Mechanisms in DNA-Replication', *Annual Review of Biochemistry*, 60, pp. 477-511.

Eckardt, N.A. (2009) 'Functions of DNA Polymerase epsilon', *Plant Cell*, 21(2), pp. 365-365.

Edgell, D.R. and Doolittle, W.F. (1997) 'Archaea and the origin(s) of DNA replication proteins', *Cell*, 89(7), pp. 995-8.

Edgell, D.R., Klenk, H.P. and Doolittle, W.F. (1997) 'Gene duplications in evolution of archaeal family B DNA polymerases', *J Bacteriol*, 179(8), pp. 2632-40.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. and Turner, S. (2009) 'Real-time DNA sequencing from single polymerase molecules', *Science*, 323(5910), pp. 133-8.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E.V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M. and Stetter, K.O. (2008) 'A korarchaeal genome reveals insights into the evolution of the Archaea', *Proc Natl Acad Sci U S A*, 105(23), pp. 8102-7.

Emptage, K., O'Neill, R., Solovyova, A. and Connolly, B.A. (2008) 'Interplay between DNA Polymerase and Proliferating Cell Nuclear Antigen Switches Off Base Excision Repair of Uracil and Hypoxanthine during Replication in Archaea', *Journal of Molecular Biology*, 383(4), pp. 762-771.

- Erzberger, J.P. and Berger, J.M. (2006) 'Evolutionary relationships and structural mechanisms of AAA+ proteins', *Annu Rev Biophys Biomol Struct*, 35, pp. 93-114.
- Feng, H., Dong, L. and Cao, W.G. (2006) 'Catalytic mechanism of endonuclease V: A catalytic and regulatory two-metal model', *Biochemistry*, 45(34), pp. 10251-10259.
- Finnegan, D.J. (2012) 'Retrotransposons', *Curr Biol*, 22(11), pp. R432-7.
- Firbank, S.J., Wardle, J., Heslop, P., Lewis, R.J. and Connolly, B.A. (2008a) 'Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography', *Journal of Molecular Biology*, 381(3), pp. 529-539.
- Firbank, S.J., Wardle, J., Heslop, P., Lewis, R.J. and Connolly, B.A. (2008b) 'Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography', *Journal of Molecular Biology*, 381(3), pp. 529-39.
- Fogg, M.J., Pearl, L.H. and Connolly, B.A. (2002) 'Structural basis for uracil recognition by archaeal family B DNA polymerases', *Nature Structural Biology*, 9(12), pp. 922-927.
- Forterre, P., Gribaldo, S., Gadelle, D. and Serre, M.C. (2007) 'Origin and evolution of DNA topoisomerases', *Biochimie*, 89(4), pp. 427-446.
- Franklin, M.C., Wang, J. and Steitz, T.A. (2001) 'Structure of the replicating complex of a pol alpha family DNA polymerase', *Cell*, 105(5), pp. 657-67.
- Franklin, R.E. and Gosling, R.G. (1953) 'Molecular configuration in sodium *M*thymonucleate', *Nature*, 171(4356), pp. 740-1.
- Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. and Steitz, T.A. (1988) 'Cocrystal structure of an editing complex of Klenow fragment with DNA', *Proc Natl Acad Sci U S A*, 85(23), pp. 8924-8.
- Frick, D.N. and Richardson, C.C. (2001) 'DNA primases', *Annual Review of Biochemistry*, 70, pp. 39-80.
- G.M, C. (2000) *The Cell: A Molecular Approach* 2nd edition edn.
- Gabellieri, E. and Strambini, G.B. (2006) 'ANS fluorescence detects widespread perturbations of protein tertiary structure in ice', *Biophysical Journal*, 90(9), pp. 3239-3245.
- Gai, D., Chang, Y.P. and Chen, X.S. (2010) 'Origin DNA melting and unwinding in DNA replication', *Curr Opin Struct Biol*, 20(6), pp. 756-62.
- Galal, W.C., Pan, M., Kelman, Z. and Hurwitz, J. (2012) 'Characterization of DNA Primase Complex Isolated from *M*the Archaeon, *Thermococcus kodakaraensis*', *Journal of Biological Chemistry*, 287(20), pp. 16209-16219.
- Gan, G.N., Wittschieben, J.P., Wittschieben, B.O. and Wood, R.D. (2008) 'DNA polymerase zeta (pol zeta) in higher eukaryotes', *Cell Research*, 18(1), pp. 174-183.
- Garg, P. and Burgers, P.M.J. (2005) 'DNA polymerases that propagate the eukaryotic DNA replication fork', *Critical Reviews in Biochemistry and Molecular Biology*, 40(2), pp. 115-128.
- Genschel, J., Litz, L., Hubert, T., Roemling, U. and Urbanke, C. (1996) 'Isolation, sequencing and overproduction of the single-stranded DNA binding protein from *Pseudomonas aeruginosa* PAO', *Gene*, 182(1-2), pp. 137-143.
- Ghadessy, F.J. and Holliger, P. (2007) 'Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes', *Methods Mol Biol*, 352, pp. 237-48.

- 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', *Journal of Molecular Biology*, 372(4), pp. 855-863.
- Gitai, Z. (2005) 'The new bacterial cell biology: moving parts and subcellular architecture', *Cell*, 120(5), pp. 577-86.
- Goff, S.P. (1990) 'Retroviral Reverse-Transcriptase - Synthesis, Structure, and Function', *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 3(8), pp. 817-831.
- Gouge, J., Ralec, C., Henneke, G. and Delarue, M. (2012) 'Molecular Recognition of Canonical and Deaminated Bases by P. abyssi Family B DNA Polymerase', *Journal of Molecular Biology*, 423(3), pp. 315-336.
- Grabowski, B. and Kelman, Z. (2003) 'Archaeal DNA replication: Eukaryal proteins in a bacterial context', *Annual Review of Microbiology*, 57, pp. 487-516.
- Greagg, M.A., Fogg, M.J., Panayotou, G., Evans, S.J., Connolly, B.A. and Pearl, L.H. (1999) 'A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil', *Proc Natl Acad Sci U S A*, 96(16), pp. 9045-50.
- Greider, C.W. and Blackburn, E.H. (1985) 'Identification of a Specific Telomere Terminal Transferase-Activity in Tetrahymena Extracts', *Cell*, 43(2), pp. 405-413.
- Greider, C.W. and Blackburn, E.H. (1987) 'The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity', *Cell*, 51(6), pp. 887-98.
- Griep, M.A. (1995) 'Primase structure and function', *Indian J Biochem Biophys*, 32(4), pp. 171-8.
- Gueguen, Y., Rolland, J.L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., Raffin, J.P. and Dietrich, J. (2001) 'Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*', *European Journal of Biochemistry*, 268(22), pp. 5961-5969.
- Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M. and DeLong, E.F. (2006) 'Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*', *Proceedings of the National Academy of Sciences of the United States of America*, 103(48), pp. 18296-18301.
- Halldorsdottir, S., Thorolfsdottir, E.T., Spilliaert, R., Johansson, M., Thorbjarnardottir, S.H., Palsdottir, A., Hreggvidsson, G.O., Kristjansson, J.K., Holst, O. and Eggertsson, G. (1998) 'Cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12', *Applied Microbiology and Biotechnology*, 49(3), pp. 277-284.
- Hamdan, S.M. and Richardson, C.C. (2009) 'Motors, switches, and contacts in the replisome', *Annu Rev Biochem*, 78, pp. 205-43.
- Hamilton, S.C., Farchaus, J.W. and Davis, M.C. (2001) 'DNA polymerases as engines for biotechnology', *Biotechniques*, 31(2), pp. 370-+.
- Hanawalt, P.C. (2004) 'Density matters: the semiconservative replication of DNA', *Proc Natl Acad Sci U S A*, 101(52), pp. 17889-94.
- Handa, P., Acharya, N. and Varshney, U. (2001) 'Chimeras between single-stranded DNA-binding proteins from *Escherichia coli* and *Mycobacterium tuberculosis* reveal that their C-terminal domains interact with uracil DNA glycosylases', *Journal of Biological Chemistry*, 276(20), pp. 16992-16997.

Haracska, L., Prakash, S. and Prakash, L. (2003) 'Yeast DNA polymerase xi is an efficient extender of primer ends opposite from 7,8-dihydro-8-oxoguanine and O-6-methylguanine', *Molecular and Cellular Biology*, 23(4), pp. 1453-1459.

Heller, R.C. and Marians, K.J. (2006) 'Replisome assembly and the direct restart of stalled replication forks', *Nature Reviews Molecular Cell Biology*, 7(12), pp. 932-943.

Hendrickson, E.L., Kaul, R., Zhou, Y., Bovee, D., Chapman, P., Chung, J., de Macario, E.C., Dodsworth, J.A., Gillett, W., Graham, D.E., Hackett, M., Haydock, A.K., Kang, A., Land, M.L., Levy, R., Lie, T.J., Major, T.A., Moore, B.C., Porat, I., Palmeiri, A., Rouse, G., Saenphimmachak, C., Soll, D., Van Dien, S., Wang, T., Whitman, W.B., Xia, Q., Zhang, Y., Larimer, F.W., Olson, M.V. and Leigh, J.A. (2004) 'Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*', *Journal of Bacteriology*, 186(20), pp. 6956-6969.

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', *Journal of Molecular Biology*, 350(1), pp. 53-64.

Henry, A.A. and Romesberg, F.E. (2005) 'The evolution of DNA polymerases with novel activities', *Current Opinion in Biotechnology*, 16(4), pp. 370-377.

Hershey, A.D. and Chase, M. (1952) 'Independent functions of viral protein and nucleic acid in growth of bacteriophage', *J Gen Physiol*, 36(1), pp. 39-56.

Heyduk, T. and Lee, J.C. (1990) 'Application of fluorescence energy transfer and polarization to monitor *Escherichia coli* cAMP receptor protein and lac promoter interaction', *Proc Natl Acad Sci U S A*, 87(5), pp. 1744-8.

Holland, P.A., Abramson, R.D., Watson, R. and Galfand, D.H. (1991) 'Detection of specific polymerase chain reaction product by utilizing the 5' - 3' exonuclease activity of *Thermus aquaticus* DNA polymerase', *Proc. Natl. Acad. Sci. USA*, 88, pp. 7276-7280.

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 *Mthermococcus gorgonarius*', *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp. 3600-3605.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C. and Stetter, K.O. (2002) 'A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont', *Nature*, 417(6884), pp. 63-7.

Hubscher, U., Maga, G. and Spadari, S. (2002) 'Eukaryotic DNA polymerases', *Annual Review of Biochemistry*, 71, pp. 133-63.

Hubscher, U., Nasheuer, H.P. and Syvaaja, J.E. (2000) 'Eukaryotic DNA polymerases, a growing family', *Trends in Biochemical Sciences*, 25(3), pp. 143-147.

Huet, J., Schnabel, R., Sentenac, A. and Zillig, W. (1983) 'Archaeobacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type', *EMBO J*, 2(8), pp. 1291-4.

Hughes, R.A., Miklos, A.E. and Ellington, A.D. (2011) 'Gene synthesis: methods and applications', *Methods Enzymol*, 498, pp. 277-309.

Igor V. Shevelev, U.H. (2002) 'The 3'-5' exonucleases', *Nature Reviews Molecular Cell Biology* 3, pp. 364-376

Imhof, P., Fischer, S. and Smith, J.C. (2009) 'Catalytic mechanism of DNA backbone cleavage by the restriction enzyme EcoRV: a quantum mechanical/molecular mechanical analysis', *Biochemistry*, 48(38), pp. 9061-75.

Isaac K. O. Cann, S.I., 1 Ikuko Hayashi, 2 Kayoko Komori, 1 Hiroyuki Toh, 3 Kosuke Morikawa, 2 and Yoshizumi Ishino, 1,* (1999) 'Functional Interactions of a Homolog of Proliferating Cell Nuclear Antigen with DNA Polymerases in Archaea', *Journal of Bacteriology*.

Ishino, Y. and Ishino, S. (2012) 'Rapid progress of DNA replication studies in Archaea, the third domain of life', *Sci China Life Sci*, 55(5), pp. 386-403.

Ishino, Y., Komori, K., Cann, I.K.O. and Koga, Y. (1998) 'A novel DNA polymerase family found in Archaea', *Journal of Bacteriology*, 180(8), pp. 2232-2236.

Ito, J. and Braithwaite, D.K. (1991) 'Compilation and Alignment of DNA-Polymerase Sequences', *Nucleic Acids Research*, 19(15), pp. 4045-4057.

Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2002) 'Clamp loaders and sliding clamps', *Curr Opin Struct Biol*, 12(2), pp. 217-24.

Jin, Y.H., Ayyagari, R., Resnick, M.A., Gordenin, D.A. and Burgers, P.M. (2003) 'Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick', *J Biol Chem*, 278(3), pp. 1626-33.

Jokela, M., Eskelinen, A., Pospiech, H., Rouvinen, J. and Syvaöja, J.E. (2004a) 'Characterization of the 3' exonuclease subunit DP1 of Methanococcus jannaschii replicative DNA polymerase D', *Nucleic Acids Res*, 32(8), pp. 2430-40.

Jokela, M., Eskelinen, A., Pospiech, H., Rouvinen, J. and Syvaöja, J.E. (2004b) 'Characterization of the 3' exonuclease subunit DP1 of Methanococcus jannaschii replicative DNA polymerase D', *Nucleic Acids Research*, 32(8), pp. 2430-2440.

Joyce, C.M. and Benkovic, S.J. (2004) 'DNA polymerase fidelity: Kinetics, structure, and checkpoints', *Biochemistry*, 43(45), pp. 14317-14324.

Kaguni, J.M. (2011) 'Replication initiation at the Escherichia coli chromosomal origin', *Current Opinion in Chemical Biology*, 15(5), pp. 606-613.

Kawasaki, Y. and Sugino, A. (2002) 'DNA polymerase epsilon', *Seikagaku*, 74(3), pp. 197-202.

Keith, B.J., Jozwiakowski, S.K. and Connolly, B.A. (2013) 'A plasmid-based lacZalpha gene assay for DNA polymerase fidelity measurement', *Anal Biochem*, 433(2), pp. 153-61.

Kelly, T.J., Simanek, P. and Brush, G.S. (1998) 'Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus jannaschii', *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), pp. 14634-14639.

Kelman, L.M. and Kelman, Z. (2004) 'Multiple origins of replication in archaea', *Trends Microbiol*, 12(9), pp. 399-401.

Kelman, Z., Pietrokovski, S. and Hurwitz, J. (1999) 'Isolation and characterization of a split B-type DNA polymerase from the archaeon Methanobacterium thermoautotrophicum Delta H', *Journal of Biological Chemistry*, 274(40), pp. 28751-28761.

Kelman, Z. and White, M.F. (2005) 'Archaeal DNA replication and repair', *Curr Opin Microbiol*, 8(6), pp. 669-76.

Killelea, T. and Connolly, B.A. (2011) 'Role of disulfide bridges in archaeal family-B DNA polymerases', *ChemBiochem*, 12(9), pp. 1330-6.

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 Isosteric Base Analogues', *Biochemistry*, 49(27), pp. 5772-5781.
- Klinge, S., Nunez-Ramirez, R., Llorca, O. and Pellegrini, L. (2009) '3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases', *Embo Journal*, 28(13), pp. 1978-1987.
- Kocalka, P., El-Sagheer, A.H. and Brown, T. (2008) 'Rapid and efficient DNA strand cross-linking by click chemistry', *ChemBiochem*, 9(8), pp. 1280-5.
- Kool, E.T. (2001) 'Hydrogen bonding, base stacking, and steric effects in DNA replication', *Annual Review of Biophysics and Biomolecular Structure*, 30, pp. 1-22.
- Kool, E.T. (2002) 'Active site tightness and substrate fit in DNA replication', *Annual Review of Biochemistry*, 71, pp. 191-219.
- Kornberg, T. and Gefter, M.L. (1971) 'Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II', *Proceedings of the National Academy of Sciences of the United States of America*, 68(4), pp. 761-4.
- Kranaster, R. and Marx, A. (2010) 'Engineered DNA Polymerases in Biotechnology', *ChemBiochem*, 11(15), pp. 2077-2084.
- Kunkel, T.A. (2004) 'DNA replication fidelity', *Journal of Biological Chemistry*, 279(17), pp. 16895-16898.
- Kunkel, T.A. and Alexander, P.S. (1986) 'The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation', *J Biol Chem*, 261(1), pp. 160-6.
- Kunkel, T.A. and Bebenek, R. (2000) 'DNA replication fidelity', *Annual Review of Biochemistry*, 69, pp. 497-529.
- Lecointe, F., Serena, C., Velten, M., Costes, A., McGovern, S., Meile, J.C., Errington, J., Ehrlich, S.D., Noirot, P. and Polard, P. (2007) 'Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks', *Embo Journal*, 26(19), pp. 4239-4251.
- Lehman, I.R. (1974) 'DNA ligase: structure, mechanism, and function', *Science*, 186(4166), pp. 790-7.
- Lehman, I.R., Zimmerman, S.B., Adler, J., Bessman, M.J., Simms, E.S. and Kornberg, A. (1958) 'Enzymatic Synthesis of Deoxyribonucleic Acid. V. Chemical Composition of Enzymatically Synthesized Deoxyribonucleic Acid', *Proc Natl Acad Sci U S A*, 44(12), pp. 1191-6.
- Leigh, J.A., Albers, S.V., Atomi, H. and Allers, T. (2011) 'Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales', *FEMS Microbiol Rev*, 35(4), pp. 577-608.
- Leipe, D.D., Aravind, L. and Koonin, E.V. (1999) 'Did DNA replication evolve twice independently?', *Nucleic Acids Research*, 27(17), pp. 3389-3401.
- Li, Z., Santangelo, T.J., Cubonova, L., Reeve, J.N. and Kelman, Z. (2010) 'Affinity purification of an archaeal DNA replication protein network', *MBio*, 1(5).
- LiCata, V.J. and Wowor, A.J. (2008) 'Applications of fluorescence anisotropy to the study of protein-DNA interactions', *Methods Cell Biol*, 84, pp. 243-62.
- Lin, T.C., Karam, G. and Konigsberg, W.H. (1994) 'Isolation, characterization, and kinetic properties of truncated forms of T4 DNA polymerase that exhibit 3'-5' exonuclease activity', *Journal of Biological Chemistry*, 269(30), pp. 19286-94.

- Lin, Y.L., Shivji, M.K.K., Chen, C., Kolodner, R., Wood, R.D. and Dutta, A. (1998) 'The evolutionarily conserved zinc finger motif in the largest subunit of human replication protein a is required for DNA replication and mismatch repair but not for nucleotide excision repair', *Journal of Biological Chemistry*, 273(3), pp. 1453-1461.
- Ling, H., Boudsocq, F., Plosky, B.S., Woodgate, R. and Yang, W. (2003) 'Replication of a cis-syn thymine dimer at atomic resolution', *Nature*, 424(6952), pp. 1083-1087.
- Ling, H., Boudsocq, F., Woodgate, R. and Yang, W. (2001) 'Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication', *Cell*, 107(1), pp. 91-102.
- Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. and Cech, T.R. (1997) 'Reverse transcriptase motifs in the catalytic subunit of telomerase', *Science*, 276(5312), pp. 561-7.
- Liu, Y., Kao, H.I. and Bambara, R.A. (2004) 'Flap endonuclease 1: A central component of DNA metabolism', *Annual Review of Biochemistry*, 73, pp. 589-615.
- Lopez-Olmos, K., Hernandez, M.P., Contreras-Garduno, J.A., Robleto, E.A., Setlow, P., Yasbin, R.E. and Pedraza-Reyes, M. (2012) 'Roles of Endonuclease V, Uracil-DNA Glycosylase, and Mismatch Repair in Bacillus subtilis DNA Base-Deamination-Induced Mutagenesis', *Journal of Bacteriology*, 194(2), pp. 243-252.
- Margulis, L. (1996) 'Archaeal-eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life', *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp. 1071-1076.
- Matsui, I., Urushibata, Y., Shen, Y.L., Matsui, E. and Yokoyama, H. (2011) 'Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from Pyrococcus horikoshii', *Febs Letters*, 585(3), pp. 452-458.
- Matsumoto, Y. and Kim, K. (1995) 'Excision of Deoxyribose Phosphate Residues by DNA-Polymerase-Beta during DNA-Repair', *Science*, 269(5224), pp. 699-702.
- Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. (2003) 'Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin', *EMBO Rep*, 4(2), pp. 154-8.
- McCulloch, S.D. and Kunkel, T.A. (2008) 'The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases', *Cell Res*, 18(1), pp. 148-61.
- Meselson, M. and Stahl, F.W. (1958) 'The Replication of DNA in Escherichia Coli', *Proc Natl Acad Sci U S A*, 44(7), pp. 671-82.
- Meyer, R.R. and Laine, P.S. (1990) 'The Single-Stranded DNA-Binding Protein of Escherichia-Coli', *Microbiological Reviews*, 54(4), pp. 342-380.
- Mi, R., Abole, A.K. and Cao, W. (2011) 'Dissecting endonuclease and exonuclease activities in endonuclease V from *Mthermotoga maritima*', *Nucleic Acids Res*, 39(2), pp. 536-44.
- Michel, B., Grompone, G., Flores, M.J. and Bidnenko, V. (2004) 'Multiple pathways process stalled replication forks', *Proceedings of the National Academy of Sciences of the United States of America*, 101(35), pp. 12783-12788.
- Mijakovic, I., Petranovic, D., Macek, B., Cepo, T., Mann, M., Davies, J., Jensen, P.R. and Vujaklija, D. (2006) 'Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine', *Nucleic Acids Research*, 34(5), pp. 1588-1596.

Miller-Coleman, R.L., Dodsworth, J.A., Ross, C.A., Shock, E.L., Williams, A.J., Hartnett, H.E., McDonald, A.I., Havig, J.R. and Hedlund, B.P. (2012) 'Korarchaeota diversity, biogeography, and abundance in Yellowstone and Great Basin hot springs and ecological niche modeling based on machine learning', *PLoS One*, 7(5), p. e35964.

Miyabe, I., Kunkel, T.A. and Carr, A.M. (2011) 'The Major Roles of DNA Polymerases Epsilon and Delta at the Eukaryotic Replication Fork Are Evolutionarily Conserved', *Plos Genetics*, 7(12).

Moe, E., Leiros, I., Smalas, A.O. and McSweeney, S. (2006) 'The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from *Deinococcus radiodurans* reveals a novel catalytic residue and broad substrate specificity', *Journal of Biological Chemistry*, 281(1), pp. 569-577.

Morita, M., Stamp, G., Robins, P., Dulic, A., Rosewell, I., Hrivnak, G., Daly, G., Lindahl, T. and Barnes, D.E. (2004) 'Gene-Targeted Mice Lacking the *Trex1* (DNase III) 3'→5' DNA Exonuclease Develop Inflammatory Myocarditis', *Mol Cell Biol*, 3, pp. 6719–6727.

Mott, M.L. and Berger, J.M. (2007) 'DNA replication initiation: mechanisms and regulation in bacteria', *Nat Rev Microbiol*, 5(5), pp. 343-54.

Murzin, A.G. (1993) 'Ob(Oligonucleotide Oligosaccharide Binding)-Fold - Common Structural and Functional Solution for Nonhomologous Sequences', *Embo Journal*, 12(3), pp. 861-867.

Muzi-Falconi, M., Giannattasio, M., Foiani, M. and Plevani, P. (2003) 'The DNA polymerase alpha-primase complex: multiple functions and interactions', *ScientificWorldJournal*, 3, pp. 21-33.

Myllykallio, H. and Forterre, P. (2000) 'Mapping of a chromosome replication origin in an archaeon: response', *Trends Microbiol*, 8(12), pp. 537-9.

Nakagawa, S., Shtaih, Z., Banta, A., Beveridge, T.J., Sako, Y. and Reysenbach, A.L. (2005) 'Sulfurihydrogenibium yellowstonense sp nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium azorense*', *International Journal of Systematic and Evolutionary Microbiology*, 55, pp. 2263-2268.

Nasheuer, H.P., Smith, R., Bauerschmidt, C., Grosse, F. and Weisshart, K. (2002) 'Initiation of eukaryotic DNA replication: regulation and mechanisms', *Prog Nucleic Acid Res Mol Biol*, 72, pp. 41-94.

Netz, D.J.A., Stith, C.M., Stumpfig, M., Kopf, G., Vogel, D., Genau, H.M., Stodola, J.L., Lill, R., Burgers, P.M.J. and Pierik, A.J. (2012) 'Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes', *Nature Chemical Biology*, 8(1), pp. 125-132.

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(9), pp. 2212-2221.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori, M., Kanai, A., Atomi, H., Takai, K. and Takami, H. (2011) 'Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group', *Nucleic Acids Research*, 39(8), pp. 3204-3223.

O'Donnell, M., Langston, L. and Stillman, B. (2013) 'Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya', *Cold Spring Harbor Perspectives in Biology*, 5(7).

Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G.C.,

- Wang, Z. and Woodgate, R. (2001) 'The Y-family of DNA polymerases', *Molecular cell*, 8(1), pp. 7-8.
- Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A. (1968) 'Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains', *Proc Natl Acad Sci U S A*, 59(2), pp. 598-605.
- Otvos, J.D. and Armitage, I.M. (1980) 'Structure of the metal clusters in rabbit liver metallothionein', *Proc Natl Acad Sci U S A*, 77(12), pp. 7094-8.
- Passerini, A., Punta, M., Ceroni, A., Rost, B. and Frasconi, P. (2006) 'Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks', *Proteins*, 65(2), pp. 305-16.
- Paytubi, S., McMahon, S.A., Graham, S., Liu, H.T., Botting, C.H., Makarova, K.S., Koonin, E.V., Naismith, J.H. and White, M.F. (2012) 'Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales', *Proceedings of the National Academy of Sciences of the United States of America*, 109(7), pp. E398-E405.
- Paz-Elizur, T., Takeshita, M., Goodman, M., O'Donnell, M. and Livneh, Z. (1996) 'Mechanism of translesion DNA synthesis by DNA polymerase II. Comparison to DNA polymerases I and III core', *The Journal of biological chemistry*, 271(40), pp. 24662-9.
- Pearl, L.H. (2000a) 'Structure and function in the uracil-DNA glycosylase superfamily', *Mutat Res*, 460(3-4), pp. 165-81.
- Pearl, L.H. (2000b) 'Structure and function in the uracil-DNA glycosylase superfamily', *Mutation Research-DNA Repair*, 460(3-4), pp. 165-181.
- Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. and Kraut, J. (1994) 'Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP', *Science*, 264(5167), pp. 1891-903.
- Petruska, J. and Goodman, M.F. (1995) 'Enthalpy-Entropy Compensation in DNA Melting Thermodynamics', *Journal of Biological Chemistry*, 270(2), pp. 746-750.
- Pospiech, H. and Syvaaja, J.E. (2003) 'DNA polymerase epsilon - more than a polymerase', *ScientificWorldJournal*, 3, pp. 87-104.
- Purnapatre, K. and Varshney, U. (1999) 'Cloning, over-expression and biochemical characterization of the single-stranded DNA binding protein from Mycobacterium tuberculosis', *European Journal of Biochemistry*, 264(2), pp. 591-598.
- Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. and Kunkel, T.A. (2007) 'Yeast DNA polymerase epsilon participates in leading-strand DNA replication', *Science*, 317(5834), pp. 127-130.
- Qiu, J.Z., Qian, Y., Frank, P., Wintersberger, U. and Shen, B.H. (1999) 'Saccharomyces cerevisiae RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease', *Molecular and Cellular Biology*, 19(12), pp. 8361-8371.
- Rangarajan, S., Woodgate, R. and Goodman, M.F. (1999) 'A phenotype for enigmatic DNA polymerase II: A pivotal role for pol II in replication restart in UV-irradiated Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), pp. 9224-9229.
- Razvi, A. and Scholtz, J.M. (2006) 'Lessons in stability from thermophilic proteins', *Protein Science*, 15(7), pp. 1569-1578.

Reha-Krantz, L.J. (2010) 'DNA polymerase proofreading: Multiple roles maintain genome stability', *Biochim Biophys Acta*, 1804(5), pp. 1049-63.

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(7), pp. 4207-4218.

Richard R. Sinden, Christopher E. Pearson, Vladimir N. Potaman and Ussery, D.W. (1998) 'DNA: Structure and Function', *Advances in genome biology*, 5(1-141).

Robinson, N.P. and Bell, S.D. (2005) 'Origins of DNA replication in the three domains of life', *FEBS J*, 272(15), pp. 3757-66.

Rothwell, P.J. and Waksman, G. (2005) 'Structure and mechanism of DNA polymerases', *Adv Protein Chem*, 71, pp. 401-40.

Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J.P. (2007) 'DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*', *Journal of Molecular Biology*, 369(2), pp. 343-355.

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 Research*, 37(22), pp. 7603-7611.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) 'Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA-Polymerase', *Science*, 239(4839), pp. 487-491.

Sarmiento, F., Mrazek, J. and Whitman, W.B. (2013) 'Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*', *Proceedings of the National Academy of Sciences of the United States of America*, 110(12), pp. 4726-4731.

Sawai, H., Nagashima, J., Kuwahara, M., Kitagata, R., Tamura, T. and Matsui, I. (2007) 'Differences in substrate specificity of C(5)-substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from thermophilic bacteria, archaea, and phages', *Chem Biodivers*, 4(9), pp. 1979-95.

Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. and Kraut, J. (1994) 'Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism', *Science*, 264(5167), pp. 1930-5.

Schroeder, G.K. and Wolfenden, R. (2007) 'Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases', *Biochemistry*, 46(47), pp. 13638-13647.

Shechter, D.F., Ying, C.Y. and Gautier, J. (2000) 'The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* Delta H minichromosome maintenance protein', *Journal of Biological Chemistry*, 275(20), pp. 15049-15059.

Shen, Y., Tang, X.F., Yokoyama, H., Matsui, E. and Matsui, I. (2004) 'A 21-amino acid peptide from the cysteine cluster II of the family D DNA polymerase from *Pyrococcus horikoshii* stimulates its nuclease activity which is Mre11-like and prefers manganese ion as the cofactor', *Nucleic Acids Res*, 32(1), pp. 158-68.

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*', *Journal of Biological Chemistry*, 276(29), pp. 27376-27383.

- Shih, Y.L. and Rothfield, L. (2006) 'The bacterial cytoskeleton', *Microbiology and Molecular Biology Reviews*, 70(3), pp. 729-+.
- Shikata, K., Sasa-Masuda, T., Okuno, Y., Waga, S. and Sugino, A. (2006) 'The DNA polymerase activity of Pol epsilon holoenzyme is required for rapid and efficient chromosomal DNA replication in *Xenopus* egg extracts', *BMC Biochem*, 7, p. 21.
- Shimizu, K., Hashimoto, K., Kirchner, J.M., Nakai, W., Nishikawa, H., Resnick, M.A. and Sugino, A. (2002) 'Fidelity of DNA polymerase epsilon holoenzyme from budding yeast *Saccharomyces cerevisiae*', *Journal of Biological Chemistry*, 277(40), pp. 37422-37429.
- Shuttleworth, G., Fogg, M.J., Kurpiewski, M.R., Jen-Jacobson, L. and Connolly, B.A. (2004) 'Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea', *Journal of Molecular Biology*, 337(3), pp. 621-634.
- Silva, Z., Borges, N., Martins, L.O., Wait, R., da Costa, M.S. and Santos, H. (1999) 'Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by *Rhodothermus marinus* and *Rhodothermus obamensis*', *Extremophiles*, 3(2), pp. 163-172.
- Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N. and et al. (1997) 'Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics', *J Bacteriol*, 179(22), pp. 7135-55.
- Steitz, T.A. (1999) 'DNA polymerases: structural diversity and common mechanisms', *J Biol Chem*, 274(25), pp. 17395-8.
- Stillman, B. (2005) 'Origin recognition and the chromosome cycle', *FEBS Lett*, 579(4), pp. 877-84.
- Sun, J., Kawakami, H., Zech, J., Speck, C., Stillman, B. and Li, H. (2012) 'Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy', *Structure*, 20(3), pp. 534-44.
- Tahirov, T.H., Makarova, K.S., Rogozin, I.B., Pavlov, Y.I. and Koonin, E.V. (2009) 'Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors', *Biology Direct*, 4.
- Tang, X.F., Shen, Y.L., Matsui, E. and Matsui, I. (2004) 'Domain topology of the DNA polymerase D complex from a hyperthermophilic Archaeon *Pyrococcus horikoshii*', *Biochemistry*, 43(37), pp. 11818-11827.
- Tori, K., Kimizu, M., Ishino, S. and Ishino, Y. (2007) 'DNA Polymerases BI and D from the hyperthermophilic Archaeon *Pyrococcus furiosus* both bind to proliferating cell nuclear antigen with their C-terminal PIP-Box motifs', *Journal of Bacteriology*, 189(15), pp. 5652-5657.
- Tye, B.K. (1999) 'MCM proteins in DNA replication', *Annual Review of Biochemistry*, 68, pp. 649-686.
- Tye, B.K. and Sawyer, S. (2000) 'The hexameric eukaryotic MCM helicase: Building symmetry from nonidentical parts', *Journal of Biological Chemistry*, 275(45), pp. 34833-34836.
- Uemori, T., Sato, Y., Kato, I., Doi, H. and Ishino, Y. (1997a) 'A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization', *Genes to Cells*, 2(8), pp. 499-512.

- Uemori, T., Sato, Y., Kato, I., Doi, H. and Ishino, Y. (1997b) 'A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization', *Genes Cells*, 2(8), pp. 499-512.
- Wang, F. and Yang, W. (2009) 'Structural insight into translesion synthesis by DNA Pol II', *Cell*, 139(7), pp. 1279-89.
- Wang, J., Sattar, A.K.M.A., Wang, C.C., Karam, J.D., Konigsberg, W.H. and Steitz, T.A. (1997) 'Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69', *Cell*, 89(7), pp. 1087-1099.
- Wardle, J., Burgers, P.M., Cann, I.K., 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(3), pp. 705-11.
- Washington, M.T., Johnson, R.E., Prakash, S. and Prakash, L. (1999) 'Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase eta', *Journal of Biological Chemistry*, 274(52), pp. 36835-36838.
- Watson, J.D. and Crick, F.H. (1953) 'Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid', *Nature*, 171(4356), pp. 737-8.
- Weiner, M.P., Costa, G.L., Schoettlin, W., Cline, J., Mathur, E. and Bauer, J.C. (1994) 'Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction', *Gene*, 151(1-2), pp. 119-23.
- White, D. (2007) *The physiology and biochemistry of prokaryotes*, Oxford University press.
- Williams, K.R., Spicer, E.K., Lopresti, M.B., Guggenheimer, R.A. and Chase, J.W. (1983) 'Limited Proteolysis Studies on the Escherichia-Coli Single-Stranded-DNA Binding-Protein - Evidence for a Functionally Homologous Domain in Both the Escherichia-Coli and T4 DNA-Binding Proteins', *Journal of Biological Chemistry*, 258(5), pp. 3346-3355.
- Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) 'Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids', *Microbiol Rev*, 47(4), pp. 621-69.
- Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) 'Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya', *Proc Natl Acad Sci U S A*, 87(12), pp. 4576-9.
- Wold, M.S. (1997) 'Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism', *Annual Review of Biochemistry*, 66, pp. 61-92.
- Wu, Y. (2012) 'Unwinding and rewinding: double faces of helicase?', *J Nucleic Acids*, 2012, p. 140601.
- Yamasaki, K., Urushibata, Y., Yamasaki, T., Arisaka, F. and Matsui, I. (2010) 'Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases', *Febs Letters*, 584(15), pp. 3370-3375.
- Yamitch, J. and Sweasy, J.B. (2010) 'DNA polymerase Family X: Function, structure, and cellular roles', *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1804(5), pp. 1136-1150.
- Yang, W. and Woodgate, R. (2007) 'What a difference a decade makes: Insights into translesion DNA synthesis', *Proceedings of the National Academy of Sciences of the United States of America*, 104(40), pp. 15591-15598.

- Yao, N.Y. and O'Donnell, M. (2009) 'Replisome structure and conformational dynamics underlie fork progression past obstacles', *Curr Opin Cell Biol*, 21(3), pp. 336-43.
- You, Z., De Falco, M., Kamada, K., Pisani, F.M. and Masai, H. (2013) 'The Mini-Chromosome Maintenance (Mcm) Complexes Interact with DNA Polymerase alpha-Primase and Stimulate Its Ability to Synthesize RNA Primers', *PLoS One*, 8(8), p. e72408.
- Yulong Shen, K.M., Madoka Hiramoto, Hisasa Kikuchi, Yataka Kawarabayashi, and Ikuo Matsui (2001) 'Invariant Asp-1122 and Asp-1124 Are Essential Residues for Polymerization Catalysis of Family D DNA Polymerase from *Pyrococcus horikoshii*', *The American Society for Biochemistry and Molecular Biology*, 276(29), pp. 27376-27383.
- Yutin, N., Makarova, K.S., Mekhedov, S.L., Wolf, Y.I. and Koonin, E.V. (2008) 'The deep archaeal roots of eukaryotes', *Mol Biol Evol*, 25(8), pp. 1619-30.
- Zhong, X.J., Garg, P., Stith, C.M., McElhinny, S.A.N., Kissling, G.E., Burgers, P.M.J. and Kunkel, T.A. (2006) 'The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins', *Nucleic Acids Research*, 34(17), pp. 4731-4742.
- Zhou, B.L., Pata, J.D. and Steitz, T.A. (2001) 'Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain', *Molecular Cell*, 8(2), pp. 427-437.
- Zhuo Lia, T.J.S., Ľubomíra Āuboňováb, John N. Reeveb, and Zvi Kelmana (2012) 'Affinity Purification of an Archaeal DNA Replication Protein Network', *American Society for Microbiology*, 1(5).

Appendix 1

Mass spectrometry results identifying endogenous *Methanococcus maripaludis* proteins.



Mascot Search Results

User : tf2013
Email :
Search title : Ultraflex_ProteinID
MS data file : 13229323905423084.mgf
Database : NCBIInr 20101130 (12348165 sequences; 4221604734 residues)
Timestamp : 12 Apr 2013 at 08:41:43 GMT
Enzyme : Trypsin
Fixed modifications : [Carbamidomethyl \(C\)](#)
Variable modifications : [Oxidation \(M\)](#)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance : ± 0.5 Da
Max Missed Cleavages : 1
Instrument type : MALDI-TOF-TOF
Number of queries : 10
Protein hits : [gi|45357946](#) S-layer protein [Methanococcus maripaludis S2]
[gi|45358932](#) elongation factor EF-2 [Methanococcus maripaludis S2]

All queries Unassigned Below homology threshold Below identity threshold

1.	gi 45357946	Mass: 58912	Score: 494	Matches: 4(4)	Sequences: 4(4)					
S-layer protein [Methanococcus maripaludis S2]										
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1	1029.5135	1028.5063	1028.5179	-11.26	0	71	0.0017	1	U	K.TFTIADYAK.L
2	1341.7606	1340.7533	1340.7738	-15.23	0	94	6.4e-06	1	U	R.IPLLQEMVVVK.L
7	1743.8859	1742.8786	1742.9203	-23.93	0	122	9.2e-09	1	U	K.FDVTDKPAAPAAATQLK.L
9	2137.0000	2135.9927	2136.0474	-25.60	0	208	1.9e-17	1	U	K.LDTDDDLAIVGTEAYEGVIK.Q
<hr/>										
2.	gi 45358932	Mass: 80653	Score: 328	Matches: 3(3)	Sequences: 3(3)					
elongation factor EF-2 [Methanococcus maripaludis S2]										
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
6	1716.8481	1715.8409	1715.8730	-18.74	0	136	3.4e-10	1	U	R.DGGIEVDVGEPIVVYR.E
8	1991.8816	1990.8743	1990.9119	-18.90	0	133	5e-10	1	U	K.ELAGDQLALDFDEEEAAR.G
10	2331.1108	2330.1036	2330.1543	-21.77	0	58	0.016	1	U	K.EYLINLIDTPGHVDFGGDVTR.A



Mascot Search Results

User : tf2013
 Email :
 Search title : Ultraflex_ProteinID
 MS data file : 13229323905423085.mgf
 Database : NCBI nr 20101130 (12348165 sequences; 4221604734 residues)
 Timestamp : 12 Apr 2013 at 08:44:34 GMT
 Enzyme : Trypsin
 Fixed modifications : [Carbamidomethyl \(C\)](#)
 Variable modifications : [Oxidation \(M\)](#)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 100 ppm
 Fragment Mass Tolerance : ± 0.5 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 10
 Protein hits : [gi|45358607](#) V-type ATP synthase subunit A [Methanococcus maripaludis S2]

Select Summary Report

Format As	Select Summary (protein hits) ▾	Help
Significance threshold p<	0.05	Max. number of hits AUTO
Standard scoring	<input checked="" type="radio"/> MudPIT scoring <input type="radio"/> Ions score or expect cut-off	0.05
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups <input type="radio"/>	Show sub-sets 0
		Require bold red <input checked="" type="checkbox"/>

All queries
 Unassigned
 Below homology threshold
 Below identity threshold

1. [gi|45358607](#) Mass: 64381 Score: 588 Matches: 6(6) Sequences: 6(6)
 V-type ATP synthase subunit A [Methanococcus maripaludis S2]

Query	Observed	Mr (expt)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3	1275.7045	1274.6972	1274.6771	15.7	0	60	0.017	1	U	K.VFWALDANLAR.R
4	1292.7732	1291.7659	1291.7612	3.64	0	66	0.003	1	U	K.QAPVIPLITGQR.V
6	1532.7966	1531.7894	1531.7842	3.36	0	107	2.9e-07	1	U	R.DQGLGVLLTADSTR.W
7	1719.8489	1718.8416	1718.8515	-5.78	0	133	6.8e-10	1		R.EASVYTGITIAEYFR.D
8	1753.8154	1752.8082	1752.8141	-3.42	0	85	4e-05	1	U	K.WSDVDVVVYIGCGER.G
10	2092.0728	2091.0655	2091.0848	-9.23	0	136	2.7e-10	1	U	K.EAELQEIVLVGPDALPDR.E



Mascot Search Results

User : tf2013
 Email :
 Search title : Ultraflex_ProteinID
 MS data file : 13229323905423086.mgf
 Database : NCBIInr 20101130 (12348165 sequences; 4221604734 residues)
 Timestamp : 12 Apr 2013 at 08:47:15 GMT
 Enzyme : Trypsin
 Fixed modifications : [Carbamidomethyl \(C\)](#)
 Variable modifications : [Oxidation \(M\)](#)
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 100 ppm
 Fragment Mass Tolerance: ± 0.5 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 10
 Protein hits : [gi|7331218](#) keratin 1 [Homo sapiens]
 [gi|28317](#) unnamed protein product [Homo sapiens]

Select Summary Report

Format As	Select Summary (protein hits) ▾	Help
Significance threshold p<	0.05	Max. number of hits AUTO
Standard scoring	<input checked="" type="radio"/> Standard scoring <input type="radio"/> MudPIT scoring <input type="radio"/> Ions score or expect cut-off 0.05	Show sub-sets 0
Show pop-ups	<input checked="" type="radio"/> Show pop-ups <input type="radio"/> Suppress pop-ups	Require bold red <input checked="" type="checkbox"/>

All queries
 Unassigned
 Below homology threshold
 Below identity threshold

1. [gi|7331218](#) Mass: 66149 Score: 414 Matches: 5(5) Sequences: 5(5)
keratin 1 [Homo sapiens]

Query	Observed	Mr (expt)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2	1179.6089	1178.6016	1178.5931	7.20	0	67	0.0036	1		K.YEELQITAGR.H
3	1277.7216	1276.7143	1276.7027	9.10	0	75	0.00047	1	U	K.LALDLEIATYR.T
6	1475.7646	1474.7574	1474.7416	10.7	0	71	0.0014	1	U	K.WELLQQVDTSTR.T
7	1716.8561	1715.8488	1715.8438	2.91	0	94	5.2e-06	1	U	K.QISNLQQSISDAEQR.G
9	1993.9739	1992.9666	1992.9693	-1.38	0	110	1.3e-07	1	U	R.THNLEPYFESFINNLR.R

Appendix 2

Mass spectrometry results identifying *Methanococcus jannaschii* DP2 protein extracted from *M*macells.



Mascot Search Results

User : tf2013
Email :
Search title : Ultraflex ProteinID
MS data file : 13229323905422739.mgf
Database : NCBI nr 20101130 (12348165 sequences; 4221604734 residues)
Taxonomy : Archaea (Archaeobacteria) (703723 sequences)
Timestamp : 12 Mar 2013 at 13:06:45 GMT
Enzyme : Trypsin
Fixed modifications : [Carbamidomethyl \(C\)](#)
Variable modifications : [Oxidation \(M\)](#)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance : ± 0.5 Da
Max Missed Cleavages : 1
Instrument type : MALDI-TOF-TOF
Number of queries : 12
Protein hits : [gi|45358576](#) carbamoyl-phosphate synthase large subunit [Methanococcus maripaludis S2]

Select Summary Report

Format As	Select Summary (protein hits) ▾	Help
	Significance threshold p < 0.05	Max. number of hits AUTO
	Standard scoring <input checked="" type="radio"/> MudPIT scoring <input type="radio"/> Ions score or expect cut-off 0.05	Show sub-sets 0
	Show pop-ups <input checked="" type="radio"/> Suppress pop-ups <input type="radio"/>	Require bold red <input checked="" type="checkbox"/>

Re-Search All queries Unassigned Below homology threshold Below identity threshold

1. [gi|45358576](#) Mass: 120047 Score: 106 Matches: 2 (2) Sequences: 2 (2)
carbamoyl-phosphate synthase large subunit [Methanococcus maripaludis S2]

Query	Observed	Mr (expt)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
6	1345.7271	1344.7198	1344.7038	11.9	0	57	0.0016	1	U	K.DGVVYIIEANPR.A
8	1419.7581	1418.7508	1418.7306	14.2	0	49	0.011	1	U	K.YHNLGFIVATR.G

Appendix 3

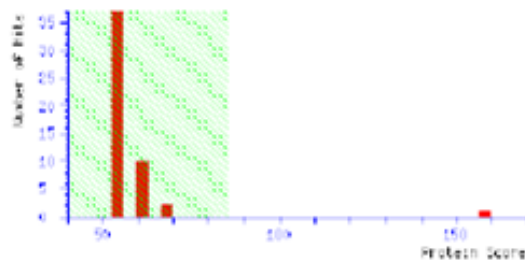
Mass spectrometry results identifying *Haloferax volcanii* PitA protein.

MASCOT Mascot Search Results

User : JoeG
Email : joe.gray@ncl.ac.uk
Search title : LG_PitA_0001.dat - SpecView
Database : NCBI nr 20120623 (19713758 sequences; 6412106995 residues)
Timestamp : 4 Jul 2012 at 13:53:25 GMT
Top Score : 158 for [gi|292656006](#), chlorite dismutase family protein [Haloferax volcanii DS2]

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 85 are significant ($p < 0.05$).



Protein Summary Report

Format As	Protein Summary	Help
Significance threshold $p <$	Max. number of hits	
0.05	50	
Re-Search All	Search Unmatched	

Index

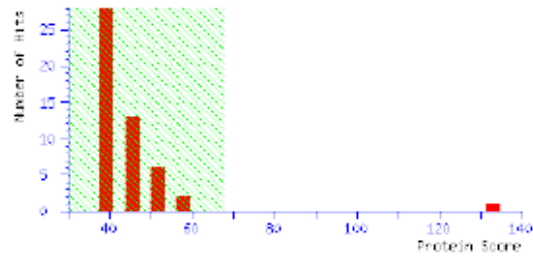
Accession	Mass	Score	Description
1. gi 292656006	56054	158	chlorite dismutase family protein [Haloferax volcanii DS2]

{MATRIX} Mascot Search Results *{SCIENCE}*

User : JoeG
Email : joe.gray@ncl.ac.uk
Search title : LG_U_0001.dat - SpecView
Database : NCBIInr 20120623 (18713758 sequences; 6412106995 residues)
Taxonomy : Archaea (Archaeobacteria) (386562 sequences)
Timestamp : 4 Jul 2012 at 14:22:19 GMT
Top Score : 133 for [gi|292656006](#), chlorite dismutase family protein [Haloferax volcanii DS2]

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 68 are significant ($p < 0.05$).



Protein Summary Report

<input type="button" value="Format As"/>	<input type="text" value="Protein Summary"/>	Help
Significance threshold p<		Max. number of hits
<input type="text" value="0.05"/>		<input type="text" value="50"/>
<input type="button" value="Re-Search All"/>	<input type="button" value="Search Unmatched"/>	

Index

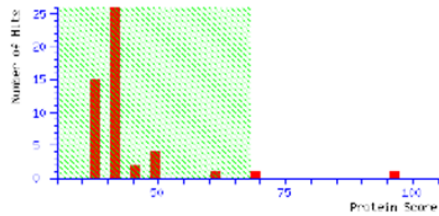
	Accession	Mass	Score	Description
1.	gi 292656006	56054	133	chlorite dismutase family protein [Haloferax volcanii DS2]

(MATRIX) Mascot Search Results
(SCIENCE)

User : JoeG
Email : joe.gray@ncl.ac.uk
Search title : LG_L_0001.dat - SpecView
Database : NCBI nr 20120623 (18713758 sequences; 6412106995 residues)
Taxonomy : Archaea (Archaeobacteria) (386562 sequences)
Timestamp : 4 Jul 2012 at 14:49:29 GMT
Top Score : 97 for [gi|292656006](#), chlorite dismutase family protein [Haloferax volcanii DS2]

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
Protein scores greater than 68 are significant ($p < 0.05$).



Protein Summary Report

Format As	Protein Summary	Help
Significance threshold p<	Max. number of hits	
0.05	50	
<input type="button" value="Re-Search All"/>	<input type="button" value="Search Unmatched"/>	

Index

Accession	Mass	Score	Description
1. gi 292656006	56054	97	chlorite dismutase family protein [Haloferax volcanii DS2]

Novel inhibition of archaeal family-D DNA polymerase by uracil

Tomas T. Richardson¹, Louise Gilroy¹, Yoshizumi Ishino², Bernard A. Connolly^{1,*} and Ghislaine Henneke^{3,4,5,*}

¹Institute for Cell and Molecular Biosciences (ICaMB), Newcastle University, Newcastle upon Tyne NE2 4HH, UK, ²Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka 812-8581, Japan, ³IFREMER, Centre de Brest, UMR 6197, LM2E, Technopole Brest-Iroise, Plouzané F-29280, France, ⁴CNRS, LM2E, UMR 6197, Technopole Brest-Iroise, Plouzané F-29280, France and ⁵Université de Brest, UEB, LM2E, UMR 6197, Technopole Brest-Iroise, Plouzané F-29280, France

Received November 29, 2012; Revised January 21, 2013; Accepted January 23, 2013

ABSTRACT

Archaeal family-D DNA polymerase is inhibited by the presence of uracil in DNA template strands. When the enzyme encounters uracil, following three parameters change: DNA binding increases roughly 2-fold, the rate of polymerization slows by a factor of ~5 and 3'-5' proof-reading exonuclease activity is stimulated by a factor of ~2. Together these changes result in a significant decrease in polymerization activity and a reduction in net DNA synthesis. Pol D appears to interact with template strand uracil irrespective of its distance ahead of the replication fork. Polymerization does not stop at a defined location relative to uracil, rather a general decrease in DNA synthesis is observed. 'Trans' inhibition, the slowing of Pol D by uracil on a DNA strand not being replicated is also observed. It is proposed that Pol D is able to interact with uracil by looping out the single-stranded template, allowing simultaneous contact of both the base and the primer-template junction to give a polymerase-DNA complex with diminished extension ability.

INTRODUCTION

Individual cells contain a varying repertoire of DNA polymerases, a subset of which is dedicated to genome replication (1). In all three domains of life, Bacteria, Eukarya and Archaea, chromosomes are copied by the replisome, a multi-protein replication machine (2,3). The bacterial replisome contains at least two molecules of DNA

polymerase III, a family-C member which is responsible for copying the genetic material (4). Although the same molecular species is used to copy both the leading and lagging strands, individual molecules of DNA Pol III are believed to be arranged in an asymmetric fashion, compatible with the requirements needed to copy two dissimilar DNA strands (5). The eukaryotic replisome contains two family-B polymerases, δ and ϵ , shown in an elegant series of experiments to be responsible for lagging and leading strand replication, respectively (6–8). Which polymerases are responsible for duplicating the archaeal genome is not currently known with the same degree of certainty as the other two domains. All Archaea contain family-B polymerases, usually present as multiple members in the Crenarchaea and as a single exemplar in the Euryarchaea (3,9). These proteins are monomeric and contain both the polymerase and 3'-5' proof-reading exonuclease active sites within the same polypeptide chain (10). In Crenarchaea, the only polymerases with properties compatible with DNA replication, i.e. interaction with proliferating cell nuclear antigen (PCNA) and rapid, accurate and processive DNA synthesis are the family-B enzymes which are widely assumed to fulfil this role (3,9,11). Whether distinct family members are individually responsible for leading and lagging strand replication or the same species copies both strands is presently unknown. Matters are more complicated in the Euryarchaea, which contain an unusual family-D enzyme (11–13) in addition to the replication-competent family-B polymerase. As well as being present in Euryarchaea, family-D polymerases are found in several emergent archaeal phyla (Thaumarchaea, Korarchaea and Nanoarchaea), which presently have few characterized members. However, this polymerase is noticeably absent in Crenarchaea.

*To whom correspondence should be addressed. Tel: +33 298224609; Fax: +33 298224757; Email: Ghislaine.Henneke@ifremer.fr
Correspondence may also be addressed to Bernard A. Connolly. Tel: +44 1912 227371; Fax: +44 1912 227424; Email: bernard.connolly@newcastle.ac.uk

Family-D polymerases exist as heterodimers, comprising a large (polymerase) and small (proof-reading exonuclease) subunit and may further assemble to give an L₂S₂ heterotetramer (11–15). The biochemical properties of the family-D enzymes are compatible with those required for a replicative polymerase: the presence of a proof-reading exonuclease activity should ensure accuracy (although fidelity has yet to be measured) and interaction with PCNA enables copying of long fragments of DNA (16–19). Both the family-B and -D polymerases have been shown to be essential for viability in a halophilic euryarchaeon using targeted gene deletion (20). Based on biochemical evidence it has been proposed that Pol D may act soon after initiation by primase and that at a later stage a switch occurs such that Pol B becomes responsible for leading strand replication, whereas Pol D continues to process the lagging strand (17,18). Both replicative DNA polymerases are also suspected to be involved in the resolution of RNA fragments in replicating cells (21). However, definitive genetic confirmation using experimental approaches such as those employed with the eukaryotic polymerases is awaited.

The archaeal family-B polymerases are unusual in recognizing uracil and hypoxanthine in DNA template strands and stalling replication when these bases are encountered (22–24). Tight and specific binding of the two deaminated bases is mediated by a pocket in the N-terminal domain (25,26). It is anticipated that stalling serves to prevent copying of uracil and hypoxanthine, which may arise by deamination of cytosine and adenine, respectively. The parent bases, cytosine and adenine, pair with guanine and thymine, respectively; however, their deamination products uracil and hypoxanthine are effective mimics of thymine and guanine and therefore code for adenine and cytosine. Thus, replication of deaminated bases results in a transition mutation (C:G → T:A, when uracil is copied; A:T → G:C, when hypoxanthine is copied) in 50% of progeny. Replicative polymerases from the bacterial and eukaryotic domains are unable to sense deaminated bases, despite the N-terminal domains of eukaryotic Pols ϵ and δ possessing considerable amino acid similarity with the corresponding region in the archaeal family-B enzymes (27). Very little is known about the response of the euryarchaeal family-D polymerases to deaminated bases. A brief report has indicated that they neither incorporate dUTP into extending strands nor copy uracil-containing templates (28). This publication presents a full characterization and shows that Pol D is indeed inhibited by template-strand uracil. While mechanistic details have not been completely characterized, it is apparent that the inhibition is markedly different from that previously observed with archaeal family-B polymerases.

MATERIALS AND METHODS

Enzymes

Wild-type and exonuclease-deficient (H451A) Pab-Pol D were produced and purified as previously described (29). These two variants of Pfu-Pol D (exo⁻ = H445A, the

equivalent amino acid change to that used in Pab-Pol D) were overexpressed in *Escherichia coli* BL21 (DE3) codon⁺ (RIL) cells using pWTD1 and pWTD2his, which encode the large and small subunits of the enzyme, respectively, with the small subunit having a (His)₆ tag (30). Cells were suspended (50 ml buffer/g of cells) in 50 mM Tris-HCl (pH 8.0), 20 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM Dithiothreitol (DTT), 10% glycerol + 1 EDTA-free protease inhibitor tablet (Roche), sonicated (on ice) and subjected to DNaseI digestion at 37°C for 30 min. The lysed cell suspension was then incubated at 75°C for 20 min and denatured protein and cell debris pelleted by centrifugation. The supernatant was filtered and applied to a 5-ml HisTrap (GE Healthcare) column, equilibrated with the above buffer containing 50 mM imidazole and lacking the protease inhibitor. The column was extensively washed with this buffer and Pol D eluted using a 30 ml linear gradient of 50–500 mM imidazole. Final purification used gel filtration on a Superdex 200 10/300 GL column (GE Healthcare) with 20 mM Tris-HCl (pH 6.5), 400 mM NaCl and 1 mM DTT.

Primer extension assays

Pab-Pol D and Pfu-Pol D assays were carried out in 10 μ l (20 μ l for experiments with U+42, 70, 102 and 134 replication fork mimics) of 10 mM Tris-HCl (pH 9), 50 mM KCl, 10 mM MgCl₂ and 200 μ M each of the dNTPs. With Pab-Pol D, 25 nM of fluorescent-labelled primed synthetic oligodeoxynucleotides and 30 nM of Pab-PolD exo⁺/exo⁻ (unless otherwise specified) were used, and the reactions were performed at 55°C for 30 min. In the case of Pfu-Pol D reactions, 20 nM of primer template and 140 nM of polymerase were used and conducted at 50°C for the times indicated. With both enzymes, the reactions were quenched by the addition of an equal volume of stop buffer [95% formamide, 10 mM EDTA, 10 mM NaOH and 1 μ M of ‘competitor oligonucleotide’ (an exact complement of the template strand under study)] (31). Samples were heated at 95°C for 5 min. The reaction products were resolved on 17% polyacrylamide, 8 M urea gels and visualized with a Mode Imager Typhoon 9400 or Typhoon FLA9500 (GE Healthcare) and quantified using Image Quant software. The percentage of extension was defined as the ratio: +1 to +*n* products (where *n* is the fully extended product)/total DNA.

DNA cutting by Pfu-Pol D

Pfu-Pol D exo⁺ or exo⁻ (80 nM) was incubated with fluorescent oligodeoxynucleotides (single and double stranded; +/- uracil) (20 nM) at 50°C for 30 min in the buffer given above. As a positive control for strand cutting, a reaction was carried out with uracil-DNA glycosylase (0.5 units, Fermentas). After 30 min, the reactions were quenched by heating with stop buffer. This buffer contains NaOH resulting in an alkaline pH and heating at 95°C for 5 min is expected to cut DNA at any abasic sites generated by the polymerase. Analysis was performed by gel electrophoresis as discussed above.

Table 1. Binding constants for the interaction of Pfu-Pol D with DNA

DNA	X location	K_D (nM) X = T	K_D (nM) X = U
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*	Single stranded	6.0 ± 1.6	3.4 ± 0.3
5' AATAGGTCCTATAGGCGAATGGTCCAGCTGGAA	-1	9.3 ± 0.6	5.2 ± 1.3
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			
5' AATAGGTCCTATAGGCGAATGGTCCAGCTGG	+1	9.1 ± 1.5	5.8 ± 0.4
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			
5' AATAGGTCCTATAGGCGAATGGTCCAGC	+4	7.3 ± 1.2	4.8 ± 0.3
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			
5' AATAGGTCCTATAGGCGAATGG	+10	8.4 ± 0.3	4.2 ± 0.5
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			
5' AATAGGTCCTATAGGCGAATGGTCCAGCTGGAACAGAAA	Double stranded	27.9 ± 6.9	29.0 ± 7.7
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			
5' AATAGGTCCTAUAGGCGAATGGTCCAGCTGGAACAGAAA	Double stranded	29.4 ± 11.1	36.6 ± 2.4
3' TTATCCAGGATATCCGCTTACCAGGTCGACCTGGTCTTT-H*			

The K_D values for the binding of Pfu-Pol D (average ± standard deviation from at least four determinations) to oligodeoxynucleotides containing uracil (thymine in controls) are given. The uracil is located in single strands, at various positions in a primer template and in double strands. H* = hexachlorofluorescein, used to determine the K_D value using fluorescence anisotropy titration (33).

Pfu-Pol D exonuclease assay and single dGTP incorporation

These were carried out as described above except that dNTPs were omitted for exonuclease assays and for dGTP incorporation only this triphosphate was added at concentrations that varied between 5 and 500 μ M. The evaluation of the rate constant for the exonuclease reaction and the K_D and k_{pol} for incorporation of a single dGTP under single turnover conditions have been described earlier (26,29,32).

Binding of Pfu-Pol D to DNA

Determination of the K_D values describing the binding of Pfu-Pol D to the oligodeoxynucleotides listed in Table 1 was performed using direct binding fluorescence anisotropy (with 5'-hexachlorofluorescein labelled DNA) as previously described (33). Titrations were carried out in 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 1 mM DTT and 0.1 mg/ml of bovine serum albumin with 1 nM DNA at 25°C. Aliquots of Pol D (1–50 nM) were added and the data analysed to yield K_D as described earlier (32).

RESULTS

Uracil in DNA template strands inhibits extension by Pol D

Throughout this publication, Pol D has been used from two different *Pyrococcus* species, *Pyrococcus abyssi* and *Pyrococcus furiosus* (Pab- and Pfu-Pol D). The amino acid sequences of the two proteins are very similar with 86 and 77% identity seen for the large and small subunits, respectively (Supplementary Figure S1). Even when different amino acids are found, in ~50% of cases the exchanges are conservative. The two polymerases behaved similarly enough in all the assays described in this publication to enable them to be used interchangeably.

When Pab-Pol D was used to extend a primer template (Cy5-labelled primer, 17 bases long, annealed to a template 87 bases in length) containing a single uracil at +16 (i.e. 16 bases ahead of the primer-template junction), a

reduction in DNA synthesis was observed as compared with a control that lacked this base (Figure 1, A1 and A2). In general, the intensity of each extended band was reduced for the uracil-containing template and, in particular, much less full length product was observed. This figure shows the results obtained with wild-type Pab-Pol D (exo⁺); however, almost identical profiles were seen with a mutant lacking 3'-5' exonuclease activity (exo⁻) (Supplementary Figure S2A). When the experiment was repeated with uracil further ahead of the primer-template junction at +59, inhibition was still observed (Figure 1, B1 and B2). However, when the uracil was situated in a double-stranded region by annealing a third 'masking' oligodeoxynucleotide to the primer template to give a gapped substrate, inhibition largely disappeared (Figure 1, B1 and B2). On this substrate, Pol D DNA synthesis was associated with pausing of the polymerase in the vicinity of the position of the nick, generating extension products in the 18–32 nucleotides (nt) size range. Although these experiments are not intended to be quantitative (kinetic parameters are determined more rigorously below), the gels were scanned to approximate the extent of polymerization. The extension percentages observed are consistent with uracil slowing polymerization by a factor of 2–5.

A parallel study was carried out with Pfu-Pol D (exo⁺) using a set of primer templates that position uracil increasingly further ahead of the primer-template junction (Figure 2A). Inspection of the extension gels obtained (Figure 2B–E) clearly shows less polymerization with uracil at +42, +70 and +102, compared with T controls. Inhibition is not obvious with uracil at +134. The impression is confirmed by Figure 2F, a summary schematic that shows the amount of primer template remaining at times of up to 30 min. In all cases, more starting material persists with U+42, 70 and 102 than with T-containing DNA. However, no difference is seen between U+134 and T+134.

A large number of extension reactions have been carried out, a selection of which are given in the Supplementary Figure S2. Strong inhibition was observed with uracil at

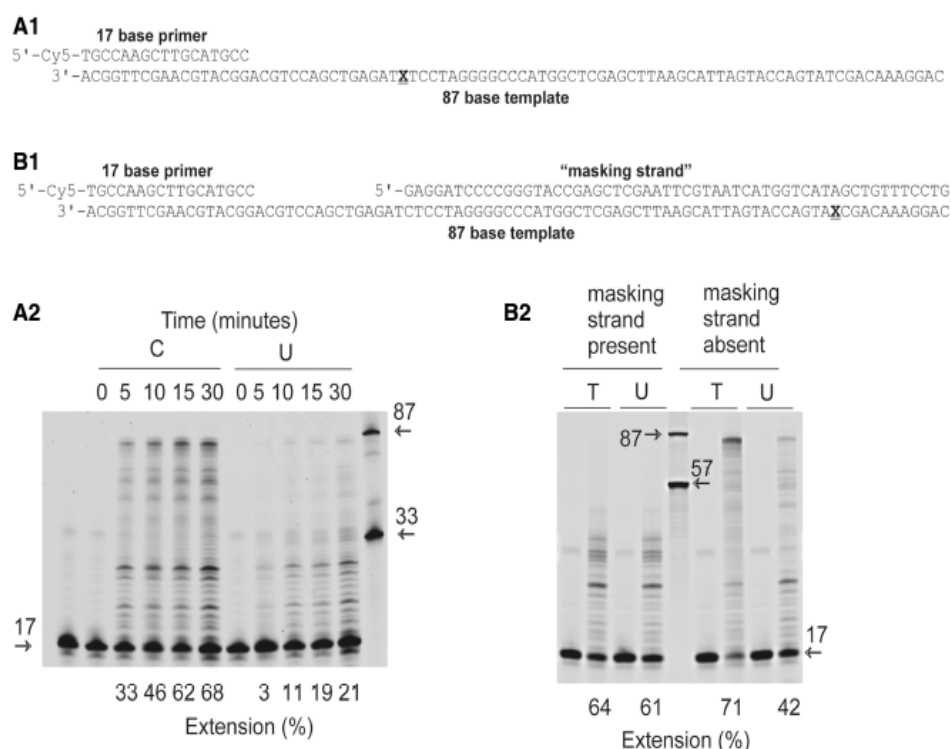


Figure 1. Inhibition of Pab-Pol D (exo^+) by template strand uracil. (**A1** and **B1**) Primer templates used for the experiments shown in panels **A2** and **B2**, respectively. (**A2**) Extension of the 17/87 primer-template A1 (X = C or U) by Pab-Pol for the times indicated. The numbers under the gel lanes represent the total percentage of extended product. Reference oligodeoxynucleotides of 17, 33 and 87 bases are indicated by the arrows. (**B2**) Influence of burying uracil in double stranded DNA on extensions by Pab-Pol D. The image shows the products formed with the 17/87 primer-template B1 (X = T or U) with either the 'masking strand' present or absent. Each reaction shows a zero and 30-min time point. The numbers under the gel lanes show the total percentage of extended product. Reference oligodeoxynucleotides of 17, 57 and 87 bases are indicated by the arrows.

+1 (Supplementary Figure S2B) and, surprisingly, a decrease in polymerization also occurred when uracil was located at -1, i.e. just within the double-stranded region, immediately behind the replication fork (Supplementary Figure S2C). In general, though, when uracil is well buried in double-stranded regions, either ahead of or behind the replication fork, no inhibition is observed (Figure 1B; Supplementary Figure S2D). However, exact boundaries, i.e. how far into a duplex uracil must be to prevent inhibition have not been mapped and a reduction in DNA synthesis was still noticeable using a derivative of the gapped substrate shown in Figure 1B, which locates uracil five bases into the duplex (Supplementary Figure S2E). Although the extensions described in the text and the Supplementary material use a very slight molar excess of Pab-Pol D and a 7-fold excess of Pfu-Pol D, inhibition by uracil with a 4-fold excess of Pab-Pol was also observed (Supplementary Figure S3).

Pol D incorporates dAMP opposite uracil

The gels shown in Figures 1, 2 and Supplementary Figure S2 indicate that although uracil inhibits primer extension by Pol D, copying beyond the deaminated base does take

place. To determine which base is inserted opposite uracil, single additions of each of the four dNTPs were made to a solution containing Pab-Pol D and a primer template with uracil at +1 (Figure 3A). As can be seen in Figure 3B, Pab-Pol D exo^+ only incorporated adenine opposite uracil, although at a reduced rate (~3% incorporation) due to uracil inhibition. In a control experiment with cytosine at +1 guanine was inserted efficiently (~52% incorporation), as expected. Similar results were seen with the exo^- variant of Pol D, although multiple additions were observed due to the inability to excise incorrectly incorporated bases (Figure 3C). In agreement with Supplementary Figure S2B, when all four dNTPs were simultaneously added, strong reduction (~6 to 13-fold depending on whether exo^+ or exo^- was used) in DNA synthesis was observed with the uracil-containing template.

'Trans' inhibition of Pol D by uracil

A replication fork mimic (Figure 4A; Supplementary Figure S4) has been used to assess the influence of translocated uracil. Following primer annealing, leading and lagging strand branches are created and uracil can be positioned at either location. Slowing of Pol D by uracil in the branch it is not copying is, strictly

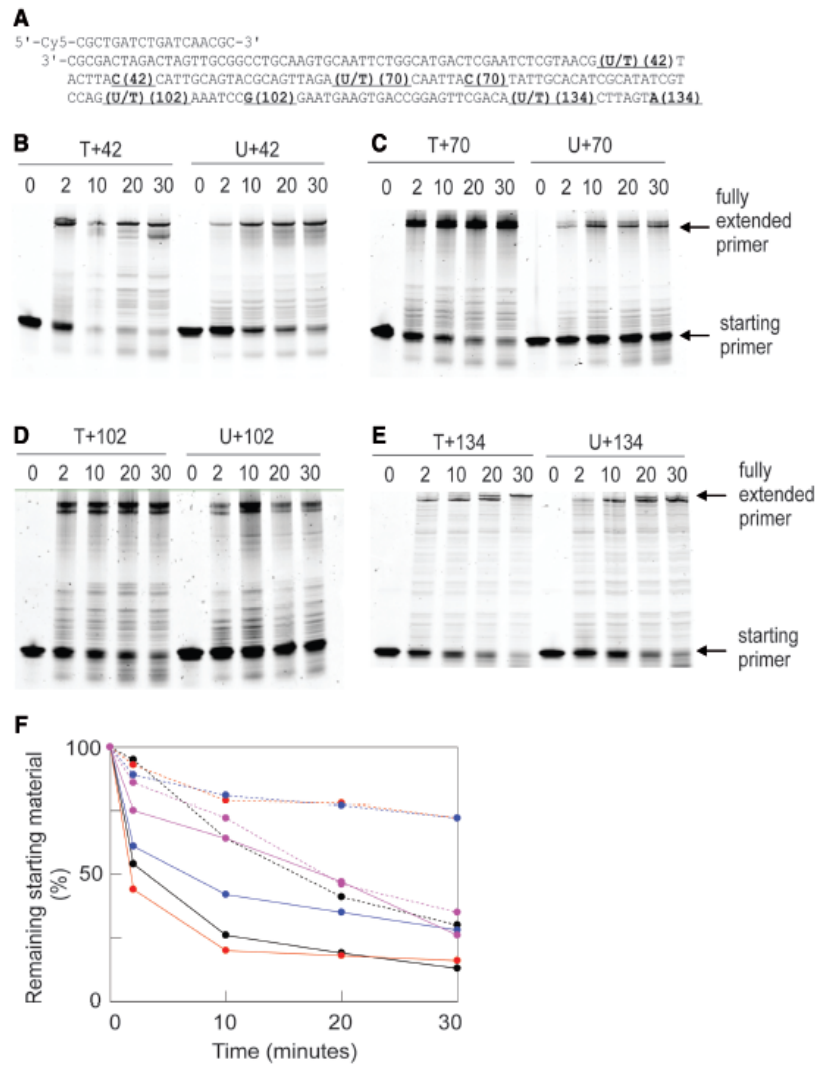


Figure 2. Extension by Pfu-Pol D (exo⁺) of primer templates that locate uracil at ever increasing distances ahead of the junction. (A) DNA substrates. Four different primer templates, consisting of a Cy5-labelled 18-mer primer annealed to four templates of increasing lengths, were used. These place uracil (control = thymidine) 42, 70, 102 and 134 bases ahead of the primer-template junction. The lengths of the templates are indicated using the same numbers, e.g. which places uracil at +42 terminates at the C labelled 42, etc. (B-E). Extension of the primer templates for the times shown above the gels (minutes) by Pfu-Pol D (exo⁺). The positions of the starting primer and full length products are indicated. (F) Summary of the data in panels (B-E) showing remaining primer template against time. Colour coding: black, +42; red, +70; blue, +102; magenta, +134. Solid lines, T; hatched lines, U.

speaking, not true trans-inhibition; the design of the mimic means that leading and lagging strands uracil are actually in the same DNA molecule (Figure 4A; Supplementary Figure S4). However, when replicating the lagging strand Pol D will clearly be unable to copy any uracil present in the leading strand. During polymerization of the leading strand, if Pol D was able to melt the double-stranded region, it could conceivably access uracil located in the lagging strand. However, under the conditions used, Pfu-Pol D only copied the leading strand up to the fork junction and no further progression into the double-stranded region was observed. Here,

therefore, trans-inhibition is defined as slowing of Pol D by uracil in a DNA region it cannot access by polymerization. As shown in Figure 4B, when Pfu-Pol D was used to copy an unmodified lagging strand, uracil in the leading strand was strongly inhibitory (~10-fold reduction in extension). When the leading strand was replicated, uracil positioned on the lagging strand also hindered progression (<2-fold reduction in extension), although to a less profound extent than observed for the opposite orientation (Figure 4C). On this substrate, Pol D DNA synthesis was likely associated with futile cycles of assembly/disassembly upon approaching the double-

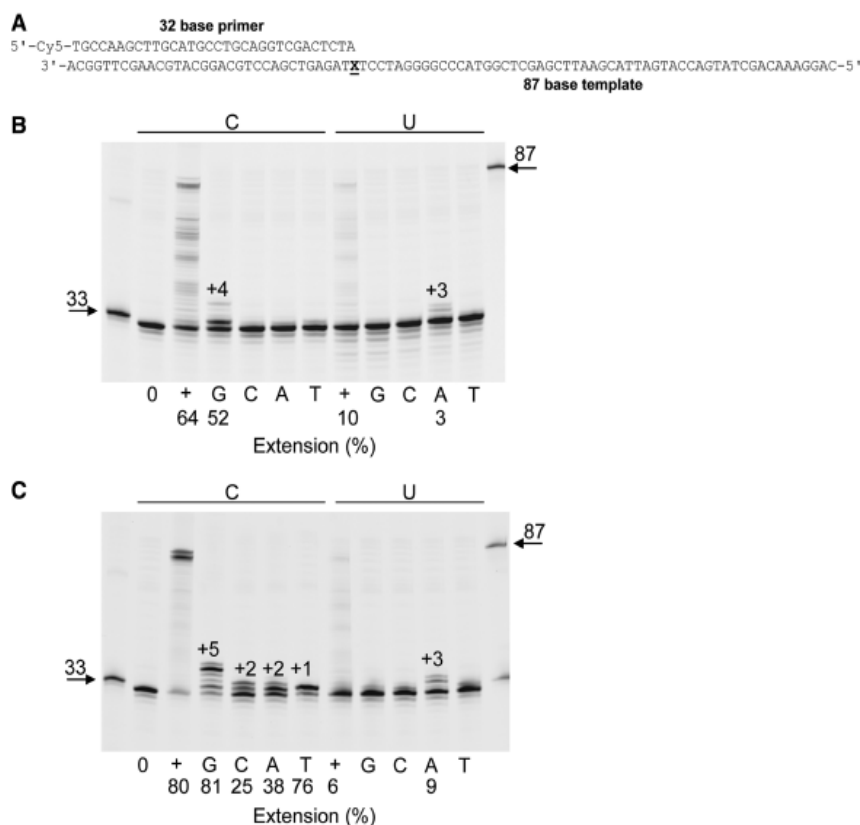


Figure 3. Base incorporated opposite template strand U by Pab-Pol D. (A) Primer templates used in these experiments (X = C or U). (B) and (C) Extensions using Pab-Pol D exo^+ (B) or exo^- (C) with either C or U at the +1 position (X), i.e. the first single-stranded template base. 0 = no dNTPs added; + = all four dNTPs added; G, C, A and T = only dGTP or dCTP or dATP or dTTP added, respectively. +1, +2, +3, +4 and +5 represent final products. The extension (%) for selected lanes is shown under the gels.

stranded DNA junction, as observed by the accumulation of multiple final products. Further information, more representative of conditions expected in the cell, can be obtained by using both primers to simultaneously copy the leading and lagging strands. The fluorophores used to label the primers (fluorescein and cyanine-5) have spectral properties that allow each strand to be individually monitored. The results (Supplementary Figure S4) are in close agreement with the studies using a single primer shown in Figure 4A; again, Pol D is inhibited by uracil situated in the opposite branch of the primer template to the strand in which polymerization is being observed.

Pol D does not cut uracil-containing DNA

The gels shown in Figures 1–4 indicate that the presence of uracil in template strands reduces DNA synthesis by Pol D. The inhibition could conceivably arise because uracil-dependent degradation of the template reduces the amount of substrate available for extension. To address this possibility, hexachlorofluorescein-labelled oligodeoxynucleotides containing uracil in either single- or double-stranded regions (Figure 5A) were incubated with Pfu-Pol

D exo^+ or exo^- . In all cases, the polymerase did not cut uracil-containing DNA in an endonucleolytic manner, i.e. by hydrolysis of the phosphodiester backbone at or near uracil (Figure 5B). The assay also scores for DNA-glycosylase action, as it involves a post-enzymatic heating step at alkaline pH which cleaves DNA at any abasic sites produced. Pfu-Pol showed no uracil-dependent DNA-glycosylase activity, in contrast to the expected positive result with uracil-DNA-glycosylase itself (Figure 5B). However, the presence of uracil does lead to a slight stimulation in 3'-5' exonuclease activity, compared with thymidine-containing controls, which results in the removal of a few bases from the 3' ends of the DNA substrates (Figure 5B). As discussed later in this publication, uracil-dependent stimulation of proof-reading exonuclease activity trims back primers and extending DNA strands and contributes to overall inhibition of polymerization. In principle, this exonuclease activity may remove bases from the 3'-termini of the template strands of the primer templates used in Figures 1–4. However, during polymerase assays, when dNTPs are present, these would be replaced. Therefore, template destruction cannot account for the inhibition of

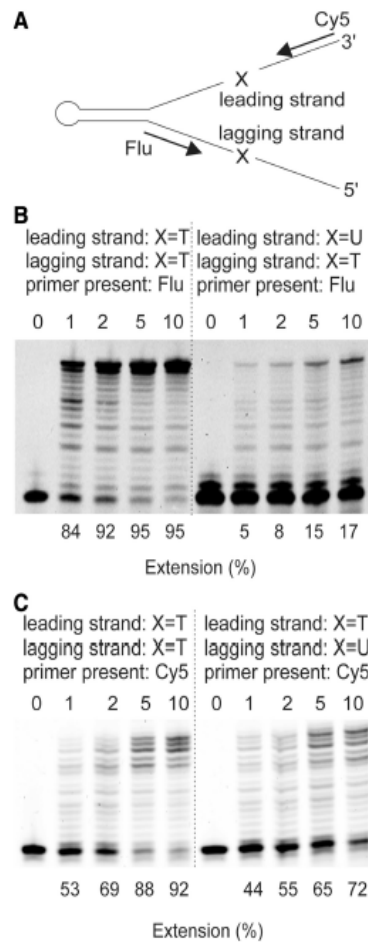


Figure 4. Inhibition of Pfu-Pol D by 'trans' located uracil. (A) Structure of the primer-template mimic (for full sequences, see Supplementary Figure S4). A long 'snap-back' oligodeoxynucleotide forms the backbone of the mimic. Annealing of primers (either singly or in pairs) produces leading and lagging strand branches which can contain a single uracil residue (indicated by X) four bases ahead of the primer-template junction. (B) Results seen when the lagging strand is copied (only Flu primer present) with uracil (thymine in controls) present in the leading strand. (C) Results seen when the leading strand is copied (only Cy5 primer present) with uracil (thymine in controls) present in the lagging strand. Here, only one primer was used per experiment. For the results observed using two primers for each experiment, see Supplementary Figure S4.

polymerization seen in the presence of uracil. A similar result was also found with Pab-Pol D (Supplementary Figure S5).

Pol D binds to uracil in single-stranded DNA

To determine if Pol D bound to uracil-containing DNA with greater affinity than to control substrates, fluorescence anisotropy binding titrations were used with hexachlorofluorescein-labelled oligodeoxynucleotides (33). A number of substrates were used (listed in Table 1), which place uracil in single-stranded DNA at a variety

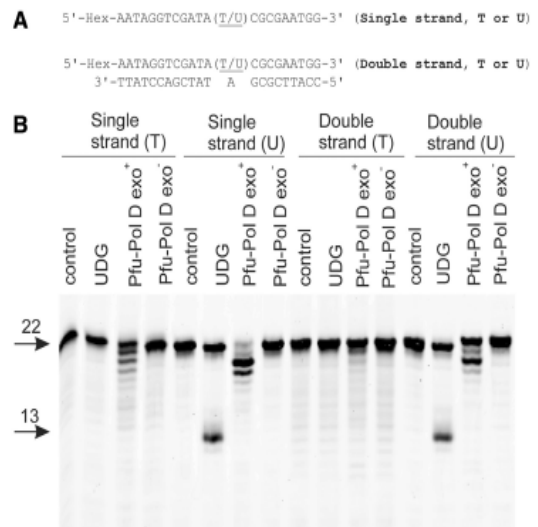


Figure 5. Pol D does not cut uracil-containing DNA. (A) Hexachlorofluorescein-labelled oligodeoxynucleotides used in these experiments. (B) Results seen when Pfu-Pol D *exo*⁻ or *exo*⁺ (80 nM) was incubated with the oligodeoxynucleotides (20 nM) at 50°C for 30 min followed by heating at 95°C for 5 min in the presence of NaOH. Control = no Pfu-Pol D added; UDG = addition of uracil-DNA glycosylase (positive control for strand cleavage at uracil). The starting oligodeoxynucleotides (22 bases) and the 13 base products expected for cleavage at uracil are shown arrowed. The ladders of products slightly reduced in length (seen with Pfu-Pol D *exo*⁺, most prominently when uracil is present) arise from 3' to 5' proof-reading exonuclease activity.

of locations in a primer template and in fully duplexed strands. In all cases where uracil was located in single strands, a slight increase, ~2-fold, in binding was observed when compared with controls lacking this base. As a typical example, the results seen when Pfu-Pol D was added to a primer template containing uracil (or thymine as a control) at +4 are shown in Figure 6. The *K_D* values found for a number of substrates are summarized in Table 1. Although the preference for uracil-DNA over controls is only about a factor of 2, the errors shown in Table 1, and the consistent results seen for a number of different DNA substrates, suggests significance. The slight selectivity for uracil is retained when the base is located at the -1 position of a primer template, just within the double-stranded region. The *K_D* values given in Table 1 are apparent binding constants that represent the measured sum of any specific interaction with uracil plus the numerous non-specific DNA binding modes possible (for uracil-containing substrates) and non-specific binding (for controls). In cases, such as that described here, where specific binding is not significantly greater than non-specific modes, it was very difficult to extrapolate the true or intrinsic binding constant for specific binding to uracil from the measured *K_D*. However, although the *K_D*s shown in Table 1 are apparent, rather than true values, a small but significant preference for uracil in single-stranded regions is clear. Table 1 also demonstrates that Pol D binds more tightly to single than double-stranded DNA and that selectivity for uracil is lost when the base is

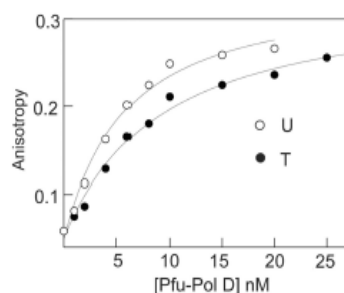


Figure 6. Binding of Pfu-Pol D to primer templates containing T or U at the +4 template position (sequences given in Table 1). The polymerase was added to the hexachlorofluorescein-labelled primer templates and the increase in fluorescence anisotropy noted. The data were fitted to a 1:1 binding stoichiometry to give the titration curves shown. Binding constants are summarized in Table 1.

buried in duplex DNA, away from the primer-template junction. Although binding constants are obtained at 25°C, preferential binding to uracil can be extrapolated to higher temperatures. This is supported by our recent study in which we show that increased temperature does not alter the binding modes of Pol D but likely accelerates the binding kinetics (association/dissociation) (34).

Uracil inhibits dNTP incorporation into primer templates

Although primer-template extensions (Figures 1–4) reveal very obviously that uracil inhibits Pol D, these experiments are, at best, semi-quantitative. In particular, it is difficult to accurately scan the multiple products formed, particularly those of low intensity. Therefore, for a more rigorous measure of the inhibitory influence of uracil, incorporation of a single dNTP under single turnover conditions has been used. This approach is widely used with DNA polymerases (35) and has previously been utilized to investigate the interaction of archaeal family-B enzymes with deaminated bases (32). For these experiments, Pol D exo^- was used to enable investigation of just the polymerase activity free from any complications arising due to proof-reading exonuclease activity. In contrast to the archaeal Pol B which requires the addition of PCNA to ensure tight binding to primer templates (32), reactions were carried out in the presence of saturated Pol D exo^- (140 nM) and primer templates (20 nM), respectively to the affinities for DNA reported in Table 1. The addition of 5–400 μ M concentrations of dGTP to Pfu-Pol D and primer template resulted in extension of the primer by a single base, as assessed by gel electrophoresis, and enabled determination of k_{obs} by fitting the amount of extended primer produced over a 30-min time course to a single exponential (Supplementary Figure S6). A secondary plot of k_{obs} against dGTP concentration, fitted using the Michaelis–Menten equation, gave the results shown in Figure 7 and the kinetic constants summarized in Table 2. From these results it is clear that the presence of uracil inhibits polymerization, almost entirely by reducing k_{pol} by a factor of ~ 5 with little change in the K_D for dGTP. This result is consistent with the slowdowns (between 2- and

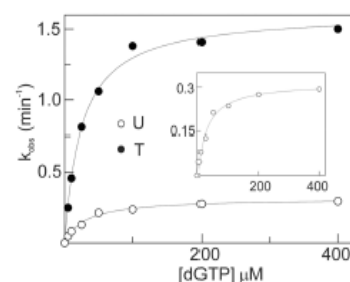


Figure 7. Incorporation of a single dGTP into primer templates containing T or U at the +4 template position (sequences given in Table 2). The primer templates (20 nM) were mixed with Pfu-Pol D (140 nM) and the reaction initiated by adding appropriate amounts of dGTP (between 5 and 400 μ M). Secondary fits to the Michaelis–Menten equation are given for both the T- and U-containing primer templates (the insert is an expansion of the U data). The kinetic parameters determined from this graph are summarized in Table 2.

Table 2. Kinetic parameters for incorporation of a single dGTP into a primer templates containing dU or dT at +4, under single turnover conditions

Base at +4 position of primer template ^a	k_{pol}^b (min ⁻¹)	K_D^b (μ M)	k_{pol}/K_D^b (s ⁻¹ M ⁻¹)
dT	1.6 \pm 0.2	25 \pm 6	3.8 \pm 1.4 $\times 10^6$
dU	0.3 \pm 0.06	31 \pm 9	0.6 \pm 0.3 $\times 10^6$

^aThe primer-template used in this experiment was:

5' Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
3' CCCCTAGGAGATCTCAGTGGACGTCCCGTTCGTXCG
AACAGAGG

where X = dU or dT.

^bKinetic parameters are the averages (\pm standard deviation) from five experiments.

13-fold) measured, less accurately, using primer-template extension.

Uracil stimulates 3'–5' proof-reading exonuclease activity

The influence of uracil on the proof-reading exonuclease activity of Pol D was determined by observing the degradation of the primer strand in a uracil-containing primer template. As with dNTP incorporation experiments, proof-reading exonuclease rates were measured with concentrations of primer template (20 nM) and Pol D (140 nM) that deliver single turnover conditions. Figure 8 shows the results obtained with two primer templates which contain either two A:T or two G:C base pairs at the primer-template junction. With the A:T primer template (identical to that used for the dNTP incorporation experiment), the presence of uracil at +4 increased the exonuclease rate slightly, by a factor of ~ 1.7 as compared with thymidine-containing controls (Figure 8A). In the case of the G:C primer template, uracil at +9 resulted in an approximate 3-fold stimulation in the rate of 3'–5' exonucleolysis (Figure 8B and C). Although the enhancements in exonuclease rates, when uracil is present in the template strand, are relatively small, they are consistently observed. The more rapid

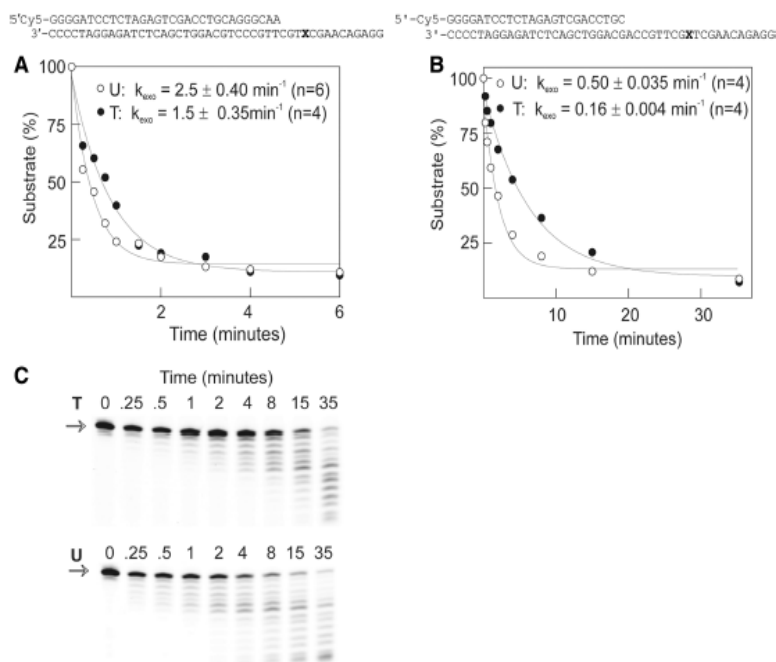


Figure 8. Proof-reading exonucleolysis of primer templates containing either T or U (sequences given below) by Pfu-Pol D. The polymerase (140 nM) was added to the primer templates (20 nM) and the degradation of the Cy5-labelled primer measured. The amount of primer remaining over time was fitted to a single exponential to give the rate constants (k_{exo}) for exonucleolysis. (A) and (B) Results found with the two primer templates illustrated at the top of each panel (X = T or U). The calculated k_{exo} values are shown on the graph and represent the means (\pm standard deviation) for the number of experiments (n) given. (C) Gels used to generate the data shown in (B). The full-length starting primers are marked with an arrow.

degradation of the A:T primer template, relative to G:C, is expected and arises from more facile strand separation; a requirement for proof-reading activity.

DISCUSSION

The data presented in this publication provide compelling evidence that the occurrence of uracil in DNA template strands diminishes polymerization by euryarchaeal family-D enzymes. Uracil must be positioned in single-stranded DNA or, if in a double-stranded region, near the primer-template junction. When the base is well embedded in duplex DNA little inhibition is observed. It is most likely that Pol D can only interact specifically with uracil in single strands but is able to capture the base in double-stranded helical DNA, following transient melting of terminal regions. Preferential binding of uracil by Pol D is also only observed when the base is in single-stranded DNA, agreeing with the same requirement for inhibition of polymerization. When uracil is present in single-stranded templates, small changes to a number of parameters are observed: binding affinity is increased by a factor of ~ 2 , polymerization decreases roughly 5-fold (as measured using k_{pol}) and 3'-5' proof-reading exonuclease rates are about two to three times faster (as measured by k_{exo}). Together these alterations result in markedly less DNA synthesis when uracil is present in

template strands, as compared with controls. As DNA polymerization involves addition of multiple dNTPs, a relatively small change in the efficiency of incorporation for each single base (such as the 5-fold reduction in k_{pol} seen in this study) may, cumulatively, lead to profound overall inhibition, especially when the accompanying increase in exonuclease activity is taken into account.

Template strand uracil has previously been observed to strongly suppress the activity of archaeal family-B polymerases, a feature assumed to reduce mutations that arise as a consequence of replicating DNA which has been subject to cytosine or adenine deamination (22,23,25,26) (uracil can also be directly incorporated into DNA from dUTP by a polymerase, but this is not mutagenic). It is most likely that the uracil-dependent inhibition of Pol D serves an identical function in protecting the genome from base deamination. The capacity of the two polymerases to detect and respond to uracil suggests they might both play a role in replication, as previously proposed (17-19). However, biochemical experiments can only tentatively identify a polymerase as replicative and the exact roles of euryarchaeal Pol B and Pol D in copying the genome await genetic investigation. The mechanisms by which the two polymerases interact with uracil appear to be different. Pol B has been well characterized and uses 'read-ahead recognition', where a running polymerase scans the template ahead of the replication fork for the

presence of uracil, sensing the base using a pocket in the N-terminal domain (22–26). The hall-marks of 'read-ahead' recognition are strong and specific binding of uracil and stalling of replication at a tightly defined position, four bases prior to uracil encounter. On the rare occasions that the polymerase manages to pass beyond uracil, inhibition ceases. Pol D shows none of these features and, therefore, must use a novel mechanism to sense uracil. Binding of the deaminated base is far from profound and a unique truncated extension product, suggesting stalling at a defined position, is not seen. Instead most of the bands that represent polymerization products are diminished, particularly those of longer products. Furthermore, Pol D is able to interact with template strand uracil at positions well beyond the replication fork and also when uracil is situated in double-stranded DNA just behind the fork. Should Pol D progress beyond template-strand uracil inhibition transiently persists until further replication positions the base deep within the duplex, when DNA melting to produce single strands becomes unlikely. Inhibition is also observed when uracil is located on a DNA strand not being copied by Pol D, a pattern never observed with Pol B. Such trans-inhibition is relevant to replication as uracil located on the lagging strand is capable of inhibiting Pol D travelling on the leading strand, or vice versa. The DNA repair processes that follow uracil sensing by both Pol B and Pol D await elucidation but must involve the accurate replacement of uracil with cytosine, the parent base from which uracil is derived by deamination. The near complete halting of replication when Pol B encounters uracil represents a very efficient method of buying time to enable downstream repair. The slowing of polymerization, characteristic of Pol D, seems at first sight to be a less effective strategy. However, the ability of the enzyme to recognize uracil well in advance of the replication fork presumably ensures success. The continuing inhibition that follows copying of uracil is counterintuitive as the damage (i.e. incorporation of adenine opposite uracil) has already been done; however, as mentioned previously, such inhibition fades as extension progresses.

The molecular mechanism that gives rise to the uracil-dependent inhibition of replication seen with Pol D is far from clear but the enzyme must have a means of sensing the presence of this pro-mutagenic base. Perhaps a uracil binding site is present either near to or overlapping the polymerase active site. Such a site would be able to interact with uracil near the primer-template junction, e.g. at the -1 and $+1$ positions, following which polymerase activity becomes attenuated. Interaction with more remotely located uracil, e.g. up to 100 bases ahead of the primer-template junction will rely on looping out of a long stretch of flexible single stranded DNA to bring uracil and its recognition site into proximity. Similarly, uracil present in single-stranded regions of a replication fork may be captured by a polymerase travelling on the opposite strand using looping, accounting for trans-inhibition. Interaction with uracil activates the 3'-5' proof-reading exonuclease activity and lowering the rate of replication. Simple partial blocking of the polymerase

active site with uracil would make interaction with the exonuclease site more probable. Alternatively, as the two sites are well separated, allosteric activation of exonuclease activity on uracil binding is also a possibility. This publication shows that Pol D can interact with single stranded uracil up to 100 bases in front of the replication fork. The lack of inhibition seen with uracil at $+134$ may arise from the greater conformational space that must be searched to locate the uracil-binding pocket. In general, excessive lengths of single-stranded DNA are not exposed during replication. In Archaea, including *P. abyssi*, Okazaki fragments are up to 120 nt long (36), implying a similar length of single stranded DNA produced during lagging strand replication. In viruses (37), bacteria (38) and presumably Archaea (3,9), the movements of the replicative helicase and polymerase are tightly coupled, mediated by protein-protein interactions. Again this suggests that relatively short lengths of single-stranded DNA are formed when the leading strand is copied. Therefore, forward scanning by Pol D for uracil at extreme distances is unlikely to be necessary for detection of this base as it moves from double- to single-stranded regions. It should be noted that we have never seen uracil-dependent complete cessation of polymerization, in contrast to an earlier brief publication which reports full inhibition of Pol D by uracil (28). Pol D is a heterodimer (11,12) but it is presently unknown if the uracil sensing apparatus is located in the large (polymerase) or the small (exonuclease) subunit. The heterodimer has been reported to further assemble into a tetramer (14); again it is unresolved whether a single heterodimer interacts with both the primer-template junction and uracil or if two separate heterodimers present in the tetramer are used for individual recognition events. At present no high resolution crystal structure is available for the entire Pol D, although information is available for both the N-terminal regions of the small and large subunits (39,40). Inspection does not reveal any obvious uracil-binding pocket as seen with Pol B (24–26). Pol D contains a large number of cysteines arranged in groups of four and may be an iron-sulphur protein as observed for two archaeal DNA repair enzymes, uracil-DNA glycosylase and XPD (41,42). Very recently it has been determined that the eukaryotic family-B polymerases (Pols α , δ , ϵ and ζ) contain an Fe-S cluster (43). Isolation of these polymerases, after heterologous expression in *E. coli*, revealed high lability of the Fe-S cluster, with anaerobic purification required to preserve its integrity. It appears that the role of the Fe-S cluster is to facilitate correct folding and interaction with other subunits necessary for the assembly of the polymerase holoenzyme, rather than direct participation in catalysis. This publication also observes that archaeal Pol D contains a similar cysteine motif and so may also be an Fe-S protein. The samples of Pol D used in this investigation were overexpressed in *E. coli* and purified under aerobic conditions, giving no indication of the olive green hue that is characteristic of iron-sulphur clusters. It is therefore possible that these putative iron-sulphur clusters may have been lost and that these elements could be involved in the correct folding of the uracil sensing region. Further studies are

currently in progress, including preparation of native Pol D under anaerobic conditions, to more thoroughly characterize the novel and enigmatic interaction of this protein with deaminated bases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

The excellent technical assistance of Pauline Heslop throughout this project is greatly appreciated.

FUNDING

French National Research Agency [ANR-10-JCJC-1501-01 to G.H.]; T.T.R. and L.G. are UK BBSRC supported PhD students. Funding for open access charge: French National Research Agency (ANR).

Conflict of interest statement. None declared.

REFERENCES

- Filee, J., Forterre, P., Sen-Lin, T. and Laurent, J. (2002) Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.*, **54**, 763–773.
- Pomerantz, R.T. and O'Donnell, M. (2007) Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol.*, **15**, 156–164.
- Barry, E.R. and Bell, S.D. (2006) DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.*, **70**, 876–887.
- Kelman, Z. and O'Donnell, M. (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.*, **64**, 171–200.
- O'Donnell, M. (2006) Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem.*, **281**, 10653–10656.
- Garg, P. and Burgers, P.M. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.*, **40**, 115–128.
- Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. and Kunkel, T.A. (2008) Division of labor at the eukaryotic replication fork. *Mol. Cell*, **30**, 137–144.
- Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. and Kunkel, T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, **317**, 127–130.
- Kelman, Z. and White, M.F. (2005) Archaeal DNA replication and repair. *Curr. Opin. Microbiol.*, **8**, 669–676.
- 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*. *Proc. Natl Acad. Sci. USA*, **96**, 3600–3605.
- Cann, I.K., Komori, K., Toh, H., Kanai, S. and Ishino, Y. (1998) A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc. Natl Acad. Sci. USA*, **95**, 14250–14255.
- Cann, I.K. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics*, **152**, 1249–1267.
- 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.
- Shen, Y., 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.
- Gueguen, Y., Rolland, J.L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., Raffin, J.P. and Dietrich, J. (2001) Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur. J. Biochem.*, **268**, 5961–5969.
- Cann, I.K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K. and Ishino, Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J. Bacteriol.*, **181**, 6591–6599.
- 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.
- Castrec, B., Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J.P. (2009) Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J. Mol. Biol.*, **394**, 209–218.
- Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J.P. (2007) DNA polymerase switching on homotrimeric PCNA at the replication fork of the Euryarchaea *Pyrococcus abyssi*. *J. Mol. Biol.*, **369**, 343–355.
- Berquist, B.R., DasSarma, P. and DasSarma, S. (2007) Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium sp. NRC-1*. *BMC Genet.*, **8**, 31.
- Henneke, G. (2012) *In vitro* reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochem. J.*, **447**, 271–280.
- Greagg, M.A., Fogg, M.J., Panayotou, G., Evans, S.J., Connolly, B.A. and Pearl, L.H. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl Acad. Sci. USA*, **96**, 9045–9050.
- Fogg, M.J., Pearl, L.H. and Connolly, B.A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.*, **9**, 922–927.
- Gouge, J., Ralec, C., Henneke, G. and Delarue, M. (2012) Molecular recognition of canonical and deaminated bases by *P. abyssi* family B DNA polymerase. *J. Mol. Biol.*, **423**, 315–336.
- 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.
- 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 isosteric base analogues. *Biochemistry*, **49**, 5772–5781.
- Wardle, J., Burgers, P.M., Cann, I.K., Darley, K., Heslop, P., Johansson, E., Lin, L.J., McGlynn, P., Sanvoisin, J., Stith, C.M. *et al.* (2008) Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res.*, **36**, 705–711.
- Sawai, H., Nagashima, J., Kuwahara, M., Kitagata, R., Tamura, T. and Matsui, I. (2007) Differences in substrate specificity of C(5)-substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from thermophilic bacteria, archaea, and phages. *Chem. Biodivers.*, **4**, 1979–1995.
- Palud, A., Villani, G., L'Haridon, S., Querellou, J., Raffin, J.P. and Henneke, G. (2008) Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance? *Mol. Microbiol.*, **70**, 746–761.
- Tori, K., Kimizu, M., Ishino, S. and Ishino, Y. (2007) DNA polymerases BI and D from the hyperthermophilic archaeon *Pyrococcus furiosus* both bind to proliferating cell nuclear antigen with their C-terminal PIP-box motifs. *J. Bacteriol.*, **189**, 5652–5657.
- 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.
- Emptage, K., O'Neill, R., Solovyova, A. and Connolly, B.A. (2008) Interplay between DNA polymerase and proliferating cell nuclear

- antigen switches off base excision repair of uracil and hypoxanthine during replication in archaea. *J. Mol. Biol.*, **383**, 762–771.
33. Reid, S.L., Parry, D., Liu, H.H. and Connolly, B.A. (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry*, **40**, 2484–2494.
34. Castrec, B., Laurent, S., Henneke, G., Flament, D. and Raffin, J.P. (2010) The glycine-rich motif of *Pyrococcus abyssi* DNA polymerase D is critical for protein stability. *J. Mol. Biol.*, **396**, 840–848.
35. Joyce, C.M. (2010) Techniques used to study the DNA polymerase reaction pathway. *Biochim. Biophys. Acta*, **1804**, 1032–1040.
36. Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.*, **4**, 154–158.
37. Delagoutte, E. and von Hippel, P.H. (2001) Molecular mechanisms of the functional coupling of the helicase (gp41) and polymerase (gp43) of bacteriophage T4 within the DNA replication fork. *Biochemistry*, **40**, 4459–4477.
38. Kim, S., Dallmann, H.G., McHenry, C.S. and Mariani, K.J. (1996) Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell*, **84**, 643–650.
39. Yamasaki, K., Urushibata, Y., Yamasaki, T., Arisaka, F. and Matsui, I. (2010) Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases. *FEBS Lett.*, **584**, 3370–3375.
40. Matsui, I., Urushibata, Y., Shen, Y., Matsui, E. and Yokoyama, H. (2011) Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*. *FEBS Lett.*, **585**, 452–458.
41. Hinks, J.A., Evans, M.C., De Miguel, Y., Sartori, A.A., Jiricny, J. and Pearl, L.H. (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. *J. Biol. Chem.*, **277**, 16936–16940.
42. Rudolf, J., Makrantonis, V., Ingledew, W.J., Stark, M.J. and White, M.F. (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol. Cell*, **23**, 801–808.
43. Netz, D.J., Stith, C.M., Stumpfig, M., Kopf, G., Vogel, D., Genau, H.M., Stodola, J.L., Lill, R., Burgers, P.M. and Pierik, A.J. (2011) Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat. Chem. Biol.*, **8**, 125–132.