Structural and functional characterisation of the Nonhomologous End-Joining proteins of the archaeon *Methanocella Paludicola*

A thesis submitted to the University of Sussex for the degree of Doctor of Philosophy

Edward J. Bartlett

August 2013

I hereby declare that this thesis has not been and will not be submitted, in whole or in part, to another University for the award of any other degree.

Edward Bartlett

Acknowledgements

I'd like to thank all members of the Doherty lab for their help over the years; Nigel, Andy, Tamsyn, Sean, Julie, Helen, Pierre, Ben, Laura, Stan, Alfie, Valia and Aidan. Special thanks must go to Aidan for the opportunity to do research in his lab, and to be part of the GDSC. Additional appreciation goes out to Nigel, who is responsible for any understanding that I have of the black magic that is crystallography. I'd also like to thank Stan for teaching me the art of getting razor sharp bands on my gels.

My parents and my brothers all deserve my appreciation for the help and encouragement that has been ongoing since long before I stepped into a lab. Natasha deserves specific thanks and apology; thanks for always believing in me, apologies for having to put up with me whilst I was writing this thesis.

My research was supported by the Biotechnology and Biological Research Sciences Council, and they too have my thanks.

University of Sussex

Edward J. Bartlett

Doctor of Philosophy Biochemistry

Structural and functional characterisation of the Nonhomologous End-Joining proteins of the archaeon *Methanocella Paludicola*

Summary

Maintenance of the genome is essential for life to prosper. Regular insults to the genome are sustained by all cellular life and can foster genetic instability if left unrepaired. The most lethal genetic damage is a double strand break (DSB), the cleavage of the phosphate backbone on both strands of the DNA double helix. Two main pathways exist which provide mechanisms for coping with DSBs; precise repair utilising the identical sister chromatid as a template to recreate the broken segment (homologous recombination; HR), and direct fusion of the broken ends in the absence of an intact template (nonhomologous end joining; NHEJ). NHEJ was first characterised in eukaryotes, and an analogous system has been found to exist in bacteria during the past decade. The bacterial NHEJ pathway is composed of four key proteins; the DNA end binding Ku homodimer, a DNA Ligase, a DNA polymerase and a phosphoesterase (PE). The first results chapter of this thesis details the identification of an orthologous set of proteins in the archaeon *Methanocella paludicola*, and their subsequent isolation and characterisation. The second results chapter expands on the individual activities of the proteins by combining them, and asserting the ability of archaeal NHEJ to join discontinuous ends *in vitro*. The role of the PE has been unclear in the bacterial system, but *in vitro* assays described here suggest that the enzyme plays a role in processing NHEJ intermediates formed by the NHEJ polymerase. The PE is found to optimise repair intermediates for ligation, and to reverse potentially genotoxic DNA strand displacements. The final results chapter investigates the structural aspects of the archaeal NHEJ enzymes. Together these studies establish a functional NHEJ system in an archaeon for the first time, and expand our knowledge of the bacterial system by proposing a standard model of archaeo-prokaryotic NHEJ.

Table of contents

Abbreviations

3’-P	3’-phosphate
5’-P	5’-phosphate
6-4PP	Pyrimidine (6-4) pyrimidone photoproduct
8-oxo-G	8-oxo-2’ deoxyguanosine
A_{280}	Absorbance at 280nm
A_{600}	Absorbance at 600nm
aa	Amino acid
A	Adenine
AEP	Archaeo-eukaryotic primase
Afu	*Archaeoglobus fulgidus*
AP	apurinic/apyrimidinic
AP-NHEJ	Archaeo-prokaryotic Nonhomologous end joining
APE-1	AP endonuclease 1
APS	Ammonium persulphate
APLF	Aprataxin and PNK-like factor
APTX	Aprataxin
ATM	Ataxia telangiectasia mutated
ATR	ATM Rad3 related
Atu	*Agrobacterium tumefaciens*
BER	Base excision repair
BIR	Break-induced replication
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
BRCT	BRCA1 C-terminus
BSA	Bovine serum albumin
C	Cytosine
C2’3’P	Cytidine 2:3’-phosphate
Cko	*Candidatus korarchaeum cryptofilum*
CSR	Class switch recombination
DDR	DNA damage response

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DnaG	Bacterial / archaeal RNA primase
dNTP	Deoxyribonucleoside triphosphate
DSB	Double strand break
DSBR	Double strand break repair
D-strand	Downstream DNA strand
EDTA	Ethylenediaminetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
FAT	FRAP/ATM/TRRAP
FEN-1	Flap structure-specific endonuclease 1
FHA	Forkhead associated
G	Guanine
G1	Growth phase 1
G2	Growth phase 2
GDSC	Genome Damage and Stability Centre
HR	Homologous recombination
Hsa	*Homo sapiens*
IDL	Insertion-deletion mismatch
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Ionising radiation
Kb	Kilobase
kDa	Kilodalton
LigD	Ligase D
LUCA	Last universal common ancestor
Mba	*Methanosarcina barkeri*
MDC1	Mediator of DNA damage checkpoint protein 1
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
Mpa	*Methanocella paludicola*
MPD	2-Methyl-2,4-pentanediol
MRN	Mre11/Rad50/Nbs1 complex
mRNA	Messenger ribonucleic acid
Msm	*Mycobacterium smegmatis*

Mtu	*Mycobacterium tuberculosis*
NER	Nucleotide excision repair
NHEJ	Nonhomologous end joining
nt	Nucleotide
NTase	Nucleotidyltransferase
OB	Oligonucleotide/oligosaccharide binding
PADT	Polymerisation across discontinuous templates
Pae	*Pseudomonas aeruginosa*
PAGE	Polyacrylamide gel electrophoresis
PARP1	Poly(ADP-ribose) polymerase 1
PCR	Polymerase chain reaction
PE	Phosphoesterase
PIKK	Phosphatidylinositol 3-kinase related kinase
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase 3’ phosphatase
pNPP	para-nitrophenyl phosphate
Pol	DNA-dependent DNA polymerase
PolDom	Polymerase domain of bacterial LigD
RNA	Ribonucleic acid
RSS	Recombination signal sequence
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing
SOB	Super optimal broth
SSB	Single strand break
SSBR	Single strand break repair
T	Thymine
T 5’ mP pNP	Thymidine 5’-monophosphate para-nitrophenyl
Taq	*Thermus aquaticus*
TB	Terrific broth
TDP1	Tyrosyl-DNA phosphodiesterase 1
TDP2	Tyrosyl-DNA phosphodiesterase 2
TdT	Terminal deoxynucleotidyltransferase
TEMED	N,N,N’,N’-Tetramethylethylenediamine
TOP I	Topoisomerase I

TOP II	Topoisomerase II
TCR	Transcription coupled repair
U	Uracil
UV	Ultraviolet
V(D)J	Variable, diverse, joining genes
vWA	van Willebrand
XLF	XRCC4-like factor
XRCC1	X-ray cross-complementing protein 1
XRCC4	X-ray cross-complementing protein 4

List of figures

List of tables

Chapter 1

Introduction

Repair of the genome is essential for life to prosper. Regular insults to the genome sustained by all cellular life either fosters instability or causes cell death. The most lethal genetic damage is a double strand break (DSB), the severance of the phosphate backbone on both strands of the DNA double helix. Two mechanisms exist for coping with DSBs; precise repair utilising the identical sister chromatid as a template to recreate the broken segment (homologous recombination), and direct fusion of the broken ends in the absence of an intact template (nonhomologous end joining). Nonhomologous end joining (NHEJ) was first characterised in mammals and then yeast, and more recently in bacteria. Putative individual components of the NHEJ pathway have been identified in archaeal genomes, but not confirmed. This thesis describes the identification of NHEJ proteins from the archaeon *Methanocella paludicola*, and their biochemical and structural characterisation. The aim of this research project was to establish if NHEJ exists in the third domain of life and, furthermore, to increase our understanding of the highly similar bacterial system. The canonical prokaryotic NHEJ complex is comprised of a DNA ligase, a DNA polymerase, a phosphoesterase and a ku dimer; this thesis aims to describe the role of the phosphoesterase in context of the other enzymes, since it has been refractory to elucidation thus far. The introduction to this thesis will offer an overview of DNA damage and the mechanisms of repair, the process of NHEJ in higher eukaryotes and bacteria, and introduce NHEJ in archaea.

1.1 DNA damage and repair

1.1.1 Introduction to DNA damage

Deoxyribonucleic acid (DNA) encodes the blueprint for life, and as such requires protection and conservation to allow organisms to flourish. Life on Earth offers a consistent challenge as natural sources, both endogenous and exogenous to cellular life, cause structural damage to DNA molecules. Proper replication and deciphering of the genome requires methods of reversing alterations and repairing damage to the DNA double helix. Unrepaired damage can prevent transcription or promote mistranscription of a gene, prohibiting correct mRNA (messenger ribonucleic acid) translation into the intended protein. Loss of fundamental proteins or production of hyperactive proteins can cause apoptosis. If the proteins are integral to the maintenance, repair or replication of DNA, genome stability can also be compromised. Such genomic instability in multicellular life poses other complications. In multi-cellular organisms an unstable genome can encourage a cell to become rampant and multiply quickly,

able to form destructive tumours that can invade neighbouring tissues. This process, cancer, can ultimately become fatal for the whole organism. Methods of repairing DNA damage are therefore essential for the survival of all cellular life.

1.1.2 Overview of DNA damage

DNA damage can be caused by factors both inside and outside of cells, and can also occur spontaneously (Figure 1.1, Figure 1.2) (Lee et al., 1999; Batra et al., 2012; Hsu et al., 2005; Takahara et al., 1995; Lu et al., 2012). Endogenous sources of DNA damage are largely byproducts of metabolic pathways. Reactive oxygen species, superoxide ($O_2^{\bullet-}$) and peroxide (O_2^{-2}), produced during oxidative phosphorylation by the electron transport chain in mitochondria are highly reactive and can cause a number of DNA lesions. The reactive oxygen species react to create hydroxyl radicals that mutate DNA by adding double bonds and abstracting hydrogen atoms from DNA bases (Cooke et al., 2003). The most thoroughly characterised of these lesions is the highly common 8-oxo-2' deoxyguanosine (8-oxo-G), which causes mispairing during subsequent replication of the damaged DNA (Cheng et al., 1992). Other endogenous causes of DNA alteration are alkylating agents and epoxyaldehydes which cause methylation and exocyclic DNA adducts respectively (De Bont & Larebeke, 2004). Spontaneous genetic malformations can arise since the glycosidic bond between deoxyribose and base is labile in specific conditions, known as depurination. Breakage of the glycosidic bond produces an abasic or apurinic/apyrimidinic (AP) site (Lindahl, 1993). DNA bases can also become deaminated by hydrolysis, most frequently cytosine to uracil (Lindahl, 1993). DNA is also sometimes damaged by elementary errors during replication. Human replicative DNA polymerases are high fidelity, and only insert one incorrect nucleotide per 10^7 bases correctly replicated (Loeb et al., 1974; Iyer et al., 2006). Even if such mismatches and other replication errors, including single nucleotide deletions and insertions, are not frequent, they constitute a type of DNA damage that requires repair.

Exogenous sources of DNA damage include ultraviolet (UV) radiation from the sun, ionising radiation (IR), and genotoxic molecules. UV radiation promotes covalent bonding between two adjacent pyrimidines, most frequently producing cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (6-4PPs) (Rastogi et al., 2010). IR comes from radioactive materials and cosmic rays, as well as from artificial sources such as X-ray generators and particle accelerators. Exposure to IR can cause disruption of hydrogen bonds and the cleavage of the DNA sugar phosphate backbone, resulting in single and double strand

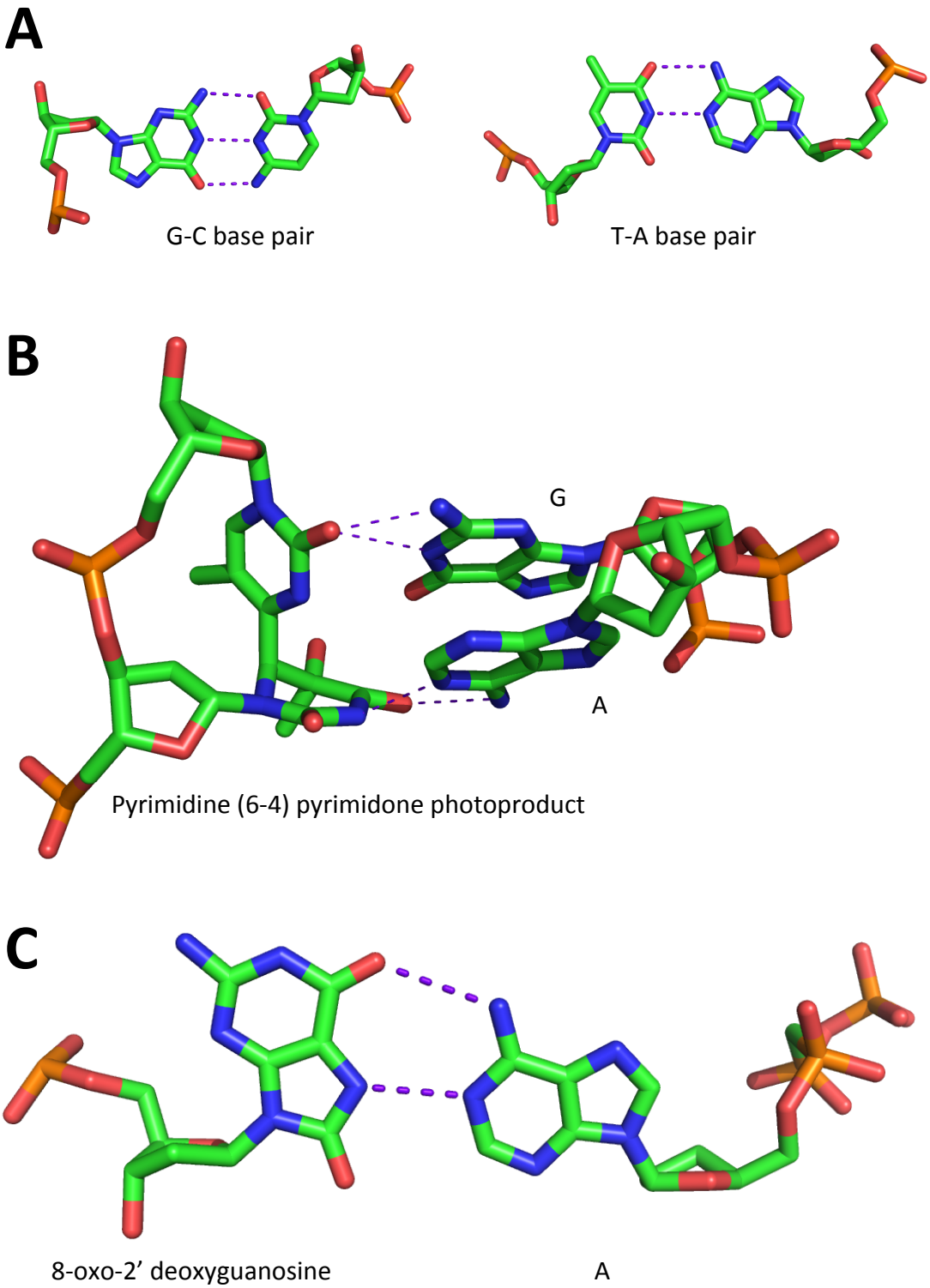


Figure 1.1 Mis-paired bases caused by DNA damage.
(**A**) Watson-Crick base paired G-C and T-A, shown for context. (**B**) Pyrimidine (6-4) pyrimidone photoproduct (6-4pp) caused by UV induced covalent bonding of two neighbouring thymidines. The 6-4pp is shown correctly base pairing with A, but also incorrectly pairing with a G. (**C**) 8-oxo-2’ deoxyguanosine (8-oxo-G) is displayed in the *syn* conformation, with the G base rotated compared to correct G-C pairing. This enables 8-oxo-G to mispair with A. 6-4pp and 8-oxo-G figures created using atomic structures from Lee (et al., 1999) PDB 1CFL, Batra (et al., 2012) PDB 3RJG respectively.

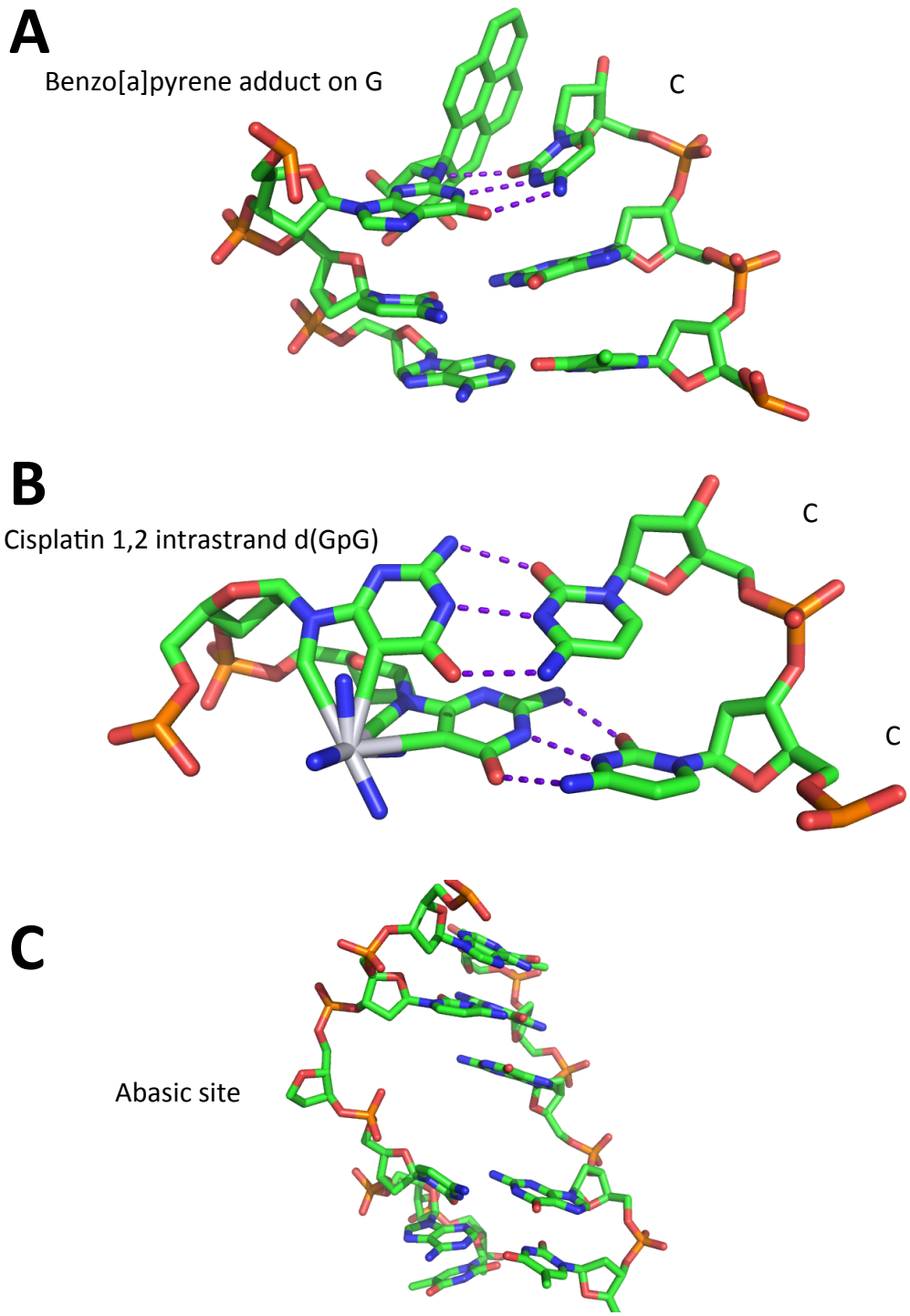


Figure 1.2 Damaged DNA.
(**A**) Benzo[a]pyrene is shown covalently bound to G. Although the benzo[a]pyrene has not prevented the correct hydrogen bonding with C, the polycyclic hydrocarbon creates a bulky lesion which intercalates into the minor groove and disrupts the DNA secondary structure. (**B**) Cisplatin can produce a 1,2 instrastrand d(GpG) crosslink. Again, like benzo[a]pyrene, correct base pairing is possible, but the crosslinking causes considerable distortion to the double helix. (**C**) The abasic or apurinic/apyrimidinic (AP) site demonstrates the result of glycolysis and loss of the base. Benzo[a]pyrene, cisplatin crosslink and abasic site figures G figures created using atomic structures from Hsu (et al., 2005) PDB 1XC9, Takahara (et al., 1995) PDB 1AIO and Lu (et al., 2012) PDB 4B5H respectively.

breaks (DSBs), of which the DSB is the most lethal DNA lesion (Jackson, 2002). DSBs are the genetic insult of interest in this thesis, and although they are predominately caused by IR, they can be produced by other processes including replication fork collapses and mechanical stress on chromosomes (Jackson, 2002). Genotoxic chemicals, both natural and artificial, are also exogenous sources of DNA damage. Chemotherapeutic agents are often specifically designed to take advantage of the unstable genomes of oncogenic cells, for example cisplatin causes the crosslinking of DNA, typically producing 1,2 instrastrand d(GpG) (Takahara et al., 1995). The polycyclic aromatic hydrocarbon benzo[a]pyrene, found in tobacco smoke, also produces a bulky DNA adduct, by intercalating and covalently binding to the DNA nucleobases (Tang et al., 2012).

1.1.3 Overview of DNA damage repair pathways

Despite the staggering array of different DNA malformations that the genome can suffer, cells have evolved multiple systems of tolerating, negating and surviving this damage (Table 1.1). In most cases, the most efficient process is to remove the genetic alteration and replicate the original DNA sequence, but some involve more drastic measures to ensure survival. There are several pathways that excise damaged segments of DNA and then use the opposite strand as a template for re-insertion. These pathways are; mismatch repair (MMR), nucleotide excision repair (NER) and base excision repair (BER). Single strand breaks (SSBs) are a common result of oxidative damage, failed excision repair or IR as noted above, and are repaired by the SSB repair pathway (SSBR) (Caldecott, 2008). DSBs are more technically demanding to repair, since the double helix is severed and bases from either terminus may be lost. There are two prominent evolutionarily conserved mechanisms for reconnecting and repairing DSBs; homologous recombination (HR) which makes use of the available sister chromatid as a template in post-replicative cells, and nonhomologous end joining (NHEJ) which can reconnect the broken ends directly, in the absence of templating DNA, in the cell cycle phase G1 or quiescence. Since repair of DSBs is the focus of the research presented in this thesis, an in-depth discussion of HR and NHEJ will follow the brief introduction to the other main DNA damage repair pathways.

1.1.3.1 Mismatch repair

Mismatch repair (MMR) is a highly conserved pathway from bacteria to eukarya which repairs mismatched bases and insertion-deletion mismatches (IDLs). The biochemical process of MMR has been robustly characterised in *Escherichia coli*; MutS binds the mismatched region along

DNA damage	Repair pathway	Key proteins
Replication error	Mismatch repair	MutS, MutL, MutH, UvrD, ExoI, Pol δ, Pol III, DNA Ligase
Bulky DNA lesions	Nucleotide excision repair	XPC, hHR23B, XPB, XPD, XPG, XPF
Single base lesions	Base excision repair	Glycosylase, AP endonuclease, Pol β, FEN-1
Single strand breaks	Single strand break repair	PARP1, XRCC1, Pol β, Pol δ, TDP1, APTX
Double strands breaks	Homologous recombination	Mre11, Rad50, Nbs1, CtIP, RPA, RAD51, PALB2, BRCA2
Double strand breaks	Nonhomologous end joining	Ku 70/80, DNA-PKcs, Artemis, DNA Ligase IV, XRCC4, XLF, Pol λ, Pol μ, PNK

Table 1.1 Pathways of DNA Repair.
A table listing broad categories of DNA damage, the title of the relevant pathway that functions to reverse the damage, and the names of the key proteins in eukaryotes. Most repair pathways can engage with a variety of different damages types, and some proteins have overlapping roles in different pathways. Double strand breaks are unique amongst DNA damage types because they have two distinct, complex repair pathways dedicated to their repair; Homologous Recombination and Nonhomologous End Joining.

with MutL, a protein that facilitates multiple steps in MMR (Hsieh & Yamane, 2008). MutH nicks the unmethylated nascent strand to indicate the correct site for subsequent helicase unwinding and exonuclease digestion. The gap in the DNA is then filled by Pol III and ligated by DNA ligase (Hsieh & Yamane, 2008). The eukaryotic pathway operates in a similar manner, with MutS and MutL orthologues, except without the methylation directed method of MutH. The human polymerase Pol δ resynthesises the excised DNA (Hsieh & Yamane, 2008; Larrea et al., 2010). MMR has not yet been properly characterised in archaea. However, some archaeal species possess MutS and MutL homologues and others appear to retain an unknown, orthologous system (Busch & DiRuggiero, 2010).

1.1.3.2 **Nucleotide excision repair**

The nucleotide excision repair (NER) pathway recognises bulky DNA lesions which distort the natural geometry of the double helix, such as CPDs and 6-4PPs formed by UV light, and polycyclic aromatic hydrocarbon adducts caused by intercalation of benzo[a]pyrene (Fuss & Cooper, 2006). The NER proteins XPC and hHR23B bind the distorted region before other factors join, and the XPB and XPD helicase components of TFIIH unwind the DNA producing a bubble structure. The damage can then be excised by the endonucleases XPG and XPF, which cleave 3' and 5' of the DNA bubble respectively (Fuss & Cooper, 2006). NER proteins can perform these functions as global genome repair, or as transcription-coupled repair (TCR) that moves along with the transcription machinery. Bacterial NER operates in a similar, yet more streamlined process (Kisker et al., 2013), whilst archaeal NER remains uncharacterised, although orthologues for XPB, XPD, XPF and XPG have been discovered (Rouillon & White, 2010; Roberts et al., 2003).

1.1.3.3 **Base excision repair**

Base excision repair (BER) conducts the role of removing lesions that do not distort DNA and are not recognised by NER, including AP sites, 8-oxo-G, alkylated and deaminated bases. BER proteins repair single base damage by hydrolysing the offending base to create an AP site, which is then excised by an AP endonuclease (Liu et al., 2007). At this juncture, repair can progress as either short-patch or long-patch. In short-patch BER Pol β inserts a single nucleotide, creating a nick that is eligible for ligation. Long-patch BER also features Pol β, except it performs strand displacement synthesis of 2 to 10 bases which creates a 5' flap from the downstream DNA, requiring flap structure-specific endonuclease 1 (FEN-1) for cleavage (Liu et al., 2007). BER orthologues have been discovered in several bacterial species

(Suvarnapunya et al., 2003) and also in archaea (Sartori & Jiricny, 2003). These equivalent processes have not been stringently characterised, but appear to be functionally homologous to the eukaryotic system, suggesting that the bacterial and archaeal BER pathways might be the evolutionary antecedents of the eukaryotic mechanism.

1.1.3.4 **Single-strand break repair**

Single-strand break repair (SSBR) is separate from the excision repair pathways discussed above since it responds to already broken DNA, rather than creating a break. However, components of SSBR can remove damaged elements of DNA at the 5' or 3' ends of the discontinuous DNA strand (Caldecott, 2008). SSBR does have a considerable overlap with the post-excision steps of BER. SSBR recognises lesions that are formed from the endogenous and exogenous sources previously discussed, and also failed repair intermediates of excision repair. SSBR also repairs lesions produced by topoisomerase I (TOP I) remaining covalently bound to DNA following an abortive attempt to relax the DNA (Yang et al., 1996). The recognition process begins with PARP1 recruiting X-ray cross-complementing protein 1 (XRCC1) and DNA end processing factors, then DNA synthesis enzymes are required to fill in the gap (Pol β and Pol δ), before the final ligation (Caldecott, 2008). SSBs with unusual damage such as 3'-TOP I fragments and 5'-AMP are resolved by tyrosyl-DNA phosphodiesterase 1 (TDP1) and aprataxin (APTX) respectively, prior to the processing described above (Clements et al., 2004; Ahel et al., 2006). Very little is known about bacterial or archaeal equivalents to the eukaryotic SSBR pathway, beyond the proteins that also function in BER.

1.2 Repair of double strand breaks

DSBs are not as frequent as other types of DNA damage, but the complexity involved in the sensing, signalling and repair and the high toxicity of the lesion makes DSBs potentially the most demanding genetic damage that a cell has to contend with (Khanna & Jackson, 2001). DSBs are associated with a variety of human disorders; neurodegenerative disorders, immune deficiencies, infertility and can cause the genomic instability that is a prerequisite for cancer (Jackson & Bartek, 2009). DSBs can be caused by a variety of different factors; IR, replication fork collapse and mechanical stress, all of which are mentioned above. DSBs can also be triggered by ROS, intentional scission during V(D)J (variable, diverse, joining genes) recombination, class switch recombination and meiosis to produce genetic variation, and

following abortive relaxation by topoisomerase II (TOP II) that leaves the enzyme covalently bound to the 5' DNA end (Shrivastav et al., 2008).

1.2.1 Homologous Recombination

Homologous recombination (HR) is a flexible system of managing DNA ends with roles well beyond DNA damage repair. HR is also tasked with telomere maintenance, rescue of replication fork collapse and the segregation of chromosome during meiosis (Sung & Klein, 2006). Of the two main methods of repairing DSBs, HR is considered the most error-averse, whereas NHEJ is error prone (Chapman et al., 2012). The mechanism of HR instigates a search for the intact copy of the broken DNA sequence in the sister chromatid. This copy is then used as a template to reproduce the lost or damaged material (Figure 1.3). Using the sister chromatid to replicate the damaged DNA is the only way to maintain the integrity of the original sequence, in the absence of the complete template strand enjoyed by the excision repair pathways and SSBR. Despite this, HR does not dominate the DSB repair relationship, since there is no template available in cycling cells until the genome is replicated. In such cases, NHEJ is the predominant repair method (Daley et al., 2005).

HR is initiated by the Mre11/Rad50/Nbs1 (MRN) complex binding the broken DNA termini. The 5' DNA ends are nucleolytically resected, producing 3' overhangs. Helicases unwind the double stranded DNA (dsDNA) to allow the 5'-3' exonucleases to resect the single stranded DNA (ssDNA) (Li & Heyer, 2008). The resection is mediated by the protein CtIP, which interacts with the MRN complex (Sartori et al., 2007). RPA then binds the ssDNA, in turn recruiting Rad51 to form a nucleoprotein filament that can then invade the intact DNA duplex with the homologous sequence. The invading nucleoprotein strand creates a displacement loop (D-loop) and allows DNA synthesis from the inserted 3' termini against the template strand. Rad51 filament formation and strand invasion is facilitated by other proteins, PALB2 and breast cancer type 2 susceptibility protein (BRCA2) (Xia et al., 2006). BRCA2 is thought to be important to successful Rad51 filament formation, and may be crucial for both filament nucleation and stabilisation (Li & Heyer, 2008). Germline mutation of BRCA2 causes a significantly increased susceptibility to breast and ovarian cancer (Xia et al., 2006).

HR can proceed on two distinct paths following strand invasion DNA synthesis; double strand break repair (DSBR) and synthesis dependent strand annealing (SDSA). In DSBR the second end of the DSB is annealed to the D-loop and a double Holliday junction structure is formed (Li &

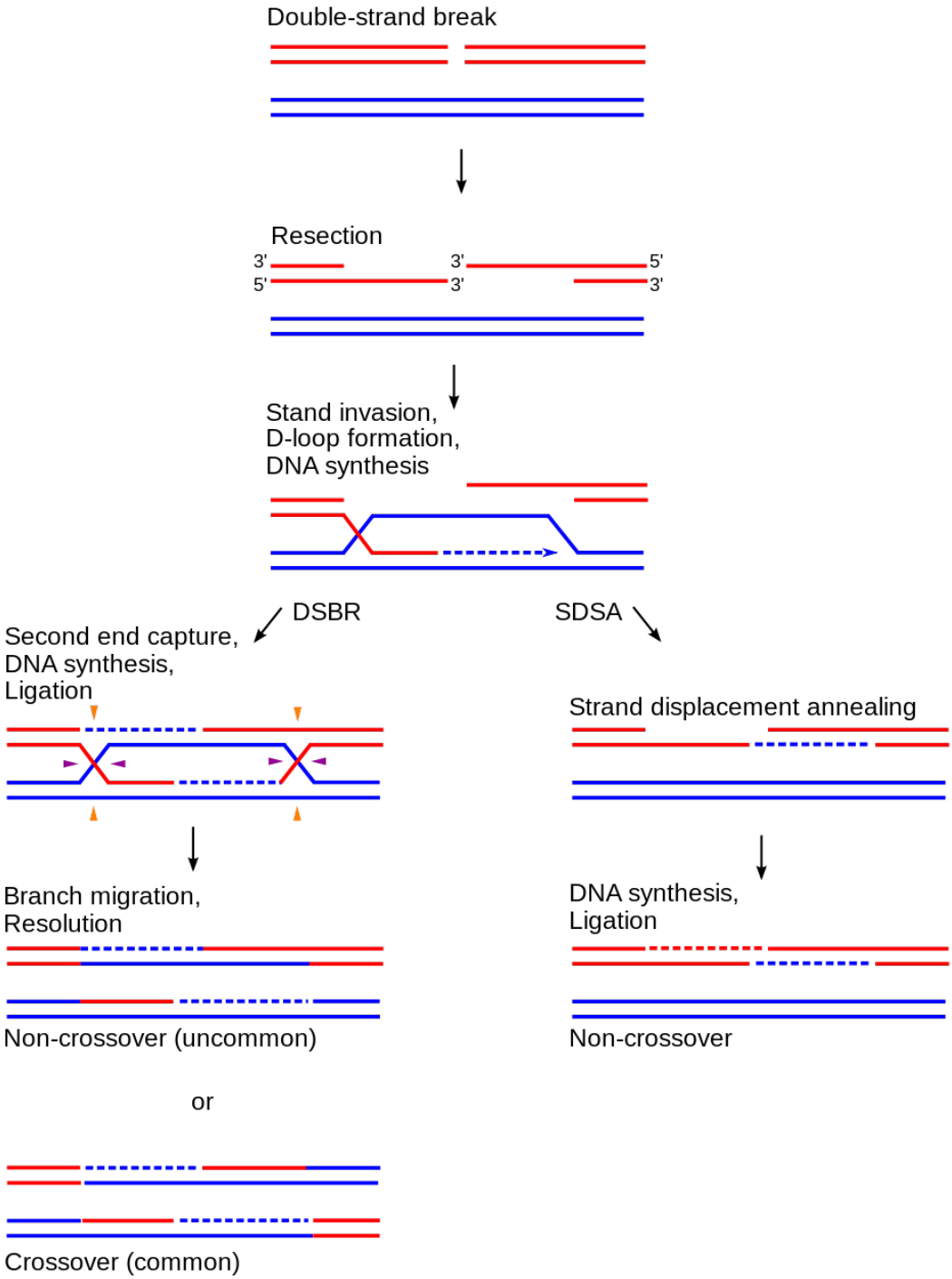


Figure 1.3 Homologous recombination repair pathway.
The main steps of homologous recominbation (HR) following a DSB. The first step is exonucleolytic resection, producing 3' overhanging ends. The free end then invades the template sister chromatid, creating a D-loop. DNA synthesis begins on the end of the invading strand. At this point HR can split down one of two pathways; double strand break repair (DSBR) or synthesis dependent strand annealing (SDSA). In SDSA the D-loop dissolves and the invading strand returns to anneal to its original duplex. The remaining gap is filled by a DNA polymerase and the nicks are ligated. In DSBR DNA synthesis also occurs opposite the D-loop strand and Holliday crossover junctions are formed. Resolution of these junctions results in either crossover or non-crossover products. Figure from Sung (et al., 2006).

Heyer, 2008). The Holliday junctions can be resolved to produce crossover or non-crossover products (Figure 1.3). In SDSA the invading strand leaves the D-loop following limited DNA synthesis and re-anneals with the second end of the DSB. This method avoids the formation of Holliday junctions. HR can also be involved in other mechanisms for managing DSBs, for example with one-ended DSBs following replication fork collapse. HR can promote break-induced replication (BIR) which creates a de novo replication fork from the invading strand, and the entire remaining chromosome arm is replicated (Li & Heyer, 2008). BIR is not yet as well understood as HR mediated DSBR or SDSA and may also play a role in telomere lengthening.

HR is conserved throughout the three main domains of life, with orthologous systems described in both bacteria and archaea. The RecA recombinase protein is a bacterial orthologue of Rad51, and is central to HR in bacteria (Sung & Klein, 2006). HR has been robustly characterised in *E. coli*, via the activities of the RecBCD pathway. RecBCD is a trimeric complex that binds DSB ends and begins to unwind the DNA with the helicase functions of RecB and RecD (Kowalczykowski et al., 1994). This activity is followed by RecA filament formation that initiates a search for a homologous sequence prior to strand invasion. The subsequent D-loop and Holliday junction formation and dissolution are approximate to the eukaryal mechanism. HR in archaea is less thoroughly delineated, however orthologues of Rad51 have been characterised in some species, labelled RadA and RadB. RadA shares the RecA/Rad51 ATPase, D-loop and strand invasion capabilities, whereas RadB is thought to play a role in resolving Holliday junctions along with the archaeal Holliday junction resolvase Hjc (Komori et al., 2000). It is clear that HR is a critical pathway for surviving DSBs, since it is conserved in all life forms. The ability to accurately repair severely damaged DNA is irreplaceable.

1.2.2 An introduction to the Nonhomologous End Joining repair pathway

NHEJ is titled to contrast with HR; no lengthy stretch of homologous DNA is required for repair, the broken ends are directly brought together and reconnected by a DNA ligase (Moore & Haber, 1996). The ends may require limited processing before ligation, and this is performed by various nucleases, polymerases, kinases and phosphatases (Figure 1.4). These modifying enzymes interact with the DNA end binding protein heterodimer Ku70/80 which, along with the DNA ligase, are the hallmark components of NHEJ (Lieber, 2010). Ku70/80 has additional functions beyond acting as a scaffolding protein, it maintains the DNA ends in proximity, and

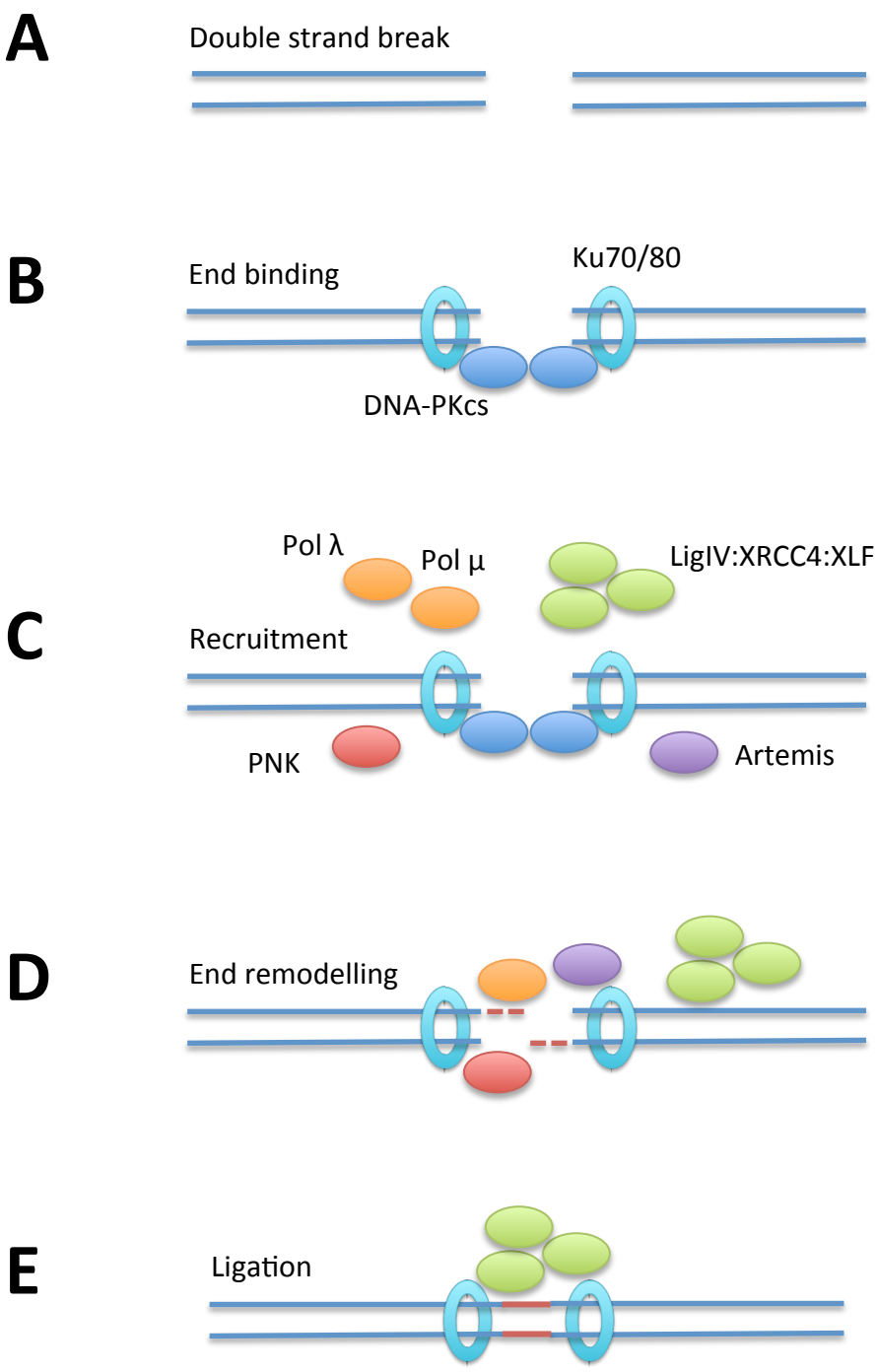


Figure 1.4 Nonhomologous end joining repair pathway.
(**A, B**) The initial step of NHEJ following a DSB is binding of the ends by the Ku70/80 heterodimer and the DNA-PKcs. This keeps the ends proximal to each other and prevents resection that could initiate HR. (**C**) DNA-PKcs then autophosphorylates itself and other NHEJ repair enzymes that are recruited to the DSB. (**D**) DNA-PKcs vacates the termini allowing the family X polymerases, Artemis nuclease and polynucleotide kinase/ phosphatase to access the broken DNA and begin remodelling it. The DNA can be remodelled in a large variety of ways and it is not yet known what process regulates the activities of the different enzymes at the DSB. (**E**) Once the breaks are restored to two DNA nicks, LigIV repairs the DNA to a complete double helix.

prevents unregulated exonucleolytic resection. The title 'NHEJ' could be considered somewhat of a misnomer, since a variety of possible repair intermediates do make use of small sequences of homology, termed microhomology. Between 1-5 bases can be paired to tether the two disconnected ends (Lieber, 2010). Since NHEJ is the focus of this thesis, it will be discussed comprehensively below. Before completing the discussion of DSB repair with the less established repair mechanisms (microhomology-mediated end joining and alternative end joining), the regulation of HR and NHEJ will be considered.

1.2.2.1 **Double strand break repair pathway choice**

Whilst a second copy of the gene sequence is a prerequisite for HR to successfully repair a DSB, HR remains active throughout the cell cycle. Likewise, NHEJ does not require any template to replicate from and is also active throughout the cell cycle, even in S/G2 when the more accurate repair option of HR is available. Regulation of NHEJ and HR is a delicate matter in eukarya, and much of the mediation is handled by the proteins 53BP1 and breast cancer type 1 susceptibility protein (BRCA1). 53BP1 locates to DSB sites and interacts with the chromatin and prevents resection of the DSB ends that would encourage HR to proceed (Chapman et al., 2012). BRCA1 is an antagonist of 53BP1, found to promote HR during the appropriate points in the cell cycle. The mechanism of this antagonism is currently unclear, but in embryonic lethal cells homozygous for *Brca1* mutations, deletion of *53BP1* rescued cell viability (Bunting et al., 2010). Competition between NHEJ and HR also occurs directly at the break, since Ku70/80 binds DNA ends with high affinity (Lieber, 2010), and the MRN complex and CtIP require access to the broken ends to begin resection for HR. It has been speculated that MRN mediated endonucleolytic cleavage of the DNA distal to the DSB could encourage loss of Ku binding and promotion of HR (Chapman et al., 2012; Garcia et al., 2011). Cell cycle CDK dependent phosphorylations regulate the resection based initiation of HR, as during G1 CtIP levels are low and the protein remains unphosphorylated until S/G2 when the genome has been copied (Chapman et al., 2012).

1.2.3 Microhomology-Mediated End Joining

Microhomology-mediated end joining (MMEJ) is pathway for repairing DSBs that was, until recently, thought of as a back up to HR and NHEJ. MMEJ is Ku70/80 independent, but requires Mre11, CtIP and DNA ligase III (Truong et al., 2013). End resection in MMEJ is limited compared to HR, only 20 bases are necessary for the formation of a microhomology synapse. This microhomology is longer than those intermediates usually detected during NHEJ. MMEJ

becomes active during S phase, contrasting with G0/G1 for NHEJ. DSB repair competition assays established that MMEJ does repair 10-20% of breaks even in the presence of functional HR and NHEJ pathways, suggesting that it is not just a back up mechanism (Truong et al., 2013).

1.2.4 Alternative End Joining

It is possible that alternative methods of Ku-independent end joining exist that are distinct from MMEJ. In Ku or DNA-PKcs deficient cell lines, end joining was shown to still occur but repair was considerably slower than NHEJ. These repaired breaks were found to rely on poly(ADP-ribose) polymerase 1 (PARP1), which is known to play a role in recruiting Mre11 and Nbs1 to DSBs for resection (Mansour et al., 2010). Although the proposed alternative end-joining pathway required resection to progress, like MMEJ, microhomologies were not utilised during repair. The average microhomology was found to be either one or no bases (Mansour et al., 2010). Alternative end joining pathways may exhibit several elements of redundancy with other repair pathways. It is conceivable that in the loss or failure of dedicated pathways, other DNA repair enzymes can be used as back ups, depending on the specific damage at the DSB.

1.3 NHEJ in eukaryotes

The core enzymes for NHEJ are Ku70/80 and DNA Ligase IV. In eukaryotes, these enzymes require an array of other repair factors to efficiently repair DSBs (Daley et al., 2005). Ku70/80 binds the DNA ends and interacts with DNA-dependent protein kinase catalytic subunit (DNA-PKcs). The Ku70/80 and DNA-PKcs complex allows for the recruitment and activation of processing enzymes that can modify the DSB termini, if necessary, before ligation (Lieber, 2010). Broken DNA ends often require DNA synthesis in order to be efficiently repaired, and it is the family X polymerases Pol μ and Pol λ that perform this task. Although resection is frequently observed in DSBs repaired by NHEJ, the steps involved are not entirely clear. Artemis bound to DNA-PKcs is reported to have a variety of exonucleolytic and endonucleolytic activities, however questions remain over the precise capabilities of Artemis (Lieber, 2010). DNA Ligase IV forms a complex with X-ray cross-complementing protein 4 (XRCC4) and XRCC4-like factor (XLF), known as LXX (Hammel, Yu, Fang, et al., 2010a; Riballo et al., 2008). The LXX complex offers a remarkable degree of flexibility in ligation of different substrates, although it does a require 3'-hydroxyl group (3'-OH) and a 5'-phosphate (5'-P) to seal the DNA.

Polynucleotide kinase/phosphatase (PNK) has two distinct domains that are capable of 3'-phosphate removal to produce a 3'-OH, and phosphorylation of 5' DNA ends to generate a 5'-P (Weinfeld et al., 2011).

1.3.1 DSB recognition

1.3.1.1 **Ataxia telangiectasia mutated**

Before the NHEJ factors described above can be recruited to a DSB, the break must first be recognised and highlighted. In higher eukaryotes, altering the structure of chromatin is a necessary first step. One of the primary DSB response proteins is Ataxia telangiectasia mutated (ATM), which, when mutated, can cause the disorder ataxia telangiectasia (Kastan & Lim, 2000). ATM is a serine/threonine kinase that exists as an inactive dimer in human cells, until IR induced DSBs promote the autophosphorylation of serine 1981 (Bakkenist & Kastan, 2003). The MRN complex is implicated in the activation of ATM, through the Nbs1 component (Lavin & Kozlov, 2007). The phosphorylation event causes the dissociation of the ATM dimer, and the active monomer may then begin to phosphorylate target proteins to initiate the DNA damage response (DDR).

1.3.1.2 **H2AX**

A key target of ATM is the histone variant of H2A called H2AX, which becomes γH2AX when phosphorylated on serine 139. Recognition of 'activated' γH2AX occurs following the subsequent dephosphorylation of tyrosine 142 (Lukas et al., 2011), when mediator of DNA damage checkpoint protein 1 (MDC1) is recruited to the histone and binds via its BRCA1 C terminus (BRCT) domain. MDC1 then becomes the target for other kinases, and along with ATM, helps to amplify the γH2AX region around the DSB (Cobb & Lees-Miller, 2010). Then ubiquitin ligases RNF8 and RNF168 are recruited, and then in turn 53BP1. The conglomeration of kinases and histone modifying enzymes at a DSB serve to relax the chromatin which allows access to the NHEJ repair factors (Goodarzi & Jeggo, 2012).

1.3.2 NEHJ repair factors at the DSB

1.3.2.1 **Ku70/80**

Once the condensation of chromatin packaging around the double strand break has been eased, the first NHEJ repair factor to interact with the DNA is the Ku70/80 heterodimer, so

named because of the relative Kilodalton (kDa) sizes of the two components estimated from protein sodium dodecyl sulphate (SDS) gel mobility (70 and 86 kDa respectively). Ku is a critical component of NHEJ, Ku deficiency causes IR sensitivity and dysfunctional DSB joining (Boulton & Jackson, 1998). The seminal 2.5 Å crystal structure of the dimer elucidated the toroidal shape of the protein, with a cradle underneath, bound to a short DNA duplex (Walker et al., 2001). The structure indicates how the Ku70/80 ring can slide onto a broken DNA end, completely encircling the helix, and attract other proteins with the interacting domains on the cradle. Ku does not make significant contacts with the DNA bases nor the phosphate back bone, instead it interacts sterically with the major and minor grooves of the double helix. This manner of interaction with DNA affords Ku a high binding affinity with a dissociation constant of $1.5\text{-}4x10^{-10}$M (Dynan & Yoo, 1998). Ku70 and Ku80 share significant sequence homology, and both share similar core domains and van Willebrand (vWA) domains, however they differ at their C-termini (Figure 1.5). Ku70 has a SAF A/B, Acinus and PIAS (SAP) domain, whereas Ku80 has an extended C-terminus, consisting of a flexible linker and then a helical bundle. The bundle is composed of 6 helices that form an overall right-handed superhelix, which is thought to be involved in protein binding (Mahaney et al., 2009; Wuttke, 2004). Ku has been demonstrated to directly interact with a large number of NHEJ repair factors, including DNA-PKcs, the LXX complex, and Pol μ and Pol λ.

An area of long standing mystery concerning Ku is how it dissociates from DNA following repair of the DSB. The ring structure of Ku should leave the protein trapped, if it is still in place after the repair of a DSB. Although the dimer can translocate along DNA, there is no evidence that it can free itself. Recent studies have suggested that Ku can be freed from DNA by proteasomal degradation, and have implicated ubiquitylation and RNF8 as a ubiquitin ligase which specifically regulates Ku80 removal (Postow, 2011; Feng & Chen, 2012).

As noted above, Ku plays a fundamental role in NHEJ, indeed loss of Ku is equated with chromosomal rearrangements and malignant transformation of cells (Difilippantonio et al., 2000). Some of these results may be related to Ku's role in telomere maintenance and silencing (Boulton & Jackson, 1998). Ku has also been reported to have other unexpected activities: 5'-deoxyribose-5'-phosphate (dRP) lyase activity (Roberts et al., 2010). This is the first evidence of Ku possessing enzymatic activities, although whether Ku makes use of the ability to remove AP sites at DSBs *in vivo* remains unknown.

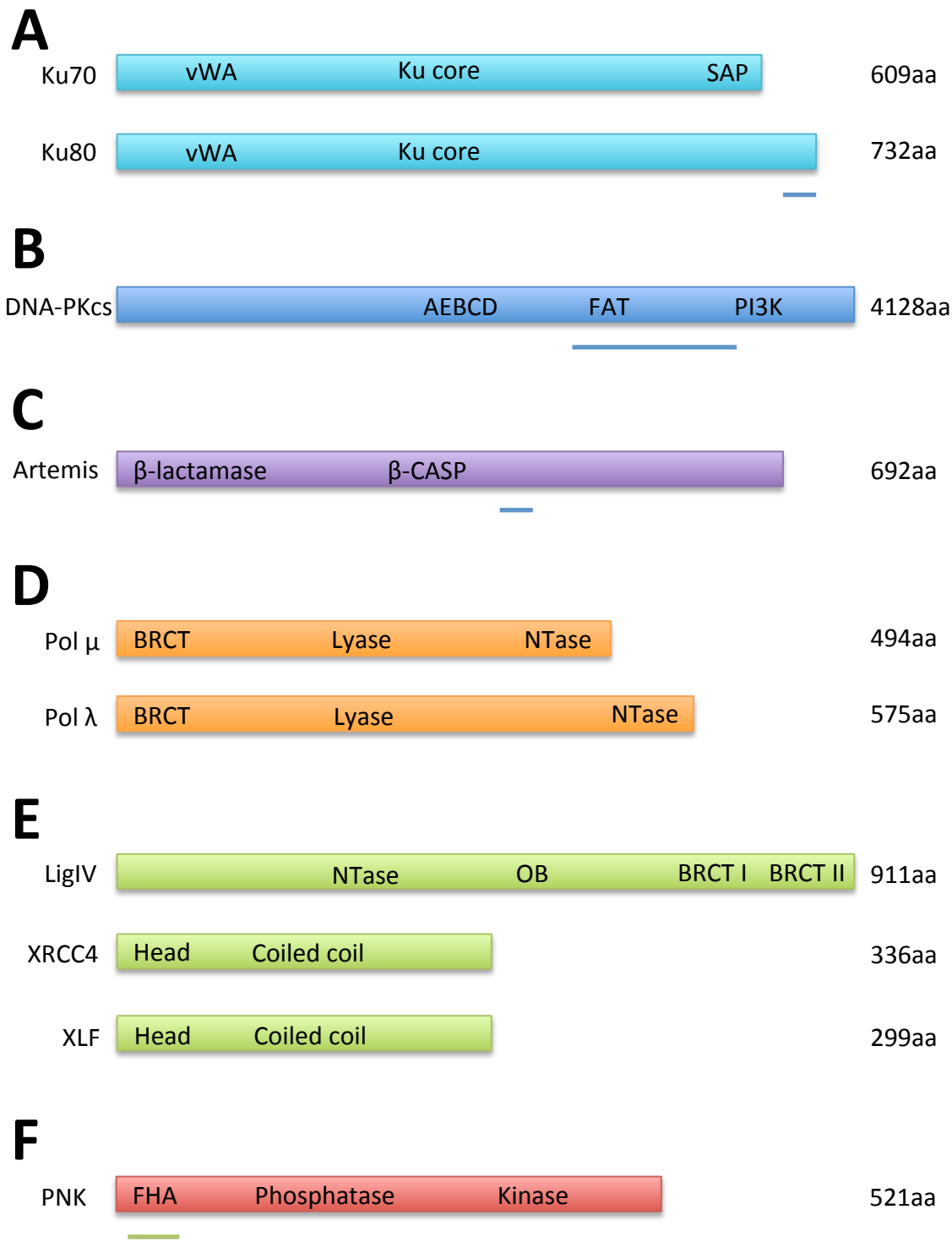


Figure 1.5 Domain organisation of *H. sapiens* NHEJ proteins.
(**A**) Ku70 and Ku80 form a heterodimer. Both Ku proteins have vWA domains and a conserved core. Ku80 has an extended C-terminal region which binds to DNA-PKcs (indicated by the blue line). (**B**) DNA-PKcs has a central autophosphorylation site (AEBCD), a FAT domain which binds Ku (indicated by the blue line), and a PI3K kinase domain. (**C**) Artemis has β-lactamase and β-CASP domains, and interacts with Ku70/80 (indicated by the blue line). (**D**) The family X polymerases μ and λ have similar domain structures with BRCT domains which recruit them to DSBs. (**E**) Ligase IV interacts with XRCC4's coiled coil domain through its BRCT domains. XRCC4 and XLF interact through their respective head domains. (**F**) The FHA domain of PNK binds XRCC4 (indicated by the green line). Adapted from Lieber (2010).

1.3.2.2 **DNA-PKcs**

DNA-PKcs is a protein kinase of the phosphatidylinositol 3-kinase related kinase (PIKK) protein family [which includes ATM and ATM Rad3 related (ATR)] and is one of the first recruitments to Ku at a DSB. The protein is relatively large, over 4000 amino acids (aa) and a molecular weight of 469 kDa (Lieber, 2010). DNA-PKcs features a FAT (FRAP/ATM/TRRAP) domain which binds the Ku70/80 heterodimer, and a C-terminal serine/threonine kinase catalytic site (Mahaney et al., 2009) (Figure 1.5). Like Ku, loss of DNA-PKcs causes severe DSB repair defects and hypersensitivity to IR. DNA-PKcs binds DNA with high affinity, but pre-loading with Ku increases this affinity further. The extended C-terminal region of DNA bound Ku80 interacts with DNA-PKcs, causing Ku to translocate further along the DNA, thus giving DNA-PKcs direct access to the termini. The formation of the DNA:Ku:DNA-PKcs complex is key to tethering the ends of the DSB, the DNA-PKcs proteins on either end of the break make contact with each and trans-autophosphorylate (Hammel, Yu, Mahaney, et al., 2010b).

The kinase activity of DNA-PKcs is critical for successful DSB repair, and small molecule inhibitors which oblate the kinase activity have been demonstrated to increase IR sensitivity. The main target of DNA-PKcs phosphorylation may well be itself, but Artemis and XLF are both found to be phosphorylated on serines or threonines that are part of SQ or TQ motifs (Dobbs et al., 2010). *In vitro* phosphorylation at non-SQ/TQ sites has also been observed on Ku70/80, XLF, XRCC4, Ligase IV Artemis, and also DNA-PKcs, although it is unclear if these events are significant *in vivo*. DNA-PKcs has a threonine 2609 ABCDE cluster which, when phosphorylated, causes a large conformational change in the protein structure and promotes dissociation from DNA. Once DNA-PKcs has vacated the broken termini, other enzymes can then access the DNA for modification and/or ligation.

1.3.2.3 **LXX complex**

The ligation complex of DNA Ligase IV:XRCC4 is conserved in all eukaryotes (Critchlow et al., 1997; Tomkinson et al., 2006). Ligase IV and XRCC4 are both core components of NHEJ, although it is Ligase IV which has the catalytic ability to seal the DSBs, whereas XRCC4 functions as a scaffold protein to recruit and interact with other partners. DNA Ligase IV is an ATP dependent ligase that has considerable flexibility concerning DNA substrates that it can ligate. Aside from being the only mammalian ligase that can seal DSBs in cells, Ligase IV can ligate across gaps, and more remarkably ligate two single stranded ends (Gu, Lu, Tippin, et al., 2007a; Gu, Lu, Tsai, et al., 2007b). Indeed, Ligase IV:XRCC4 and Ku are able to ligate fully

incompatible DNA ends without any microhomology present. Whilst Ligase IV is catalytically competent alone, its activity is stimulated by binding XRCC4, which occurs between the DNA ligase's two C-terminal BRCT domains (Figure 1.5) (Grawunder et al., 1997; Critchlow et al., 1997). The enzymatic activity of Ligase IV:XRCC4 is further stimulated by recruitment to the DSB by binding Ku70/80 (Nick McElhinny et al., 2000).

Mammalian cells have an additional protein, XLF (also known as Cernunnos), which combines with DNA Ligase IV:XRCC4 to create the LXX complex (Ahnesorg et al., 2006; Ropars et al., 2011). XLF is a recently discovered component of NHEJ, and has been shown to further stimulate ligation of non compatible DSBs by DNA Ligase IV:XRCC4 (Gu, Lu, Tsai, et al., 2007b; Hentges et al., 2006). XLF is very structurally similar to XRCC4, a homodimer with a globular head domain and a coiled-coil C-terminus (Figure 1.5). Crystallographic studies have shown that XRCC4 and XLF form filament structures which are proposed to wrap around the DNA ends, encouraging ligation by DNA Ligase IV (Hammel, Yu, Fang, et al., 2010a).

1.3.2.4 **Pol μ**

A subset of polymerases is able to synthesise DNA at a DSB break if the configuration of the ends requires it. Although it has been established that the LXX complex can ligate ends which are non-complementary and even across gaps, these solutions might not be preferred *in vivo*, and having the option to resynthesise DNA adds extra flexibility to an already multifaceted repair toolkit. Pol μ is a family X polymerase, along with pol β, pol λ, pol σ and terminal deoxynucleotidyltransferase (TdT) (Burgers et al., 2001). Pol μ has an unusual set of nucleotidyltransferase capabilities, including tolerance to ribonucleoside triphosphate (NTP) and deoxyribonucleoside triphosphate (dNTP) incorporation opposite a DNA template (Nick McElhinny & Ramsden, 2003). Pol μ is found to favour RNA synthesis over DNA *in vitro*, when the concentrations of NTPs to dNTPs is equivalent to those of cellular pools during G1. The significance of RNA incorporation at DSBs is not yet fully understood, and may just reflect prudent use of available material for repair. Pol μ is capable of not only template dependent synthesis, but also template independent synthesis (Moon, Garcia-Diaz, Batra, et al., 2007a). This ability allows Pol μ to add nucleotides to the end of a break, potentially generating a short microhomology that could stimulate repair and ligation. Perhaps the most remarkable ability of Pol μ is being able to polymerise across a discontinuous template, essentially across two single stranded ends that have been bridged (Nick McElhinny et al., 2005; Davis et al., 2008).

Pol μ is recruited to the DSB via a C-terminal BRCT domain, and has been shown to interact with Ku, DNA Ligase IV:XRCC4.

1.3.2.5 **Pol λ**

Pol λ does not enjoy the same flexibility of activities as Pol μ, and instead is more likely the polymerase of choice if a small region of complementarity is established between the ends of a break, and the resulting gaps either side required gap filling (Lee et al., 2004). Pol λ lacks the structural elements that allow Pol μ to stably incorporate a nucleotide to primer in the absence of a template, but is still able to deal with the irregular DNA templates that it might encounter at a DSB (Moon, Garcia-Diaz, Batra, et al., 2007a). Pol λ is considered a processive polymerase, unlike Pol μ which is distributive. This means that Pol λ is able to efficiently synthesise a stretch of DNA without pausing.

1.3.2.6 **Artemis**

Artemis is a nuclease with roles in NHEJ and V(D)J recombination, and is composed of an N-terminal metallo β-lactamase domain and a β-CASP domain (Figure 1.5) (Mahaney et al., 2009). The C-terminus is highly phosphorylated although the precise importance of these events is not yet clear. Artemis has a 5'-3' DNA exonuclease function *in vitro* when alone, but forms a complex with DNA-PKcs which endows Artemis with 5' and 3' endonuclease functions, and also the ability to nick hairpin DNA (Ma et al., 2002). The 5' endonuclease function produces a blunt dsDNA end, whilst the 3' endonuclease and hairpin nicking functions create a 4 nucleotide (nt) 3'-overhang. These abilities confer further flexibility to the range of damaged ends that NHEJ can repair, in fact Artemis is thought to facilitate NHEJ by removing problematic DNA structures around the break, such as loops (Ma et al., 2005). Artemis, Pol μ and Pol λ have somewhat overlapping roles when creating ends that can form a synapsis, but it is unlikely that all of these factors are active at the same time at a DSB. The regulation of Artemis was previously thought to reside in the C-terminal domain that is highly phosphorylated by DNA-PKcs and ATM *in vitro*. However, more recent experiments have demonstrated that although DNA-PKcs and Ku are necessary for Artemis' endonuclease activities, phosphorylation of Artemis may not be (Goodarzi et al., 2006). Instead, autophosphorylation of DNA-PKcs changes the conformation of the kinase, allowing Artemis access to the DNA ends whilst still in contact with DNA-PKcs.

1.3.2.7 **Polynucleotide kinase**

Polynucleotide kinase (also known as polynucleotide kinase 3' phosphatase, PNK and PNKP respectively) is another unusual enzyme in the NHEJ toolkit. The protein consists of distinct kinase, phosphatase and FHA (forkhead associated) domains, and possesses both 5'-phosphotransferase and 3'-phosphatase abilities (Bernstein et al., 2005; Karimi-Busheri et al., 1998). PNK was demonstrated to interact with members of the BER and SSBR pathways; XRCC1, DNA ligase III and Pol β, and was initially considered only as a factor in repairing single strand breaks. However subsequent studies have shown that PNK also plays an important role in the repair of DSBs in NHEJ; the loss of 5'-phosphotransferase activity was observed after depletion of XRCC4 and DNK-PKcs (Chappell et al., 2002). The role of PNK in NHEJ is two-fold, but both activities enable production of DNA termini that are receptive to ligation. Successful ligation is contingent upon the presence of 5'-P for adenyltransfer and a 3'-OH, however DSBs are often the result of compound damage and might have incorrect ends that cannot be chemically modified by a ligase. PNK provides a resolution to both a 5'-OH and 3'-P (3'-phosphate) which would otherwise impede repair, and creates the 5'-P and 3'-OH necessary. Recent structural studies have enlightened discussion on how exactly PNK interacts with DNA, demonstrating that it is able to interact with both sides of a gap in a single stranded break, and that it could potentially receive DNA termini into both active sites simultaneously (Garces et al., 2011; Coquelle et al., 2011).

1.3.2.8 **Aprataxin and APLF**

There are a number of other enzymes that are implicated in NHEJ, and whilst they perhaps do not have functions as critical as the enzymes discussed above, they offer further repair options for the diverse types of DNA damage present at DSBs. Aprataxin is a member of the histidine triad (HIT) family of nucleotide hydrolases and transferases which catalyses the removal of 5'-adenylate groups (5'-AMP) from DNA ends (Ahel et al., 2006). 5'-AMP is a product of an aborted ligation attempt that can interfere with future ligation reactions. Aprataxin interacts with XRCC4, which implicates that it has a function in NHEJ, along with other pathways (Clements et al., 2004).

Aprataxin and PNK-like factor (APLF, also known as PALF) shares an FHA domain with its titular proteins, and is a recent addition to the NHEJ repair family of proteins. APLF interacts with XRCC4 in the same manner as PNK and Apraxatin, via the FHA domain, and is therefore also recruited to DSBs (Iles et al., 2007). APLF is reported to possess both endonuclease and 3'-

exonuclease functions, the latter of which is interesting given that Artemis lacks 3' directionality of nucleolytic degradation, although the biochemistry requires further investigation (Kanno et al., 2007).

1.3.3 NHEJ proteins also function to generate genetic variation

As discussed above, DSBs repaired by NHEJ often have insertions or deletions that do not match the original sequence. These may have been caused by the damage directly, however the polymerases and nucleases are likely the main cause of the insertions and deletions. This error prone nature can make NHEJ a risky choice, since the repaired DNA could be mutated from the original sequence. However even a single DSB can cause apoptosis, therefore it is often favourable to repair the damage in any way possible. Evolution has found uses for this method of creating genetic diversity and many of the NHEJ proteins play key roles in vertebrate immune systems.

1.3.3.1 V(D)J recombination

V(D)J recombination is a process of generating almost randomised sequences of DNA from segments of three gene sections; variable, diverse and joining. V(D)J occurs only in early B and T cells and makes use of intentionally formed DSBs to enable the genetic segments to be recombined in a new order. These new gene combinations create immunoglobulins and T-cell receptor proteins which can respond to vast array of antigens (Gellert, 2002). In V(D)J it is both the small variations created by NHEJ and the function to join ends which are used to recombine the genes. The DSBs are created by the RAG1/2 protein complex at specific sites, which are called recombination signal sequences (RSS). Cleavage only occurs on one strand at the RSS, which causes the formation of a DNA hairpin at the ends of the sequence. Following the excision of a segment of DNA, the NHEJ proteins become involved. DNA Ligase IV, XRCC4, XLF, DNA-PKcs, Ku and Artemis are all implicated, along with another family X polymerase TdT (Gellert, 2002). TdT is similar to Pol μ, but is only expressed in early B and T-cells, and is capable of extended template independent DNA synthesis (Moon, Garcia-Diaz, Batra, et al., 2007a). The role of Artemis is perhaps better defined in V(D)J recombination than in radiation induced DSB repair. The hairpin nicking function of Artemis serves to create short overhanging DNA flaps that can form microhomologies (Goodarzi et al., 2006). Loss of any of the core NHEJ proteins causes radiosensitive severe combined immunodeficiency (RS-SCID), in which V(D)J recombination is impaired (Goodarzi et al., 2006; Critchlow et al., 1997).

1.3.3.2 **Class switch recombination**

Class switch recombination (CSR) is the second of the only two somatic processes to intentionally cause DSBs. CSR occurs in mature B cells that have already progressed through V(D)J recombination. The mechanism of CSR allows the B cells to change the 'constant' region antibody heavy chain that is produced, whilst the variable segment altered in V(D)J remains unchanged during CSR (Lieber et al., 2006). A B cell is able to switch from producing Igμ, IgM, Igγ, Igα and Igε to IgG, IgA and IgE respectively. CSR is achieved by generating DSBs following cytidine deamination by the enzyme AID, from C to U, on ssDNA regions. These class switch regions are formed by lengthy R-loops, sequences of RNA which base pair with the DNA. RNase HI resects the R-loops which allows AID access to the ssDNA to initiate CSR (Yu et al., 2003). The uracil bases are then converted to abasic sites, which become substrates for APE-1 (AP endonuclease 1). Further enzymatic activity causes the nicks produced by APE-1 to become DSBs, at which point the NHEJ pathway is used to join the DNA ends (Lieber et al., 2006).

1.4 NHEJ in prokaryotes

1.4.1 Discovery of NHEJ in bacteria

DNA repair pathways are often first discovered in bacteria, the genetic malleability of bacteria and rapid multiplication rates make them a highly tractable model system. Homologous recombination was initially described in bacteria, before further characterisation in yeast and vertebrates (Lusetti & Cox, 2002). In contrast, NHEJ was initially discovered in higher eukaryotes and considered an advanced mechanism for genetic repair that only vertebrate members use (Roth & J. H. Wilson, 1986). However, within a decade NHEJ was established in the yeast species *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, with both shown to possess the hallmark NHEJ protein homologues of the Ku dimer and DNA Ligase IV (Goedecke et al., 1994; Kramer et al., 1994). Despite the growing body of knowledge on NHEJ, and its importance in repairing DSBs in the absence of a genetic template, the NHEJ pathway was thought of as a 'recent' evolutionary adaption that had developed since the three domains of life (bacteria, archaea and eukarya) descended from the last universal common ancestor (LUCA) (Figure 1.6). The postulation that bacteria did not feature an NHEJ pathway was reinforced by research in *E. coli*, easily the most widely studied bacteria, that found that the bacterium does not possess any recognisable NHEJ orthologues (Chayot et al., 2010).

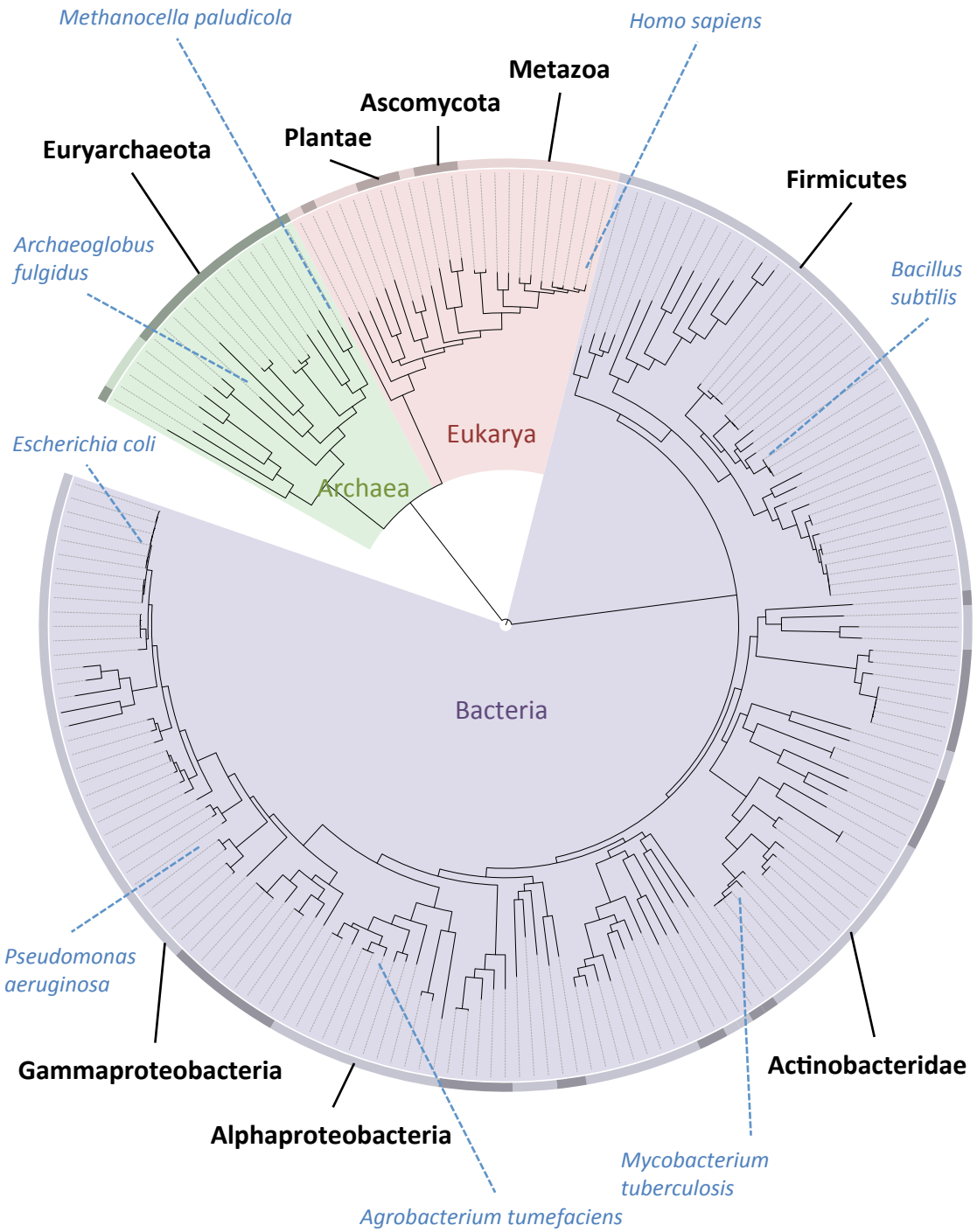


Figure 1.6 Tree of life.
A tree of life showing the three domains of life, with archaea in green, eukarya in red, and bacteria in purple. Phylum are denoted on the outer circle by black text and black links, species are indicated by blue text and blue dashed links. This tree of life was generated by automated comparison of 31 orthologues across 191 species as described in Ciccarelli et al., 2006, using the interactive Tree of Life 2.0 web program (Letunic and Bork, 2011).

At the turn of the millennium, comprehensive *in silico* sequence homology searches revealed the presence of a Ku-like gene in several bacterial species (including *Mycobacterium tuberculosis, Bacillus subtilis, Streptomyces coelicolor*) and a single archaeal species (*Archaeoglobus fulgidus*) (Aravind & Koonin, 2001; Doherty et al., 2001). The pivotal papers also detailed the discovery of ATP-dependent ligase, archaeo-eukaryotic primase (AEP) and potential DNA nuclease genes in an operonic arrangement with the Ku gene (Aravind & Koonin, 2001; Weller & Doherty, 2001). The authors proposed that these operons contained the basic members of a double strand break repair pathway, although the archaeal repair operon in *A. fulgidus* comprised of only a DNA ligase and Ku (Figure 1.7). The speculation that these proteins composed a bacterial end-joining system analogous to that of eukaryotic NHEJ was soon confirmed after experiments in *B. subtilis* and *M. tuberculosis* showed that the bacterial Ku was also a DNA end binding protein, and that it recruited the ATP dependent DNA ligase to broken DNA ends and stimulated ligation (Weller et al., 2002). Loss of either of the ligase or Ku caused hypersensitivity to IR, equivalent to that seen in the vertebrate and yeast systems. The establishment of NHEJ in prokaryotes came with some curiosities, for example the lack of an equivalent protein to XRCC4, fundamental for NHEJ in eukaryotes. Another interesting facet of the prokaryotic repair operon was that the DNA ligase (Ligase D or LigD) was often found as a gene fusion with the AEP and predicted nuclease, expressed as a single contiguous peptide. A subsequent study in *M. tuberculosis* demonstrated that these domains were indeed functional, and incredibly that the AEP possessed an array of nucleotidyltransferase capabilities; gap-filling, template independent synthesis and primase; and that the nuclease domain had 3'-5' DNA exonuclease activity (Della et al., 2004). The prokaryotic NHEJ pathway appeared to be the paradigm of a streamlined repair process; Ku to bind broken ends, a primase/polymerase for DNA synthesis, an exonuclease to produce single stranded DNA for microhomologies, and a DNA ligase to seal the ends following processing. As the number of sequenced bacterial genomes increased, bacteria proved to be a fruitful medium for manipulation and exposition of a NHEJ model system. In depth studies of each component confirmed some of the early conclusions, and others confused the matter, suggesting exciting new possibilities in NHEJ.

1.4.2 Ku

The bacterial Ku protein shares distinct similarities with its eukarotic counterpart, yet is also significantly different. All known bacterial Ku proteins are homodimers, and they lack the vWA and SAP domains that eukaryotes have (Aravind & Koonin, 2001; Doherty et al., 2001). The

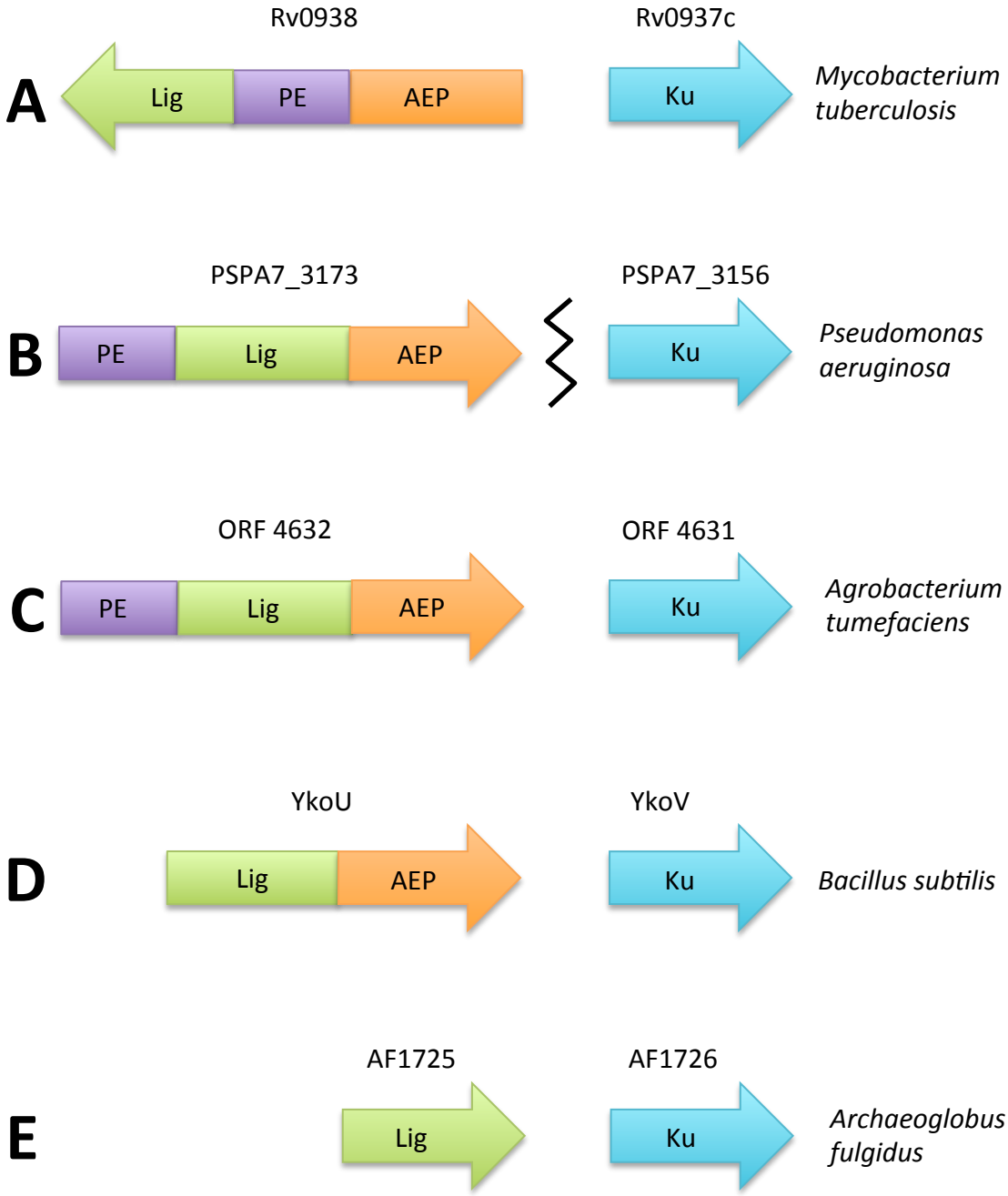


Figure 1.7 Operonic arrangement of bacterial and archaeal Ligase D and Ku.
The domain organisation of multifunctional LigD genes in bacteria (**A**, **B**, **C** & **D**) and archaea (**E**). The gene name is listed above each cartoon representation. The black jagged line between LigD and Ku seen in the *P. aeruginosa* (**B**) cartoon represents an omitted genetic sequence, indicating that the LigD and Ku are not operonic in this case.

cradle observed on eukaryotic Ku, composed of seven-stranded anti-parallel β-barrels, is likely still present in prokaryotes and is predicted by sequence threading analysis (Figure 1.8) (Walker et al., 2001; Doherty & Jackson, 2001; Pitcher, Brissett & Doherty, 2007a). No crystal structure of bacterial Ku has been solved to date, but the conservation of the core of the protein indicates that it probably behaves in a very similar fashion to the eukaryotic equivalent, forming a ring-like structure and sliding over DNA ends. *B. subtilis* (Bsu) Ku was observed to bind dsDNA fragment ends with high affinity in electrophoretic mobility shift assays (EMSA), and increasing the length of the DNA substrate allowed the binding of additional Ku dimers. This suggests that, once bound, Ku can translocate along DNA molecules (Weller et al., 2002). Ku was also proven to stimulate the ligation activity of the associated DNA ligase in both *B. subtilis* and *M. tuberculosis* (Mtu), which was speculated to be caused by recruitment of the ligase to the broken ends by Ku (Della et al., 2004). Yeast two hybrid experiments with *Mycobacterium smegmatis* (Msm) LigD and Ku showed a clear protein-protein interaction (Gong et al., 2005). EMSAs using full length Mtu Ku, LigD and truncated individual domains highlighted the NHEJ polymerase domain as the binding partner of Ku (Pitcher et al., 2005). Msm Ku was also observed to regulate the exonuclease activity of the nuclease domain of Msm LigD; an increasing titration of Ku led to a decrease in resected 3' ssDNA or flap DNA ends (Della et al., 2004). These results supported the theory that Ku also functions in protecting DNA ends from nucleolytic degradation. Ku and LigD have been characterised as the essential members of bacterial NHEJ, depletion of either or both results in hypersensitivity to IR and desiccation in both Bsu and Msm (Pitcher, Green, et al., 2007c; Moeller et al., 2007).

1.4.3 Ligase D

DNA ligases are members of one of two clades; NAD^+ -dependent ligases, or ATP-dependent ligases (Doherty & Suh, 2000). Although the families require different co-factors for the ligation reaction, both feature three otherwise identical nucleotidyltransfer steps during catalysis. In the first step the ligase attacks the α phosphorus of ATP (or NAD) and forms a phosphoramidate bond with the ε amine group of an essential lysine residue in the catalytic site. This results in a covalent bond between the ligase and adenylate group and loss of PPi (or NMN). The second step is transfer of the AMP to the 5' phosphate on the 5' end of the DNA, adenylating the DNA. In the final step of the ligation reaction the ligase facilitates nucleophilic attack of the 3'-OH group on the adenylate group, forming a phosphodiester bond and freeing AMP (Doherty & Suh, 2000; Gong et al., 2004).

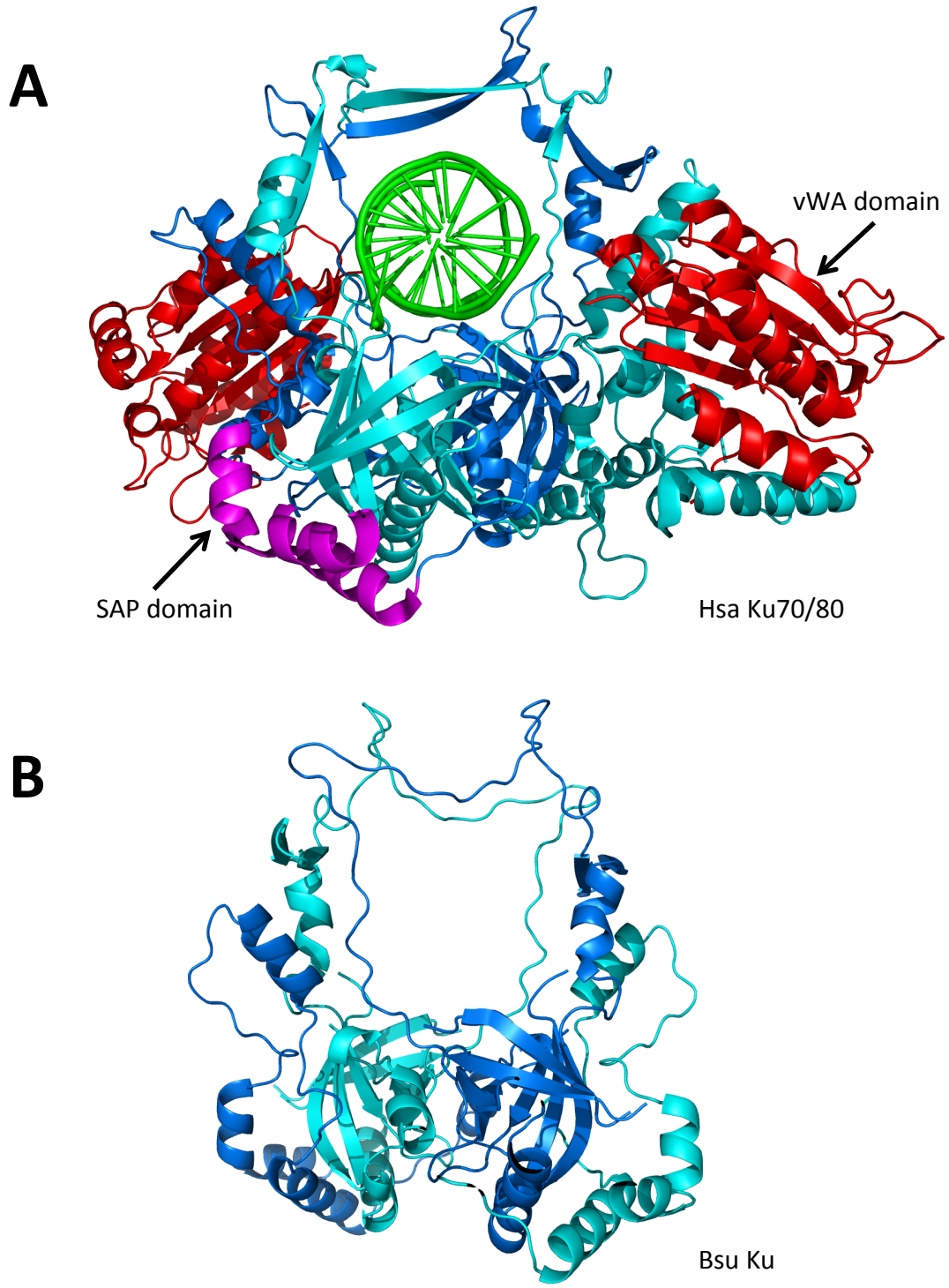


Figure 1.8 A comparison of eukaryotic and bacterial Ku structures.
(**A**) Hsa Ku70/80 heterodimer bound to DNA. Ku70 is cyan and Ku80 is blue. Both have a vWA domain, shown in red, and Ku70 has a C-terminal SAP domain, shown in magenta. (**B**) A threaded model of Bsu Ku. The two individual peptides are coloured teal and blue. The structure prediction shows the conserved β-barrel core of each unit in the cradle beneath the ring. The bacterial Ku is predicted to maintain a highly similar structure to the human protein, except without the auxillary SAP and vWA domains. Hsa Ku70/80 figure created using the atomic structure from Walker et al., 2001 PDB 1JEQ. Bsu Ku figure created from model made by threading the Bsu sequence onto the Hsu Ku70 structure using Phyre (adapted from Pitcher et al., 2007).

The NAD+ dependent ligase is an essential enzyme utilised in the ligation of Okazaki fragments during replication, and was thought to be the only variety of bacterial ligase until the discovery of an ATP-dependent ligase in *Haemophilus influenzae* (Cheng & Shuman, 1997). *M. tuberculosis* is one of the most widely studied bacteria for NHEJ, and it possesses four DNA ligases; NAD^+-dependent LigA, and three ATP-dependent ligases; LigB, LigC and LigD. Of these, LigD is the most comprehensively characterised and is the only ATP-dependent ligase that is composed of three distinct catalytic domains; ligase, polymerase and nuclease (Gong et al., 2004; Wilkinson et al., 2001). As a result, multifunctional DNA ligases in other species that share a similar domain arrangement are also share the appellation 'LigD', even in the absence of ligases B and C.

1.4.4 Structure and function of bacterial NHEJ Ligase D

Initial structural predictions of the mycobacterial LigD ligase domain made from the primary amino acid sequence indicated an N-terminal nucleotidyltransferase (NTase) sub domain and a C-terminal OB (oligonucleotide/oligosaccharide binding) sub domain (Pitcher et al., 2005; Murzin, 1993). The subsequent crystal structure of Mtu LigD ligase domain concurred with the predictions, and displayed a lysine-conjugated AMP in the active site of the NTase domain (Akey et al., 2006). Comparison of Mtu LigD (Figure 1.9, C) and *Homo sapiens* (Hsa) LigI bound to nicked DNA (Figure 1.9, A) revealed a considerable shift in positions of the NTase domains relative to the OB domain (Pascal et al., 2004). Hsa LigI completely encircles the DNA substrates, and the NTase domain is enclosed around the nick, whereas the NTase domain of Mtu LigD does not make significant contacts with the OB domain. This most likely reflects the flexible nature of the ligase protein, with the hinge region between the NTase and OB domains allowing the ligase to engage the nick once it is bound to DNA. A recent structure of bacterial *E. coli* (Eco) LigA bound to an adenylylated DNA nick (Figure 1.9, B) reveals a similar encirclement of the DNA, albeit with different domain organisation (Nandakumar et al., 2007). As of yet there is no crystal structure of a LigD bound to DNA, and it would be very interesting to see if the NTase domain latches around the DNA in a similar way to the essential replicative Hsa LigI and Eco LigA structures.

Mycobacterial LigD possesses the key conserved motifs that confer membership to the polynucleotide ligase and RNA capping enzyme superfamily. Motifs I, III, IIIa, IV and V are found in the NTase domain, along with the recently identified Ia motif, whilst motif VI is found

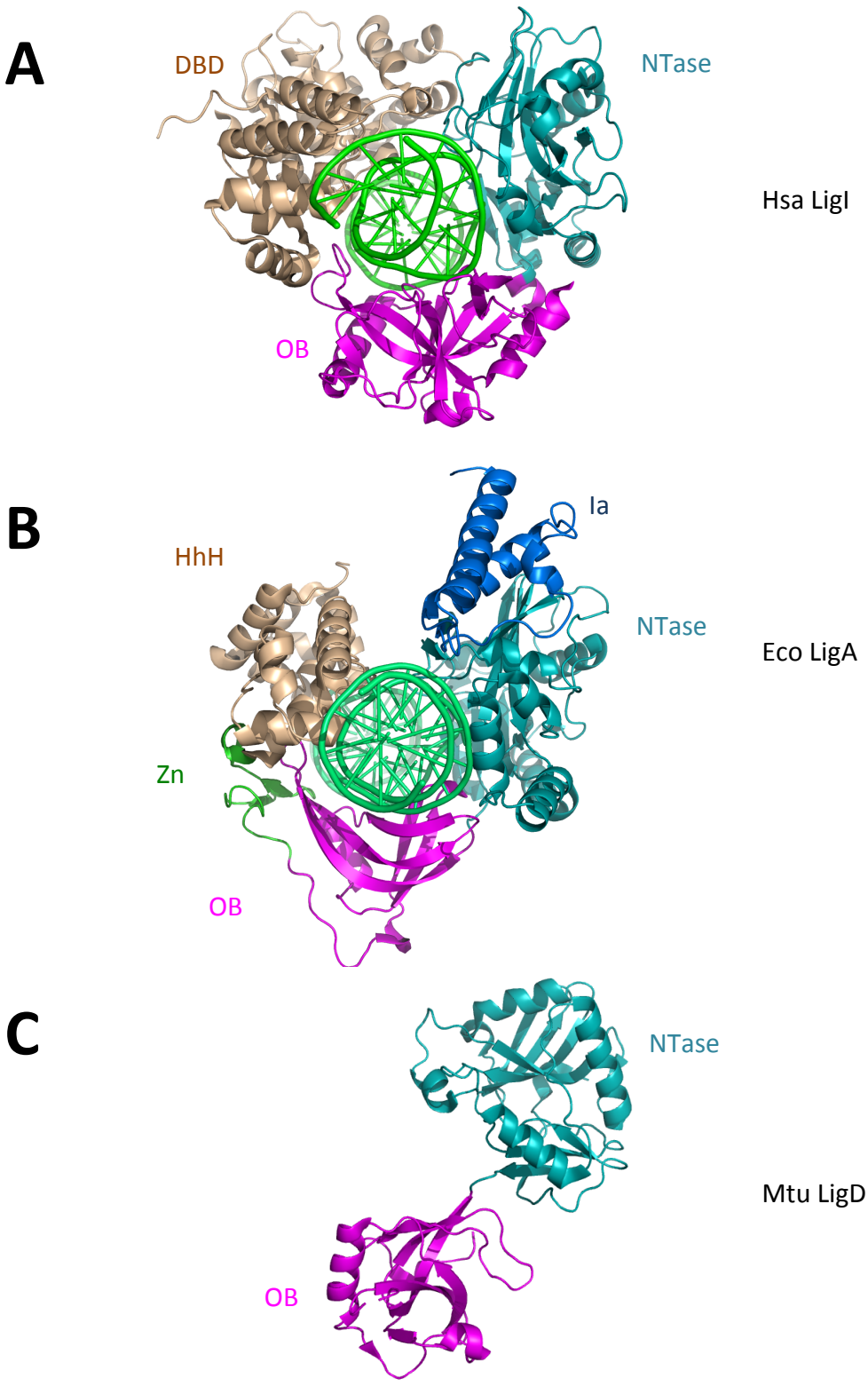


Figure 1.9 A structural comparison between Hsa LigI, Eco LigA and Mtu LigD.
Ribbon diagrams are shown of *H. sapiens* LigI (**A**), *E. coli* LigA (**B**) and *M. tuberculosis* LigD (**C**). The conserved NTase and OB domains are shown in teal and magenta respectively, whilst the DNA binding DBD and HhH regions are wheat coloured, and the Eco Ia and Zn domains are blue and green. Adapted from Nandakumar (et al. 2007), figures of Hsa LigI, Eco LigA and Mtu LigD were created using the structures from Pascal (*et al*. 2004) PDB 1X9N, Nandakumar (*et al*. 2007) PDB 2OWO and Akey (*et al*. 2006) PDB 1VSO respectively.

in the OB domain which coordinates the pyrophosphate leaving group and is essential for adenylylation of the ligase (Figure 1.10) (Doherty & Suh, 2000; Shuman & Schwer, 1995; Shuman et al., 1994; Shuman & Lima, 2004; Smith & Shuman, 2012). The lysine and arginine of motif I contact the α-phosphate and ATP ribose 3'-OH respectively, whilst the conserved arginine of Ia makes ionic interactions with the β and γ phosphates of ATP. The ribose 2'-OH of ATP and AMP is contacted by the glutamate of III, whilst the adenine base is stacked against the aromatic hydrocarbon of either the phenylalanine or tyrosine of motif IIIa. The aspartate or glutamate found in motif IV coordinates the metal ion, and the lysines present in motif V make contacts with the α and γ phosphates of the ATP.

A puzzling aspect of early LigD investigations was the weak capability to seal 3'-OH/5'-P DNA nicks (Gong et al., 2004; Pitcher et al., 2005; Della et al., 2004). Despite being determined as the central ligase for joining DSBs, *in vitro* LigD did not assert a convincing capacity to repair even DNA nicks, let alone bona fide broken ends. Initially, the relative *in vitro* inactivity of LigD was thought to be caused by the proposed structural inability of the protein to completely wrap around DNA (Akey et al., 2006). Observations were also made that showed the propensity of LigD to dissociate from the substrate mid-ligation, after having transferred the adenylyl group to the 5'-P DNA (Gong et al., 2004; Zhu & Shuman, 2007). It was not until investigations in the activities of the other domains of LigD informed the solution to the concerning lack of *in vitro* ligation: LigD preferentially ligates DNA nicks with a 3'-OH monoribonucleoside (Zhu & Shuman, 2008). Surprisingly, the presence of a single hydroxyl group on the 2' position of the ribose of the 3'ultimate nucleotide at the nick conferred a robust ligation activity. The necessity for an RNA-DNA nick for LigD was first demonstrated in *Pseudomonas aeruginosa* (Pae) and then in *Agrobacterium tumefaciens* (Atu) and seemed the likely preferred substrate for Mtu and Msm LigD (Zhu & Shuman, 2007). Without any structural information, it has been difficult to ascertain why LigD has such precise substrate requirements. The available DNA-ligase crystal structures demonstrate how the OB domain requires a strict B-form DNA secondary structure on the 5'-P side of the nick (Pascal et al., 2004). Incorporation of RNA into DNA causes the change of the secondary structure from B-form to A-form (Figure 1.11) (Watson & Crick, 1953; Franklin & Gosling, 1953; Fedoroff et al., 1993). The 3'-OH side of the nick in the LigI structure adopts an A-form structure, caused by direct interactions from the DNA ligase, it is possible that LigD is unable to replicate these contacts and requires the DNA to adopt A-form by itself in order to seal the nick. Therefore the necessity for a 3'-monoribonucleoside at a nick might be caused by LigD's inability to manipulate the DNA, rather than actively selecting for it.

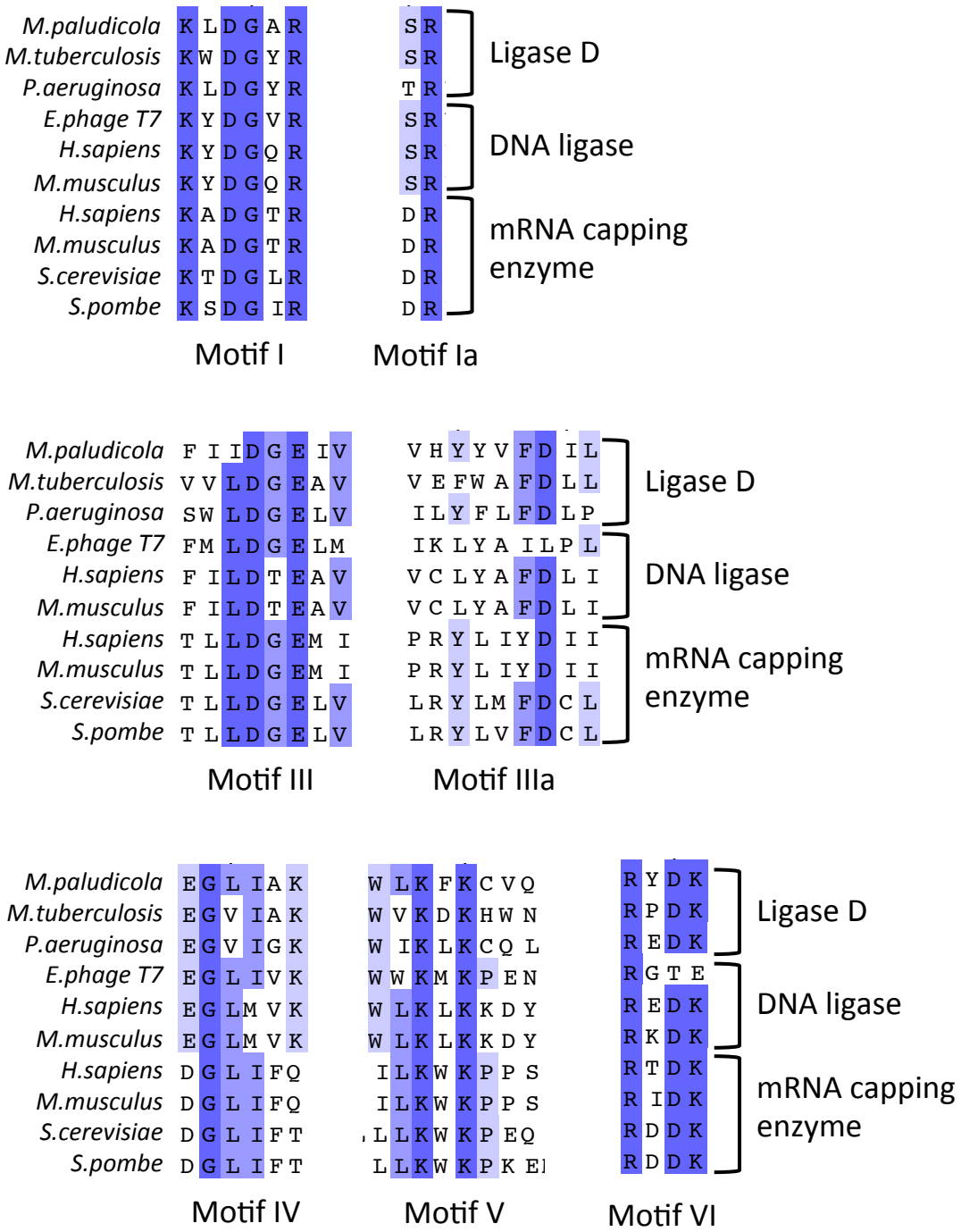
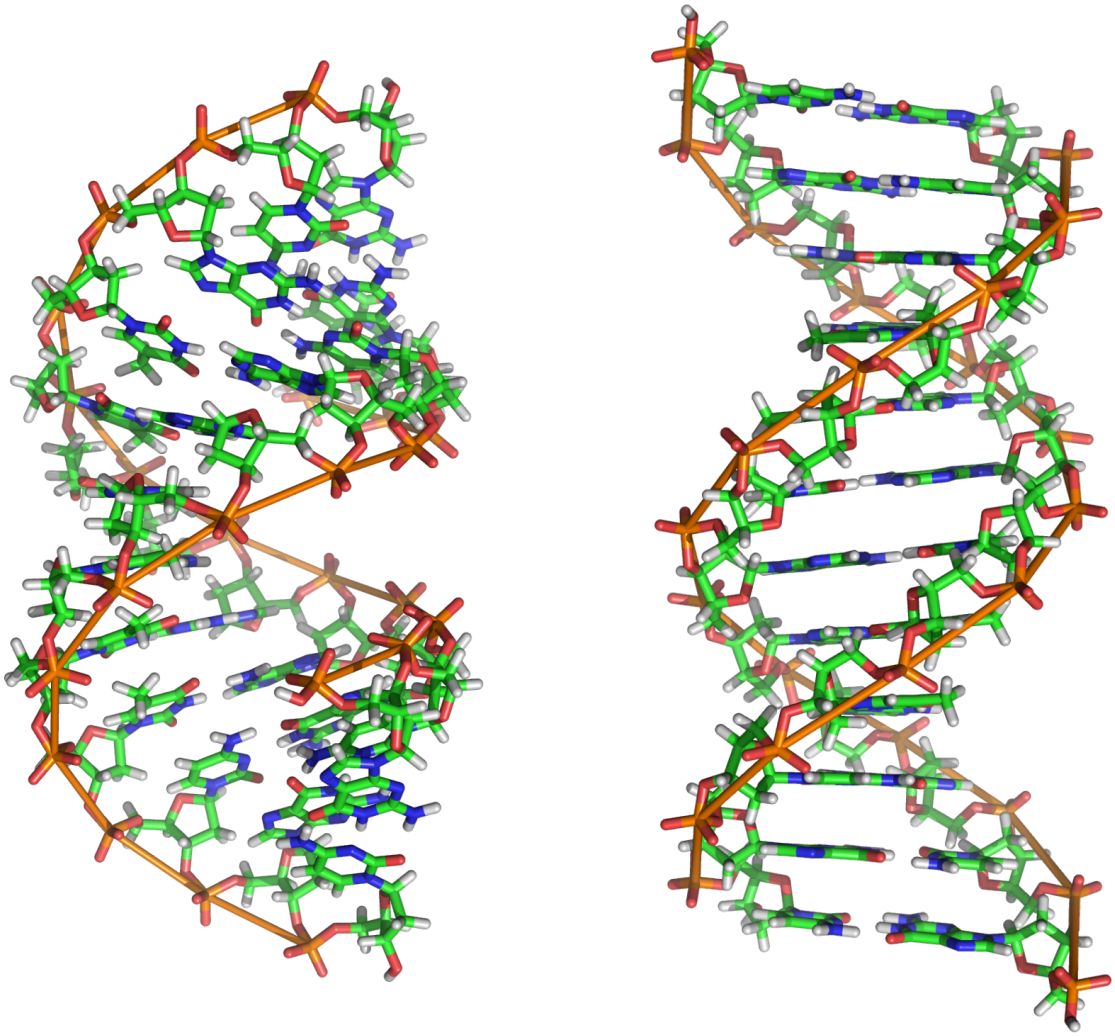


Figure 1.10 The polynucleotide ligase and RNA capping superfamily motifs.
Six key conserved motifs (I, Ia, III, IIIa, IV and V) confer membership to the polynucleotide ligase and mRNA capping superfamily. Motif VI is also included with highly conserved residues. Here Mpa Lig is aligned with *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* Ligase D; Enterobacterio phage T7, *Homo sapiens*, *Mus musculus* DNA ligase; and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* mRNA capping enzymes.



A-form DNA

B-form DNA

Figure 1.11 A structural comparison between A-form and B-form DNA.
A-form DNA has marginally more bases per turn than B-form (11, rather than 10.5), causing a deeper major groove and a shallower minor groove. Addition of RNA bases to a DNA double helix can make it assume A-form. B-form DNA composes the recognisable double helix, with a right handed turn, shared by A-form (Watson & Crick, 1953; Franklin et al., 1953).

Biochemical *in vitro* tests of a range of RNA-DNA substrates have shown that Pae LigD is most efficient at sealing nicks with just a single 3'-ribonucleoside (Zhu & Shuman, 2008). Whilst Pae LigD was capable of ligating RNA-DNA nicks with up to three ribonucleosides on the 3' end of the nick, the efficiency of ligation dropped appreciably, to 10% of that of a monoribonucleotide nick when three ribonucleotides were present (Zhu & Shuman, 2008). Another interesting finding is that the ligation activity of Pae LigD is more efficient using manganese as a cofactor, rather than magnesium, although limited ligation was still observed in the presence of the latter metal ion (Zhu & Shuman, 2008). Understanding the somewhat unusual preferences of the ligase domain of LigD is a target of this thesis, although they begin to make more sense when considered in the context of the other domains.

1.4.5 Structure and function of bacterial NHEJ polymerase

The polymerase domain of the multifunctional LigD was the first component of bacterial NHEJ to be crystallised, and remains the most intensively studied (Zhu et al., 2006). The interest around the bacterial NHEJ polymerase domain of LigD was inspired by its myriad nucleotidyltransferase activities, its membership of the archaeo-eukaryotic primase (AEP) superfamily and more recently the ability to manipulate DNA DSB termini and create a synapse (Aravind & Koonin, 2001; Brissett et al., 2007; Pitcher, Brissett, Picher, et al., 2007b; Zhu & Shuman, 2005a).

1.4.5.1 **Archaeo-eukaryotic primase superfamily**

DNA primases are enzymes with a unique nucleotidyltransferase activity; the ability to synthesise *de novo* short tracts of RNA opposite a ssDNA template. DNA polymerases are unable to synthesise DNA unless an existing DNA or RNA primer is available to offer a 3'-OH to extend from, and they require primases for this task (Frick & C. C. Richardson, 2001). Bacterial primases are classified as DnaG-like primases, so called because of the prior studies of the *E. coli* DnaG protein which typified the homologues of this class of enzyme (Aravind et al., 1998). DnaG primases are essential for replication and often found associated with DNA helicases. All DnaG primases feature the TOPRIM (topoisomerase/primase) motif, which consists of a 5 stranded β-sheet which is surrounded by 6 α-helices (Aravind et al., 1998; Keck et al., 2000). Archaea and eukaryotes have a different class of primase, a heterodimeric protein composed of a small and a large subunit, dubbed PriSL (Frick & Richardson, 2001). Members of the AEP superfamily share homology with the small subunit of the DNA primase, PriS, and have since

been found to populate even bacterial and viral species (Aravind & Koonin, 2001; Iyer et al., 2005).

Structurally, an AEP features an N-terminal $(\alpha\beta)_2$ unit and a RNA recognition motif-like (RRM) fold that is present in the 'palm' domain of DNA and RNA polymerases. Three distinct motifs in the active site are highly conserved; motif I 'hhhDhD' where 'h' is a hydrophobic residue, motif II 'sxH' where 's' is a small residue, and motif III 'h-' where '-' is an acidic residue (usually aspartate) (Figure 1.12) (Iyer et al., 2005; Pitcher et al., 2005). The aspartate triad present in motifs I and III coordinate the metal ions in the active site, whilst the histidine in motif II helps to position the incoming nucleotide correctly (Figure 1.13). Crystal structures of the LigD polymerase domain (referred to as PolDom, LigD Pol or NHEJ Pol) bound to both deoxyribonucleotides and ribonucleotides, and alone, confirm that the NHEJ Pol is indeed an AEP (Zhu et al., 2006; Pitcher, Brissett, Picher, et al., 2007b). An intriguing observation arises from comparison with the NHEJ Pol active site to that of the eukaryotic family X NHEJ polymerases (μ, λ, TdT and β); shared geometry of the active site aspartate triad (Figure 1.13). The family X polymerases are not members of the AEP superfamily and they do not possess the RRM fold, and instead have an 8kDa phosphate binding domain which PolDom does not have. The family X polymerases do not appear to be genetic descendants from the bacterial AEP, which itself is likely a product of horizontal gene transfer, and yet the NHEJ polymerases have convergently evolved towards the same mechanism of chelating metal ions in the active site (Pitcher, Brissett, Picher, et al., 2007b).

There are several structural elements that are absolutely conserved throughout the NHEJ AEP sub-family that are not found in replicative AEPs; two surface loop insertions, Loop 1 and Loop 2, and a positively charged 5' phosphate binding pocket (Figure 1.14) (Brissett et al., 2007; Brissett et al., 2011). Loop 2 is adjacent to the active site, and upon binding of the correct incoming nucleotide the loop moves out of the active site, allowing the aspartate triad to rotate and coordinate the metal ions properly with the nucleotide phosphates (Brissett et al., 2011). EMSAs have demonstrated that Mtu PolDom binds specifically to DNA ends with a 5'-P, and this binding is lost upon mutation of a conserved lysine residue (K16) in the positively charged pocket (Brissett et al., 2007). Whilst earlier binding studies implicated Ku as the factor that recruits LigD to the DSB via the PolDom, it is likely that once localised, PolDom interacts with a 5'-P if present. Loop 1 is involved in the formation of microhomology synapses, and this ability of the NHEJ AEP will be detailed after discussion of the nucleotidyltransferase activities.

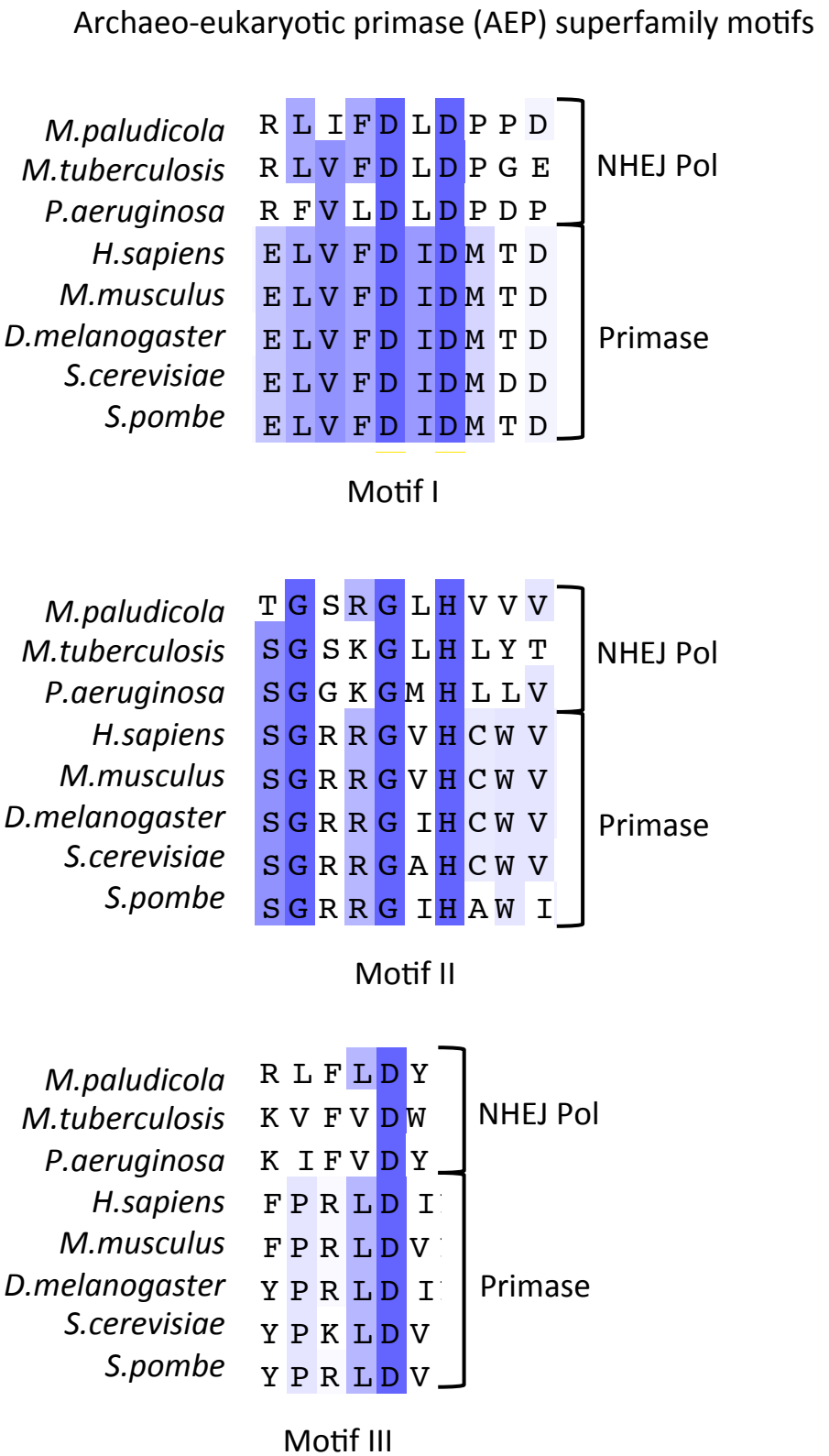


Figure 1.12 Archaeo-eukaryotic primase (AEP) superfamily motifs.
Three key conserved motifs (I, II and III) confer membership to the AEP superfamily. Here Mpa Pol is aligned with *Mycobacterium tuberculosis* and *Pseudomonas Aeruginoa* NHEJ Pols, and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* primase small sub-units.

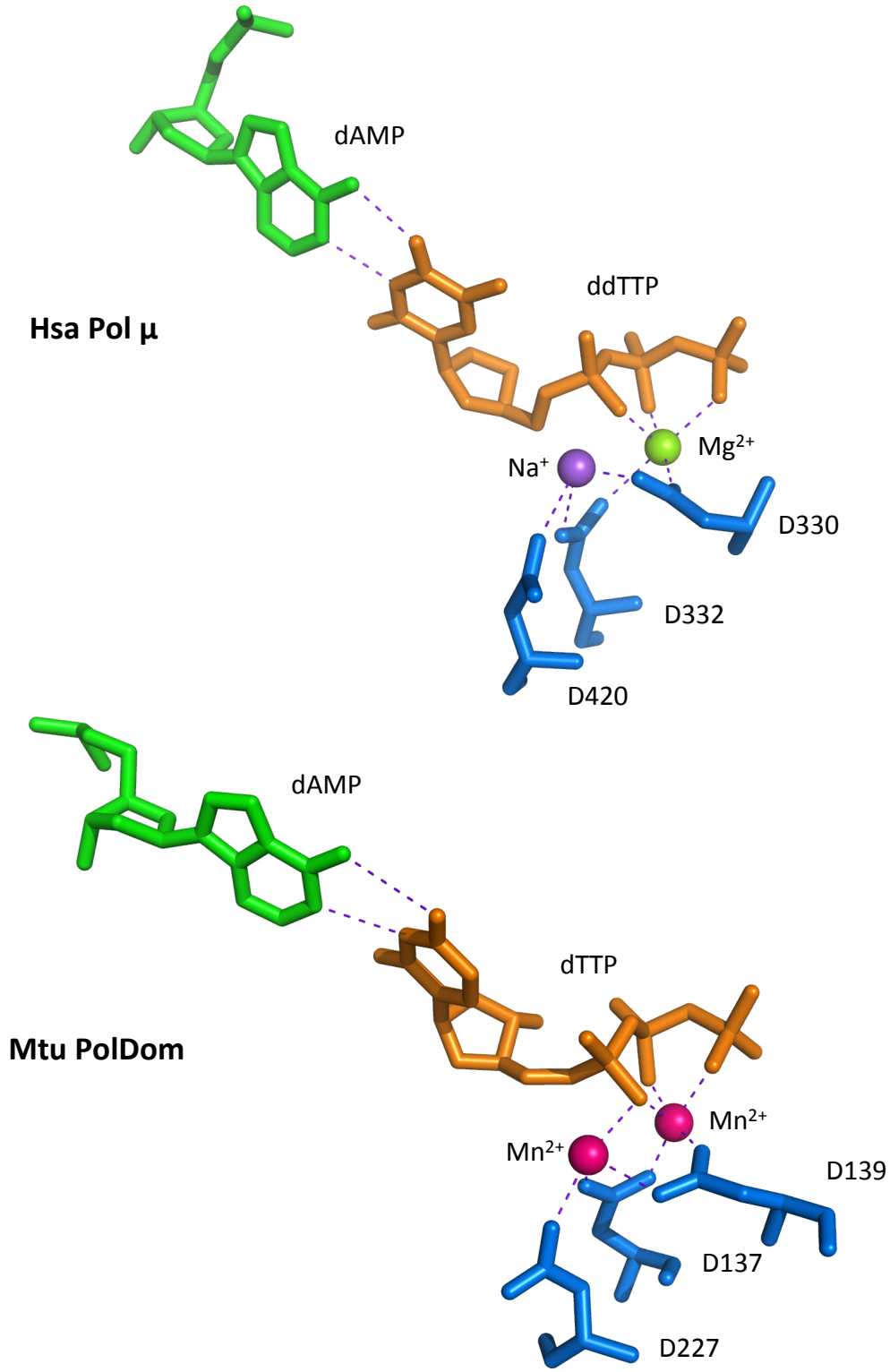


Figure 1.13 Hsa Pol μ and Mtu PolDom aspartate triad in the active site.
The aspartate catalytic triads present in the family X *H. sapiens* Pol μ and the *M. tuberculosis* LigD AEP domain. The aspartate triads are coloured blue, the templating base is green, and the incoming nucleotide is orange. Other residues in the active site help to coordinate the triphosphate tails of the incoming nucleotide, but they have been excluded for clarity. The figures were made using the structures from Moon et al., 2007, PDB 2IHM and Brissett et al., 2011, PDB 3PKY respectively.

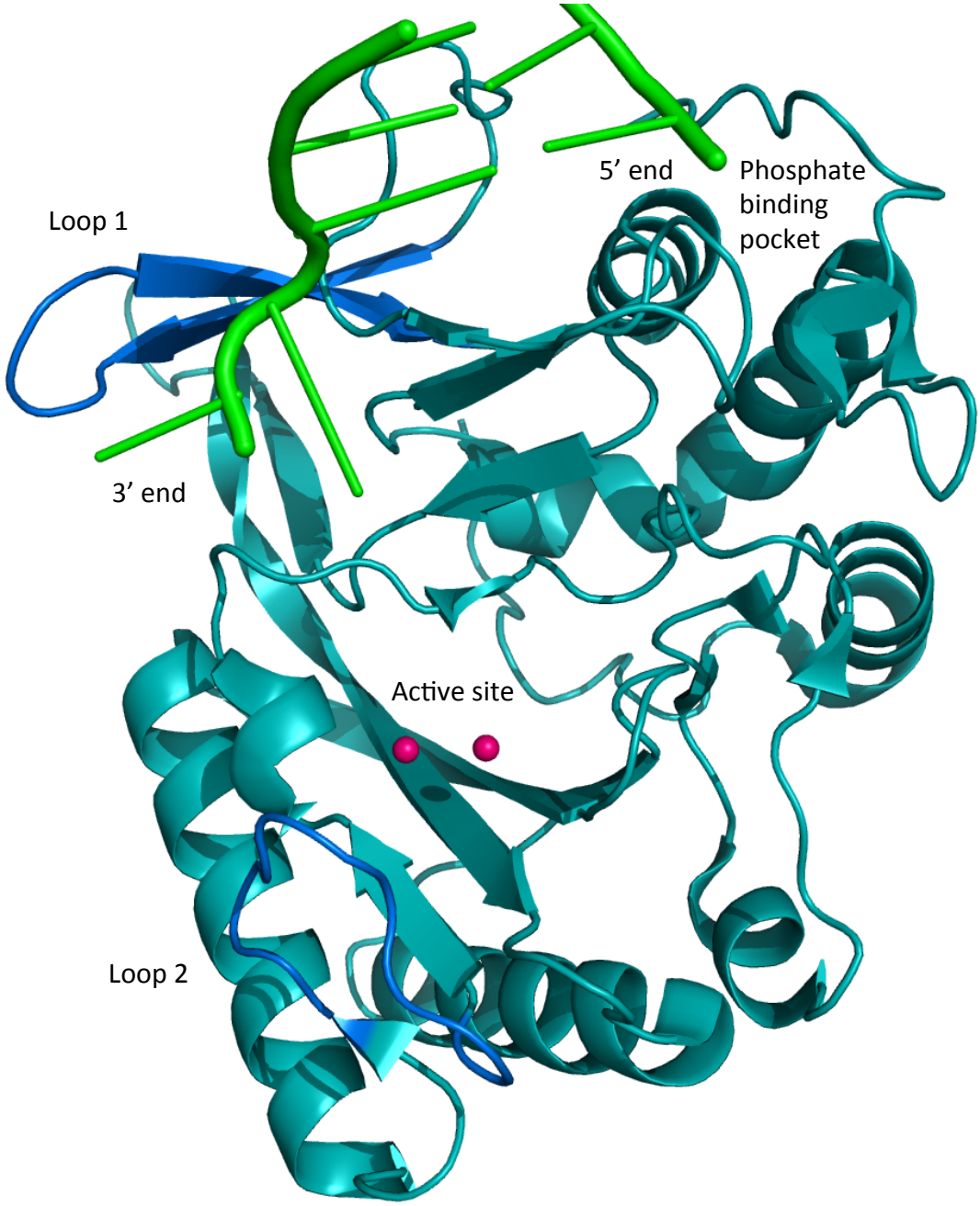


Figure 1.14 Structural features of the NHEJ Pol.
Mtu PolDom loop 1 (blue) interacting with 3’ DNA overhang (green). The active site position is indicated by the presence of manganese ions in pink. Loop 2 (blue) is arranged opposite the active site. The pphosphate binding pocket is noted interacting with the 5’-phosphate (5’-P) DNA end (green). Figure made using structure from Brissett et al., 2011 PDB 3PKY.

1.4.5.2 **NHEJ Pol has a variety of nucleotidyltransferase activities**

Whilst the primary study of the LigD polymerase described it as a DNA dependent DNA polymerase, the research that followed discovered that the PolDom had not only a considerable array of nucleotidyltransferase activites, but also a flexibility in nucleotide usage; both deoxynucleotides and ribonucleotides could be incorporated opposite a DNA template (Della et al., 2004; Pitcher, Brissett, Picher, et al., 2007b; Zhu & Shuman, 2005a). It was quickly established that a variety of LigD PolDoms were not only capable of RNA synthesis, but infact preferred incorporating NTPs to dNTPs. The ability to utilise NTPs in repairing DNA was somewhat unexpected, and may reflect the ability of the polymerase to make use of the most abundant source of repair materials at the point in a cell cycle that NHEJ is prominent. NTP pools are significantly higher than that of dNTP pools during G1 or senescence, and therefore it makes sense for the NHEJ polymerase to make use of them (Nick McElhinny & Ramsden, 2003; Pitcher, Brissett & Doherty, 2007a). The NHEJ Pol, like the ligase domain, is optimally active with manganese as the divalent metal ion co factor. Cobalt offers a similar level of activity, but incubation with magnesium delivered significantly less catalytic activity (Zhu & Shuman, 2005a; Pitcher et al., 2005).

The most surprising aspect of the NHEJ Pol is that they catalyse a variety of reactions *in vitro*; template dependent synthesis, template dependent gap-filling, template independent synthesis, and lesion bypass synthesis (Figure 1.15) (Zhu & Shuman, 2010; Zhu & Shuman, 2005a; Pitcher, Brissett, Picher, et al., 2007b). Neither Mtu nor Pae PolDoms were found to be highly processive polymerases, although easily capable of extending up to a dozen bases. Whilst the template dependent capability was expected, the terminal transferase-like template independent synthesis was not. The only other polymerases to display this activity are the family X polymerases Pol μ and TdT. The ability to insert nucleotides at blunt ends is both a boon in the process of NHEJ, and a risk. If a DNA helix is broken cleanly, with blunt ends, any insertion is a guaranteed alteration to the sequence. TdT activity is limited to V(D)J recombination, which intends to generate new genetic sequences. It is likely that the NHEJ polymerases only possess this capability because of the severity of DSBs, if even one goes unrepaired the cell might undergo programmed death. The ability of the NHEJ Pol to synthesise DNA or RNA on a large amount of substrates is highly similar to that of Pol μ. Both enzymes are also proficient at gap filling, which is thought to be an essential tool for NHEJ polymerases following the formation of a microhomology at a DSB (Moon, Garcia-Diaz, Bebenek, et al., 2007b). In most cases a gap would be present at either side of the synapse, and the ability to efficiently fill the gap stabilises the break and allows repair to proceed

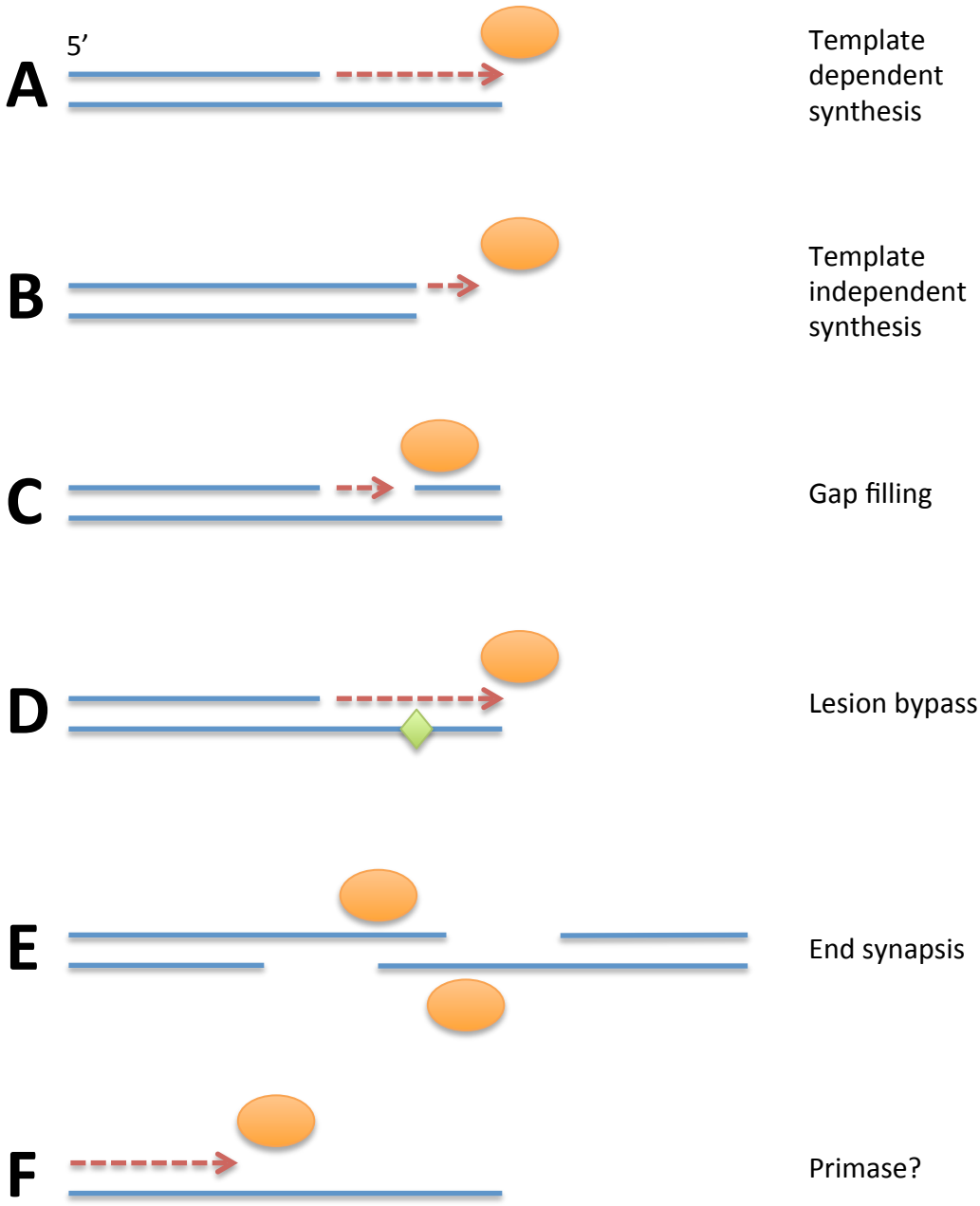


Figure 1.15 Catalytic capabilities of the NHEJ Pol.
The AEP domains of LigD from *M. tuberculosis, M. smegmatis* and *P. aeruginosa* have a large array of nucleotidyltransferase activities. (**A**) The NHEJ Pol preferentially synthesis short tracts of RNA, and does so in a template dependent manner. (**B**) The Pol can also incorporate a nucleotide at a blunt dsDNA end. (**C**) NHEJ AEPs are proficient at gap filling. (**D**) Mtu PolDom has demonstrated the ability to correctly incorporate a nucleotide opposite an 8-oxo-G lesion. (**E**) The NHEJ Pol can configure two ssDNA termini into synapsis, creating stability at a DSB, and gaps that can be filled. (**F**) Early studies suggested that the NHEJ Pol had retained the ability to prime from its primase evolutionary ancestors, although this remains unconfirmed.

towards ligation. Although Pol μ shares with the NHEJ Pol a tendency towards NTP incorporation, gap-filling and template independent synthesis, it does not possess the limited lesion bypass activity that is observed in Mtu PolDom. Mtu PolDom can incorporate the correct nucleotide opposite an 8-oxo-G lesion, although it does not have to ability to extend onwards from this (Pitcher, Brissett, Picher, et al., 2007b). The ability of PolDom to achieve this feat probably rests in the large, open active site of the enzyme, able to encompass a distorted template. This capability could prove very useful at DSBs that have compound damage, and are not amenable to simple reconnection and ligation. The last of a lengthy roster of nucleotidyltransferse abilities that have been reported for the NHEJ Pol is DNA priming. Assays using ss M13 DNA incubated with Mtu PolDom generated lengthy, labelled primers (Pitcher et al., 2005). Whilst PolDom is an AEP, none of the other NHEJ Pols have been reported to possess this primase function, and in the absence of a more comprehensive assay it is difficult to discern if this class of enzyme is a true primase, as well as polymerase.

1.4.5.3 **NHEJ Pol bridges DNA ends**

In addition to a range of polymerase activities, the LigD PolDom also has non-catalytic functions important to the repair of DSBs. The groundbreaking crystal structure of a pair of PolDoms bound to separate DNA termini revealed the ability of the protein to bridge the single stranded ends (Brissett et al., 2007). PolDom binds the DNA at the double stranded region, with the negatively charged 5'-P moiety locking into the positively charged lysine pocket. The 3' ssDNA end is then directed into interaction with the opposite DNA termini by the surface anti-parallel β-stranded loop, Loop 1. Mutation of the apical Loop 1 residues to alanine caused the loss of synapsis formation *in vitro* (Brissett et al., 2007). *In vivo* support of this data agreed that the polymerase domain plays a significant role in NHEJ that is unrelated to its catalytic abilities. Plasmid repair assays with Msm LigD with either PolDom catalytic mutants or PolDom deletion had contrasting results; loss of polymerase activity reduced fidelity of repair, presumably another polymerase took the role of PolDom, but loss of the domain entirely severely reduced DSB repair efficiency (Aniukwu et al., 2008). This body of data supports the concept of the NHEJ Pol functioning in bringing the ends together in certain situations of DSBs. It is unclear whether any of the eukaryotic homologues also possess this function, it may be that Pol μ is able to polymerise DNA across discontinuous ends using a similar methodology (Nick McElhinny & Ramsden, 2003).

A subsequent crystal structure showed that PolDom can bind a recessed 5'-P with a 3' DNA overhang and accept an incoming nucleotide in the absence of a primer (Brissett et al., 2011). This complex is known as 'pre-ternary', since it ready for catalysis, and only needs the primer to extend from. Prior to this data, it was unclear how PolDom might prepare for extension of the DNA after having bound the 5'-P. It seems that the enzyme is capable of assembling the ternary complex and preparing for gap filling before the incoming *in trans* end arrives. It remains uncertain if this model is viable for all DNA ends that the NHEJ Pol might encounter, but it highlights an interesting resemblance to primer synthesis by a DNA primase. The initial step of assembling a primer is the formation of a dinucleotide, and this is performed in the 3'-5' direction, unlike all other DNA synthesis (Frick & C. C. Richardson, 2001). When PolDom assembles the pre-ternary complex, it is essentially mimicking this activity, albeit in the presence of downstream DNA bound by the 5'-P (Brissett et al., 2011).

Two key conserved residues play an important part in the positioning of the incoming nucleotide to the active site, and the incoming primer; both phenylalanines (Figure 1.16). In Mtu PolDom, F63 is observed to base stack with the aromatic ring of a nucleobase of the templating strand. This causes causes a kink in the structure of the DNA, splaying it open over the active site, allowing access to the templating base (Brissett et al., 2007). The neighbouring residue, F64, projects in the opposite direction from the β-strand. F64 first base pairs with the incoming nucleotide, welcoming it into the active site. Then as the templating DNA arrives, Loop 2 moves out, allowing the priming of the active site, which repositions the incoming nucleotide and F64 rotates around to base stack with the templating base (Brissett et al., 2011). The pre-ternary complex is then formed, and awaits the incoming primer for nucleotidyltransfer.

1.4.6 Structure and function of bacterial phosphoesterase

The phosphoesterase (PE) component of LigD has been the most recalcitrant to structural and functional dissemination. The PE domain was first identified as a putative nuclease via *in silico* sequence homology searches when the AEP was classified (Aravind & Koonin, 2001; Weller & Doherty, 2001). The study detected conserved histidine and aspartate residues around an all β-strand core, which were proposed to chelate a metal ion. Given the other predicted enzymatic abilities of the product of the *LigD* gene and the neighbouring *Ku* gene, the authors posited that the likely function of this novel domain was that of DNA exonuclease. The first functional assays of the Msm LigD suggested that the enzyme did possess DNA exonuclease

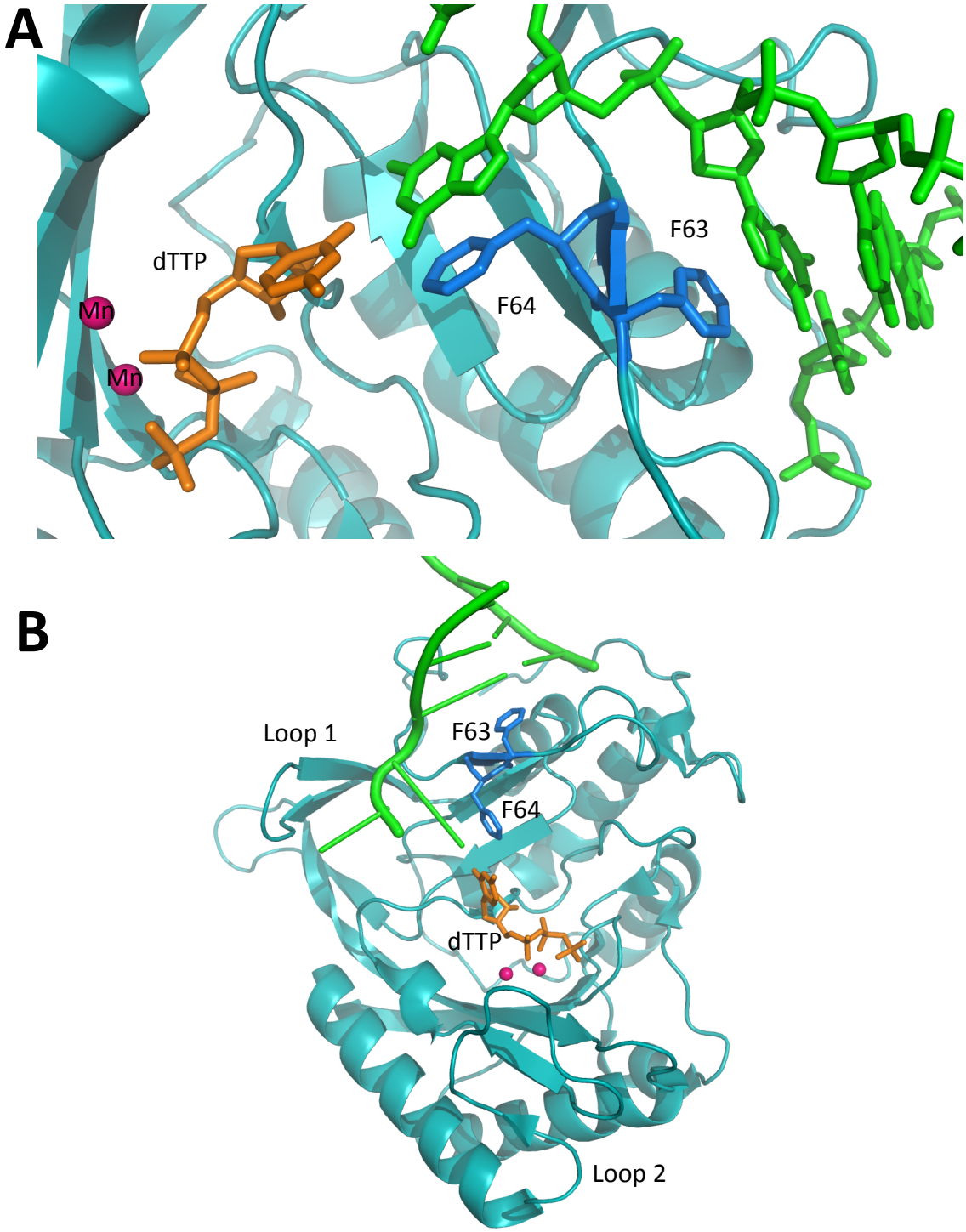


Figure 1.16 The role of the phenylalanines near the active site of the NHEJ Pol. (**A**) The two phenylalanines found on the cusp of the active site of Mtu PolDom are highlighted in blue. F63 can be seen base stacking with upstream DNA (green) and creating a kink in the phosphate backbone. This repositioning of the DNA allows F64 to base stack with the templating base (green) and align it with the incoming nucleotide (orange). The manganese ions (pink) suggest the position of the catalytic residues which are omitted for clarity. (**B**) The full structure of Mtu PolDom bound to DNA shows how the dsDNA region meets F63 and the ssDNA is kinked to allow the templating base to orient towards the active site. Figures made using the structure from Brissett et al., 2011, PDB 3PKY.

activity, and that the ligase, polymerase and DNase activities compared neatly to those of the eukaryotic NHEJ system (Della et al., 2004). However attempts to isolate the nuclease domain (termed NucDom) did not yield any soluble or active protein (Pitcher et al., 2005). The groundbreaking discovery on the nuclease component of LigD came from studies in Pae LigD. The full length Pae LigD was challenged with a variety of substrates to discern the ability of the polymerase to extend from a primer with RNA at the 3' end (Zhu & Shuman, 2005b). The control sample without NTPs or dNTPs yielded a product that was smaller than the initial substrate, indicting resection. A series of experiments quickly established that the nuclease domain alone was capable of catalysing this resection, and that it was actually ribonuclease activity. The ribonuclease had a number of interesting preferences, and two distinct activities; phosphodiesterase and phosphomonoesterase. For this reason, the enzyme was dubbed a phosphoesterase or PE (Zhu et al., 2005).

The phosphoesterase activity of PE is similar to a typical ribonuclease activity, but has a key difference. The PE has a necessity for a 2'-OH group on the penultimate ribonucleobase, not the ultimate base actually being removed (Figure 1.17) (Zhu et al., 2005; Zhu & Shuman, 2005b; Zhu & Shuman, 2006). Therefore, PE can technically remove a terminal dNMP, but only if it is proceeded by a ribonucleobase. The 2'-OH was substituted with a number of other groups, including amine, fluoride and methoxy, but none enabled the catalytic activity similar to that of the hydroxyl (Zhu & Shuman, 2005b). This specificity renders PE inert when challenged with an all DNA substrate, and intriguingly also when only a monoribonucleoside is present at the 3' end. This substrate is the preferred target of the ligase domain of LigD, and suggests some interplay between the two. Indeed, it was the discovery of the RNA based preference of both the PolDom and the PE that lead to the unearthing of LigD's own RNA preference (Zhu & Shuman, 2008). The PE phosphomonoesterase activity involves the scission of a phosphomonoester bond; the removal of a phosphate group from the 3' termini of a break. Unlike the phosphodiesterase activity, the phosphomonoesterase activity has no requirement for a 2'-OH on the penultimate or any other base. The PE can remove a phosphate from an all DNA 3' end, or a DNA-RNA hybrid end (Zhu et al., 2005; Zhu & Shuman, 2005b; Zhu & Shuman, 2006). If the RNA at the 3' end of the DNA substrate exceeds two bases, the phosphodiesterase and phosphoesterase activities operate sequentially, first removing the ribonucleobase, then the 3'-P, before the next ribonucleobase. The PE loses efficiency rapidly when encountering longer stretches of RNA to resect, and has minimal activity when confronted with an all RNA 'primer' (Zhu & Shuman, 2005b).

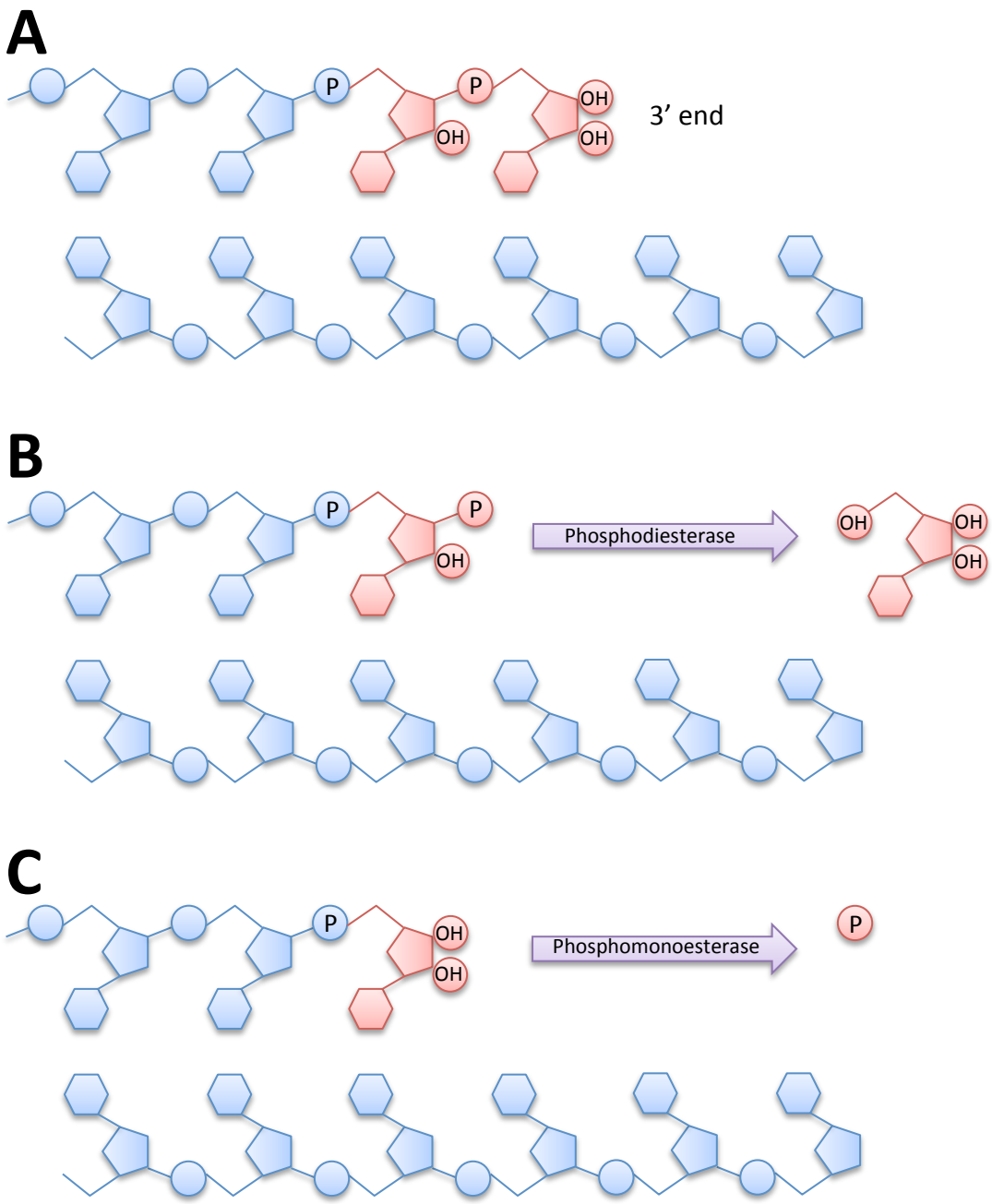


Figure 1.17 Catalytic capabilities of the NHEJ PE.
The NHEJ PE possesses both phosphodiesterase and phosphomonoesterase activities, and is thus labelled a phosphoesterase. (**A**) The substrate for PE is a DNA-RNA hybrid sequence with only a short tract of RNA at the 3’ end. DNA is coloured blue, RNA is red. (**B**) PE recognises the 2’-OH on the penultimate base and engages the substrate. PE then catalyses the removal of the ribonucleoside by cleaving the phosphodiester bond. This leaves a 3’-phosphate group. (**C**) The phosphate is removed by cleavage of the phosphomonoester bond, rendering a 3’-OH group. The PE cannot remove the terminal ribonucleoside since there is no 2’-OH on the penultimate deoxyribonucleoside.

The PE domain of LigD shares the metal ion preferences of its catalytic partners, both of the PE's catalytic activities are optimal in the presence of manganese (Zhu & Shuman, 2005b). This institutes an interesting enzymatic trio with preferences to catalyse reactions dependent on the 2'-OH group of RNA and in the presence of manganese. The significance of these features is not yet clear, beyond that NHEJ in bacteria follows a theme. The residues involved in the chelation of manganese were identified from conserved sequences, and alanine mutations identified six proposed catalytic site amino acids (Zhu & Shuman, 2006). Three histidines, an aspartate, a tyrosine and an arginine were all found to be critical to both of PE's phosphoesterase capabilities. Four residues are also classified as critical for the function of phosphomonoesterase activity alone; an arginine, an aspartate and two glutamates. Interestingly, three of these residues are located at the N-terminus of the PE (Zhu & Shuman, 2006). It is believed that both of PE's reactions occur at the same active site, and it appears that each has some unique residues that are key for each catalytic reaction. N-terminal truncations of Pae PE yielded a protein still capable of phosphodiesterase activity, but not phosphomonoesterase activity (Zhu et al., 2005). Later studies of an archaeal orthologue of PE in *Candidatus koracharchaeum cryptofilum* (Cko), which is missing the N-terminal segment of the Pae PE, also produced similar results. A truncated PE that is capable of removing NMPs, but not a 3'-P, lends credence to the theory that the phosphomonoesterase activity is attributable to residues in the N-terminal sequence of the protein (Smith et al., 2011).

The recent structures of the phosphoesterase apo enzyme have confirmed that earlier predictions of the composition of the active site were correct; two histidines and an aspartate coordinate the manganese ion, whilst an arginine and a third histidine residue contact the phosphate (presumed by positioning of a phosphate anion) and a tyrosine coordinates a water molecule also contacting the phosphate (Figure 1.18) (Nair et al., 2010; Smith et al., 2011). The active site is situated in a β-stranded cleft in a crescent shape, on the side of an eight-stranded β barrel that forms the hydrophobic core. The β-barrel is surrounded by two α-helices, and a 3^{10} helix. The fold composition of the bacterial (Pae) and two archaeal [Cko, and *Methanosarcina barkeri* (Mba)] PE proteins is novel, and confirms that the PE is a distinct class of putative DNA repair enzyme (Figure 1.18) (Nair et al., 2010; Smith et al., 2011). Neither structure of the PEs that possessed the full length N-terminal region (Pae, Mba) had any electron density for the first 30 amino acids, suggesting that the region is highly flexible, and probably does not adopt a rigid structure, at least in the absence of a substrate. Analysis of the structure reveals that the N-terminus protrudes from the other side of the β-barrel core to the active site, and that if the residues previously identified in the phosphomonoesterase activity

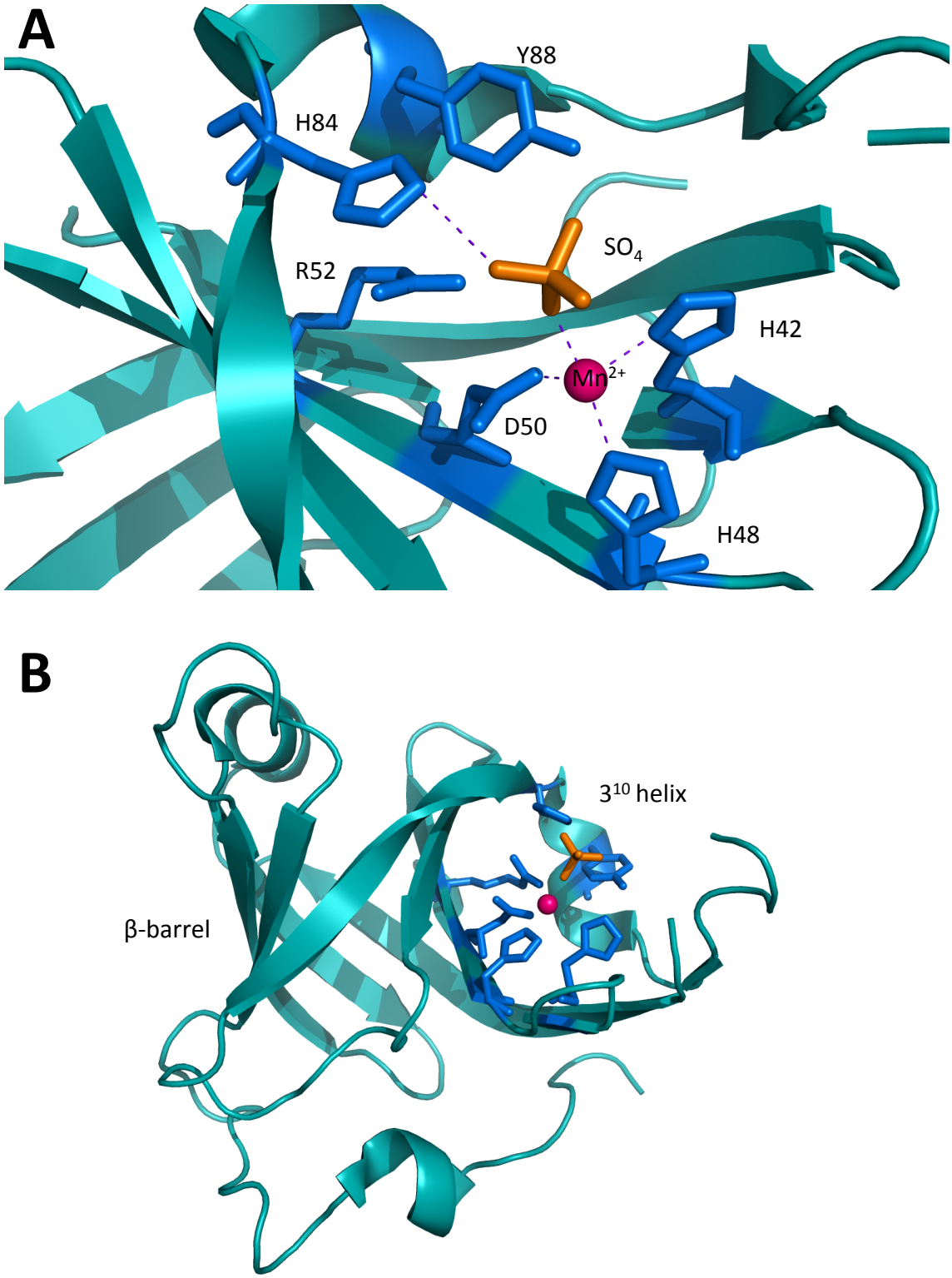


Figure 1.18 Catalytic site arrangement and structural composition of the PE.
(**A**) The catalytic site of the Pae PE shows the three histidines and aspartate coordinating the manganese and the sulphate ions directly. The arginine and tyrosine residue contact water molecules which also help to coordinate the sulfate. (**B**) The structure of PE contains a distinct β-barrel with a hydrophobic core, and an active site cleft on the outside. The 3^{10} helix borders one side of the active site. Figures made using the structure from Nair et al., 2010, PDB 3N9D.

play a role at the active site, then the N-terminus must be flexible enough to wrap around and engage the catalytic centre. Otherwise it is possible that the conserved phosphomonoesterase residues play a role in substrate recognition and positioning, and that without them phosphate cleavage is inefficient.

The structural features that distinguish between the PE's abilities to cleave phosphodiester and phosphomonoester bonds remain a mystery, and vigourous attempts have been made to solve a structure which informs on this matter. Since the N-terminal is too flexible to gather reliable electron density in X-ray crystallography, NMR (nuclear magnetic resonance) solution structures of the PE were obtained (Dutta et al., 2011; Natarajan et al., 2012). Despite a variety of DNA substrates used, and several different metal ions, no significant spectral perturbations were witnessed for the N-terminal segment of the enzyme. Dynamic changes were observed around the active site however, with β-strands shifting to adopt 'open' and 'closed' formations around the catalytic residues, depending on the relevance of the substrate (Natarajan et al., 2012). The studies revealed that the crystal structures had captured the enzyme in a somewhat 'closed' formation, whereas large spectral changes occurred when a DNA primer with 3' RNA was added to the solution with PE during NMR. The results suggest a dynamic active site that opens to accommodate the 'correct' substrate when manganese is present.

1.4.6.1 **The unknown role of PE in NHEJ**

Despite a strong grasp on the substrate specificity of the PE, the role of the enzyme in NHEJ repair remains unclear. Whilst the PE can produce a DNA end with a monoribonucleoside and a 3'-OH that is suitable for ligation by LigD, the logic of also maintaining a NHEJ polymerase which preferentially incorporates NTPs to the same substrate appears counter intuitive. On the face of it, the NHEJ Pol and PE would engage in a futile cycle of ribonucleotide addition and removal, and would require additional regulation to operate in a productive manner. *In vivo* plasmid repair assays concluded that both ligase and polymerase domains of LigD had a significant role in NHEJ, and loss of either severely impacted the DSB repair efficiency (Aniukwu et al., 2008). However, inactivation of the PE did not reduce the incurrence of deletions, and did not significantly alter the constituents of the repair of blunt or 5' or 3' overhang ends. The authors conclude that the main role of the PE, if any in NHEJ, is to provide 3'-P removal which was not challenged in the plasmid repair assay (Aniukwu et al., 2008). It is tempting to conclude that PE has no significant role in NHEJ, since it certainly is not accountable for the excessive 3' deletions observed in some NHEJ plasmid repair assays *in*

vivo. However, the enzyme is highly conserved throughout bacteria, most often as a component of the LigD protein, and is even apparent in several archaeal species along with the other LigD domains (Nair et al., 2010). None of these species are known to possess Ku, an essential partner of a functional NHEJ apparatus, and therefore it is not possible to rule out other metabolic functions for the PE. One of the main objectives of this thesis is elucidating the role of the PE in NHEJ.

1.5 NHEJ in archaea

1.5.1 Identification of NHEJ factors in archaeal genomes

Unlike other DNA repair pathways, NHEJ was thought absent in archaea. Although LigD-like ligase and Ku genes were identified in an operonic arrangement in the archaeon *Archaeoglobus fulgidus*, the AEP and PE were not present and it was presumed not to be a completely functional homologous system of NHEJ (Aravind & Koonin, 2001; Doherty et al., 2001; Weller & Doherty, 2001). At the time *A. fulgidus* was the only archaeal species in which any NHEJ components were identified, although this most likely relates to the relative scarcity of sequenced archaeal genomes. Also it seemed unusual that an archaeal DNA repair pathway would mirror that of the bacterial system, since the other DNA repair pathways observed in archaea are more similar to eukarya than bacteria (Komori et al., 2000; White, 2003). Archaea share a complexity of information systems with eukarya that is not present in bacteria. Genetic maintenance, replication and transcription in archaea are all more akin to eukarya than bacteria (Kelman & White, 2005).

It seemed unlikely that archaea possessed no distinct method of joining DSBs outside of HR though, especially given that even bacteria maintained a core system (Weller et al., 2002). Other components of NHEJ have since been noted in archaeal genomes, notably both NHEJ-AEP and PE-like genes in the species *Methanocella paludicola*. In other archaea from the phylum *euryarchaeota* and also in the fungal phylum *Ascomycota,* the PE domain was observed as a single gene (Nair et al., 2010). However none of these sequence homology searches revealed the presence of Ku, essential in NHEJ, leaving it unclear what functions these collections of enzymes possess.

The Doherty lab was the amongst the first to identify the existence of NHEJ-like genes in bacteria, and pioneered the research in proving that the system was functionally analogous to

that of the established eukaryotic NHEJ pathway (Doherty et al., 2001; Aravind & Koonin, 2001; Weller et al., 2002). Whilst the roles of the LigD ligase and polymerase domains are now clear, that of the phosphoesterase domain is not. Comprehensive BLAST (Basic Local Alignment Search Tool) homology searches performed in the Doherty lab have revealed that the archaeon *Methanocella paludicola* (Mpa) not only possesses LigD-like ligase, polymerase and phosphoesterase as individual genes in an operonic arrangement, but also a Ku-like gene in a distal region of the genome (Sakai et al., 2008). The aim of this thesis was to establish if these genes are functionally analogous to the bacterial NHEJ system (Chapter 3) and proving so, argue for the presence of a functional NHEJ pathway in archaea for the first time. Secondly, the discrete nature of the Mpa NHEJ-like genes offered a tantalising opportunity to interrogate the individual components of the system, then reconstitute them one enzyme at a time, revealing what function the PE has in relation to its neighbours (Chapter 4). Finally, building on functional observations from the prior studies, structural analyses of several components was attempted, aiming to discover the specific interactions with the enzymes and their preferred DNA-RNA substrates.

Chapter 2

Materials and Methods

All chemicals used were purchased from Sigma Aldrich or Fisher Scientific unless stated otherwise.

2.1 Molecular cloning

2.1.1 Polymerase chain reaction and site directed mutagenesis

Polymerase chain reaction (PCR) was performed with Phusion (Thermo Scientific) following the manufacturer's recommended procedures. Standard primers were designed with a Tm of ~60°C. All primers were analysed using Oligo Analyzer (Integrated DNA Technologies). A typical PCR of 50µL contained 10ng plasmid DNA or 200ng genomic DNA, 0.5µM forward and reverse primers, and 1 U Phusion. PCR was optimised by altering annealing temperatures, including DMSO or additional $MgCl_2$. PCR products were purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol. All primers used for PCR as listed in Table 2.1.

2.1.1.1 Site directed mutagenesis

PCR for site directed mutagenesis and primer design was conducted using the QuikChange protocol (Stratagene). PCR products were incubated with 10-20 U Dpn1 (New England Biolabs) for 1 hour at 37 °C to digest methylated DNA in order to remove the unmodified template. 1 µL of the sample was transformed in to competent *E. coli* (section 2.1.5.2) and any colonies that had grown overnight were checked by sequencing (2.1.6).

2.1.2 Electrophoresis of DNA

DNA samples for size separation by electrophoresis were mixed with loading buffer [2.5% (w/v) Ficoll 400, 11 mM EDTA, 3.3 mM Tris pH 8, 0.017% (w/v) SDS, 0.015% (w/v) bromophenol blue] supplied in a 6X stock (New England Biolabs). Samples were typically analysed on a 1% (w/v) agarose TAE (0.4 M Tris-acetate pH 8, 1 mM EDTA) gel containing 0.3µg/mL ethidium bromide. Samples were electrophoresed at 100 V for 20 minutes and resolved alongside 1 kB ladder (New England Biolabs). The gels were documented using a UV illuminator (Syngene InGenius Gel Documentation System) and DNA images analysed using GeneSnap (Syngene).

Name	Sequence 5’-3’	Restriction site
MpaPE_F	CGGAGTTCATATGTCCCCGAAAGAACCG	NdeI
MpaPE_R	CGTTCCAGCAGAGGATCCATCTCATCCCTCTTTAGCG	BamHI
MpaPol_F	CGGGAGGTGCCCATATGACCGAAGTATTGCACATCGAAGGC	NdeI
MpaPol_R	GCATAACCTGAAGGGATCCTTAATTATCAAGGTTTTTTATGCTCG	BamHI
MpaKu_F	CGTATTTAGGGGCTAGCATGCAAGCCATCTGGTCC	NheI
MpaKu_R	GTTTATATAAGAGGATCCTTAAACGAGCATCCTTTCAAGCTCG	BamHI
MpaLig_F	GGTGCCGCGCGGCAGCCATATGGACCCTCTGCTGGAACGCC	NdeI
MpaLig_R	GGCGTTCCAGCAGAGGGTCCATATGGCTGCCGCGCGGCACC	NotI

Table 2.1 PCR primers.
A list of PCR primers used to amplify sequences from genomic DNA for cloning into expression vectors. The forward and reverse primer for each gene is listed, and the restriction site that was designed into each primer.

2.1.3 Restriction digest

Restriction endonucleases were used to digest plasmid DNA and PCR products to produce matching sticky ends for subsequent ligation. Typically 20 µL of purified plasmid (2 µg) or purified PCR product were digested with 20 U of the chosen restriction endonuclease and relevant buffer (New England Biolabs) in a final volume of 40 µL. The samples were incubated at 37 °C for 1-8 hours before a 1 µL fraction was electrophoresed (2.1.2) to observe digestion. DNA was then extracted from the agarose gel and purified using the Gel Extraction Kit (Qiagen) as per the manufacturer's recommendations.

Restriction endonucleases were also used to screen potentially successful clones. Endonucleases that cleaved only once within the new insert were selected from a list that was determined by Webcutter 2.0 or NEBcutter web tools (Vincze et al., 2003). Digests consisted of 3 µL sample DNA and 10 U restriction endonuclease of a final volume of 10 µL with the relevant buffer (New England Biolabs). Half of the sample was analysed by agarose gel electrophoresis (2.1.2) and contrasted with unmodified plasmid DNA to observe a size difference that would indicate a successful clone.

2.1.4 Construct ligation

The final step of generating a construct was ligating the digested PCR insert and plasmid together using T4 DNA ligase (New England Biolabs). The concentration of both insert and vector plasmid were determined by the Nanodrop spectrophotometer (Thermo Scientific). Ligation reactions were calculated at 3:1 insert to vector using the equation I(ng) = 3 x [I(bp)/V(bp)] x V(ng). Reactions contained 200 U (0.5 µL) T4 DNA ligase and T4 DNA ligase buffer (New England Biolabs) in a final volume of 20 µL. Ligation reactions were incubated at room temperature for 1 hour before a 5 µL was transformed into *E. coli* DH5α.

2.1.5 Preparation of plasmid DNA

2.1.5.1 Preparation of competent *E. coli* DH5α

Plasmid DNA was produced using *E. coli* DH5α cells. First a stock of competent DH5α cells was made. 3 mL of Luria Bertani (LB) medium [1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl, pH 7] was inoculated with a colony of DH5α selected from an agar plate, and then placed over night in a shaking incubator at 37 °C. The next morning the culture was diluted in 250 mL super optimal broth (SOB) [2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 10 mM NaCl,

2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$] and was placed in a shaking incubator at 18 °C until the culture grew to an optical density of 0.4 at 600nm (OD_{600}). The cells were cooled for 10 minutes on ice and were pelleted by centrifugation (4000 rpm, 10 minutes, 4 °C). The cells were then resuspended in pre-chilled transformation buffer (100 mM PIPES pH 6.7, 15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2$) and chilled on ice for another 10 minutes. The cells were pelleted by centrifugation again, and resuspended in 20 mL transformation buffer with 7 % (v/v) DMSO (Dimethyl sulfoxide). Following another 10 minutes on ice, the cells were pipetted into 50-200 µL aliquots and snap frozen in liquid nitrogen and stored at -80 °C.

2.1.5.2 **Transformation of DH5α**

A typical transformation reaction consisted of 1 µL MiniPrepped plasmid DNA (~100 ng) or 5 µL ligation reaction, and 50 µL competent DH5α cells which had been thawed on ice (2.1.5.1). The DNA and cells were incubated on ice for 10 minutes, then heatshocked at 42 °C for 45 seconds, before returning to ice for a further 10 minutes. 1 mL of LB medium was added to mixture which was then placed in a shaking incubator at 37 °C for 1 hour. 100 µL of culture was then pipetted onto an agar plate [LB medium with 1.5 % (w/v) agar] that contained the relevant antibiotic to enable selection for successfully transformed bacteria. The plate was then incubated at 37 °C overnight and checked the following morning for colonies. Concentrations of ampicillin and kanamycin in agar plates were 100 µg/mL and 30 µg/mL respectively.

2.1.5.3 **Plasmid DNA amplification and purification**

To amplify plasmid DNA a plasmid was first transformed into competent cells (2.1.5.2) and grown on an agar plate. One colony was selected and 3 mL of LB medium was inoculated and incubated at 37 °C overnight in a shaking incubator. The following day the DNA plasmid was purified from the culture using the QIAprep Spin MiniPrep kit (Qiagen) according to the manufacturer's instructions and stored at -20 °C. Plasmid DNA was usually eluted in 50 µL of distilled water, of a typical yield of 100 ng/µL.

2.1.6 Sequencing

Sequencing of plasmid DNA was performed by GATC biotech and Lark Technologies DNA Sequencing. Standard universal primers were used for sequencing reactions. Sequencing chromatograms were read using Chromas (Technelysium Pty. Ltd).

2.2 Expression of recombinant proteins in *E. coli*

2.2.1 Protein expression in competent B834S

2.2.1.1 Preparing competent B834S

The competent cells that were used for protein expression were the *E. coli* strain B834 (DE3) pLysS (B834S) (Novagen). The cells were prepared in a TSS medium which negates the need for a heat shock step as described with DH5α cells in section 2.1.5.2, which is beneficial for B834S cells given that they are designed to lyse easily for target protein separation (Chung et al., 1989). An overnight culture of B834S was diluted into a 1:100 into LB medium and placed in a shaking incubator at 37 °C until an OD of ~0.4 was observed by measurement at 600 nm wavelength with a Spectrophotometer (Cary 50 UV-Vis). The culture was diluted 1:2 with 2X TSS [transformation and storage solution; LB medium with 20 % (w/v) PEG 3350, 10 % (v/v) DMSO, 100 mM $MgSO_4$, pH 6.5], gently mixed and aliquoted in volumes of 50 – 200 µL, before being snap frozen in liquid nitrogen for storage at -80 °C.

2.2.1.2 Transformation of B834S

A standard transformation of B834S cells consisted of 1 µL of MiniPrepped plasmid DNA (~100 ng) added to 50 µL competent cells that had been thawed on ice (2.2.1.1). The reaction was incubated on ice for 10 minutes before addition of 1 mL of LB medium. The TSS prepared B834S cells did not require heat shock to increase plasmid DNA uptake. The mixture was placed directly in a shaking incubator at 37 °C for a typical duration of 3 hours to ensure successful culture growth. The culture was gently pelleted by centrifugation at 3000 rpm for 2 minutes and then 900 µL of supernatant was removed. The cells were gently resuspended in the remaining 100 µL medium and pipetted into an agar plate that contained the relevant antibiotics (usually kanamycin and chloramphenicol) to enable selection for successfully transformed bacterial colonies. The plate was then incubated overnight at 37 °C and checked the next morning for colonies. Concentrations of chloramphenicol and kanamycin in agar plates were 34 µg/mL and 30 µg/mL respectively.

2.2.2 Protein expression trials

Following the successful cloning (2.1.3, 2.1.4) and sequencing (2.1.6) of a target gene into an inducible expression vector (typically pET28a), the plasmid DNA was transformed in B834S cells for protein expression (2.2.1.2). A single colony was selected and an inoculated in a 3 mL

overnight starter culture containing chloramphenicol and kanamycin. The 3 mL was incubated overnight in a shaking incubator at 37 °C to produce a saturated culture. The next morning 3 x 100 mL of LB medium was mixed in 250 mL flasks with both chloramphenicol and kanamycin, and inoculated with 500 µL of overnight culture. The three flasks were placed in a shaking incubator at 37 °C until the culture reached ~0.6 at OD_{600}. A sample of each culture was taken and protein expression was then induced by addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) to a concentration of 1mM. The three flasks were then incubated at different temperatures for induction, typically 20, 30 and 37 °C. The cultures were induced for three hours and the cells were pelleted by centrifugation and snap frozen in liquid nitrogen and stored at -20 °C for processing the following day. The cells were thawed and mixed with Lysis Buffer [50 mM Tris pH 7.5, 250 mM NaCl, 10 % (v/v) glycerol, 1mM DTT (Dithiothreitol)] and sonicated for 30 seconds, 10 seconds on and off at 20 % amplitude. A sample of each lysate was taken and resolved by SDS-PAGE (Sodium dodecylsulphate polyacrylamide gel electrophoresis) described in 2.2.4.

2.2.3 Protein expression and B834S pellet storage

Once the conditions for protein expression have been optimised by protein expression trials (2.2.2), a large-scale growth was performed to maximize the protein yield. An expression vector (usually pET28a) containing the target gene was transformed into B834S cells (2.2.1.2). A single colony was selected and an inoculated in a 100 mL overnight starter culture containing chloramphenicol and kanamycin. 8 x 1 L flasks were filled with 500 mL TB [Terrific Broth: 1.2% tryptone, 2.4% yeast extract, 72 mM K_2HPO_4, 17 mM KH_2PO_4 and 1% (v/v) glycerol], 1:1000 dilutions of chloramphenicol and kanamycin and 5 mL overnight B834S culture. The cultures were grown at 37 °C until the OD_{600} was approximately 0.6 (typically 3-4 hours). Then the cultures were cooled on ice for 10 minutes before the addition of IPTG to 1 mM and returned to the incubator at the temperature determined during the expression trials, for a set amount of time (usually 3-4 hours). Cells were collected by centrifugation (5000 rpm, 10 minutes at 4 °C), and the supernatant was removed. The cell pellets were re-suspended in 5 mL Lysis Buffer and stored in 50 mL falcon tubes. The tubes were then snap frozen in liquid nitrogen and stored at -80 °C until required for processing.

2.2.4 Electrophoresis of proteins in sodium dodecyl sulphate polyacrylamide gels

Protein samples were resolved using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), and were first prepared with Laemmli sample buffer [2 % (w/v)

SDS, 10 % (v/v) β-mercaptoethanol, 20 % (v/v) glycerol, 0.002 % (w/v) bromophenol blue, 0.125 M Tris pH 6.8] (Laemmli, 1970). The samples were then boiled for 10 minutes. The XCell SureLock Mini-Cell Electrophoresis System (Invitrogen) was used for SDS-PAGE. The gels were prepared in 1 mM cassettes (Invitrogen) and typically consisted of 15 % resolving mix [acrylamide/bisacrylamide 30 % (37.5:1) mix (National diagnostics), 375 mM Tris pH 8.8, 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulphate (APS), 0.04 % (v/v) TEMED (N,N,N',N'-Tetramethylethylenediamine)]. Water was added to the top of the gel to allow it to set, and then poured off before the addition of 5 % stacking mix [5% acrylamide mix, 125 mM Tris pH 6.8, 0.1 % (w/v) SDS, 0.1 % (w/v) APS, 0.1 % (v/v) TEMED] and a comb to create wells. When set, the wells were washed using SDS Running Buffer [25mM Tris, 250 mM glycine, 0.01 % (w/v) SDS] before protein samples were loaded alongside a Precision Plus molecular weight marker (Biorad). Samples were electrophoresed at 100 V in SDS Running Buffer until the bromophenol blue dye reached the bottom of the gel (typically 1 hr).

2.2.4.1 **Coomassie blue staining**

SDS polyacrylamide gels were stained with Coomassie blue solution [50 % (v/v) methanol, 10 % (v/v) acetic acid, 0.5 % (w/v) Coomassie blue] for ~5 minutes on a rocker. The Coomassie blue solution was removed and the gel was gently washed with water to remove any excess solution. Destain [10 – 20 % (v/v) methanol, 10 % (v/v) acetic acid] was added to the gel on a rocker, for ~3-4 hours, until the protein bands could be properly visualised by eye. The gel was then stored in water with 1 % (v/v) glycerol until they were air-dried, using the Gel Air Drying System (Biorad) as per the manufacturer's instructions.

2.3 Protein purification

2.3.1 Lysate preparation

Cell pellets were prepared for lysis by removal from -80 °C and defrosting on ice. The cells were diluted in IMAC Buffer A [50 mM Tris pH 7.5, 500 mM NaCl, 30 mM Imidazole, 10 % (v/v) glycerol, 17 µg/mL PMSF (phenylmethylsulfonyl fluoride), 34 µg/mL Benzamidine], and stirred for ~30 minutes on ice with an addition of 0.1 mg/mL lysozyme. The cells were then disrupted (on ice) by sonication using a Vibra-Cell sonicator, with 6 rounds of 10 second pulses of 30 % amplitude, resting for 10 seconds in between pulses. Sonicated lysates were then separated by centrifugation (18000 rpm, 1 hr, 4 °C) in a Sorvall RC26 Plus with a SS-34 rotor. The soluble

fraction was removed from the pelleted cell debris, filtered using a 0.45 µm Millipore filter, and stored on ice before purification by chromatography.

2.3.2 Immobilised metal affinity chromatography

Immobilisied metal affinity chromatography (IMAC) was used as the first step of purifying each of Mpa PE, Pol, Ku and Lig. The column's nickel resin binds proteins with histidine tags, which can then later be eluted. 25 mL Ni^{2+}-NTA (nickel-nitrolotriacetic) agarose resin IMAC columns (Qiagen) were run using the ÄKTAprime system (GE Healthcare). All buffers and protein samples were kept on ice. Mpa PE, Pol and Lig were purified with the same protocol; Mpa Ku was purified with an extra wash step, detailed below. The columns were pre-equilibrated with IMAC Buffer A, and the cell lysates were loaded onto the columns at a flow rate of 3 mL/min. The column was then washed with IMAC Buffer A until all non-specifically bound proteins were eluted, measured by the absorbance at 280 nm (A_{280}) returning to pre-loading levels. The Mpa Ku loaded column was then washed extensively with IMAC Buffer C (as Buffer A, with 1 M NaCl), to remove any fragments of DNA that might have bound to Ku during lysis. The Mpa Ku IMAC was then equilibrated with Buffer A again, and all purification protocols followed the same procedure once more. IMAC Buffer B (as Buffer A, with 300 mM Imidazole) was added at 100 % to elute the bound proteins. Samples from the load, flow through, wash and elution steps were analysed by SDS-PAGE (2.2.4).

2.3.3 Ion exchange affinity chromatography

Ion exchange affinity chromatography was always used as the second step of purification for the Mpa NHEJ proteins, as it was highly successful at removing exonuclease contaminants. 5 mL fast-flow Q and S Ion exchange affinity chromatography columns (GE Healthcare) were run using the ÄKTAprime system (GE Healthcare). Pol and Lig were purified using both Q and S columns simultaneously, whilst PE and Ku were purified using just the Q column. Ion exchange columns were pre-equilbrated with IEx Buffer A [50 mM Tris pH 7.5, 10% (v/v) glycerol]. Protein samples from the 100 % B elution of the IMAC were diluted 1 in 10 with IEx Buffer A to reduce the concentration of imidazole. The protein samples were then loaded on to the Q and S columns at 5 mL/min and were separated by ionic charge. Mpa PE, Pol and Lig were eluted during the flow through of this step. After a wash step with IEx Buffer A, Mpa Ku was eluted in 10 % IEx Buffer B [50 mM Tris pH 7.5, 2 M NaCl, 10% (v/v) glycerol]. All protein samples were analysed for purity by SDS-PAGE (2.2.4).

2.3.4 Size exclusion chromatography

Since X-ray crystallography was a target for each of the Mpa NHEJ proteins, they all required extensive purification to offer the best possibility of a homogenous solution to grow crystals. The final step of protein purification was size exclusion chromatography. 300 mL Superdex S200 Size exclusion chromatography columns (GE Healthcare) were run using the ÄKTApurifier system (GE Healthcare). The S200 column was pre-equilbrated with SEC Buffer [25 mM Tris pH 7.5, 500 mM NaCl, 10% (v/v) glycerol]. Protein samples were concentrated to ~3 mL and loaded onto the S200 column using a 5 mL loop which was pre-filled with SEC Buffer. The S200 column was run at 2.2 mL/min and samples were collected after 100 mL of flow through. Potentially highly concentrated protein samples were isolated by comparison to the A_{280} at elution. All relevant samples were analysed by SDS-PAGE to estimate purity (2.2.4).

2.3.5 Storage of purified proteins

Proteins purified by size exclusion chromatography were concentrated using a Vivaspin sample concentrator (GE Healthcare) to ~5 mL or less. The protein concentration was determined by use of the Nanodrop spectrophotometer (Thermo Scientific). Once the protein reached a desired concentration, the sample was aliquoted into small volumes (typically 200 µL) and snap frozen in liquid nitrogen for storage at -80 °C.

2.4 Primer extension assay

Primer extension assays were based on the protocols described in (Jozwiakowski & Connolly, 2011).

2.4.1 Primer, template and downstream substrates

The DNA oligomers used for primer extension, phosphoesterase and ligation assays were manufactured by Fisher Scientific and Sigma Aldrich. Any oligomer containing a ribonucleoside was ordered from Integrated DNA Technologies. The DNA oligomer containing the thymine-thymine pyrimidine (6-4) pyrimidone photoproduct (H2) was a gift from Prof. Alan Lehmann (Sussex, UK). All primers contained a 5'-fluorescein label, except those shown in figure 3.17 (H1, H2 and H3), of which H1 and H3 were designed by Dr Stanislaw Jozwiakowski, and have 5'-hexachlorofluorescein labels. The sequences of the primers, templates and downstream (D-strand) oligomers used are detailed in Tables 2.2, 2.3 and 2.4.

Name	Sequence 5'-3'	Component
F16	CTATGAGCGAATCGCC	Primer
FD14R2	CTATGAGCGAATCGrCrC	Primer
FD12R2D19	CTATGAGCGAATCGrCrCCCGACTACACTATGCGACT	Primer
P37	P-CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACA	Primer
F37	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACA	Primer
F36	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTAC	Primer
FD10R6	CTATGAGCGArArUrCrGrCrC	Primer
FD8R8	CTATGAGCGArArUrCrGrCrC	Primer
FD36R1	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACrA	Primer
FD35R1	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTArC	Primer
FD16P	CTATGAGCGAATCGCC-P	Primer
FD15R1	CTATGAGCGAATCGCrC	Primer
F42	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACAACGCG	Primer
F48	CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACAACGCG ATCGCG	Primer
D35	AGTCGCATAGTGTAGTCGGTCCGATTCGCTCATAG	Template
D35b	AGTCGCATAGTGTAGTCGGGGCGATTCGCTCATAG	Template
D37	GTAGTCACTATCGGAATAAGGGGCGACACGGATATGT	Template
D37b	TGTAGTCACTATCGGAATAAGGGGCGACACGGATATG	Template
P36	P-GGTAGTCACTATCGGAATAAGGGGCGACACGGATAT	Template
P39	P-CATGGTAGTCACTATCGGAATAAGGGGCGACACGGATAT	Template
D35c	AGTCGCATAGTGTAGTCACGGCGATTCGCTCATAG	Template
D35g	AGTCGCATAGTGTAGTCGTGGCGATTCGCTCATAG	Template

Table 2.2 5'-fluorescein labelled primers, templates and downstream strands.
A list of the DNA sequences of primers, templates and D-strand used in experiments for this thesis. All primers are 5' labelled with fluorescein, except for primers H20 P, H16 P, H27 P which are 5' labelled with hexachlorofluorescein. The lengths of all sequences are indicated by the numbers in the name of the oligomer, eg: F16 is 16 bases long. Ribonucleosides are indicated by a lower case 'r', eg: rC. A 'P' indicates a phosphate, and is either 5' or 3' as depicted by which end of the sequence it is appended to. P19ddC has a 3' dideoxycytidine base. H30 T has a 6-4 PP as indicated by the T^T in the sequence. H47 CPD T has a CPD as indicated by the T=T in the sequence.

Name	Sequence 5’-3’	Component
D35h	AGTCGCATAGTGTAGTCGGTGCGATTCGCTCATAG	Template
P37b	P-TGTAGTCACTATCGGAATAAGGGGCGACACGGATATG	Template
P41	CATGCATATCCGTGTCGCCCCTTATTCCGATAGTGACTACA	Template
D19	CCGACTACACTATGCGACT	D-strand
P19	P-CCGACTACACTATGCGACT	D-strand
D18	CGACTACACTATGCGACT	D-strand
D22	CTACCGACTACACTATGCGACT	D-strand
P18	P-CGACTACACTATGCGACT	D-strand
P17	P-GACTACACTATGCGACT	D-strand
P16	P-ACTACACTATGCGACT	D-strand
P20	P-CCCGACTACACTATGCGACT	D-strand
P19ddC	P-CCGACTACACTATGCGACTddC	D-strand
H20 P	TGTCGTCTGTTCGGTCGTTC	Primer
H16 P	CACTGACTGTATGATG	Primer
H27 P	TGTCGTCTGTTCGGTCGTTCAATCTTC	Primer
H47 T	CGCGCAGGGCGCACAACAGCCTTGAAGACCGAACGACCGAACAGACGACA	Template
H30 T	CTCGTCAGCATCT^TCATCATACAGTCAGTG	Template
H47 CPD T	CGCGCAGGGCGCACAACAGCCT=TGAAGACCGAACGACCGAACAGACGACA	Template

Table 2.3 5’-fluorescein labelled primers, templates and downstream strands.
A list of the DNA sequences of primers, templates and D-strand used in experiments for this thesis. All primers are 5’ labelled with fluorescein, except for primers H20 P, H16 P, H27 P which are 5’ labelled with hexachlorofluorescein. The lengths of all sequences are indicated by the numbers in the name of the oligomer, eg: F16 is 16 bases long. Ribonucleosides are indicated by a lower case ‘r’, eg: rC. A ‘P’ indicates a phosphate, and is either 5’ or 3’ as depicted by which end of the sequence it is appended to. P19ddC has a 3’ dideoxycytidine base. H30 T has a 6-4 PP as indicated by the T^T in the sequence. H47 CPD T has a CPD as indicated by the T=T in the sequence.

Name	Primer	Template	D-strand
F1	FD14R2	D35	D19
F3	FD14R2	D35b	
F4	FD14R2	D35b	P19
F5	FD14R2	D35b	D18
F6	D12R2D19	D35b	
F9	F16	D35b	
F10	FD14R2	D35b	P19
F11	F16	D35b	P19
F12	FD14R2	D35b	
F13	F16	D35b	D18
F19	P37	D37	
F20	FD37	D37b	
F21	P37	D37b	
F22	F36	P36	
F25	FD14R2	D35b	D22
F27	FD10R6	D35b	
F31	FD8R8	D35b	
F37	F16	D35b	P18
F38	FD14R2	D35b	P18
F40	FD36R1	D37b	
F41	F16	D35b	P17
F42	F16	D35b	P16

Table 2.4 Composition of 5'-fluorescein labelled DNA substrates.
A list of the primers, templates and D-strands that compose the substrates used in this thesis. Primers and D-strands have a complementary sequence to the template strand so that they can anneal.

Name	Primer	Template	D-strand
F43	FD14R2	D35b	P17
F45	FD35R1	P39	
F48	FD10R6	D35b	P19
F56	F16P	D35b	
F57	F16P	D35b	D19
F60	FD15R1	D35b	P19
F61	FD15R1	D35b	P20
F62	FD14R2	D35b	P20
F63	F16P	D35b	P18
F68	FD36R1	P37b	
F69	FD15R1	D35c	P19
F72	FD15R1	D35g	P19
F74	F36	P41	
F76	F42	P37b	
F77	F48	P37b	
F78	F16	D35b	P19ddC
F79	FD15R1	D35	P19
F80	FD15R1	D35h	P19
H1	H20 P	H47 T	
H2	H16 P	H30 T	
H3	H27 P	H47 CPD T	

Table 2.4 continued. Composition of 5'-fluorescein labelled DNA substrates.
A list of the primers, templates and D-strands that compose the substrates used in this thesis. Primers and D-strands have a complementary sequence to the template strand so that they can anneal.

2.4.2 Annealing of primer-template substrates

Primers were annealed to templates and mixed 1:1.5, with the non-labelled oligomers in excess, in a buffer of 50 mM NaCl. The samples were incubated at 95 °C for 3 minutes, before slowly cooling to room temperature. Annealed substrates were diluted to 600 nM and stored at -20 °C.

2.4.3 Primer extension

Reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese (5 mM Magnesium was also used as a comparison metal ion). 30 nM DNA substrate was added and either 250 µM NTP mix or 62.5 µM individual NTPs (dNTPs were used in some cases, at the same concentrations) and finally 300 nM Mpa Pol (Mtu PolDom was present in some assays, at the same concentration). Reactions were incubated at 37 °C for the required time (typically 1 hr) and were then quenched by addition of 20 µL 2x Stop Buffer [95 % (v/v) formamide, 0.09 % (w/v) xylene cyanol, 40 mM EDTA and 300 nM competitor oligonucleotide]. The competitor oligonucleotide is complementary to the template strand, and in such excess prevents any reannealing of the labelled primer to the template strand under cool gel conditions. This prevents confusion of the analysis of the polyacrylamide gel electrophoresis, which manifests as smearing of the DNA bands. Quenched reactions were boiled for 10 minutes before brief centrifugation (30 s, 5000 rpm) and left to cool before being loaded on a polyacrylamide gel for separation. Only 20 µL of the sample was used per gel, the remaining sample was saved and stored at -20 °C.

2.4.4 DNA polyacrylamide gel electrophoresis

The results of the primer extension reactions were resolved by electrophoresis with a denaturing urea-polyacrylamide gel. Gels consisted of 15 % polyacrylamide bis-acrylamide (19:1) solution (National Diagnostics), 7 M urea, TBE (89 mM Tris, 2 mM EDTA, 0.89 M Boric Acid, pH 8.3), 0.1 % (w/v) APS, 0.03 % (v/v) TEMED. Gels were cast between 16.5 x 28 cm glass plates with 0.75 mm spacers and a 20-well comb, and resolved on a Dual Vertical Slab Gel Kit (CBS Scientific). Gels were pre-run at a constant wattage of 15 watts for 30-60 minutes in TBE, before the samples were loaded. Gels were run for 2 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.4.5 Gap filling and strand displacement

Gap filling and strand displacement assays were performed using the same protocol as the primer extension assay (2.4.3). The DNA substrates used defined whether the reaction was gap filling or strand displacing.

2.4.6 Template independent extension

Template independent extensions were performed using the same protocol as the primer extension assay (2.4.3). Blunt dsDNA substrates were used to make the assays template independent.

2.5 Ligation assay

Ligation reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese (or 5 mM Magnesium for metal ion comparison). 30 nM DNA substrate was added and initially 62.5 µM ATP (until it was established that Mpa Lig was pre-adenylated and did not require additional ATP to catalyse ligation reactions) and finally 300 nM Mpa Lig (Afu Lig was present in some assays, at the same concentration). Reactions were incubated at 37 °C for the required time (typically 1 hr) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded on a 15 % urea-polyacrylamide gel as described in 2.4.4. Gels were run for 2 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.5.1 Mismatch ligation assays

Mismatch ligations were performed as described for a standard ligation reaction (2.5), except with 400 nM Mpa Ku and DNA substrates with nicks mismatched on the 3’ or 5’ sides.

2.6 Phosphoesterase assay

2.6.1 Para-nitrophenyl phosphate assay

Para-nitrophenyl phosphate (pNPP) was used as a substrate to test the phosphodiesterase and phosphomonoesterase activities of the PE. Reactions contained 20 mM pNPP, 50 mM Tris pH 7.5 and 10 mM manganese or magnesium, and were made to a final volume of 150 µL. A titration of Mpa PE was added up to 10 µg and reactions were incubated at 37 °C for 1-3 hours.

Colourimetric changes were observed by spectrophotometer (Varian Cary 50 UV-Vis spectrophotometer) at 405 nm wavelength. Positive control assays used 0.0001 units of PDE 1 (phosphodiesterase 1). Other similar substrates were also used including; bis-para-nitrophenyl phosphate (bis-pNPP), para-nitrophenyl phosphonate (pnPPP), and thymidine 5'-monophosphate para-nitrophenyl (T 5' mP pNP). All were successfully converted into para-nitrophenyl by PDE1, but not by Mpa PE.

2.6.2 Cytidine 2':3' phosphate assay

Cytidine 2':3' phosphate (C2'3'P) reactions contained 50 mM C2'3'P, 50 mM Tris pH 7.5 and 5 mM manganese or magnesium. Reactions were incubated at 37 °C for 1 hr. No colourimetric change was observed by spectrophotometer (Varian Cary 50 UV-Vis spectrophotometer) at 280 nm wavelength, using the methods described by (Crook et al., 1960). Positive control samples used Ribonuclease.

2.6.3 Phosphodiesterase and phosphomonoesterase assay

Phosphoesterase reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese (or 5 mM Calcium, Cobalt, Copper, Magnesium, Nickel, Zinc or EDTA for metal ion comparison). 30 nM DNA substrate was added and lastly 300 nM Mpa PE. Reactions were incubated at 37 °C for the required time (typically 1 hr) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded to a 15 % urea-polyacrylamide gel as described in 2.4.4. Gels were run for 2 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.6.4 Phosphomonoesterase assay

Phosphomonoesterase reactions were performed essentially the same as full 'phosphoesterase' reactions except that the DNA substrates used only had 3'-P that could be cleaved, and no ribonucleosides. These assays allowed for the independent measurement of 3'-phosphatase activity.

2.6.5 DNA gap filling, phosphoesterase and ligations assays

DNA gap filling, phosphoesterase and ligations reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese. 30

nM DNA substrate was added with 250 µM NTP mix. Some samples were pre-incubated with 300 nM Mpa Pol for 15, 30 and 60 minutes. Following this 300 nM Mpa Lig, 300 nM Mpa PE and 400 nM Mpa Ku were added. Reactions that did not have a Pol pre-incubation also had 300 nM Mpa Pol added. Reactions were incubated at 37 °C for the required time (typically 1 hr 30 minutes) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded on a 15 % urea-polyacrylamide gel as described above (2.4.4). Gels were run for 2 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.7 End joining assay

2.7.1 Ligase and Ku dependent end joining

Mpa Lig and Ku dependent end-joining reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese. 30 nM DNA substrate was added along with 300 nM Mpa Lig and 200, 400, 800 nm Mpa Ku (Afu Lig and Afu Ku were present in some assays, at the same concentrations). No NTPs were necessary because of the ribonucleosides present in the DNA substrates used, and pre-adenylated Mpa Lig did not require additional ATP. Reactions were incubated at 37 °C for the required time (typically 18 hrs) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded to a 15 % urea-polyacrylamide gel as described above (2.4.4). Gels were run for 2-3 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.7.2 Polymerase catalysed extension of DNA in end joining

Polymerase dependent end-joining reactions were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5 and 5 mM Manganese. 30 nM DNA substrate was added and either 250 µM NTP mix or 62.5 µM of ATP, CTP, GTP or UTP. Finally 300 nM Mpa Lig, 300 nM Mpa Pol and 400 nM Mpa Ku were added. Reactions were incubated at 37 °C for the required time (typically 18 hrs) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded on a 15 % urea-polyacrylamide gel as described above (2.4.4). Gels were run for 2-3 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.7.3 Ku, Lig, Pol and PE end-joining reaction

End-joining reactions with the complete ensemble of Mpa NHEJ proteins were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese. 30 nM DNA substrate was added with 250 µM NTP mix. Some samples were pre-incubated with 300 nM Mpa Pol for 30 minutes. Following this 300 nM Mpa Lig, 300 nM Mpa PE and 400 nM Mpa Ku were added. Reactions that did not have a Pol pre-incubation also had 300 nM Mpa Pol added at this time. Reactions were incubated at 37 °C for the required time (typically 18 hr) and were then quenched by addition of 20 µL 2x Stop Buffer. The samples were boiled and loaded to a 15 % urea-polyacrylamide gel as described in 2.4.4. Gels were run for 2-3 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.8 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSAs) were assembled at room temperature in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5, 5 mM Manganese and 5 % (v/v) glycerol. 60 nM DNA substrate was added, and between 200 – 800 nM of the protein being shifted was typically added (Mpa Lig, Pol, PE or Ku). Reactions were incubated at room temperature for 30 minutes, and were not quenched before loading onto a native 5 % polyacrylamide gel. No loading dye was added to EMSA samples.

2.8.1 Native DNA polyacrylamide gel electrophoresis

The EMSA reactions were resolved by electrophoresis with a native polyacrylamide gel. Gels consisted of a 5 % mix [40 % (v/v) polyacrylamide bis-acrylamide (37.5:1) ProtoGel solution (National Diagnostics), 0.5x TBE, 0.1 % (w/v) APS, 0.03 % (v/v) TEMED]. Gels were cast in 1 mM cassettes (Invitrogen) plates with a 16-well comb, and resolved on an XCell SureLock Mini-Cell Electrophoresis System (Invitrogen). Gels were pre-run at a constant voltage of 75 V for 30 minutes in TBE, before the samples were loaded. Gels were run for 1 hour at 75 V. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.9 Hexa-histidine tag cleavage

The hexa-histidine tag on Mpa Pol was removed to see if it affected the ability of the protein to bind a 5’-phosphate. The histidine tag linker in pET28a includes a Thrombin site. Thrombin was

used to cleave the tag from 1 mg of the purified sample of Mpa Pol. 1.5 U of Thrombin (1 µL) was mixed with 40 µL of concentrated Mpa Pol and incubated at 25 °C over night. The cleaved sample was compared to non-cleaved Mpa Pol using SDS-PAGE.

2.10 Primase assay

Primase assays were performed according to the protocol designed by Dr. Stanislaw Jozwiakowski, and were assembled on ice in a total volume of 20 µL. Buffer conditions were 50 mM Tris pH 7.5 and 5 mM manganese. 1 µM ssDNA, conjugated to biotin at the 3' end, was added along with 500 µM dNTPs or NTPs, and finally 1 µM Mpa Pol. The reaction was incubated at 37 °C for 2 hours. 5 µM fluorescein linked dATP and 1 µM Klenow Taq (*Thermus aquaticus*) (kindly donated by Dr. Stanislaw Jozwiakowski) were added before a further incubation at 37 °C for 1 hour. A 50 µL sample of magnetic beads, Dynabeads M-280 Streptavidin (Invitrogen), was separated from the storage buffer using a magnetic rack. 500 µL Binding Buffer (10 mM Tris pH 7.5, 250 mM Nacl) was added to the beads, and then 20 µL of the primase reaction. The beads were resuspended and mixed on a rotation wheel for 1 hour, at 4 °C. This process allowed the Streptavidin magnetic beads to bind the biotin linked DNA template and any primer that was annealed, so that the sample could be separated from the free fluorescein labelled dATP. The beads were then gently washed three times using Wash Buffer (10 mM Tris pH 7.5, 250 mM NaCl, 25 mM EDTA) and the magnetic rack. The beads were then pelleted by centrifugation (3000 rpm, 30 s, room temperature). The beads were resuspended in 20 µL water, and mixed with an equal volume of Stop Buffer (without inhibitor). The samples were boiled and loaded on a 15 % urea-polyacrylamide gel as described above (2.4.4). Gels were run for 2 hours at 17 watts. Fluorescently labeled oligos were detected by scanning using a Fujifilm FLA-5100 fluorescent image analyser.

2.11 Protein crystallization

2.11.1 Buffer exchange of protein

Purified proteins were snap frozen in a glycerol rich buffer [50 mM Tris pH 7.5, 250 mM NaCl, 10% (v/v) glycerol] and stored at -80°C. The glycerol buffer had to be removed in order to try and crystallise the proteins, so the frozen protein aliquots were thawed on ice. Micro Bio-Spin Chromatography columns (Biorad) were loaded with the relevant buffer for crystallisation screening, prior to the addition of the protein samples, as per the manufacturer's instructions.

Buffer exchanged samples were analysed by Nanodrop to re-assess protein concentration for accuracy in subsequent screening.

2.11.2 Pre-crystallisation trials

Pre-crystallisation trials were always performed for each protein prior to screening, in order to estimate a suitable protein concentration to begin screening with. Trials were conducted using the Pre-Crystallization Test (Hampton Research) according the manufacturer's instructions.

2.11.3 Crystallographic screening

Crystallographic screening was conducted using a series of pre-made crystallisation screens; Crystal Screen I, Crystal Screen II (Hampton Research), PEG/Ion suite, PEGS suite, MPD suite, Ammonium Sulphate suite, Nucleix suite (Qiagen), Structure Screen I, Structure Screen 2, Mini screen, PACT screen (Molecular Dimensions). Crystallisation screens were typically set using a 96-well tray with sitting drop format, with 0.7 µL protein solution mixed 1:1 with 0.7 µL reservoir condition (from a reservoir of 70 µL). Screens were sealed and incubated at 12 °C.

2.12 Determination of protein atomic structure by X-ray diffraction

2.12.1 Protein crystal cryogenic protection

Prior to mounting crystals on the goniometer of the Rigaku MicroMax 007-HF the samples were mixed with a cryo-protectant material. Usually this was performed immediately before mounting the crystals. The protein crystal was looped out of the drop using 100-200 µM aperture Micro Mounts (MiTeGen) and wafted into a drop of reservoir buffer and 17 % (v/v) ethylene glycol for ~60 seconds. In cases of fragile crystals during data collection, periods of cryo-protection were extended to 5 or 10 minutes. Cryo-protected crystals were then mounted on to the goniometer for data collection.

2.12.2 Crystallographic analysis

Single wavelength diffraction data were collected for Mpa PolI and Mpa PE using a Rigaku MicroMax 007-HF (University of Sussex), and for Mpa PE vanadate at station I03 Diamond Light Source, Didcot, UK. The diffraction data were processed with iMOSFLM (Battye et al., 2011) and programs from the CCP4 suite (Collaborative Computational Project, Number 4, 1994; Potterton et al., 2002; Potterton et al., 2004).

2.12.2.1 **Mpa Pol crystallographic analysis**

The structure of Mpa Pol was determined by molecular replacement, with apo Mtu PolDom (PDB 2IRU) as the search model, using the program PHASER (Pitcher, Brissett, Picher, et al., 2007b; McCoy et al., 2005; McCoy et al., 2007). Initial refinement was carried out against 95% of the data with REFMAC5 (Vagin et al., 2004; Murshudov et al., 2011; Murshudov et al., 1997). The remaining 5%, which were randomly excluded from the full data set, was used for cross validation by calculating the Rfree to follow the progress of the refinement. The same subset of reflections was used throughout the refinement. Each cycle of refinement was accompanied by manual rebuilding using the program COOT (Emsley & Cowtan, 2004). All structural images were prepared using PyMol (Schrodinger, 2010).

2.12.2.2 **Mpa PE and Mpa PE vanadate crystallographic analysis**

The structures of Mpa PE and Mpa PE vanadate were determined by Dr. Nigel Brissett, using molecular replacement with Pae PE (PDB 3N9D) as the search model for both, with the program PHASER (McCoy et al., 2007; Nair et al., 2010). Initial refinement was carried out against 95% of the data from each set with REFMAC5 (Vagin et al., 2004; Murshudov et al., 2011; Murshudov et al., 1997). The remaining 5%, which were randomly excluded from both full data sets, was used for cross validation by calculating the Rfree to follow the progress of the refinement. The same subset of reflections was used throughout the refinement. Each cycle of refinement was accompanied by manual rebuilding using the program COOT (Emsley & Cowtan, 2004). All structural images were prepared using PyMol (Schrodinger, 2010).

2.13 Bioinformatics

Gene sequences were retrieved using the National Centre for Biotechnology Information (NCBI) from GenBank (Sayers et al., 2010; Benson et al., 2010). Multiple sequence alignments were created using ClustalW2 and TCoffee via the program JALview (Larkin et al., 2007; Waterhouse et al., 2009). Phylogenetic trees were generated using the interactive tree of life (iTOL) (Letunic & Bork, 2011). The program ProtParam, from the ExPASy server, was used for gathering physical and chemical properties of proteins estimated from their primary sequence (Gasteiger et al., 2003; Wilkins et al., 1999).

Chapter 3

Identification and initial characterisation of a complete set of bacterial Ligase D–like NHEJ proteins in the archaeon *Methanocella paludicola*

3.1 Introduction

NHEJ has been established in bacteria for a decade, but questions about of the existence of an equivalent system in archaea have remained unanswered (Weller et al., 2002). The bacterial mechanism centres around the multifunctional Ligase D and Ku proteins, operating in a streamlined, analogous manner to the eukaryotic system (Della et al., 2004). When the bacterial NHEJ genes were initially identified, orthologues were detected in the genome of *Archaeoglobus fulgidus* (Aravind & Koonin, 2001). Notably, the Afu NHEJ-like genes consisted only of Ku and the ligase domain of LigD, whilst no NHEJ polymerase or phosphoesterase (PE) were present in an operonic arrangement, or elsewhere in the genome. Ligase and Ku alone would only be sufficient to repair precise breaks, and they did not seem to constitute a complete NHEJ repair complex. Therefore the conclusion on the presence of an NHEJ pathway in archaea remained open; NHEJ-like genes had been identified yet it seemed unlikely that they constituted a functional DSB repair pathway in the way that they were currently understood.

The Doherty lab has been a leader in the research of NHEJ in prokaryotes, and has sought to fully understand the role of each protein and domain in repairing DSBs. The mycobacterial NHEJ system has been the most thoroughly studied, in the Doherty lab and others, and yet several aspects were difficult to elucidate. The PE domain alone did not exhibit activity and the full length LigD protein was unstable and difficult to purify in large amounts. Both were also targets for crystallographic studies but seemed unlikely to be tractable. Searches for orthologous genes in similar species that might be more amenable to biochemical characterisation using BLAST produced an interesting result; an archaeon that appeared to hold three individual genes in an operonic arrangement that shared a high sequence similarity with that of Mtu LigD's three domains. This species, *Methanocella paludicola* (Mpa), was chosen as the subject for interrogation of the individual proteins and subsequent crystallographic study.

The aspiration for the research described in this chapter was to clone each of the Mpa NHEJ-like genes, purify their protein products and biochemically assay them. Hopefully, the study of each protein would demonstrate similar capabilities to the bacterial NHEJ proteins, and implicate a complete NHEJ mechanism in an archaeon for the first time. Ideally, the *M.*

paludicola NHEJ proteins would prove a fruitful resource, and help us to delineate some of the aspects of the NHEJ mechanism that have previously been recalcitrant.

3.2 *Methanocella paludicola* has operonic NHEJ genes

BLAST searches for the individual domains of the *M. tuberculosis* LigD protein revealed that the archaeon *Methanocella paludicola* has three sequential genes MCP_2125, MCP_2126 and MCP_2127 that share significant sequence similarity with the Mtu LigD NHEJ polymerase, ligase and phosphoesterase domains respectively (Figure 3.1). Although this finding was later published by another research group, they had not identified the Ku gene that was also present in the genome (Nair et al., 2010). Unlike almost all of the discovered LigD and Ku genes, *M. paludicola* Ku (MCP_0581) was not found to be operonic, and is quite distant from LigD in the genome.

Methanocella paludicola is a recently described archaeon of the Euryarchaeotia phylum (Figure 1.6) that thrives in rice paddy clusters in Japan (Sakai et al., 2008). It is methanogenic, and grows well in moderate temperatures between 20 and 45 °C, and thrives at 35 to 37 °C. The pH range for growth of the microbe was observed between 6.5 and 7.8, and was optimal at pH 7.0. The temperature and pH range that *M. paludicola* prefers seemed to be ideal experimental conditions to assay proteins in under laboratory conditions. Since *M. paludicola* also appeared to possess a complete repertoire of NHEJ proteins, and that the LigD polymerase, ligase and phosphoesterase were discrete genes, the system became a target for characterisation of the individual components both biochemically and structurally.

3.3 Cloning Mpa NHEJ-like genes into expression vectors

To begin experiments on the NHEJ proteins of *M. paludicola*, the genomic DNA was required to clone the genes into expressions vectors, and it was kindly supplied by Dr Imachi (Sakai et al., 2011). Dr Imachi's group had previously made the genomic sequence of *M. paludicola* available; therefore it could be used to design PCR primers to amplify the target genes. The archaeal genes did not have intervening intronic sequences, therefore the cloning desing was relatively simple, as with cloning bacterial genes. The sequence of each gene was checked with Webcutter 2.0, a web tool that identifies which restriction endonucleases could cleave the query sequence. Two enzymes which did not cut the sequence, but could both cut the pET28a vector in the multiple cloning site were chosen for each gene (Figure 3.2). PCR Primers were

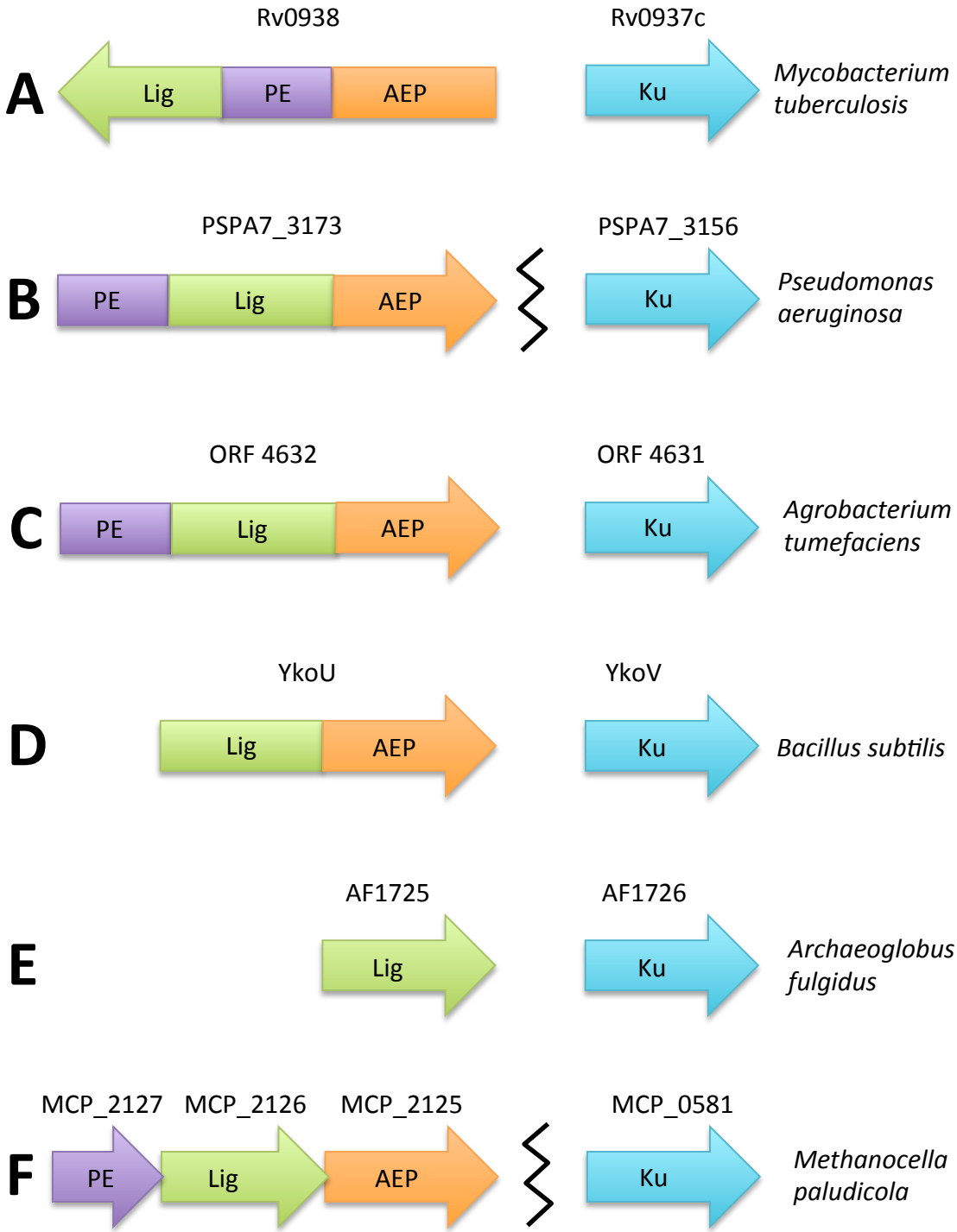


Figure 3.1 Identification of a LigD-like operon in *Methanocella paludicola*.
The operonic arrangement of the previously established bacterial NHEJ operons (**A**, **B**, **C** & **D**) and the putative NHEJ operon from the archaeon *A. fulgidus* (**E**) compared to Mpa NHEJ-like genes (**F**). The black jagged line in the *P. aeruginosa* and *M. paludicola* (**B** & **F**) cartoons represent an omitted genetic sequence, indicating that the LigD and Ku are not operonic in these cases. *M. paludicola* does posses a Ku-like gene (MCP_0581) within its genome, however it is not adjacent to the LigD-like genes.

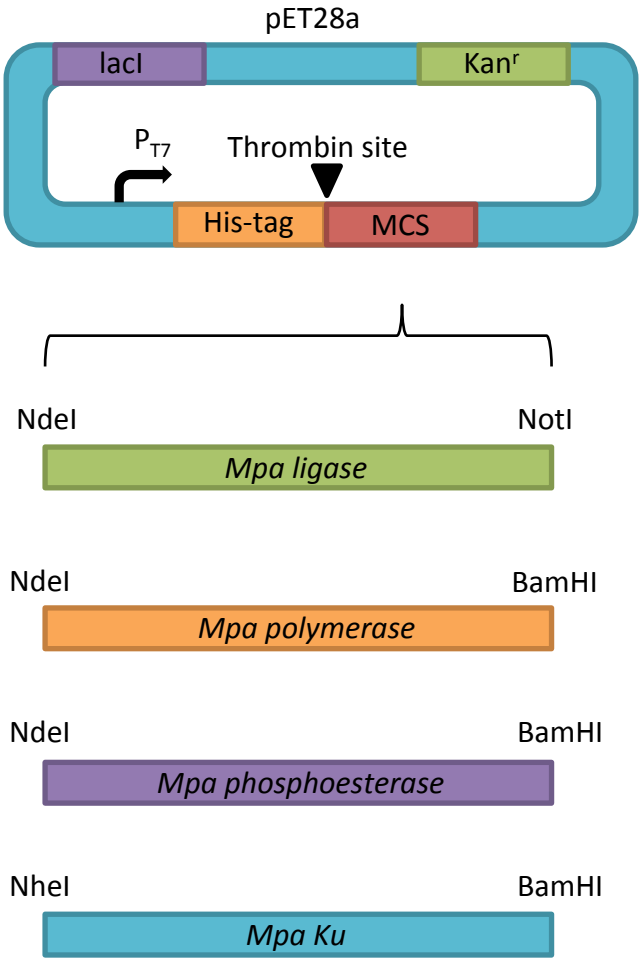


Figure 3.2 Cloning *Mpa Lig, Pol, PE* and *Ku* genes into pET28a expression vector. pET28a overexpression vector with hexahistidine tag in orange, multiple cloning site (MCS) in red, and kanamycin resistance gene in green. The restriction endonucleases used to generate sticky ends on the PCR products and pET28a vector are listed next to the relevant gene. Each gene was then ligated into the pET28a vector, amplified and sequenced to check for accuracy. Primers used for PCR of NHEJ genes from genomic DNA are listed in Table 2.1

designed, using the genomic sequence, which included the restriction endonuclease recognition sequences, and each gene was successfully PCR amplified out of the genomic sequence. The genes and pET28a plasmids were separately digested by the endonucleases, and the purified products were ligated together. Samples from each ligation were transformed in competent DH5α *E. coli* cells, and colonies that grew had their plasmid DNA purified and the plasmids sent for sequencing. Each of the Mpa Lig, Pol, PE and Ku genes were successfully cloned into pET28a.

3.4 Sequence alignment, expression and purification of Mpa PE

3.4.1 Mpa PE protein sequence alignment

The protein sequence of Mpa PE was aligned with Mtu, Msm and Pae LigD, the most comprehensively studied of the bacterial Ligase D proteins (Figure 3.3). The predicted key catalytic residues were strictly conserved in Mpa (Figure 3.3, red and green stars), and the residues that were argued to be involved in the phosphatase activity, but not ribonuclease were also present in strictly conserved regions (Figure 3.3, purple stars) (Zhu & Shuman, 2006). The protein sequence alignments revealed a strikingly high sequence homology, with many absolutely conserved amino acid patches.

3.4.2 Mpa PE expression

Following the successful sequencing of Mpa PE inside the pET28a vector, we initiated expression trials. The newly sequenced plasmid was transformed in B834S *E. coli* cells, suitable for large growths and are easily lysed to yield over expressed proteins. 3 x 100 mL of LB was inoculated with an overnight culture, and were grown at 37 °C until an OD of ~0.6 at 600 nm was reached. A final concentration of 1 mM IPTG was then added to each of the cultures, and each flask was then incubated at a different temperature for protein induction; 20, 30 and 37 °C for three hours. The cells were collected and lysed, and incubated with nickel resin to sequester any his-tagged protein. The pET28a vectors encode a hexa-histidine tag prior to the inserted protein sequence, allowing for such separation by metal ion binding to the aromatic hydrocarbon rings. The resin was then washed and boiled, freeing the bound protein (Figure 3.4 A). Soluble protein was observed in each of the samples, and 4 L growths proceeded, which were induced with 1 mM IPTG at 30 °C. The cultures were collected by centrifugation after 3 hours of induction and snap frozen in liquid nitrogen. The frozen pellets were stored at -80 °C until they were required for processing.

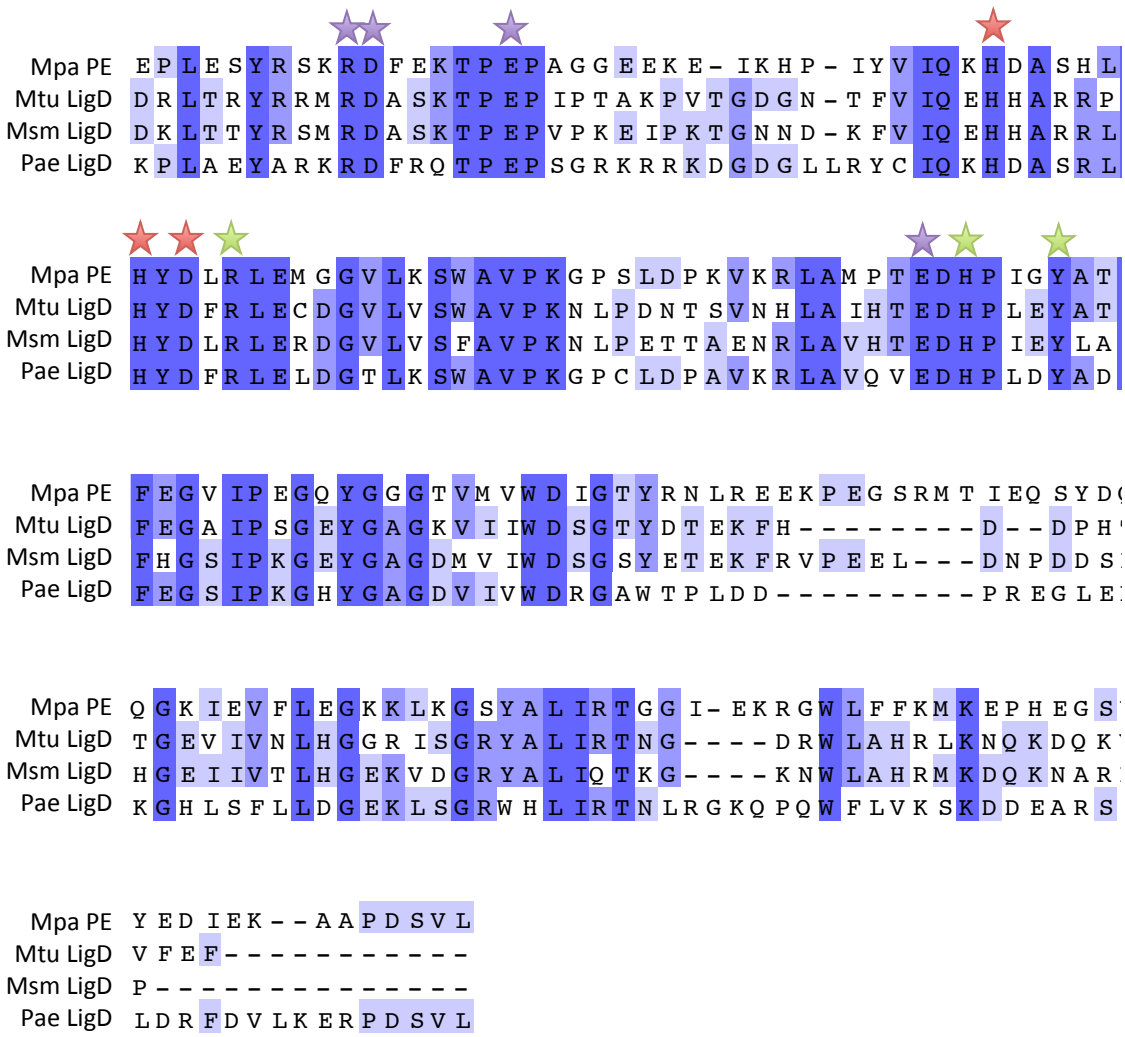


Figure 3.3 Protein sequence alignment of Mpa PE with Mtu, Msm and Pae LigD phosphoesterase domains.

The protein sequence alignments indicate identical residues highlighted in dark blue, and conserved residues in lighter shades of blue. The red stars indicate the presence of the two histidines and aspartate residues that coordinate the divalent metal ion in the active site. The green stars indicate the position of the arginine, histidine and tyrosine residues which contact the scissile phosphate in the active site. The purple stars indicate the positions of the residues that are reported to affect the phosphomonoesterase, but not phosphodiesterase activity.

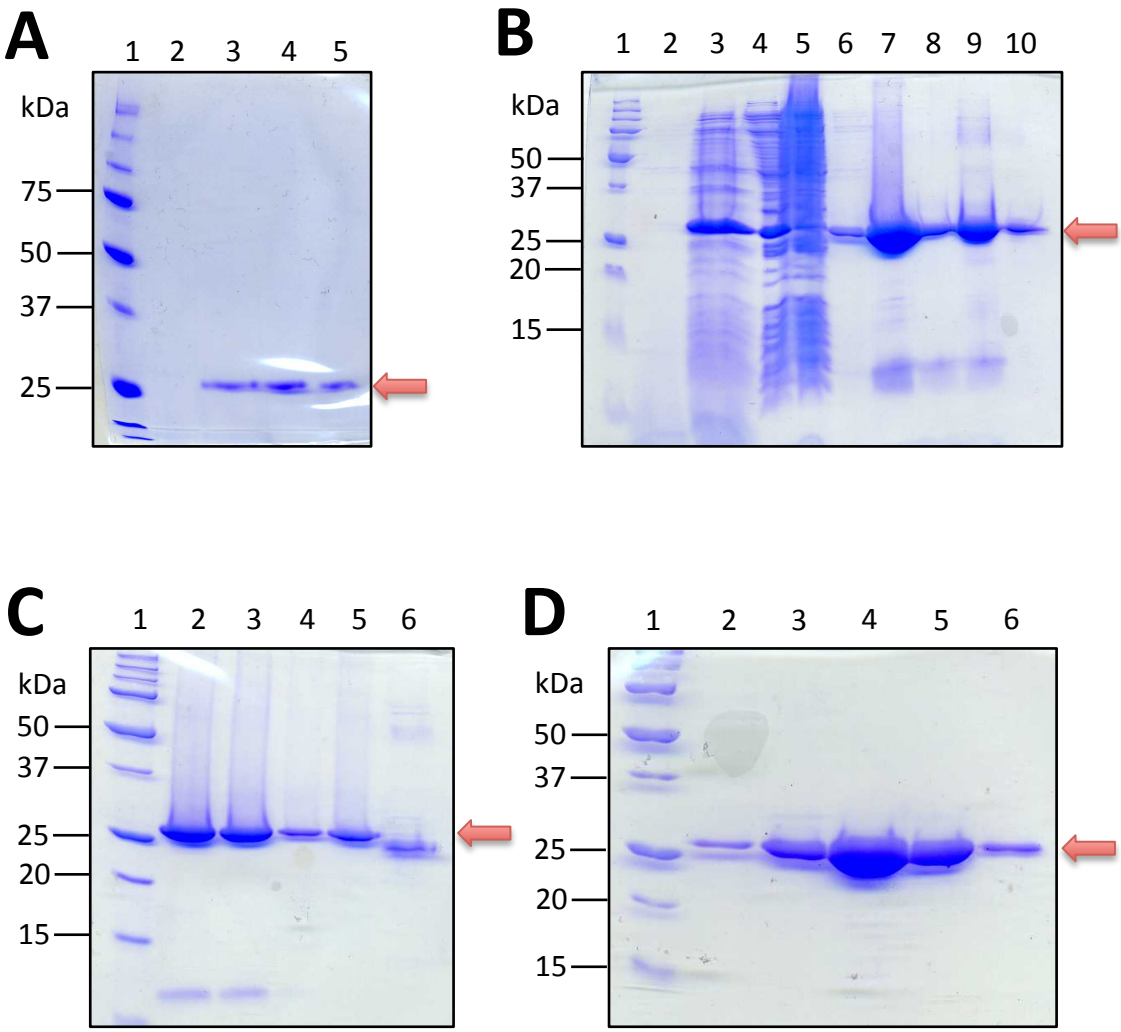


Figure 3.4 Expression and purification of Mpa PE

Coomassie stained protein gels. Presence of Mpa PE protein indicated by the red arrow at ~24.5 kDa. (**A**) Expression of Mpa PE in B834S cells, induced by 1 mM IPTG. 1= Molecular weight marker, 2= Uninduced lysed cells, 3,4,5= Induced cells grown at 20, 30 and 37 °C, lysed and mixed with nickel beads. (**B**) Purification of Mpa PE by IMAC by nickel resin column. 1= Molecular weight marker, 2= Uninduced lysed cells, 3= B834S cells induced with 1 mM IPTG, 4= IMAC load, 5= Flow through, 6= Wash, 7= 90 mM imidazole, 8= 90 mM imidazole fraction II, 9= 300 mM imidazole, 10= 300 mM imidazole fraction II. (**C**) Purification of Mpa PE by mono-Q column. 1= Molecular weight marker, 2= Mono-Q load, 3= Flow through, 4= Wash, 5= 100 mM NaCl, 6= 360 mM NaCl. (**D**) Purification of Mpa PE by size exclusion chromatography. 1= Molecular weight marker, 2= Fraction D11, 3= Fraction D7, 4= Fraction D5, 5= Fraction D3, 6= Fraction D2.

3.4.3 Purification of Mpa PE

The snap frozen pellets were defrosted, mixed with ~50 mL lysis buffer and lysozyme. The lysed cells were further disrupted by sonication, and the soluble cell lysate was separated from the insoluble debris by centrifugation. The cell lysate was loaded onto an IMAC column that was pre-equilibrated with IMAC Buffer A. Following the load and wash steps, IMAC Buffer B was added to the column at concentrations of 90 to 300 mM imidazole to elute the proteins bound to the nickel resin by their histidine tags (Figure 3.4 B). Samples from the load, wash and elutions were resolved by coomassie blue staining of a polyacrylamide gel, and a protein band of the estimated PE molecular weight of 24.5 kDa was observed in both the 90 mM and 300 mM imidazole elutions. Typically, 300 mM imidazole elutions were collected for further purification.

Anion exchange was selected as the next purification step, based on the theoretical isoelectric point of Mpa PE (5.93), estimated by the Expasy web tool, Prot Param (Wilkins et al., 1999). The protein eluted from the IMAC column was diluted 1 in 10 in IEx Buffer A, and loaded onto the pre-equilibrated 5 mL mono-Q column. Following the wash step, a gradient of IEx Buffer B was added to column, from 2 mM to 1 M NaCl to elute any bound protein (Figure 3.4 C). The protein, however, was observed in the flow through, and much of the contaminants were found to bind to the mono-Q column.

To remove the last remaining protein contaminants, a size exclusion chromatography step was selected. Mpa PE that was not retained on the mono-Q column was concentrated and loaded onto the pre-equilibrated Superdex 200 column. Mpa PE eluted in a sharp peak and was highly purified (Figure 3.4 D). The protein was aliquoted and snap frozen in liquid nitrogen for storage at -80 °C. A typical yield of recombinant Mpa PE was 70 mg from a 4 L culture.

3.5 Mpa PE has phosphodiesterase and monophosphoesterase activites

3.5.1 Mpa PE has strict substrate requirements and cannot cleave para-nitrophenyl phosphate

The PE domain of LigD had first been described as a DNA exonuclease, but has since been reclassified as a phosphodiesterase and phosphomonoesterase (Della et al., 2004; Pitcher et al., 2005; Zhu & Shuman, 2005b). In order to test the extent of Mpa PE's substrate specificity

the enzyme was assayed with para-nitrophenyl phosphate (pNPP), a compound that can be hydrolysed to pNP, which is a bright yellow colour, and production of which can be observed in a spectrophotometer at 405 nm wavelength (Keppetipola & Shuman, 2007). pNPP is potentially a substrate for both phosphodiesterase and phosphomonoesterase activities, since although the compound has only a phosphomonoester bond, most phosphodiesterases will still liberate the phosphate from the phenol ring (Figure 3.5 A).

PDE1 (phosphodisterase 1) was used as a positive control, and addition of 0.001 units to 20 mM pNPP caused a near instant colourimetric change, visible by the naked eye. Increasing titrations of Mpa PE up to 10 µg and lengthy incubations at 37 °C did not produce any product formation measurable by spectrophotometer (data not shown). A variety of other, similar substrates were tested with Mpa PE to see how rigid the substrate requirements were. Bis-pNPP and para-nitrophenyl phosphonate (pnPPP) were used to assess whether a strict phosphodiester bond was required to elicit enzymatic activity (Figure 3.5, B and C, respectively). Again, the PDE1 positive control, but not Mpa PE, cleaved both susbtrates. A substrate that mimicked DNA was tested next, thymidine 5'-monophosphate para-nitrophenyl (T 5' mP pNP) (Figure 3.5 D). It was thought that the complete nucleic acid with a phosphodiester bond might be a viable substrate for Mpa PE, but again no activity was discerned. A final spectrophotometric based assay was attempted, this time using cytidine 2':3' phosphate (C2'3'P) which has a cyclic phosphate group (Figure 3.5 E). It has been shown that C2'3'P is substrate for standard ribonucleases which can hydrolyse the cyclic phosphate (Crook et al., 1960). The product cytidine 3'-phosphate has a different absorption spectrum to C2'3'P, and thus the conversion of the substrate can be monitored by spectrophotometer. Mpa PE was also unable to cleave C2'3'P (data not shown).

3.5.2 3'-ribonucleotide and 3'-phosphate removal by Mpa PE

Mpa PE was tested with DNA-RNA hybrid substrates that had proved successful substrates for Pae PE, and indeed was able to replicate similar results (Figure 3.6 A). To assess the activity of Mpa PE a 5'-fluorescein labelled DNA-RNA primer, with just two 3' ribonucleosides, annealed to a longer DNA template strand to produce a 3' recessed end, was used. In Figure 3.6 A, and all subsequent gel figures in this thesis, the control 'ctrl' lane consists of the labelled substrate only. Loss of a 3' nucleoside resulted in a marginally faster migration through a polyacrylamide gel because of increased electrophoretic mobility. However, loss of the subsequent 3'-P caused a loss of electronegativity, generating a product with a lower electrophoretic mobility

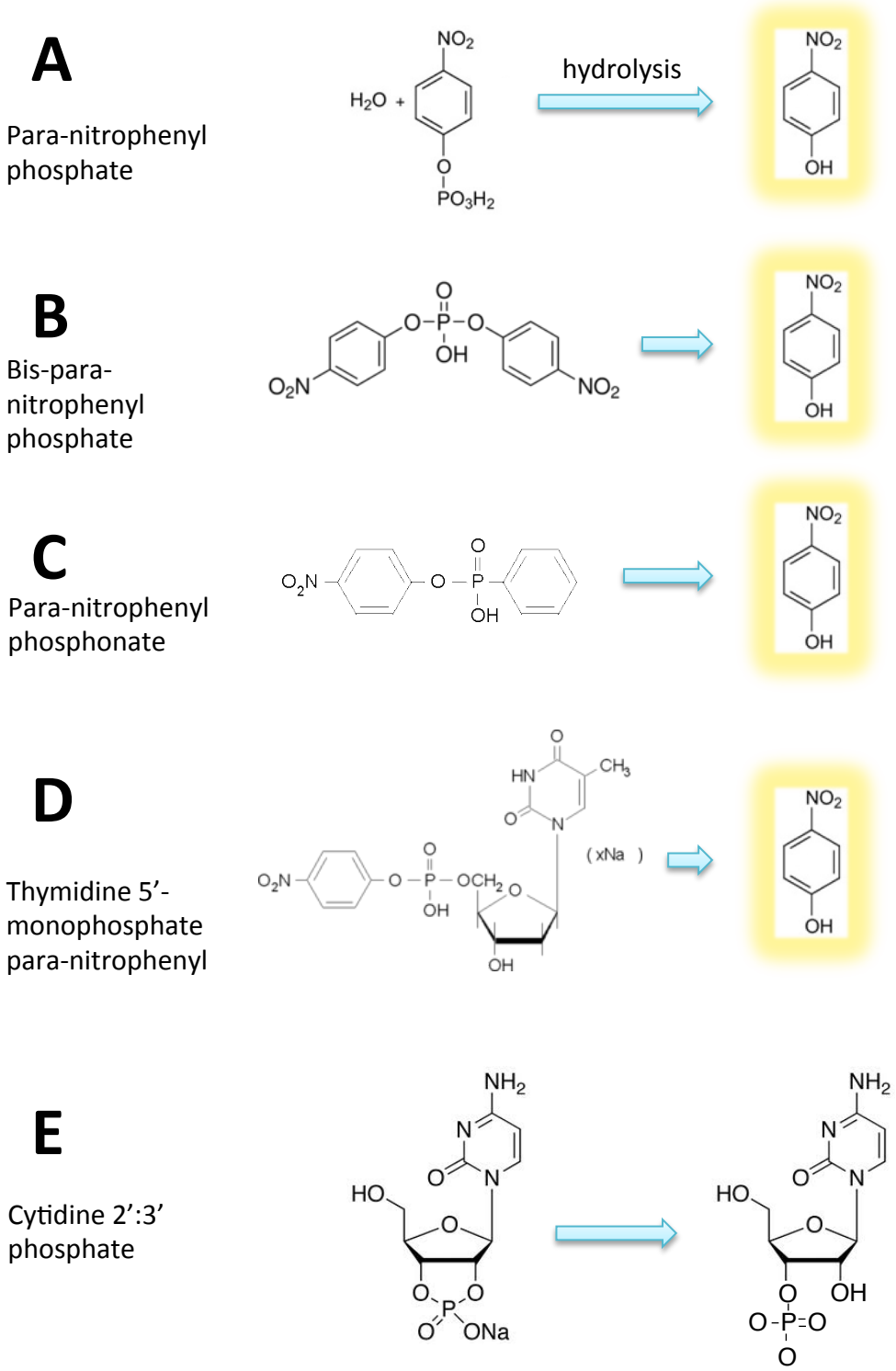


Figure 3.5 Mpa PE cannot cleave para-nitrophenyl phosphate-like substrates.
A variety of compounds were used to define the limits of Mpa PE substrate specificity. Substrates **A** – **D** can be hydrolysed to produce para-nitrophenyl which creates an intense colourimetric change from clear to bright yellow. PDE1 was used as a positive control, Mpa PE could not cleave any of the substrates. The hydrolysis of C2'3'P (**E**) creates cytidine 3'-monophosphate which has a different absorption spectrum and can be monitored by spectrophotometer. RNase A was used as a positive control, Mpa PE could not hydrolyse the cyclic phosphate.

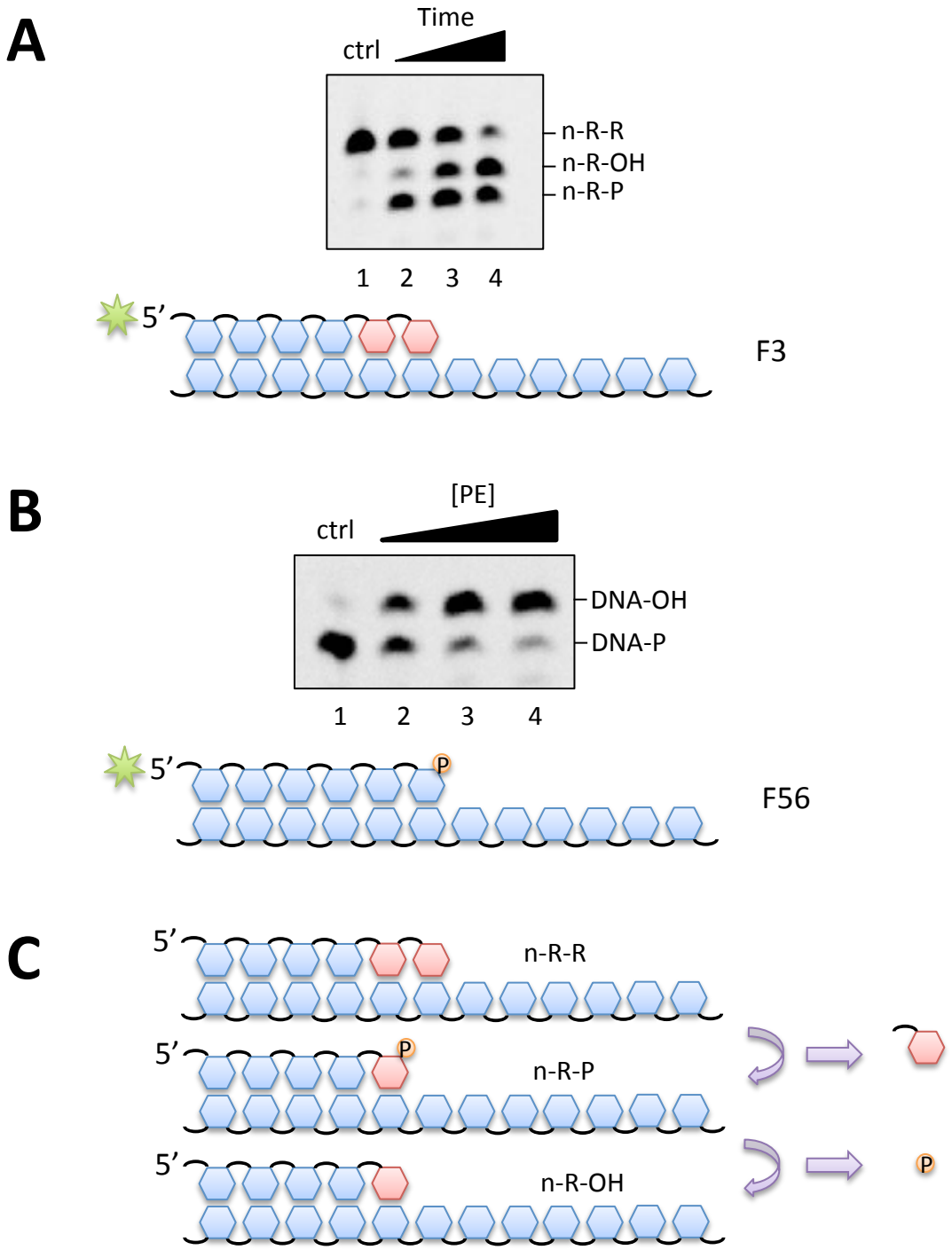


Figure 3.6 Mpa PE catalyses the cleavage of phosphodiester and phosphomonoester bonds.

A schematic of the DNA substrate used in the phosphoesterase and phosphatase assays is shown below the scan of each gel. The red bases indicate rNMPs and the orange circle is a phosphate. 'ctrl' lanes contain labelled DNA substrates and no enzymes in all instances throughout this thesis. (**A**) Phosphoesterase reactions contained 300 nM Mpa PE and 30 nM 5'-fluorescein labelled substrate F3 (primer 16-mer, template 35-mer) and 5 mM Mn. The reactions were incubated at 37 °C for 30, 60 and 90 minutes. (**B**) Gap filling reactions contained 300 nM Mpa PE and 30 nM 5'-fluorescein labelled substrate F56 (primer 16-mer, template 35-mer) and 5 mM Mn. The reactions were incubated at 37 °C for 30, 60 and 90 minutes. (**C**) A schematic detailing the intermediate and product of the phosphoesterase reaction, showing the phosphodiesterase and phosphomonoesterase steps respectively.

than the DNA-RNA-P primer. This caused the 'upwards' phosphate shift observed (Figure 3.6, A, lanes 3 and 4, Figure 3.6, B, lanes 2, 3 and 4).

Mpa PE performed sequential phosphodiesterase and phosphomonoesterase cleavage steps, removing the terminal ribonucleoside and then 3'-phosphate (Figure 3.6 C). As with Pae PE, the archaeal PE could not remove the final ribonucleoside, owing to the absence of a 2'-OH group on the penultimate nucleoside (Zhu & Shuman, 2005b). Mpa PE therefore prefers to generate DNA products with a 3'-monoribonucleoside.

3.5.3 Mpa PE can remove a 3' phosphate from a DNA primer strand

The phosphomonoesterase ability of Mpa PE was tested separately, against an all DNA primer with a 3'-phosphate. Upon incubation with the PE, a phosphate shift was observed, demonstrating the ability to cleave phosphomonoester bonds in the absence of a 2'-OH on the penultimate nucleoside (Figure 3.6 B). The product remaining was an all DNA primer with a 3'-OH group, and no degradation of this was observed, agreeing with prior results that the PE domain has no DNA exonuclease activity.

3.5.4 Mpa PE has preferential nucleolytic activity with manganese

The prior establishment of PE activity was performed in the presence of manganese, based on the published preference for other PE proteins. A divalent metal ion test was used to see if other ions could elicit more efficient or comparable activity in Mpa PE. A DNA-RNA substrate was incubated with Mpa PE and a variety of different divalent metal ions; manganese, magnesium, cobalt, copper, calcium nickel and zinc (Figure 3.7). The metal chelating agent EDTA was also included in one sample, to observe activity, if any, in the absence of metal ions. All samples were incubated for 90 minutes and resolved on a polyacrylamide gel. The sample incubated with manganese sample was the only one to remove the ribonucleoside and phosphate from all of the substrate, whilst cobalt elicited a similar but not quite as efficient reaction (Figure 3.7, lanes 2 and 4). Magnesium showed a moderate ability to enable removal of the ribonucleoside, but was poor at enabling phosphatase activity (Figure 3.7, lane 3). Addition of nickel and EDTA almost completely abrogated PE activity, and zinc was also a poor cofactor (Figure 3.7, lanes 7, 8 and 9). The results concur with those found with Pae PE, that manganese offers optimal activity for the PE (Zhu & Shuman, 2005b).

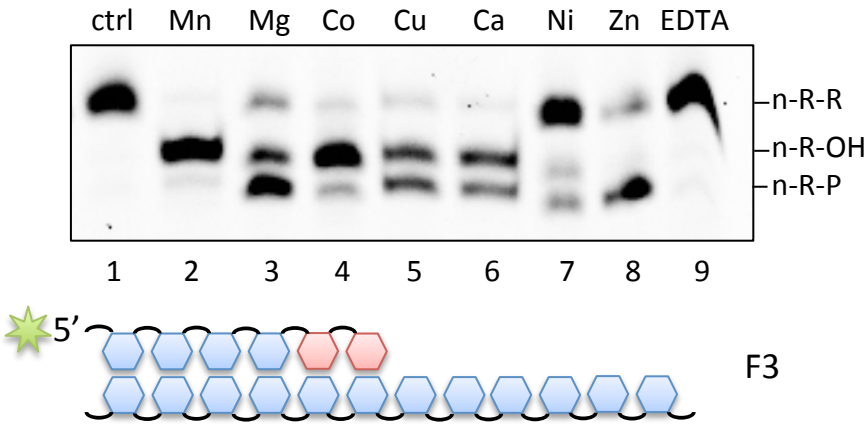


Figure 3.7 Mpa PE preferentially removes ribonucleosides and phosphates in the presence of manganese.

The schematic of the DNA substrate used in the phosphoesterase and phosphatase assay is shown below the scan of the gel. Phosphoesterase reactions contained 300 nM Mpa PE and 30 nM 5’-fluorescein labelled substrate F3 (primer 16-mer, template 35-mer). Reactions also contained 5 mM Mn, Mg, Co, Cu, Ca, Ni, Zn or EDTA as indicated.

3.5.5 Mpa PE is not a processive ribonuclease or phosphatase

To explore the limits of Mpa PE RNA resection, a variety of fluorescein substrates were designed with increasing lengths of RNA at the 3' end. An assay that compared four of these substrates was performed, with 16-mer DNA-RNA primers with 3' RNA tracts of 2, 4, 6 and 8 ribonucleosides, annealed to a 35-mer template. The samples were incubated for 30, 60 and 90 minutes each. Mpa PE efficiently catalysed the cleavage of the primer with just 2 RNA bases, as previously observed, but was entirely inefficient at cleavage of any of the longer sequences (Figure 3.8). This interesting result shows that despite being capable of removing RNA bases, and despite necessity of the 2'-OH to remove any kind of base, Mpa PE is inefficient at resection of even a moderate length of RNA. The results suggest that Mpa PE may not resect such lengths of RNA physiologically, preferring to resect one ribonucleoside. Pae PE was shown to be capable of sequential cleavage of two ribonucleosides, in a step wise manner, and removing both the intermediate and terminal 3'-P, but was unable to resect RNA from an all RNA primer (Zhu & Shuman, 2005b). It is unclear whether PE recognises the DNA-RNA juncture point, and then adopts an active interaction with the substrate, or whether PE recognises the general secondary structure of the helix. An extended stretch of RNA in DNA can cause it to assume A-form structure, and the PE may only bind B-form with reasonable efficiency.

3.5.6 Residue H82 is essential for removal of annealed ribonucleosides

A proposed catalytically null mutant of Mpa PE was generated in the Doherty lab by Tom Carlberg, and was kindly offered for use in experiments. The conserved active site residue H82 was selected as a likely candidate for being essential for activity, and data from Pae PE supported this choice (Zhu & Shuman, 2005b). Site directed mutagenesis was used to alter the histidine to an alanine. When compared with wt (wild type) Mpa PE against a standard DNA-RNA substrate, Mpa PE H82A was completely inactive (Figure 3.9, lane 5). A similar substrate with two ribonucleosides that were non-complementary to the template, essentially a short 3' flap, was also tested. Wt (wild type) Mpa PE was slightly less efficient at removing the 3'-phosphate, but the phosphodiesterase activity was not reduced (Figure 3.9, lanes 7-9). Interestingly, Mpa PE H82A was able to remove the terminal ribonucleoside from much of the substrate over a 90-minute incubation, but was unable to remove the phosphate. All previous experiments with Pae PE found the equivalent residue, H84, to be essential for activity (Zhu et al., 2005; Zhu & Shuman, 2006). The implication from the results with Mpa PE H82A is that the residue is essential for removing the terminal ribonucleoside if it is properly annealed to the

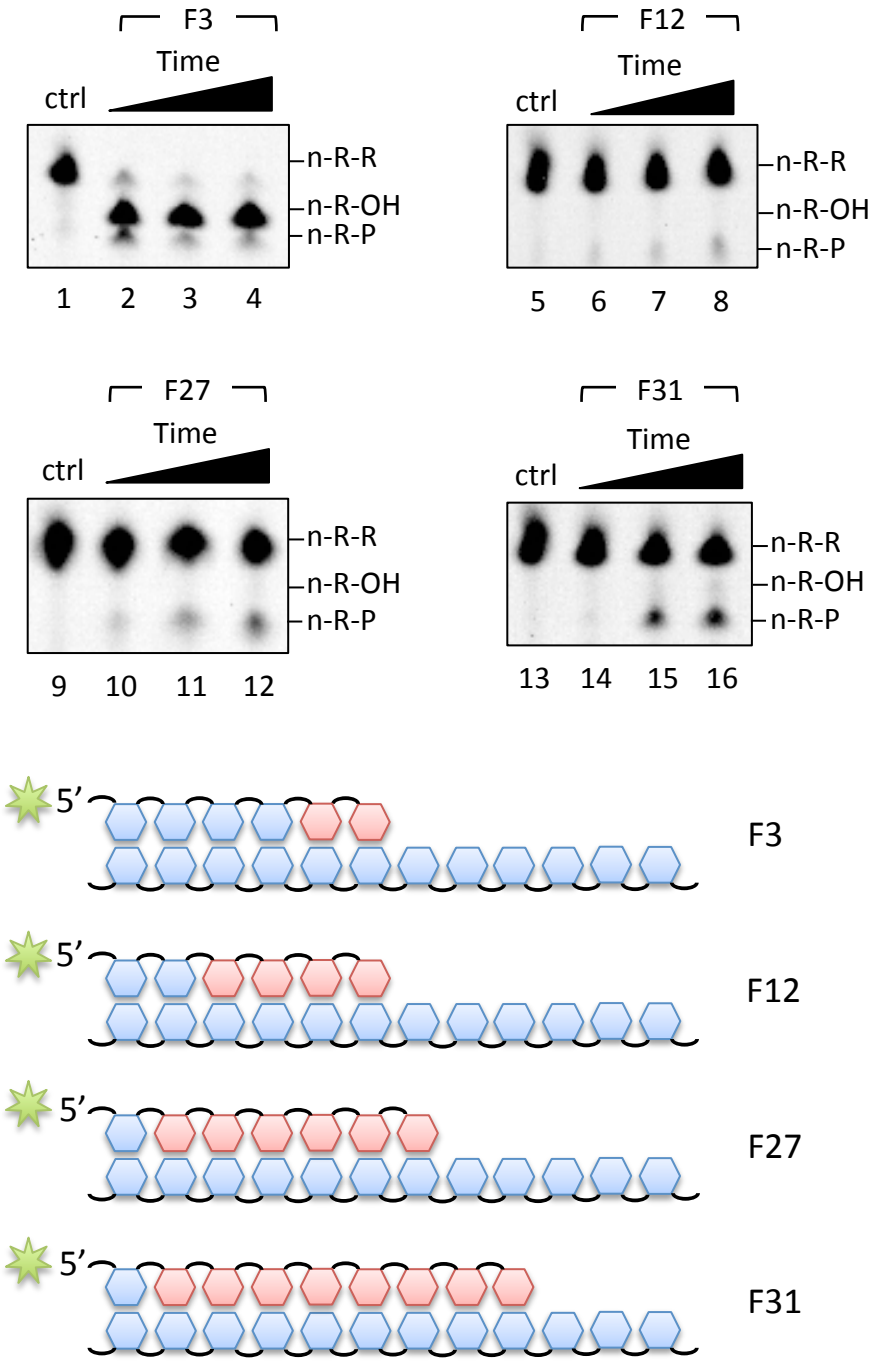


Figure 3.8 Mpa PE is inefficient at removing 3' RNA of 4 or more bases.
Schematics of the DNA substrates used in the phosphoesterase assays are shown below the scan of each gel. Phosphoesterase reactions contained 300 nM Mpa PE, 30 nM of the indicated 5'-fluorescein labelled substrate F3, F12, F27 and F31 (each with primer 16-mer, template 35-mer) and 5 mM Mn. The reactions were incubated at 37 °C for 30, 60 and 90 minutes.

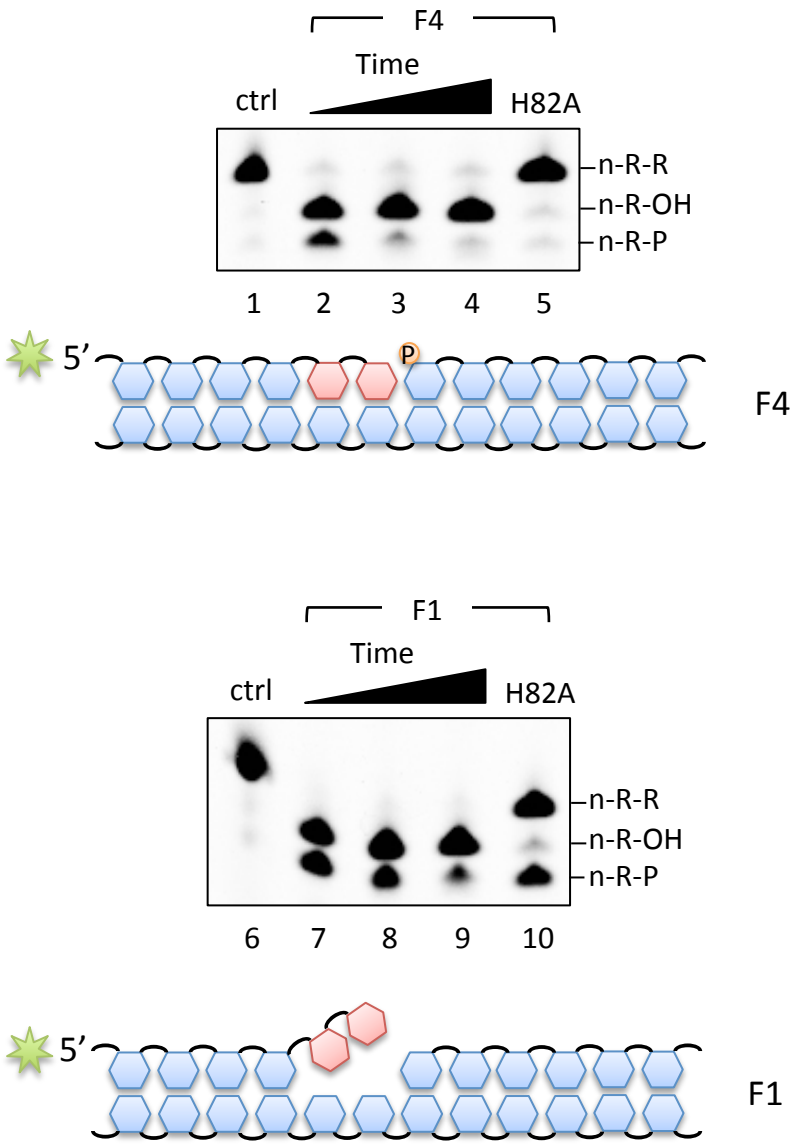


Figure 3.9 Mpa PE residue H82 is essential for removal of annealed ribonucleosides.

Schematics of the DNA substrates used in the phosphoesterase assays are shown below the scan of each gel. Phosphoesterase reactions contained 300 nM Mpa PE and 300 nM Mpa PE H82A mutant where indicated. Reactions also contained 30 nM 5’-fluorescein labelled substrate F4 or F1 (each with primer 16-mer, template 35-mer, and downstream 19-mer) and 5 mM Mn. The wild type reactions were incubated for 30, 60 and 90 minutes, and the H82A mutants were incubated for 90 minutes, all at 37 °C. The two 3’ RNA bases of substrate F1 are non-complementary with the template strand, and create a 3’ flap.

template strand. It is tempting to conclude that the hisitidine residue plays a role in interrupting the base pairing of the incoming nucleoside and the DNA template, allowing the correct positioning of the scissile phosphate in the active site, however structural information of the PE bound to a substrate is required to be certain.

3.6 Sequence alignment, expression and purification of Mpa Pol

The polymerase domain (PolDom) of bacterial LigD is by far the most thoroughly characterised of all of the LigD domains. PolDom is amenable to truncation to a single domain and maintains solubility and activity (Pitcher, Brissett, Picher, et al., 2007b; Brissett et al., 2011). The identification of the first potential archaeal NHEJ Pol offered an opportunity to establish whether it shared the myriad functions with the bacterial PolDom. The limits of the archaeal NHEJ Pol could be assayed with substrates used to test eukaryotic AEP primase-polymerases that were also studied in the Doherty lab. A further goal for the Mpa Pol project was to crystallise it in coordination with one of the other Mpa NHEJ proteins, namely Ku, in the presence of DNA. This target would hopefully shed light on the specific interactions between Ku and its known LigD interacting domain, and will be discussed further in Chapter 5.

3.6.1 Mpa Pol protein sequence alignment

The protein sequence of Mpa Pol was aligned against Mtu, Msm and Pae LigD PolDoms (Figure 3.10). The motifs that confer membership to the AEP superfamily are highly conserved between the polymerases, and the catalytic aspartate triad is strictly conserved in all of the proteins (Figure 3.10, red stars) (Iyer et al., 2005). The phenylalanine pair that sits on the cusp of the bacterial PolDom active site is also conserved, along with two N-terminal lysines and an asparagine residue which form the positively charged phosphate binding pocket (Figure 3.10, green and purple stars for the phenylalanines and phosphate binding residues respectively). The Loop 2 region is also well conserved with the active lysine and arginine residues in place in both mycobacteria and *M. paludicola*. Interestingly, much of Loop 1 from Mtu PolDom is not conserved in Mpa Pol. Subsequent structural studies could demonstrate whether the archaeal polymerase has a similar appendage, and whether it retains the bacterial polymerase's finer control of DNA ends attributed to Loop 1 (Brissett et al., 2007; Brissett et al., 2011).

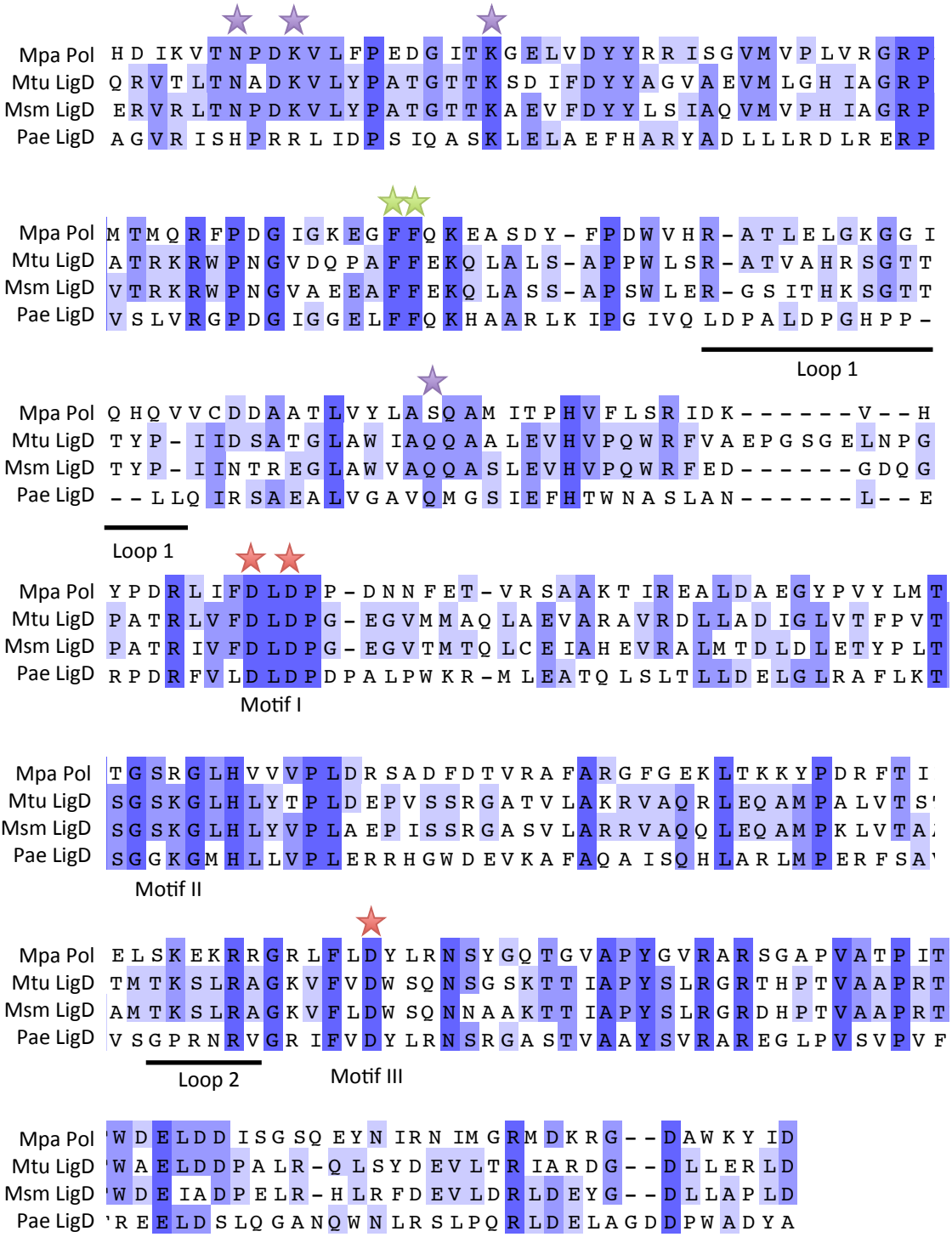


Figure 3.10 Protein sequence alignment of Mpa Pol with Mtu, Msm and Pae LigD NHEJ polymerase domains.

The protein sequence alignments indicate identical residues highlighted in dark blue, and conserved residues in lighter shades of blue. The positions of the archaeo-eukaryotic primase superfamily motifs and Loops 1 and 2 are noted beneath the sequence. The red stars indicate the positions of the aspartate triad which coordinates the divalent metal ion in the active site. The green stars indicate the positions of the two phenylalanine residues that orient the DNA template strand towards the active site. The purple stars indicate the position of the conserved residues which form the 5'-phosphate binding pocket.

3.6.2 Mpa Pol expression

Mpa Pol expression trials were carried out under the same conditions as Mpa PE, described above in section 3.4.2. The greatest amount of soluble Mpa Pol (~35.7 kDa) was observed in the lysate from cells induced at 37 °C (Figure 3.11, A lane 8). The 4 litre growth of Mpa Pol cultures were therefore incubated at 37 °C for both growth and induction. Cells were collected by centrifugation and snap frozen in liquid nitrogen for storage at -80 °C.

3.6.3 Chromatography of Mpa Pol

The Mpa Pol frozen pellets were defrosted and lysed, as described above for Mpa PE. The soluble lysate was loaded onto an IMAC nickel resin column pre-equilibrated in IMAC Buffer A. The protein bound to the nickel resin and, following a wash step, was eluted in 300 mM imidazole by IMAC Buffer B. The coomassie blue stained polyacrylamide gel showed a significant enrichment for Mpa Pol, although several other contaminants also remained (Figure 3.11, B, lane 5). Applying 300 mM imidazole to the IMAC column as the intial elution means that any other proteins with affinity for the nickel resin also elute with the target protein. However, this method also ensures that as much of the intended protein is rescued as possible. The concept is then to remove the persisting contaminants with later purification steps.

The next step for Mpa Pol was combined cation and anion exchange, a technique that was developed after it was observed that Mpa Pol did not bind either mono-Q or mono-S columns. Many of the contaminants did bind to the columns, which resulted in a more highly purified Mpa Pol in the flow through solution (Figure 3.12, A). The Ion exchange flow through was concentrated and loaded onto a Superdex 200 size exclusion chromatography column, which was pre-equilibrated with SEC Buffer A. The polymerase eluted with a distinct peak and only trace amounts of contaminating protein could be observed in a very overloaded polyacrylamide gel (Figure 3.12, B). The protein was aliquoted and snap frozen in liquid nitrogen for storage at -80 °C. The average yield of highly purified Mpa Pol was 60 mg from a 4 litre growth in Terrific Broth.

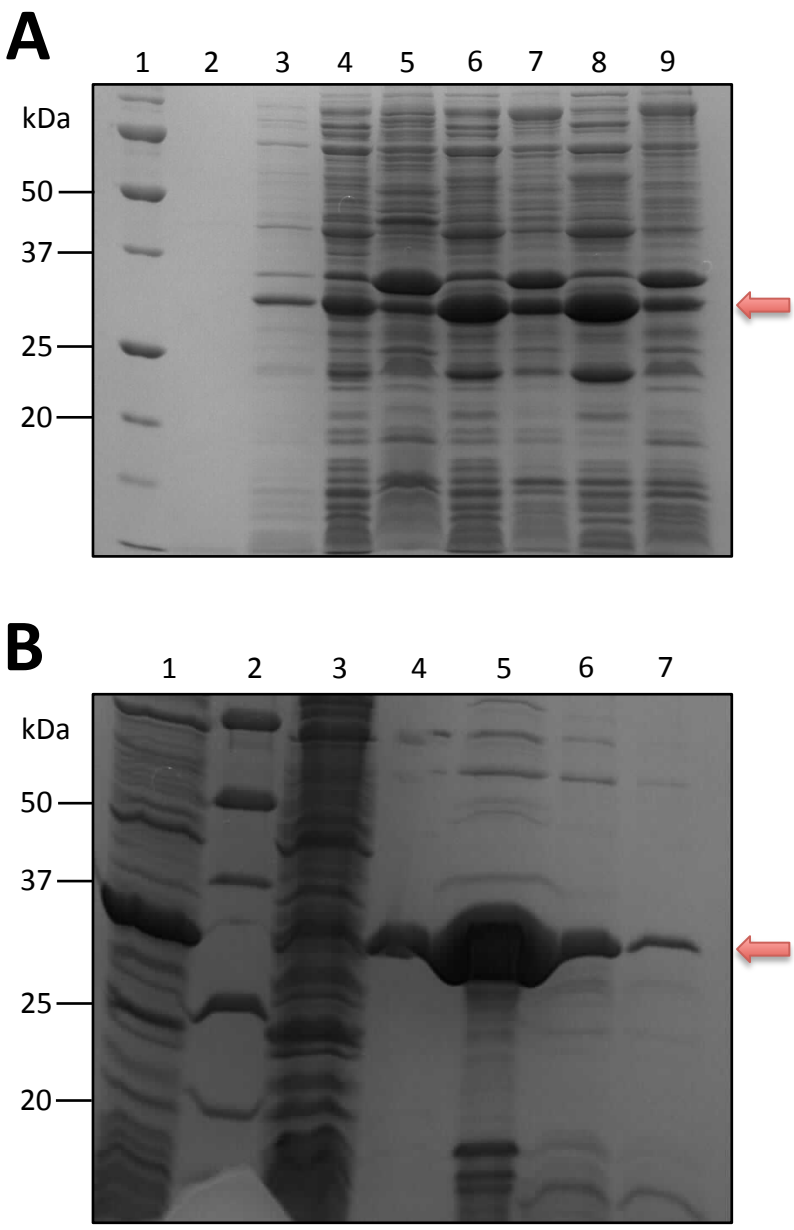


Figure 3.11 Expression and purification of Mpa Pol.
Coomassie stained protein gels. Presence of Mpa Pol protein indicated by the red arrow at ~35.7 kDa. (**A**) Expression of Mpa Pol in B384S cells, induced by 1 mM IPTG. 1= Molecular weight marker, 2= Uninduced cells, 3= Induced cells after 1 hr, 4= Soluble lysate 20 °C induction, 5= Insoluble pellet 20 °C induction, 6= Soluble lysate 30 °C induction, 7= Insoluble pellet 30 °C induction, 8= Soluble lysate 37 °C induction, 9= Insoluble pellet 37 °C induction. (**B**) Purification of Mpa Pol by IMAC nickel resin column. 1= Load, 2= Molecular weight marker, 3= Flow through, 4= Wash, 5= 300 mM imidazole, 6= 300 mM imidazole fraction II, 7= 300 mM imidazole wash.

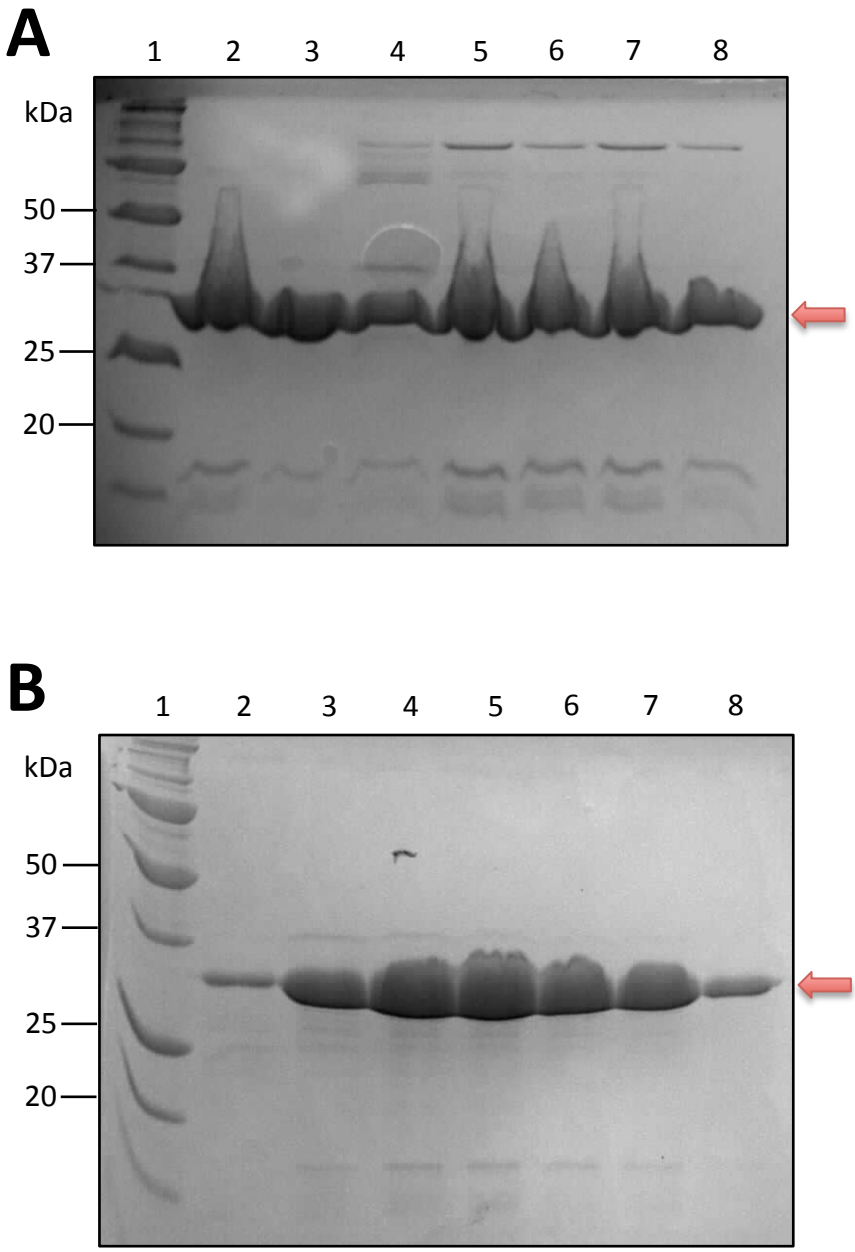


Figure 3.12 Purification of Mpa Pol.
Coomassie stained protein gels. Presence of Mpa Pol protein indicated by the red arrow at ~35.7 kDa. (**A**) Purification of Mpa Pol by mono-Q and mono-S columns. 1= Molecular weight marker, 2= Flow through fraction 2, 3= Flow through fraction 3, 4= Flow through fraction 5, 5= Flow through fraction 8, 6= Flow through fraction 11, 7= Flow through fraction 14, 8= 200 mM NaCl. (**B**) Purification of Mpa Pol by size exclusion chromatography. 1= Molecular weight marker, 2= Fraction 14, 3= Fraction 16, 4= Fraction 17, 5= Fraction 18, 6= Fraction 19, 7= Fraction 20, 8= Fraction 21.

3.7 Mpa Pol is a DNA-dependent DNA/RNA polymerase

3.7.1 Mpa Pol DNA primer extension

Mpa Pol was incubated with a 5'-fluorescein labelled DNA substrate with a recessed 3' end to assess the template dependent DNA synthesis capabilities. The assay tested 4 standard New England Biolabs buffers, and two custom buffers with 50 mM Tris pH 7.5 and 5 mM manganese or magnesium. Each buffer condition was tested with Mpa Pol in the presence of either dNTPs or NTPs (Figure 3.13, A). Mpa Pol was unable to extend the DNA primer with dNTPs except in the presence of manganese, in which case only 1 base was efficiently added (Figure 3.13, A, lanes 1-7). However when using NTPs Mpa Pol was able to extend the primer by at least a single base in all buffers, and in the presence of manganese was able to extend the primer by 6 bases (Figure 3.13, A, lanes 8-14). This experiment established that Mpa Pol, like all of the bacterial NHEJ Pols characterised thus far, preferentially incorporates NTPs to a DNA template in the presence of manganese (Zhu & Shuman, 2005a; Pitcher, Brissett, Picher, et al., 2007b). The NHEJ Pol is a distributive polymerase, like human Pol μ, and cannot extend primers in a processive manner. The DNA template used in this assay allowed for a full extension of 19 bases, but Mpa Pol did not incorporate more than 6 or 7 bases, even with an extended time course (data not shown). Mpa Pol shares similar substrate specificity as Mpa PE; it is unable to interact with a DNA-RNA primer over a certain length. Both sets of data suggest that these enzymes prefer to operate with only short RNA insertions at DNA ends.

3.7.2 Gap filling activity of Mpa Pol

Mpa Pol was also shown to efficiently fill DNA gaps, using a similar substrate as described above, except with downstream DNA creating a single nucleotide gap (Figure 3.13, B). Mpa Pol also filled the single gap accurately, selecting the correct base pairing nucleotide (Figure 3.13, B, lane 2). Gap filling is a common function of NHEJ polymerases, as it allows them to repair gaps either side of a microhomology synapse (Pitcher, Brissett, Picher, et al., 2007b). As such, a DNA gap substrate is often used as an analogue of a tethered break in an *in vitro* assay. In this assay, Mpa Pol was observed to incorporate two cytidine nucleotides into a one nucleotide gap, and this will be discussed further in Chapter 4.

3.7.3 Template independent synthesis activity of Mpa Pol

As well as template dependent synthesis, Mpa Pol can catalyse the extension of a blunt dsDNA end (Figure 3.14). Mpa Pol incorporated a single nucleotide only, but could add any of the

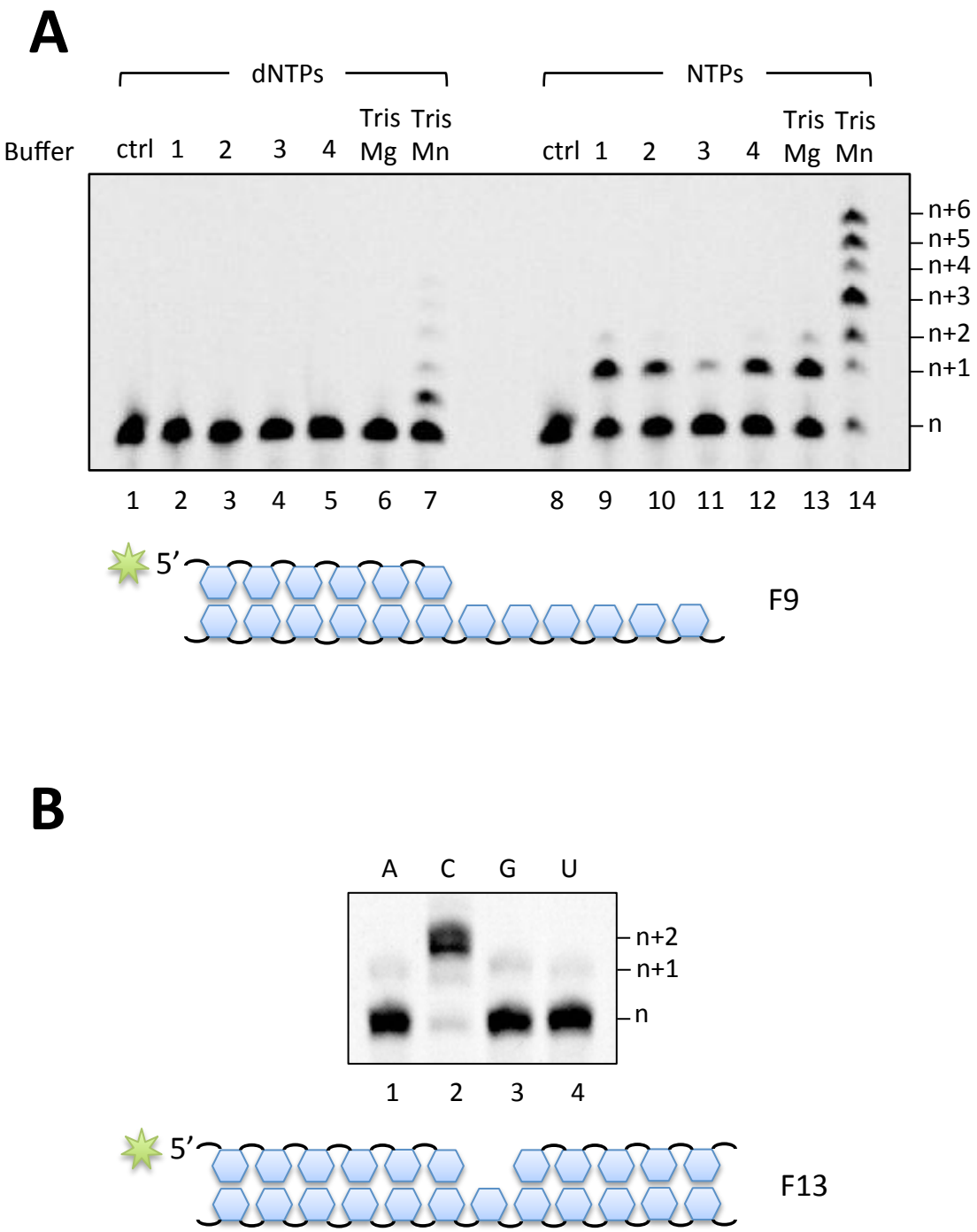


Figure 3.13 Mpa Pol is a DNA-dependent DNA/RNA polymerase and is capable of gap filling.

Schematics of the DNA substrates used in the primer extension and gap filling assays are shown below the scan of each gel. (**A**) Primer extension reactions contained 300 nM Mpa Pol, 30 nM 5'-fluorescein labelled substrate F9 (primer 16-mer, template 35-mer), and 250 µM NTPs or dNTPs. Reaction buffers were NEB 1-4, or 50 mM Tris pH 7 and 5 mM Mn or Mg as indicated. (**B**) Gap filling reactions contained 300 nM Mpa Pol and 30 nM 5'-fluorescein labelled substrate F13 (primer 16-mer, template 35-mer, downstream 15-mer). Reactions also contained 62.5 µM of either ATP, CTP, GTP or UTP, and 5 mM Mn.

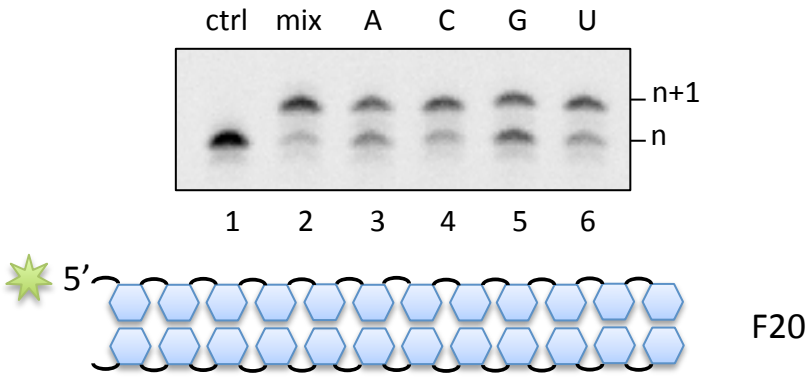


Figure 3.14 Mpa Pol can catalyse template independent synthesis.
A schematic of the DNA substrate used in the template independent extension assay is shown below the scan of the gel. Template independent extension reactions contained 300 nM Mpa Pol and 30 nM 5’-fluorescein labelled substrate F20 (primer 35-mer, template 35-mer). Reactions also contained 62.5 µM of either ATP, CTP, GTP, UTP or a 250 µM mix, and 5 mM Mn.

nucleotides independent of a DNA template. This activity could be essential if a DSB produced blunt ends, especially given that the bacterial NHEJ ligases require a monoribonucleoside for ligation. This means that blunt dsDNA ends cannot be directly ligated by the prokaryotic NHEJ ligase, and requires NHEJ Pol input before it can reliably repair the break. This was observed during *in vivo* plasmid repair assays of blunt ends, when two of the metal ion chelating aspartate residues in the polymerase domain were mutated to alanine (Aniukwu et al., 2008). The LigD double Asp (aspartate) mutant was catalytically dead, and the efficiency of repair dropped by 71 % from that of wild type. Interestingly, in the plasmids that were still repaired (presumably by a back-up system) the fidelity of repair improved significantly, suggesting that there may be methods for blunt DNA ends to be directly ligated. Even so, it seems the bacterial, and now archaeal, NHEJ pathway prefers to make use of the NHEJ Pol ability to add nucleotides independently of templates at blunt dsDNA ends.

3.7.4 Extension from blunt and 3' overhanging ends by Mpa Pol

Addition of a non-labelled DNA substrate with a single 3'-base overhang to the blunt ended extension assay enabled a demonstration that Mpa Pol can use the single 3'-base as a template. In the previous assay, and repeated here, Mpa Pol extended a moderate amount of blunt dsDNA ends with ATP (Figure 3.14, lane 3, Figure 3.15, A, lane 3). Addition of the single templating base substrate increased the amount of ATP incorporated to almost 100 % of the substrate (Figure 3.14, A, lane 5). This intriguing result supports the findings that the bacterial NHEJ Pol can interact with two DNA ends simultaneously, and supports the data on Mtu PolDom forming a pre-ternary complex (Brissett et al., 2011). A crystal structure and associated biochemistry revealed that Mtu PolDom could bind a DNA end with a 3' overhang and direct a nucleotide into a templated position in the absence of a primer to extend from. It was shown that this pre-ternary complex formation enabled a more efficient extension reaction when the primer was added.

A similar experiment used a single DNA substrate with a 5 base 3' overhang, of which the last four bases were self-complementary. This would allow for the formation of a microhomology synapse with a single nucleotide gap either side (Figure 3.14, B). In this case the unpaired base was adenine, and Mpa Pol accurately incorporated UTP opposite (Figure 3.14, B, lane 5). This exposed the ability of Mpa Pol to encourage microhomology synapse formation, to anneal the break, and then accurately fill a gap either side.

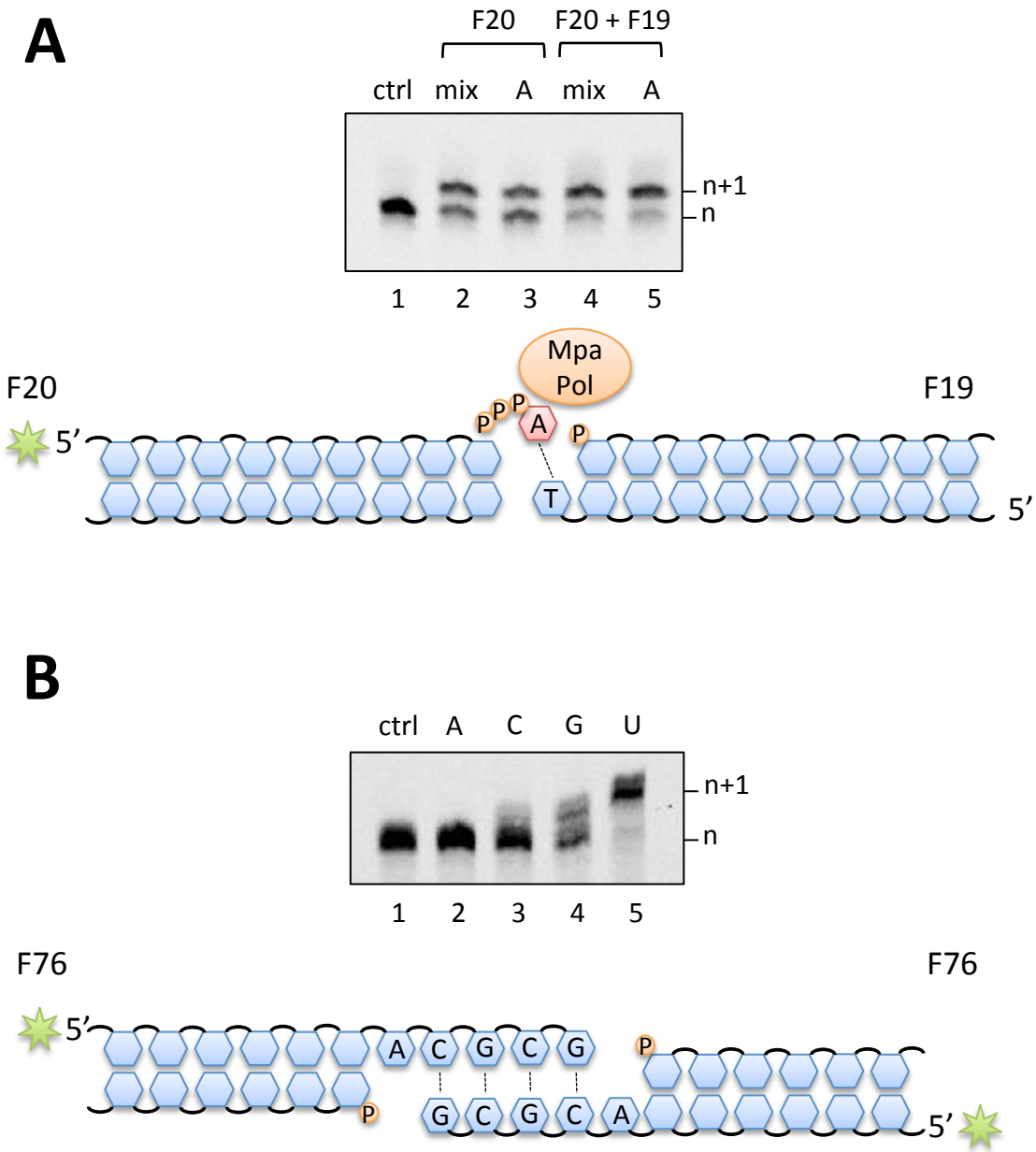


Figure 3.15 Mpa Pol can extend from blunt and 3’ overhanging DNA ends using a separate DNA end as a template.

Schematics of the DNA substrates used in the blunt end and 3’ overhanging end extensions are shown below the scan of each gel. (**A**) Blunt end extension reactions contained 300 nM Mpa Pol and 30 nM 5’-fluorescein labelled substrate F20 (primer 35-mer, template 35-mer) with unlabelled substrate F19 (primer 35-mer, template 36-mer) where indicated. Reaction mixtures also contained 62.5 µM ATP or a 250 µM NTP mix, and 5 mM Mn. (**B**) 3’ overhanging end extensions contained 300 nM Mpa Pol and 30 nM 5’-fluorescein labelled substrate F76 (primer 42-mer, template 37-mer). F76 has a self-complementary 4 base 3’ overhanging end which can form a microhomology when brought together, which essentially forms a double *in trans* gap-filling substrate. Reaction mixtures also contained 62.5 µM ATP, CTP, GTP or UTP, and 5 mM Mn.

3.7.5 5'-phosphate binding promotes DNA binding by Mpa Pol

The previous assays suggest that Mpa Pol has the ability to bind a 5'-phosphate of a recessed DNA end. The protein sequence alignment of Mpa Pol with the mycobacterial LigD PolDoms showed that it retains the N-terminal positively charged lysine residues and the conserved asparagine residue, and therefore probably also maintains this phosphate binding pocket (Figure 3.10, purple stars). When challenged with three different substrates; primer and template, gap DNA, gap DNA with a 5'-P, Mpa Pol could only form a stable complex with the latter substrate (Figure 3.16, lanes 7-9). Mpa Pol does not seem to bind to a 5'-P with the same affinity as Mtu PolDom which does so with high affinity (Pitcher, Brissett, Picher, et al., 2007b; Brissett et al., 2011). Nevertheless, this suggests that Mpa Pol can recognise and bind to a 5'-P, which is a conserved feature of NHEJ Pols. Structural information on the architecture of the Mpa Pol phosphate-binding pocket will be discussed in Chapter 5.

3.7.6 Mpa Pol can not incorporate NTPs opposite bulky UV lesions or prime *in vitro*

Investigation of other AEPs in the Doherty lab have discovered translesion synthesis capabilities, namely by the eukaryotic PrimPol (Bianchi et al., 2013). PrimPol is a eukaryotic AEP and can synthesis DNA opposite 6-4 photoproducts, and continue with a limited amount of extension. Since Mtu PolDom was reported to possess the ability to correctly insert CTP opposite 8-oxo-G, Mpa Pol was tested for the ability to incorporate NTPs opposite bulky DNA lesions (Pitcher, Brissett, Picher, et al., 2007b). Mpa Pol was not found to incorporate opposite either CPD or 6-4pp lesions (Figure 3.17). Although the NHEJ Pols boast an open active site, Mpa Pol does not have the flexibility to facilitate the bulky UV lesions.

To determine if Mpa Pol has primase activity we used a novel primase assay. In the assay the potential primase is incubated with dNTPs or NTPs and a single-stranded template with a biotin tag. Following an incubation at 37 °C, allowing for the synthesis of short DNA or RNA primers, a fluorescently labelled dATP was added along with Taq (*Thermus aquaticus*) DNA polymerase, which would then add the label to the end of the new primer. The DNA template had a 3'-dideoxynucleoside terminus preventing any terminal transferase addition of a fluorescent label. The sample was added to streptavidin-coated beads that bind biotin with high affinity. Following a gentle wash to remove all excess reaction components except DNA, the beads were boiled to liberate any fluorescently labelled primers. Mpa Pol did not produce any primers under any experimental conditions that were favourable for a positive control of Tgo primase (data not shown). These results suggest that despite possessing AEP structural

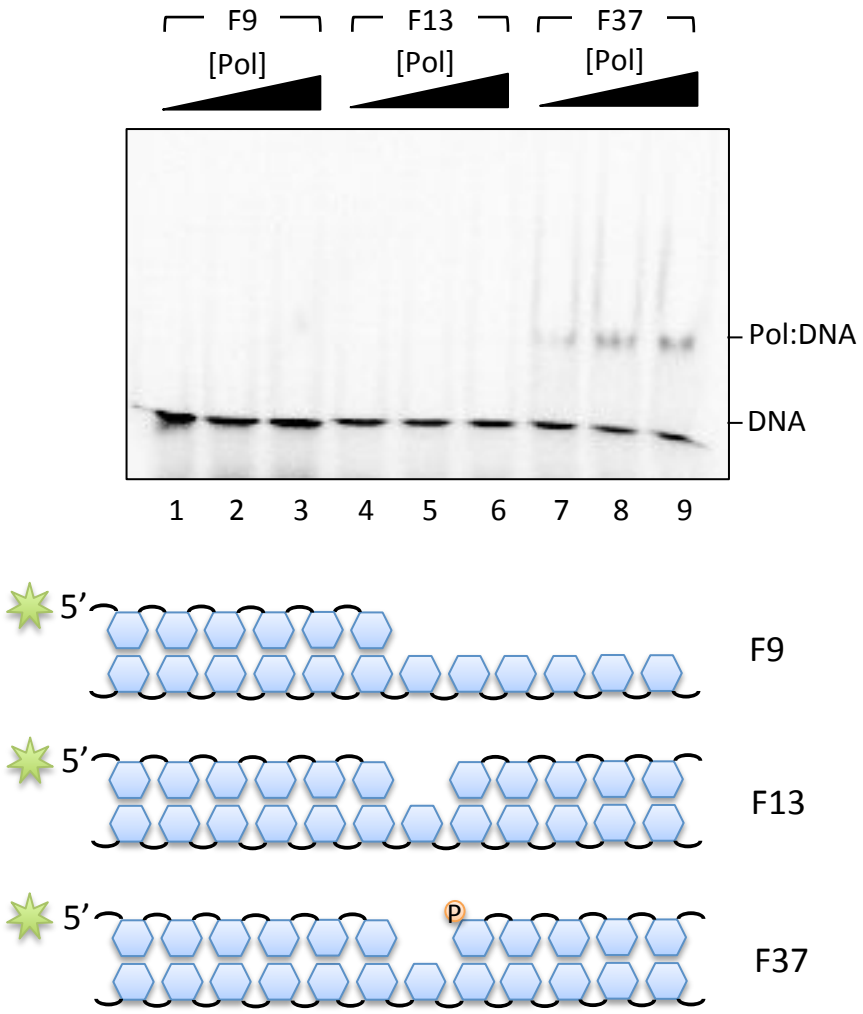


Figure 3.16 Mpa Pol stably binds DNA via 5’-phosphate.
Schematics of the DNA substrates used in the EMSA are shown below the scan of the gel. EMSA reactions contained 200, 400 and 800 nM Mpa Pol where indicated, 60 nM 5’-fluorescein labelled substrates F9 (primer 16-mer, template 35-mer), F13 (primer 16-mer, template 35-mer, downstream 18-mer) or F37 (primer 16-mer, template 35-mer, downstream 18-mer). Reaction mixtures also contained 5 mM Mn.

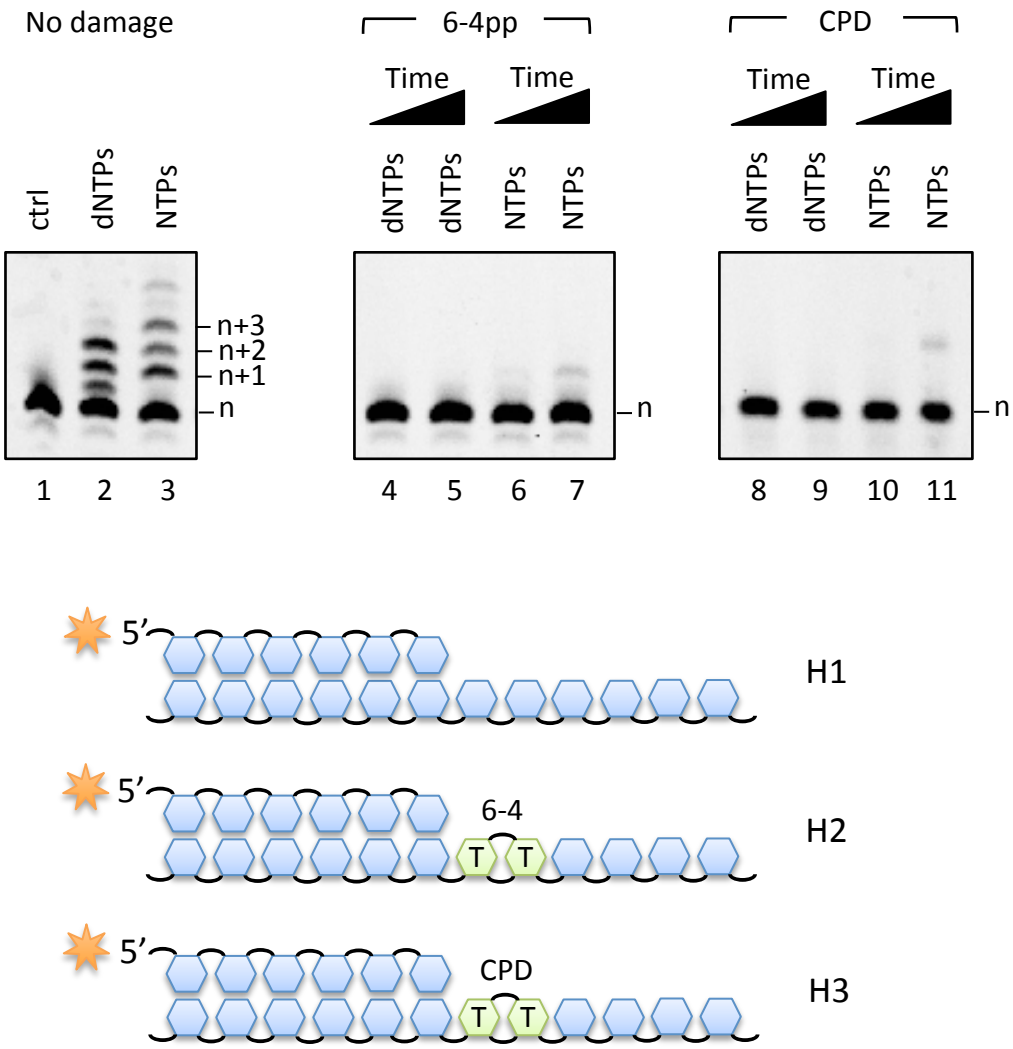


Figure 3.17 Mpa Pol cannot incorporate NTPs or dNTPs opposite 6-4pp or CPDs.
A schematic of the DNA substrate used in the primer extension translesion synthesis assays is shown below the scans of the gels. Translesion synthesis reactions contained 100 nM Mpa Pol and 20 nM 5'-hex labelled substrate undamaged substrate H1 (primer 27-mer, template 50-mer), 6-4pp substrate H2 (primer 16-mer, template 30-mer), or CPD substrate H3 (primer 27-mer, template 50-mer) as indicated. Reactions also contained 5 mM Mn and 200 µM NTPs or dNTPs as indicated.

motifs, Mpa Pol does not have RNA or DNA primase capacities and is therefore only a polymerase.

3.8 Sequence alignment, expression and purification of Mpa Ku

Prior to the identification of a Ku gene in *M. paludicola*, Ku had only been noted in one archaeal genome; that of *Archaeoglobus fulgidus* (Aravind & Koonin, 2001; Weller & Doherty, 2001; Doherty et al., 2001). Therefore, no archaeal Ku has never been characterised, and the Mpa NHEJ system offered an excellent opportunity to rectify this. The crystal structure of bacterial Ku has never been solved, either alone or bound to DNA. Structural studies of Mpa Ku, which was also a target of this research, will be discussed in Chapter 5.

3.8.1 Mpa Ku protein sequence alignment

The protein sequence of Mpa Ku was aligned with Mtu, Msm and Pae Ku (Figure 3.18). The N-termini of the sequences are well conserved, although Mpa Ku has a slightly truncated C-terminus. Given the overall level of amino acid conservation, it is likely that the archaeal Ku adopts a similar fold to the bacterial proteins. Extensive protein sequence alignments of bacterial and eukaryotic Ku revealed conserved putative zinc binding motifs (Krishna & Aravind, 2010). The two motifs in bacterial Ku are HxxC and CxxC (where 'x' is any residue) and are proposed to chelate zinc at the loop region of two β-hairpins, one from each monomer, forming interlocking 'knuckles' (Figure 3.18, red stars indicate the positions of the conserved residues). The mycobacteria possess one motif that is strictly conserved (CxxC), but the other, and both in *M. paludicola*, has conservative substitutions with an aspartate replacing the cysteine. It is possible that the acidic aspartate could still coordinate a zinc ion. Interestingly the *P. aeruginosa* sequence appears to have completely lost both of these sites. It is unclear what, if any, role these sites play in the activity of Ku during NHEJ.

3.8.2 Mpa Ku expression

Mpa Ku expression trials were conducted in the same way as Mpa PE and Mpa Pol (sections 3.4.2 and 3.6.2 respectively). Over-expression of Mpa Ku produced the greatest amount of soluble protein after a 3 hour induction at 37 °C (Figure 3.19, A, lane 7), although a noticeable amount of Ku was visible in the insoluble fractions (Figure 3.19, A, lanes 4 and 6). Mpa Ku was expressed in a 4 L culture with Terrific Broth medium with the same protocol as Mpa Pol

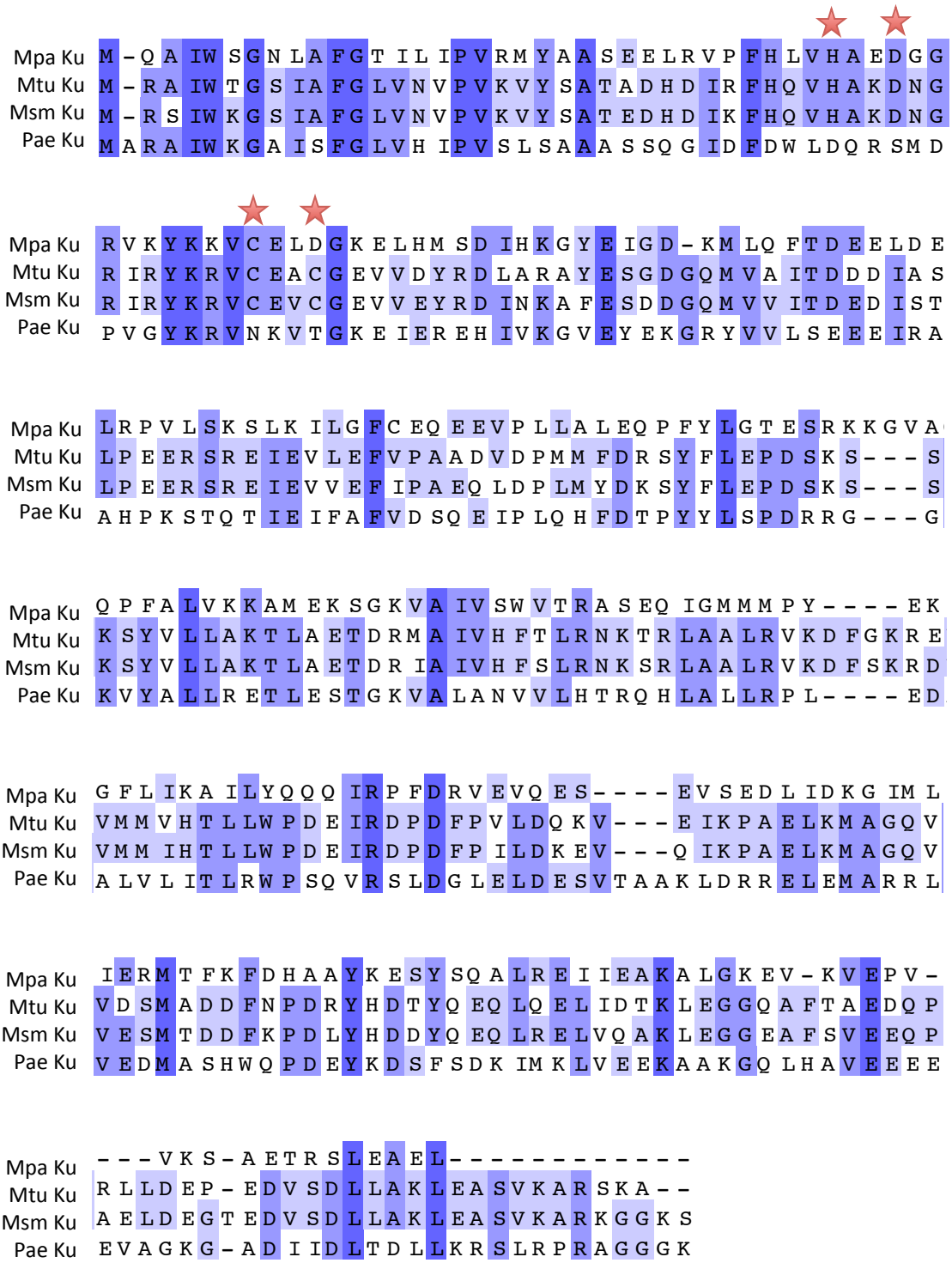


Figure 3.18 Protein sequence alignment of Mpa Ku with Mtu, Msm and Pae Ku.
The protein sequence alignments indicate identical residues highlighted in dark blue, and conserved residues in lighter shades of blue. The red stars indicate the presence of putative zinc binding motifs that are predicted to appear at the loop regions of β-hairpins. The conserved motifs are HxxC and CxxC, where 'x' is any residue. Both the *M. paludicola* and mycobacterial sequences have conserved substitutions where an acidic aspartate replaces a cysteine.

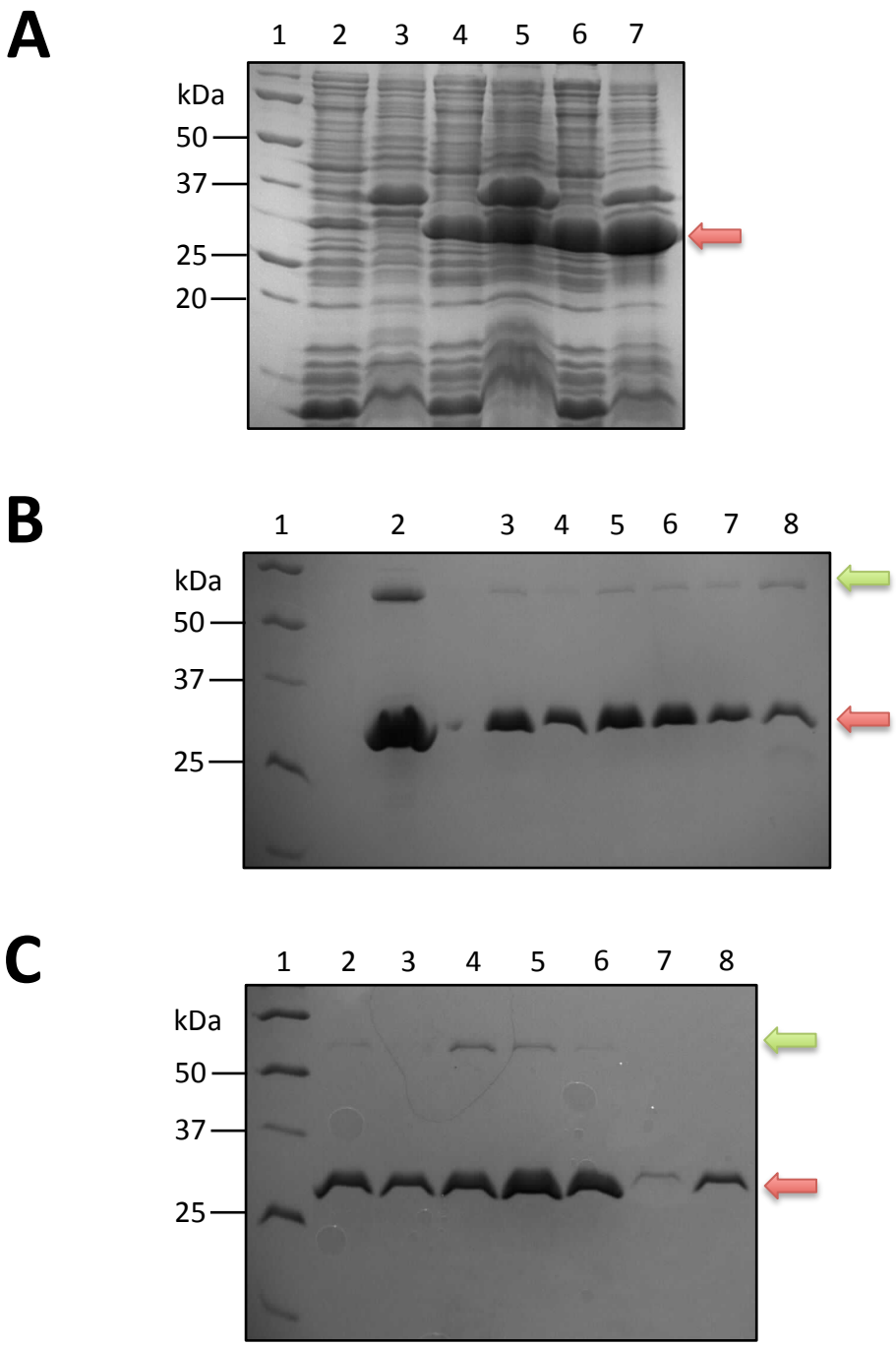


Figure 3.19 Expression and purification of Mpa Ku.
Coomassie stained protein gels. Presence of Mpa Ku protein indicated by the red arrow at ~30.9 kDa. The green arrow indicates dimerised Ku. (**A**) Expression of Mpa Ku in B384S cells, induced by 1 mM IPTG. 1= Molecular weight marker, 2= Soluble lysate 20 °C induction, 3= Insoluble pellet 20 °C induction, 4= Soluble lysate 30 °C induction, 5= Insoluble pellet 30 °C induction, 6= Soluble lysate 37 °C induction, 7= Insoluble pellet 37 °C induction. (**B**) Purification of Mpa Ku by IMAC nickel resin column and mono-Q column. 1= Molecular weight marker, 2= IMAC 300 mM imidazole, 3= Q flow through fraction 3, 4= Q flow through fraction 6, 5= Q 50 mM NaCl, 6= Q 100 mM NaCl, 7= Q 200 mM NaCl, 8= Q 600 mM NaCl. (**C**) Purification of Mpa Ku by size exclusion chromatography. 1= Molecular weight marker, 2= Fraction 8, 3= Fraction 9, 4= Fraction 10, 5= Fraction 11, 6= Fraction 12, 7= Fraction 13, 8= Fraction 14.

(section 3.6.2). The B834S cells were collected by centrifugation and snap frozen in liquid nitrogen, then stored at -80 °C.

3.8.3 Purification of Mpa Ku

The Mpa Ku cell pellets were defrosted and lysed using the standard protocol described in section 3.4.3. The soluble lysate was loaded on a nickel resin IMAC column that was pre-equilibrated with IMAC Buffer A. Following a wash step, the immobilised protein was washed a second time, using a high salt buffer of 1 M NaCl in IMAC Buffer C. This step was intended to remove any contaminating DNA that the Ku proteins might have bound to after the cells were disrupted during sonication. During this second wash period the optical density of the eluent was measured at 260 nm wavelength to ascertain when most of the free DNA had been removed. The column was then re-equilibrated with IMAC Buffer A before IMAC Buffer B was used to liberate Mpa Ku from the nickel resin with 300 mM imidazole. A significantly enriched portion of Mpa Ku was observed in this fraction on a polyacrylamide gel (Figure 3.20, B, lane 2).

The isoelectric point of Mpa Ku (5.58) was estimated using the Expasy web tool Prot Param, and a mono-Q column was selected to further purifiy Ku (Wilkins et al., 1999). The IMAC eluent was diluted 1 in 10 in IEx Buffer A, to reduce the imidazole concentration to prevent it from affecting the anion exchange step. Samples of the flow through taken during the load step showed that most of the Mpa Ku protein did not bind the mono-Q column (Figure 3.20, B, lanes 3 and 4), although some of it did and was eluted by an increasing gradient of NaCl from 50 mM to 600 mM (Figure 3.20, B, lanes 5-8). Interestingly, all samples had a protein product that migrated at roughly twice the size of the Mpa Ku monomer on a polyacrylamide gel. It is likely that these bands represent Ku dimers that have somehow resisted denaturisation by SDS treatment and boiling. Since the vast majority of the Ku protein did not bind the mono-Q column, the flow through was used for further purification.

The final step of the Mpa Ku purification protocol was size exclusion chromatography using a Superdex 200 column pre-equilibrated in SEC Buffer A. The mono-Q flow through was concentrated and loaded into the SEC column. Mpa Ku eluted in two moderate peaks (Figure 3.19, C, lanes 2-4 and lanes 5-8 respectively) most likely as dimer and a possibly a tetramer, since it seems unlikely that Ku would be monomeric in solution. The fractions from the two

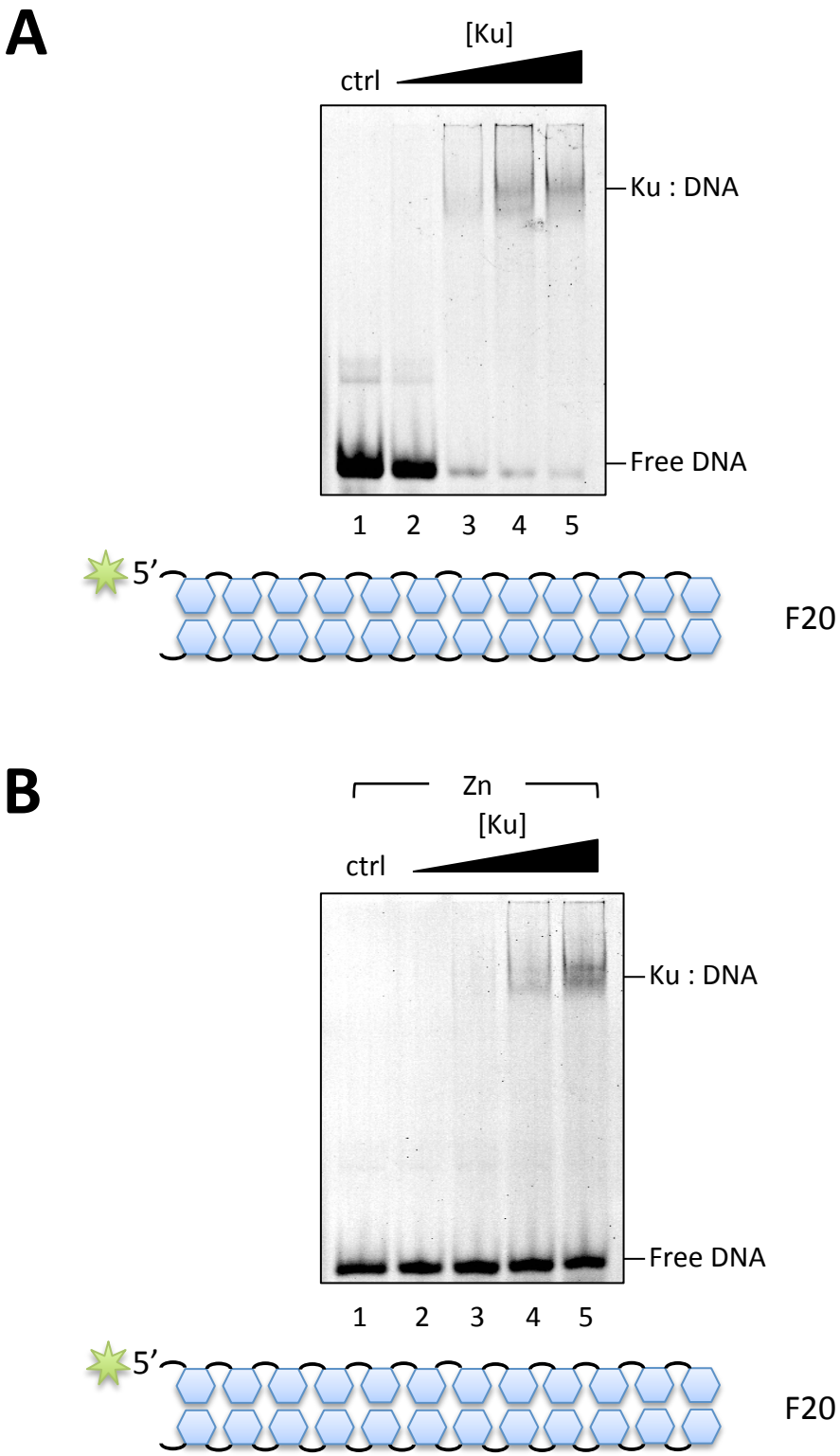


Figure 3.20 Mpa Ku binds dsDNA ends, zinc does not increase the binding affinity. A schematic of the DNA substrate used in the DNA binding EMSAs is shown below the scan of each gel. (**A**) The EMSA reactions contained 60 nM 5’-fluorescein labelled DNA substrate F20 (primer 36-mer, template 36-mer), and 0.2, 0.4, 0.8 or 1.6 µM Mpa Ku. (**B**) The EMSA reactions contained 60 nM 5’-fluorescein labelled DNA substrate F20 (primer 36-mer, template 36-mer), and 0.2, 0.4, 0.8 or 1.6 µM Mpa Ku and 5 mM zinc.

peaks were not pooled, but aliquoted separately and snap frozen in liquid nitrogen for storage at -80 °C. The yield of purified Ku was ~15 mg from a 4 litre culture.

3.9 Mpa Ku is a homodimeric DNA end binding protein

Mpa Ku was incubated with 5'-fluorescein labelled dsDNA to observe any effect on the electrophoretic mobility of the DNA. An increasing titration of Mpa Ku protein concentration was used to retard the migration of the dsDNA through the polyacrylamide gel in native conditions (Figure 3.20, A). A concentration of 400 nM Mpa Ku was sufficient to shift the entire labelled DNA, although this accounts for just 200 nM Mpa Ku dimer. Mpa Ku did not shift single stranded DNA in an assay under the same conditions (data not shown).

Since analysis of the Mpa Ku protein sequence suggested conservation of zinc binding motifs (Figure 3.18, red stars), Mpa Ku was tested for dsDNA binding affinity whilst in the presence of 5 mM zinc. The metal ion did not improve the ability of Mpa Ku to associate with DNA, and instead a decrease in binding efficiency was observed when compared to the metal ion free buffer used in the previously described experiment (Figure 3.20, B). A titration of zinc was also tested to see if less saturating amounts had a different effect on Mpa Ku's DNA binding, but no significant changes were visible (data not shown). Based on these results it is impossible to conclude if the putative zinc binding motifs have any significant effect on the DNA binding activities of Mpa Ku.

3.10 Sequence alignment, expression and purification of Mpa Lig

The functions of the bacterial LigD ligase domain (LigDom) are reasonably well understood and LigDom has been crystallised as an apo structure (Akey et al., 2006). An end joining ligase has not been previously described in the kingdom of archaea, and study of the LigD-like ligase is essential to establish that an NHEJ pathway is active in an archaeal species. Crystallising Mpa Lig bound to nicked DNA was also a target of this research, and a structure could elucidate the motifs that mediate the strict requirements for RNA-DNA nicks. Crystallisation screening of Mpa Lig will be discussed in further detail in Chapter 5.

3.10.1 Mpa Lig protein sequence alignment

The protein sequence of Mpa Lig was aligned with the Mtu, Msm and Pae LigD LigDoms. The seven motifs discussed in section 1.4.4 are well conserved in Mpa Lig (Figure 3.21, noted

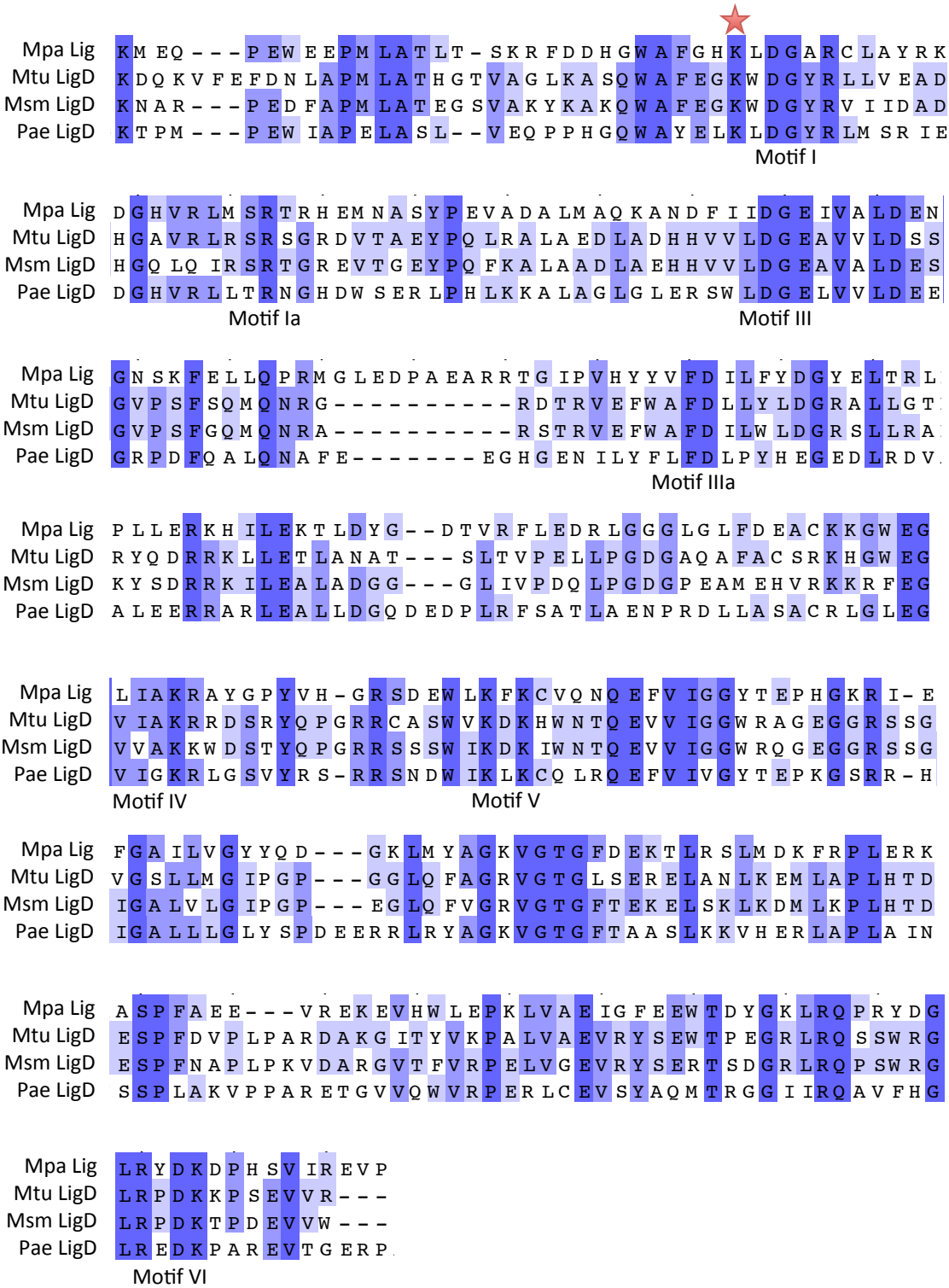


Figure 3.21 Protein sequence alignment of Mpa Lig with Mtu, Msm and Pae LigD ligase domains.

The protein sequence alignments indicate identical residues highlighted in dark blue, and conserved residues in lighter shades of blue. The conserved motifs found in the members of the polynucleotide ligase and RNA capping superfamily are indicated below the relevant position in the sequence alignments. The red star above motif I indicates the lysine residue which forms the lysyl-adenylate intermediate during a ligation reaction.

beneath the sequence). The lysine that becomes adenylated during the ligation reaction is strictly conserved (Figure 3.21, red star). Mpa Lig shares a high sequence identity to the bacterial ligases, and was predicted to retain their key functions.

3.10.2 Mpa Lig expression

Mpa Lig was subjected to expression trials under the same conditions as Mpa PE, Pol and Ku (sections 3.4.2, 3.6.2 and 3.8.2 respectively). Overexpression of Mpa Lig was moderate when compared to the other components of the archaeal NHEJ system, and there was little difference between the expression levels at different induction temperatures (Figure 3.22, A, lanes 2, 4 and 6). An induction temperature of 30 °C was chosen, and a 4 litre culture was grown and induced in Terrific Broth as described previously. After three hours of induction the B834S cells were collected by centrifugation and snap frozen in liquid nitrogen for storage at -80 °C.

3.10.3 Purification of Mpa Lig

Defrosting and lysis of the Mpa Lig pelleted cultures was conducted as standard with the other NHEJ components. The soluble lysate was loaded on a nickel resin IMAC column that was pre-equilibrated in IMAC buffer A. Following a wash step the proteins bound to the nickel resin were eluted by 300 mM Imidazole in IMAC Buffer B (Figure 3.22, B). The eluate was enriched for the ~40.1 kDa band of Mpa Lig (Figure 3.22, B, lane 5).

Mono-Q and mono-S ion exchange columns were selected for the second stage of Mpa Lig purification. The columns were connected together and pre-equilibrated in IEx Buffer A. The IMAC Buffer B elution was diluted 1 in 10 in IEx Buffer A to lower the relative concentration of imidazole. The protein solution was loaded onto the ion exchange columns, and a lengthy peak of high absorbance was observed to flow through, which resolved at the expected molecular weight of Mpa Lig on a polyacrylamide gel stained by coomassie blue (Figure 3.22, B, lane 6). A wash step followed the loading, and then subsequent stepwise increases of NaCl concentration using IEx Buffer B was performed, from 200 mM NaCl to 600 mM NaCl (Figure 3.22, B, lanes 7-9). A large amount of contaminating protein was noted to have bound to the ion exchange columns, along with a small amount of Mpa Lig. The Mpa Lig that had flowed through the columns was therefore separated from the contaminants which bound the Q and S columns.

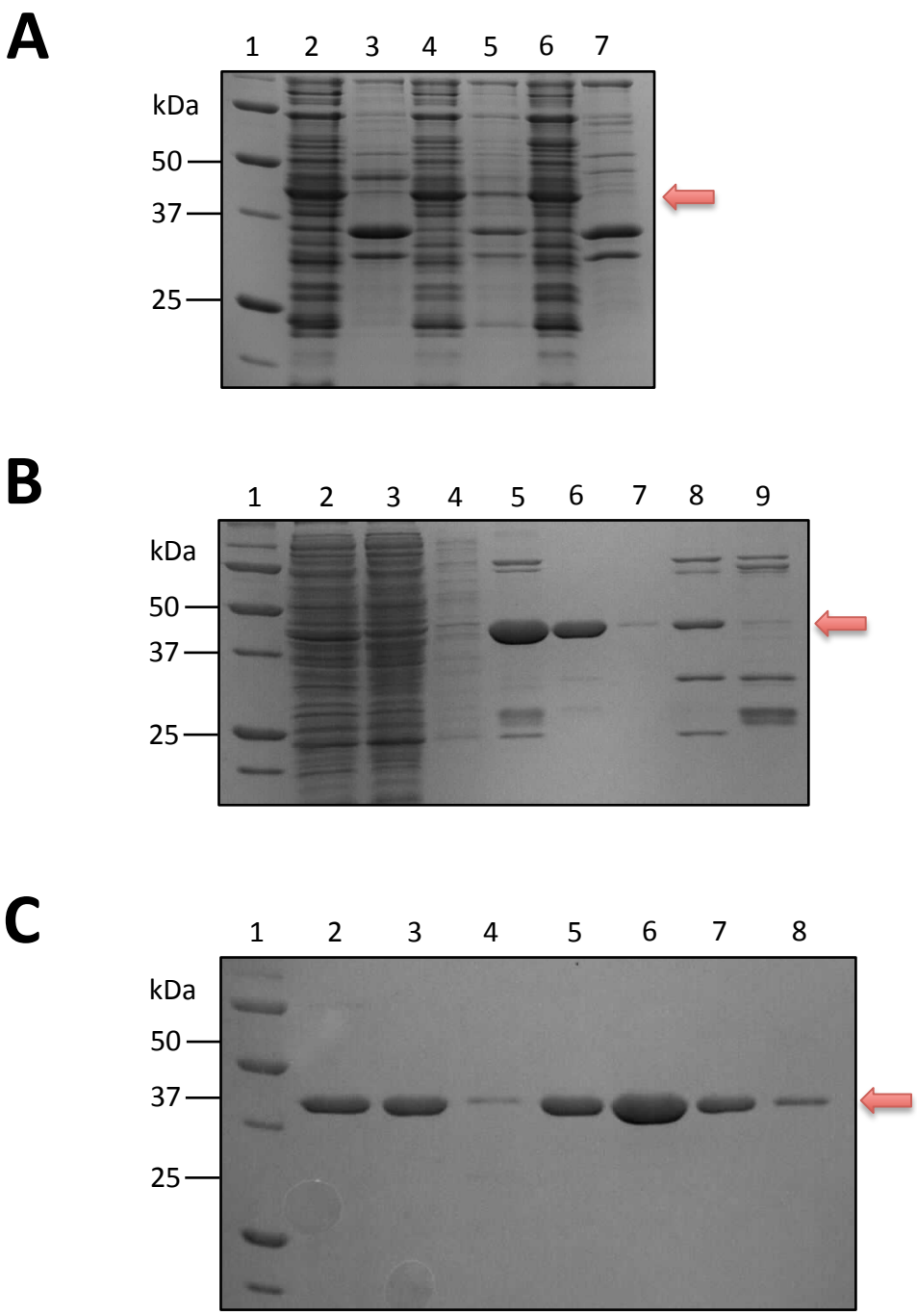


Figure 3.22 Expression and purification of Mpa Lig.
Coomassie stained protein gels. Presence of Mpa Lig protein indicated by the red arrow at ~40.1 kDa. (**A**) Expression of Mpa Lig in B384S cells, induced by 1 mM IPTG. 1= Molecular weight marker, 2= Soluble lysate 20 °C induction, 3= Insoluble pellet 20 °C induction, 4= Soluble lysate 30 °C induction, 5= Insoluble pellet 30 °C induction, 6= Soluble lysate 37 °C induction, 7= Insoluble pellet 37 °C induction. (**B**) Purification of Mpa Lig by IMAC nickel resin column and mono-Q and mono-S columns. 1= Molecular weight marker, 2= IMAC load, 3= IMAC flow through, 4= IMAC wash, 5= IMAC 300 mM imidazole, 6= QS flow through, 7= QS wash, 8= QS 200 mM NaCl, 9= QS 600 mM NaCl. (**C**) Purification of Mpa Ku by size exclusion chromatography. 1= Molecular weight marker, 2= IMAC 300 mM imidazole, 3= QS flow through, 4= Fraction 13, 5= Fraction 14, 6= Fraction 15, 7= Fraction 16, 8= Fraction 17.

The flow through from the mono-Q and mono-S columns was concentrated to ~3 mL for loading on a size exclusion Superdex S200 column that was pre-equilibrated in SEC Buffer A. Mpa Lig eluted in a sharp peak from the size exclusion column and was highly purified (Figure 3.22, C, lanes 5-7). The concentration of Mpa Lig was measured by use of a Nanodrop spectrophotometer, and the overall yield was ~30 mg of protein from a 4 litre TB culture. The protein was aliquoted in small volumes and snap frozen in liquid nitrogen and stored at -80 °C.

3.11 Mpa Lig is an RNA-DNA ligase

3.11.1 Mpa Lig preferentially ligates nicks with a 5'-monoribonucleoside

The initial assay conducted with Mpa Lig defined whether it possessed specific substrate requirements for ligation, like Pae and Atu Ligase D (Zhu & Shuman, 2008; Zhu & Shuman, 2007). Mpa Lig was incubated with one of two 5'-fluorescein labelled nicked DNA substrates; first an all DNA nick, the second with a monoribonucleoside at the 3' side of the nick. Both substrates were tested separately with magnesium and manganese metal ions (Figure 3.23). The results supported the published data on bacterial Ligase D; Mpa Pol was capable of ligating an RNA-DNA nick, but not an all-DNA nick, and did so preferentially in the presence of manganese (Zhu & Shuman, 2008). The necessity for the ribonucleoside at the 3' side of the nick appeared more stringent than the metal ion requirement, since magnesium did elicit a feeble turnover with the RNA-DNA nick, but any ligation of the all-DNA nick was barely discernable, even with manganese. This result confirmed a developing pattern with the Mpa NHEJ enzymes, they are all optimal in the presence of manganese, and they all share a preference for catalysing reactions involving ribonucleotides or ribonucleosides.

Following the substantiation that Mpa Lig requires RNA at the 3' side of the nick, a series of substrates were designed to find the limits of the length of RNA tract that Mpa Lig could interact with. Pae LigD was demonstrated to possess a dwindling ligation activity as the length of 3' RNA tract increased from 1 to just 3 bases (Zhu & Shuman, 2008). Mpa Lig was incubated with substrates with 0, 1, 2, 4 and 6 3'-ribonucleoside tracts and was observed to efficiently ligate all but the DNA only nick (Figure 3.24). The monoribonucleoside was ligated the most efficiently, but surprisingly all of the other RNA substrates were also ligated with reasonable levels of efficiency. This is perhaps unexpected given the inability of Mpa PE and Mpa Pol to interact with or synthesise lengths of RNA more than 2 or 3 bases long, and especially when considering that Pae LigD has rather strict substrate ligation abilities.

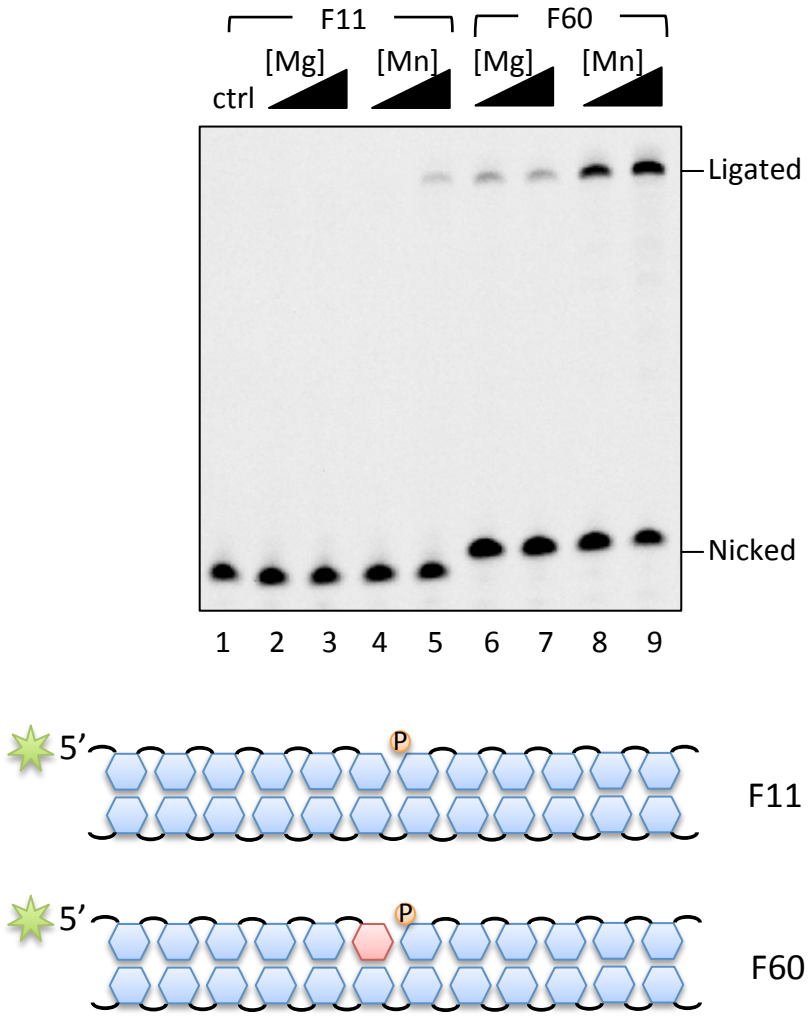


Figure 3.23 Mpa Lig cannot ligate all DNA nicks, but can seal DNA with a 3’-monoribonucleoside.

Schematics of the substrates used are included beneath the scan of the gel. Ligation reactions contained 300 nM Mpa Lig, 30 nM 5’-fluorescein labelled substrates F11 or F60 (both primer 16-mer, template 35-mer, downstream 19-mer) as indicated, and 0.5 or 5 mM magnesium or manganese as indicated. The red base in substrate F60 indicates a ribonucleoside.

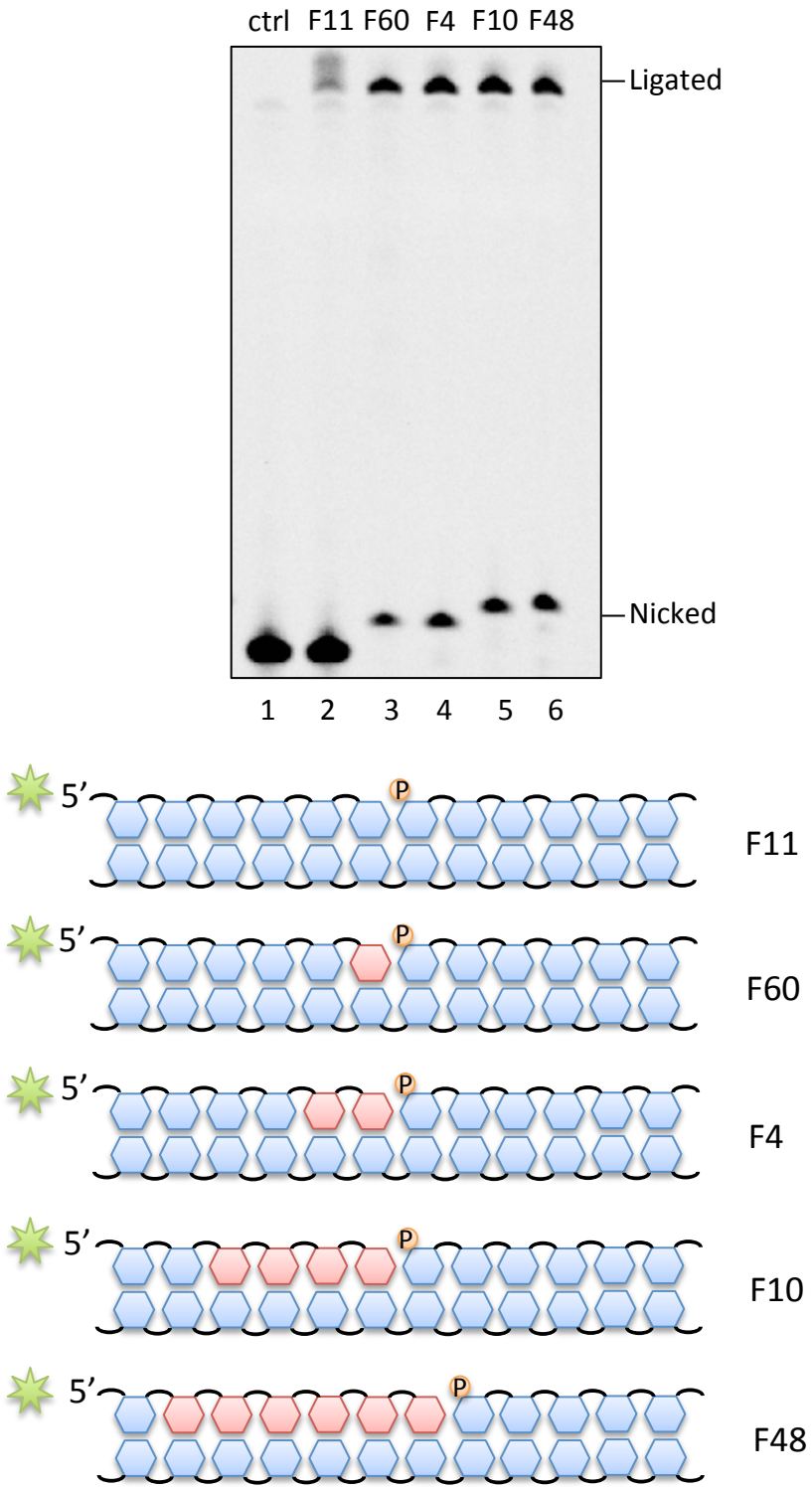


Figure 3.24 Mpa Lig is most efficient at ligating 3'-monoribonucleoside nicks, but can ligate longer tracts of 3'-RNA.

Schematics of the substrates used are included beneath the scan of the gel. Ligation reactions contained 300 nM Mpa Lig, 30 nM 5'-fluorescein labelled substrates F11, F60, F4, F10 or F48 (all primer 16-mer, template 35-mer, downstream 19-mer) as indicated, and 5 mM manganese.

3.11.2 Mpa Lig is pre-adenylated

Pae and Atu LigD have been characterised as being pre-adenylated prior to purification, and as such do not require addition of ATP to ensure proper ligation (Zhu & Shuman, 2007). Mpa Lig was also assayed with a 3'-monoribonucleoside nicked substrate in the presence and absence of ATP, and the ligation profiles were indistinguishable (Figure 3.25, lanes 1-7). Addition of an increased concentration of ATP did not reduce the efficiency of ligation.

The range of ligation substrates of Mpa Lig was tested by two oligomers designed with a 1-nucleotide gap and a 1 base (+1) overhang instead of a nick. Human LigIV can ligate across a gap in the presence of XRCC4, and it was unknown whether bacterial or archaeal DNA ligases could also replicate this remarkable flexibility (Gu, Lu, Tippin, et al., 2007a). Increasing titrations of Mpa Lig were unable to ligate the gap or overhang substrate, suggesting that Mpa Lig may not be able to mobilise and manipulate DNA in the way the human enzyme is capable of.

3.12 Discussion

The primary targets of this chapter were to produce suitable amounts of each protein for structural and biochemical characterisation. Fortunately each of the *M. paludicola* NHEJ proteins expressed well in bacterial cultures, and remained soluble throughout multiple purification steps. A high enough yield of each protein was obtained for crystallographic screening, and more than enough for extensive biochemical assays; between 15 and 70 mg of highly purified protein from 4 litres of culture for each protein. Furthermore, the three enzymatic components of the NHEJ repair pathway were active, and had very similar activites to those established for the bacterial systems from *Mycobacterium*, *Pseudomonas* and *Agrobacterium*.

Archaea occupy an interesting academic niche, often overlooked in favour of their similar microbial relatives bacteria. One reason for the wealth of research in bacteria and dearth of it in archaea is that there are no known pathogenic species amongst the latter domain, whilst the former present many threatening infectious human diseases that research can hope to prevent or cure. Undoubtedly, the relative lack of research in archaea also stems from the fact that they were considered a division of bacteria little more than twenty years ago. Since

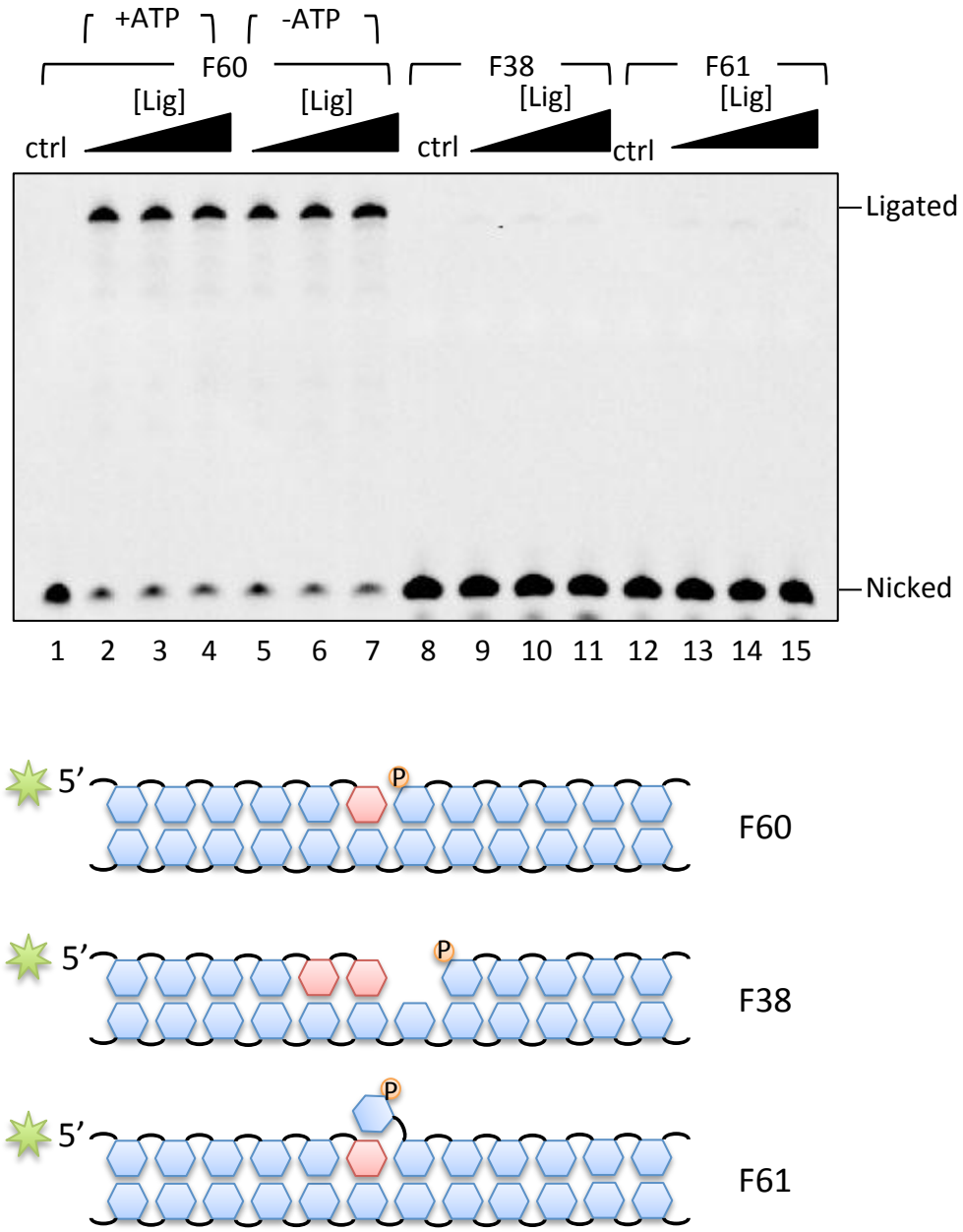


Figure 3.25 Mpa Lig is partially pre-adenylated and cannot ligate across gaps or overhangs.

Schematics of the substrates used are included beneath the scan of the gel. Ligation reactions contained 250, 500 or 1000 nM Mpa Lig, 30 nM 5’-fluorescein labelled substrates F60 (primer 16-mer, template 35-mer, downstream 19-mer), F38 (primer 16-mer, template 35-mer, downstream 18-mer) or F61 (primer 16-mer, template 35-mer, downstream 20-mer) as indicated, and 5 mM manganese. Some of the F60 ligation reactions contained 250 µM ATP where indicated.

Woese introduced the three domain taxonomic system, archaea have gained traction as a valuable tool in understanding aspects of DNA replication and repair that are highly similar to that of eukarya (Woese et al., 1990). From being regarded only as extremophiles that dwell in the harshest environments on the planet to being accepted as an abundant constituent of the biosphere, archaeal research has been quickly gathering pace. A fascinating aspect of studying NHEJ in archaea is that many of the other DNA repair pathways are analogous to eukaryotic mechanisms, whereas the NHEJ system in *M. paludicola* is firmly entrenched in the bacterial method. The evolutionary division of distinct sets of NHEJ proteins is expanded in the next chapter and the conclusion.

The initial characterisation of each component of the *M. paludicola* NHEJ complex has revealed that they share the same pivotal activities as the established bacterial systems. The phosphoesterase shares the same preference for removing one or two ribonucleosides from 3' RNA-DNA hybrids, the polymerase preferentially incorporates NTPs to a DNA primer, the ligase is optimal when ligating an RNA-DNA nick, and Ku binds to dsDNA ends with high affinity. A shared necessity for RNA usage makes sense, if all of the enzymes in the pathway share the same requirements. When viewed disparately, the need for interaction with 2'-OH is not clear. Additionally, the NHEJ system presents itself as one with opposing functions; a polymerase that creates short tracts of RNA at 3' DNA ends, and a phosphoesterase that specialises in removing them. Such a futile cycle cannot be efficient, and yet the Pol and PE are well conserved, and most often found in operonic arrangements in microbial genomes. Only when the enzymes are examined in concert do their roles become coherent, and this is the focus of Chapter 4.

A revealing trait of bacterial NHEJ, and now apparently archaeal NHEJ, is that the three repair enzymes all use manganese as a cofactor. Mpa Lig is almost completely inert in the presence of magnesium, and the polymerase can only extend a DNA primer by a single nucleotide, whilst the phosphoesterase cannot properly catalyse phosphate removal. A dependence on manganese is surprising, since the metal ion is often cited as a cofactor that 'relaxes' the specificity of polymerases and reduces the fidelity of base-templated incorporation, and has been described as a 'non-physiological cation' for human enzymes (Joyce, 1997; Nick McElhinny & Ramsden, 2003). The 'proper' metal ion for DNA or RNA polymerases was considered to be magnesium, which is the most abundant divalent cation in the cell at concentrations of 5-20 mM (Jahnen-Dechent & Ketteler, 2012). These assertions have been challenged recently, with experiments on Hsa Pol μ showing that manganese can elicit

accurate synthesis comparable to magnesium at physiological concentrations (Martin et al., 2013). It is interesting to note that Pol μ actually has a preference for catalysis in the presence of manganese, and that it shares many of the catalytic activities of the bacterial NHEJ Pol. Another human polymerase, Pol ι, a Y family polymerase, also has a strong preference for manganese. Pol ι opts for coordinating a manganese cation even when magnesium is present at a 10 fold concentration (Frank & Woodgate, 2007). Acceptance of Mn^{2+} as a physiologically relevant cation for certain enzymes is growing, although little is known about what benefits these enzymes might reap from evolving activities that are optimal with manganese.

Beyond the specificities for Mn^{2+} and RNA, these experiments show that the *in vitro* activities of the archaeal NHEJ proteins concur with those from the bacterial systems. The limited nucleotide insertions seen from Mpa Pol agree with *in vivo* data of plasmid repair assays, between 1 and 4 four nucleotides added, depending on the DNA ends (Aniukwu et al., 2008). Along with Mpa Pol, the activities of Mpa Lig and PE are so similar to the bacterial LigD domains, that it is clear that the archaeal NHEJ system from *M. paludicola* offers an excellent chance to interrogate the individual enzymatic functions and apply the results to generate a broader understanding of how many microbial NHEJ repair mechanisms work.

Chapter 4

Understanding the orchestration of NHEJ repair in *M. paludicola*

4.1 Introduction

Having established the individual catalytic capabilities of Mpa PE, Pol and Lig, and detailed their substrate specificities and co-factors in the previous chapter, the task remained to delineate how these proteins co-operate. Based on assessment of prior *in vivo* plasmid repair assays on the bacterial NHEJ proteins, a plan of *in vitro* experiments was devised. In the plasmid repair assays catalytic mutation and complete deletion of the PE domain did not cause any significant alterations to repair efficiency or fidelity, and the role of the PE remained ambiguous (Aniukwu et al., 2008). The objective of the research described in this chapter was to discover exactly what benefit the PE offers to the NHEJ mechanism via direct *in vitro* experiments by addition of specific substrates and by contrasting the results in the presence and absence of the partner proteins. Additional objectives were to discover the limits of substrate variety that they might be able to repair. The initial aspiration, following the purification of the complete collection of NHEJ proteins, was to reconstitute NHEJ *in vitro*. This would demonstrate that the Mpa NHEJ system was capable of repairing DSBs, and consolidates the presence of this repair mechanism in archaea. A method was designed to achieve this, using fluorescently labelled DNA substrates rather than DNA plasmids, so that the ends could be configured precisely, allowing for the creation of different microhomologies, with and without ribonucleotides, some with gaps.

Further studies of the LigD and Ku pathway would hopefully lead to new mechanistic insights into the bacterial equivalent, as similarities and differences develop. Comparison of the individual NHEJ proteins of *M. paludicola* to those of the multifunctional LigD of *M. tuberculosis* might suggest something about the benefits of either discrete proteins or a system with all of the proteins fused into a single molecule. One of the larger remaining questions about NHEJ in bacteria, and now archaea, is why is RNA used in preference to DNA to repair DSBs? The necessity to utilise RNA for repair is cryptic. Although there are advantages of NTP usage, it was hoped that further study of the NHEJ enzymes might clarify the issue. Given that evidence is emerging that some eukaryotic NHEJ repair polymerases can make use of NTPs also, answers to this question may have a broader relevance.

4.2 Mpa Ku is necessary for ligation of DNA-RNA ends by Mpa Lig

In order to classify the putative Mpa NHEJ enzymes as *bona fide* DSB repair enzymes, they must be able to reconnect and ligate two discontinuous DNA molecules. In order to achieve this, DNA substrates were designed and synthesised with 5'-fluorescein labels and ends configured for either blunt end-joining, or 5'-overhang end-joining. In order to keep the first assays as simple as possible the oligomers were designed to include a 3'-monoribonucleoside so that theoretically the ligase would be capable of sealing the break, should the ends be brought into proximity. The 5'-overhanging DNA substrate featured a four base region that was self-complementary and could form a microhomology, bringing the 3'-ribo-OH into postion with the 5'-phosphate of the 'downstream' strand (Figure 4.1, A). Initial attempts at encouraging Mpa Lig and Ku to join the DNA ends had limited success after a 3 hour incubation at 37 °C (data not shown). The enzymes were capable of joining the ends, but product levels were barely visible. A large variety of buffer conditions and enzyme concentrations were tested (data not shown), however extending incubation times produced the most reliable results. Overnight incubations at 37 °C yielded reasonable levels of ligated product (Figure 4.1, A). Despite the lengthy incubation, Lig alone was insufficient to catalyse any end joining, but co-incubation with an increasing titration of Ku stimulated ligation. Presumably the DNA end binding protein Ku is essential for bridging the broken ends *in vitro*, and possibly for recruiting the ligase to the ends. Blunt end-joining was less efficient than the reaction with a microhomology region (Figure 4.1, B). This result was expected, since the microhomology provides additional stability, and creates a direct synapse between the ends. Again, Lig alone was unable to ligate the ends, but in the presence of Ku, end-joining was observed. The conclusion from these initial experiments is that Mpa Lig and Mpa Ku can rejoin and ligate nonhomologous ends, and that Ku is essential for the ligation to occur.

4.2.1 Mpa Pol is not sufficient for end bridging in the absence of Mpa Ku

Mtu PolDom is capable of creating a synapse between two DNA ends with a small region of microhomology (Brissett et al., 2007). In order to assess whether Mpa Pol could bridge 5'-overhang ends and stimulate ligation, Mpa Lig was incubated with a 5'-overhang DNA substrate with Pol in the absence and presence of Ku, and with no NTPs (Figure 4.2). Whilst Mpa Lig and Ku were able to elicit ligation of the DNA ends, Lig alone and Lig with Pol produced no visible ligated DNA. These results suggest that *in vitro* Mpa Pol does not play a significant role in bridging short 5'-overhanging DNA ends during ligation reactions, or at the very least, if Pol is bridging the ends it does not improve the proficiency of ligation.

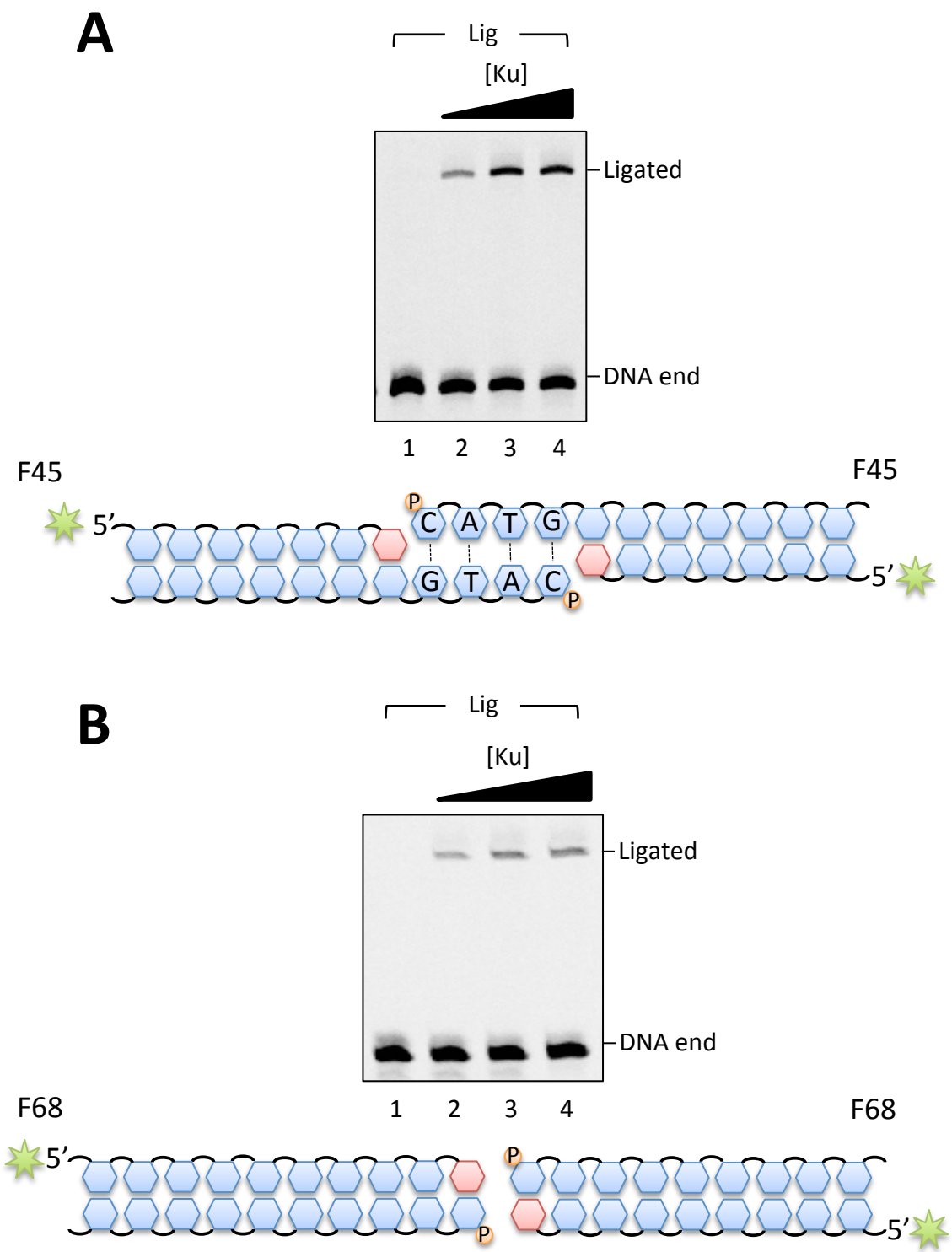


Figure 4.1 Mpa Ku is necessary for the ligation of DNA-RNA ends by Mpa Lig.
Schematics of the DNA substrates used in the 5' overhanging end and blunt end ligations are shown below the scan of each gel. (**A**) 5' overhanging end ligations contained 30 nM 5'-fluorescein labelled DNA substrate F45 (primer 36-mer, template 39-mer) which has four self-complementary bases that can form a microhomology. Ligation reactions also contained 300 nM Mpa Lig and 200, 400 and 800 nM Mpa Ku with 5 mM Mn. (**B**) Blunt end ligations contained 30 nM 5'-fluorescein labelled DNA substrate F68 (primer 37-mer, template 37-mer) and 300 nM Mpa Lig and 200, 400 and 800 nM Mpa Ku with 5 mM Mn.

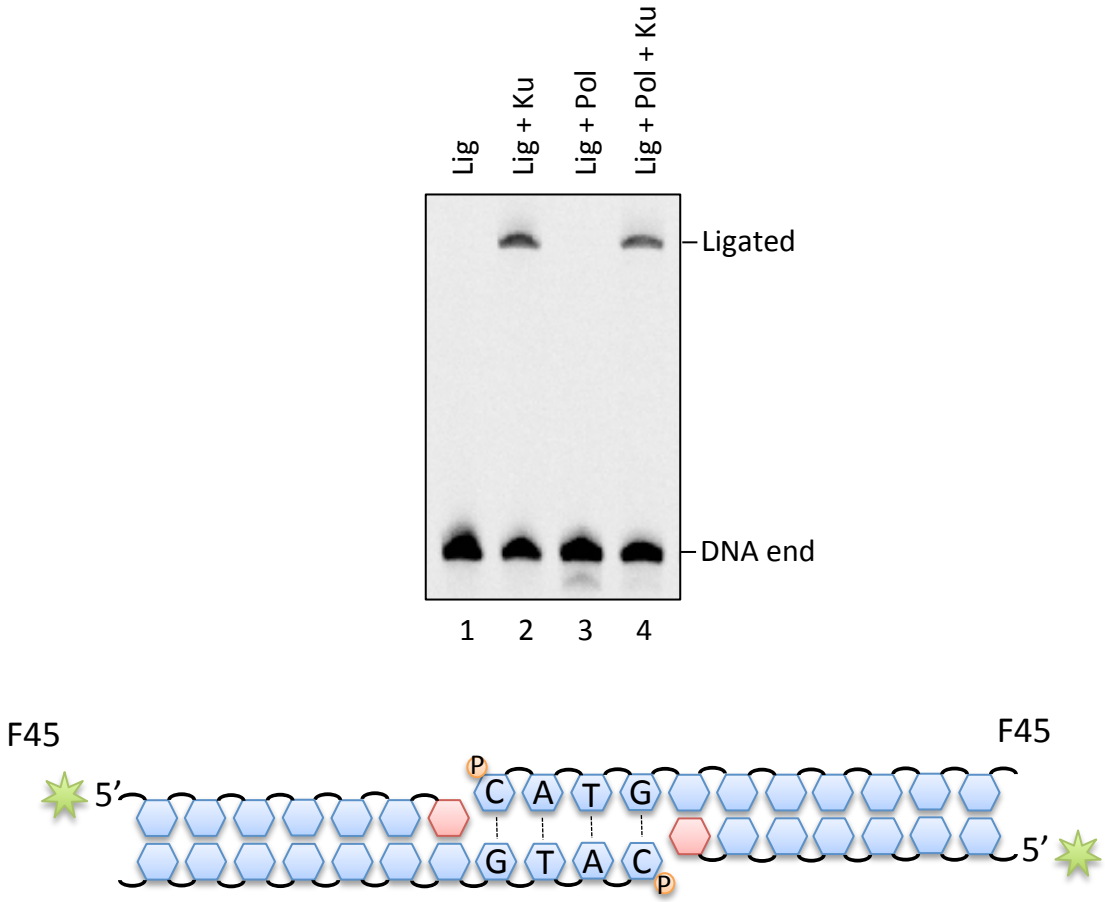


Figure 4.2 Mpa Pol alone is not sufficient to bridge ends for ligation.
Schematics of the DNA substrate used in the 5’ overhanging end ligations is shown below the scan of the gel. 5’ overhanging end ligations contained 30 nM 5’-fluorescein labelled DNA substrate F45 (primer 36-mer, template 39-mer) which has four self-complementary bases that can form a microhomology. Ligation reactions also contained 300 nM Mpa Lig, 300 nM Mpa Pol and 400nM Mpa Ku (where indicated) with 5 mM Mn.

Interestingly, when Mpa Pol and Ku were co-incubated with Lig, the ligation reaction was marginally less efficient than just Mpa Lig and Ku (Figure 4.2, lanes 2 and 4). This may reflect the fact that Mpa Pol and Lig compete for interaction with the 5'-P, and since the polymerase is not required for any synthesis the ligation is slowed down whenever it is occupying the phosphate.

Earlier, results with Mpa Pol indicated that it is able to interact with two discontinuous DNA ends *in vitro*, in the absence of Ku (Figure 3.15, A and B). This implies that whilst the polymerase might be able to bridge the certain end configurations, it may not be directly beneficial for ligation in this situation. This conclusion however does contradict the *in vivo* plasmid repair work on *M. smegmatis,* which found that loss of the polymerase domain was substantially worse for DSB repair than just catalytic inactivity with short 5'-overhang ends (Aniukwu et al., 2008). However, the plasmid repair assay presents difficulties when interpreting results for several reasons. Firstly, the PolDom deletion may have aversely affected the whole fold of LigD since the ligation efficiency is even worse than the LigDom deletion. Secondly, *M. smegmatis* has three additional stand-alone AEP polymerases besides the LigD PolDom (MSMEG_6301, MSMEG_0597 and MSEMG_2105). The activities of these polymerases are not accounted for in these experiments, and may have affected the results. More recent studies have identified that two of these AEPs share similar activities with those of Msm PolDom; template dependent and independent synthesis, and preferential incorporation of ribonucleotides (Zhu et al., 2012). The role of these polymerases in supporting NHEJ or other processes remains unknown.

4.2.2 Mpa Ku stimulates limited mismatch ligation ability in Mpa Lig

Human LigIV and XRCC4 can ligate mismatched ends and across gaps (Gu, Lu, Tippin, et al., 2007a). Results described in the previous chapter showed that Mpa Lig does not possess the abilities to ligate across gaps, or seal a nick with an overhang. In further assays, Mpa Lig was challenged with a variety of substrates with mismatched nicks to assess whether it could ligate them, and whether Mpa Ku could stimulate any additional activity. The substrates had either 1 or 2 nucleoside flaps, on either the 3' or 5' side of the nick and were compared against a properly annealed nick (Figure 4.3). Mpa Lig alone was unable to seal any of the mismatched substrate efficiently, and ligated a very small amount of a nick with a 5' single base mismatch (Figure 4.3, lane 11). Mpa Ku stimulated this ability and created ~25% product compared to a normal nick (Figure 4.3, lane 12). It is interesting to note that Mpa Ku did not stimulate

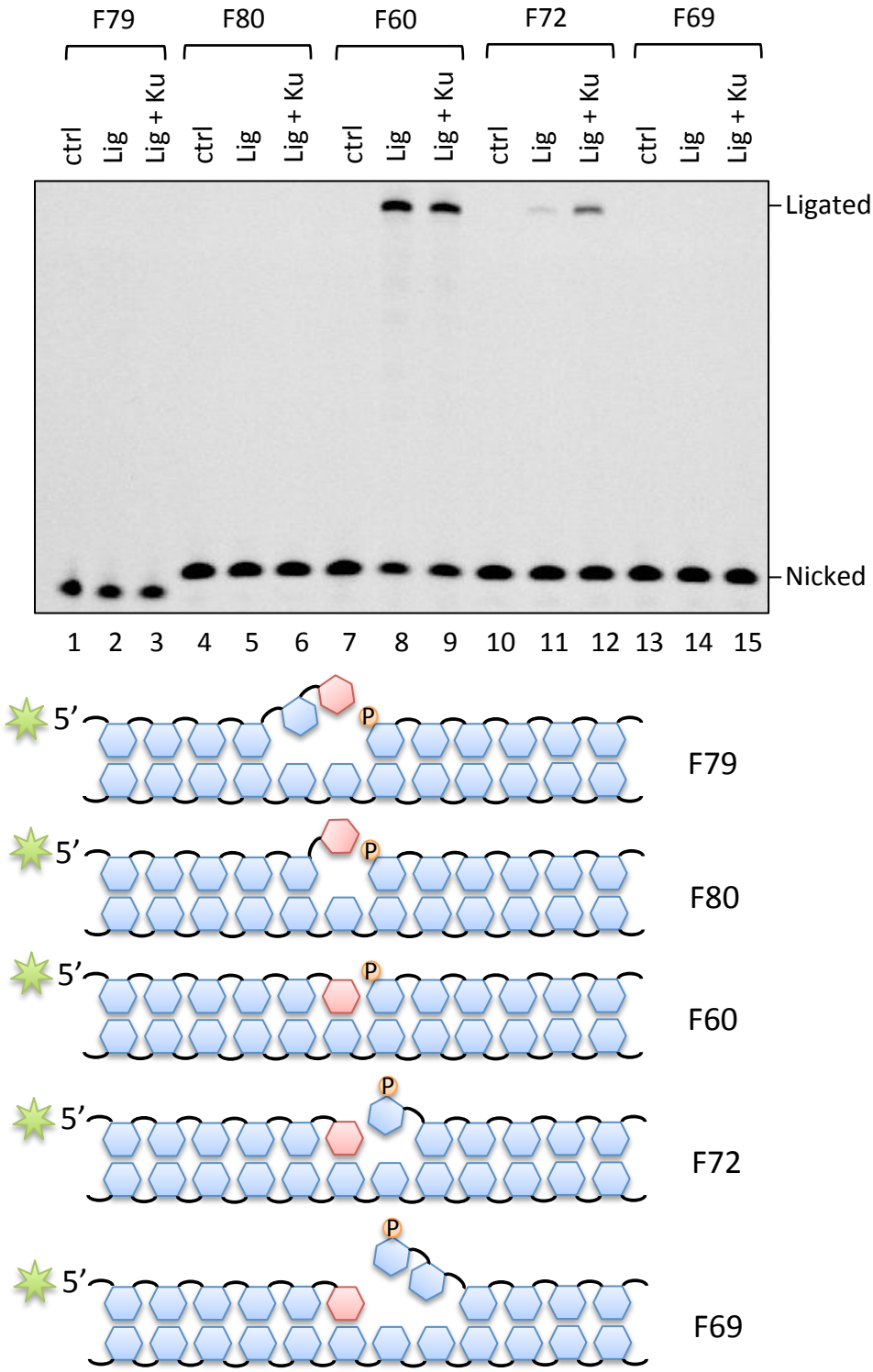


Figure 4.3 Mpa Ku stimulates limited mismatch ligation ability in Mpa Lig.
Schematics of the DNA substrates used in mismatch ligations are shown below the scan of the gel. Mismatch ligations contained 30 nM 5'-fluorescein labelled DNA substrate F79, F80, F60, F72 or F69 (all primer 16-mer, template 35-mer, downstream 19-mer) as indicated. Ligation reactions also contained 300 nM Mpa Lig and 400 nM Mpa Ku (where indicated) with 5 mM Mn.

mismatch ligation for any of the other substrates, suggesting that Mpa Lig is not very flexible in engaging substrates that are not correctly base paired. The data suggest that the 3'-ribo must be correctly base paired, as no mismatches on the 3' end were ligated. The position of the 3'-ribo is likely critical to the ligation reaction, and this agrees with the stringency of Mpa Lig in RNA-DNA nick sealing preferences. It is curious that the addition of Mpa Ku does not greatly stimulate nick ligation, but does benefit sealing of the 5'-mismatch. One interpretation of this result is that Mpa Ku disorts the DNA in a favourable manner allowing the ligase to correctly position itself, despite the mismatch. Mpa Lig and Ku were also tested with gap and overhang DNA substrates as described in the previous chapter, but ligation was not stimulated in either case (data not shown).

4.3 Mpa Pol facilitates end joining via ribonucleotide incorporation

3'-overhang DNA substrates that did not include 3' -ribos were designed to expand the *in vitro* end-joining assays. These substrates required NTP incorporation to enable ligation, and featured a short microhomology region that would terminate in a gap. Mpa Lig alone, or with Pol, was insufficient to ligate two of these substrates together (data not shown) but following addition of Mpa Ku, the ligation reaction was successful (Figure 4.4). Correct incorporation of a single TTP can be observed, recreating the results seen in Figure 3.15, showing again that the polymerase can use a discontinuous DNA end as a template to extend a DNA primer, in this case a 3'-overhanging end. The implication is that a microhomology is formed, and the annealed break forms a double gap substrate, one on each strand at either end of the microhomology region. Mpa Pol then can fill the gap in a template dependent manner and Mpa Lig can seal the remaining nick. This key result determines that the Mpa NHEJ enzymes are capable of joining DNA ends in solution, presumably a difficult task given that there are no structural proteins to stabilise the DNA or limit its movement. Mpa Lig, Pol and Ku can be described as functional repair proteins that can cooperate to rejoin two DNA ends, each essential for the reaction to be completed.

4.4 Mpa PE phosphatase activity enables gap filling and ligation

When considering the role of the PE in context of its partner enzymes, the importance of the phosphomonoesterase activity appeared most obvious. The phosphomonoesterase activity is essentially a 3'-phosphatase activity that removes a phosphate moiety leaving a 3'-OH. In the framework of NHEJ, the benefit of maintaining a 3'-phosphatase is clear; both polymerase and

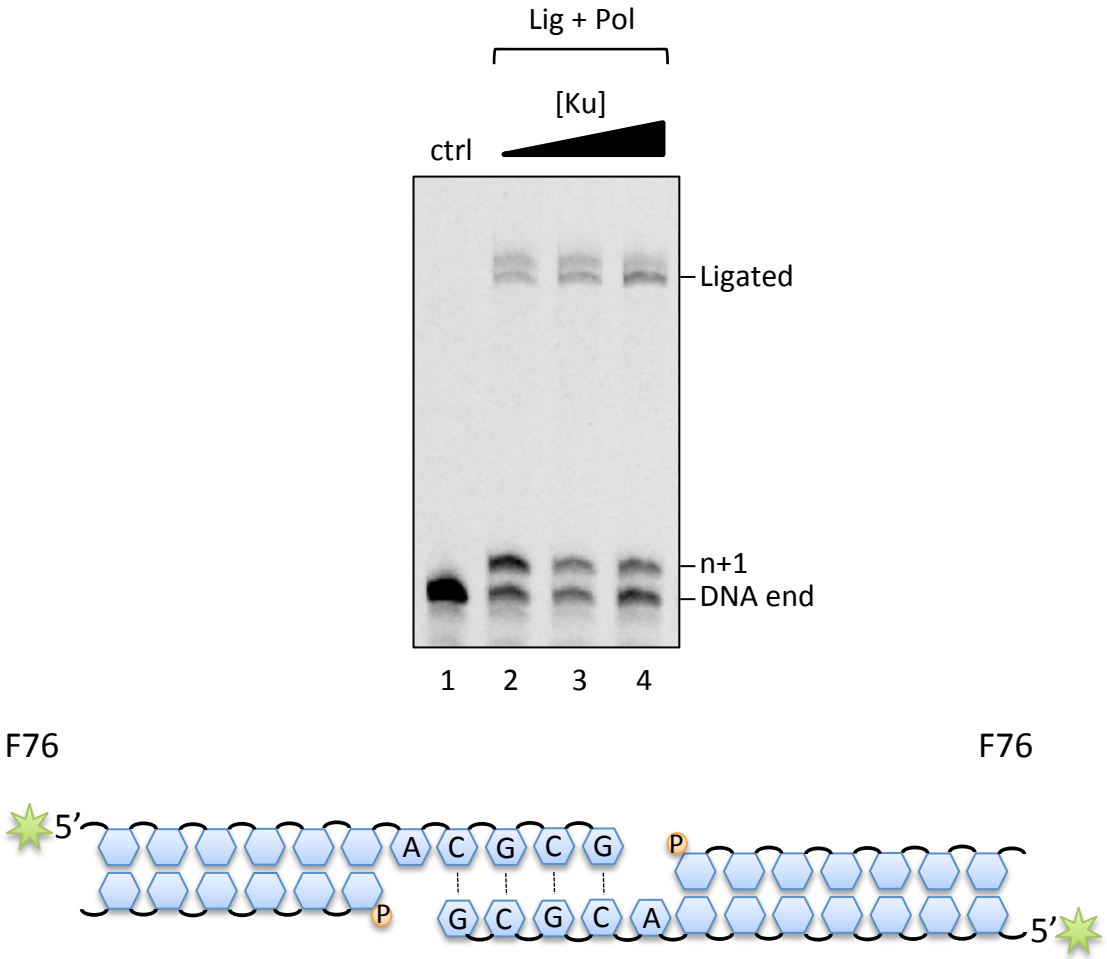


Figure 4.4 Mpa Pol facilitates end joining via ribonucleotide incorporation.
A schematic of the DNA substrate used for 5’ overhanging end microhomology gap filling and ligation is shown below the scan of the gel. End joining ligations contained 30 nM 5’-fluorescein labelled DNA substrate F76 (all primer 42-mer, template 37-mer) as indicated. Ligation reactions also contained 300 nM Mpa Lig, 300 nM Mpa Pol and 200, 400 and 800 nM Mpa Ku with 62.5 µM TTP and 5 mM Mn. The additional band seen above the ligated band is the product of +1 terminal transferase activity by Mpa Pol.

ligase require a 3'-OH for modification. DSBs can produce a variety of damaged ends as well as lost nucleotides, amongst which is the 3'-P. This would prevent a broken end from being extended or ligated unless a 3'-phophatase could repair it. A substrate was designed to test whether the Mpa PE could fulfill this role, featuring a single nucleotide gap and phosphate groups at both 3' and 5' sides. The substrate was intended to mimic a break that had been annealed, and was prepared for gap filling. As expected, Mpa Lig and Pol were unable to modify the substrate (Figure 4.5, lane 2). Mpa PE catalysed removal of the 3'-phosphate, which can be observed by the upwards phosphate shift, and then Mpa Pol filled the gap, stimulating ligation by Mpa Lig (Figure 4.5, lane 3). This shows that the phosphomonoesterase function of Mpa PE is valuable to NHEJ in its own right, and that it probably occurs at an early stage during the DSB repair to allow for ensuing polymerase activity.

4.5 Mpa PE preferentially removes ribonucleosides from dsDNA substrates with D-strands

Mpa PE is capable of removing a terminal ribonucleoside from a di-ribo 3' end. This activity has previously been demonstrated at 3' recessed ends. Given that the phosphodiesterase activity of PE is contingent on the presence of RNA at a DSB, and that Mpa Pol is almost certainly the source of the ribo incorporation, it was reasoned that the PE's ribonuclease function most likely operates during or after Pol has accessed the ends. Bearing in mind that the work on Mpa Pol and prior research in the Doherty lab on Mtu PolDom have suggested that the NHEJ Pol is adept at gap filling after the ends have been brought together, it was reasoned that the PE might be effective on a gap filled substrate. To address this query, Mpa PE was tested with a substrate that was identical the to the di-ribo 3' recessed end, except with a downstream sequence of DNA (D-strand) that converted the substrate into a RNA-DNA nick. Both substrates were compared over a time course with Mpa PE (Figure 4.6, A). Strikingly, Mpa PE was much more efficient at the combined phosphodiesterase activity in the presence of a D-strand than in its absence. The ribonuclease activity was complete after the 15 minutes using the nicked substrate, whereas after the full time course with the recessed end there was still ~25% starting substrate remaining. This stimulation of activity infers that the PE prefers double-stranded DNA to be present downstream of the RNA, which it is attempting to resect. Additionally it confirms the theory that the Mpa PE can operate after the broken ends have been annealed, and that the ribonuclease activity is actually more efficient in this scenario.

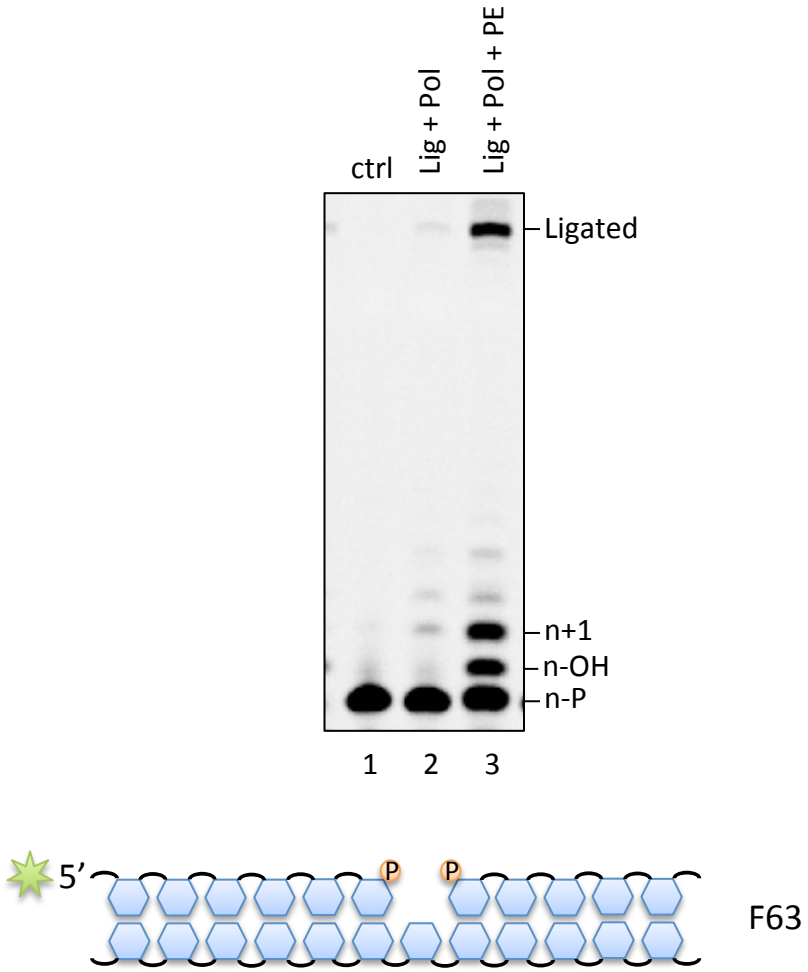


Figure 4.5 Mpa PE 3’-phosphatase activity enables gap filling and ligation.
A schematic of the DNA substrate used for 3’-phosphatase, gap filling and ligation is shown below the scan of the gel. Ligation reactions contained 30 nM 5’-fluorescein labelled DNA substrate F63 (primer 16-mer, template 35-mer, downstream 18-mer) and 300 nM Mpa Lig, 300 nM Mpa Pol and 300 nM Mpa PE (where indicated) with 250 µM NTPs and 5 mM Mn.

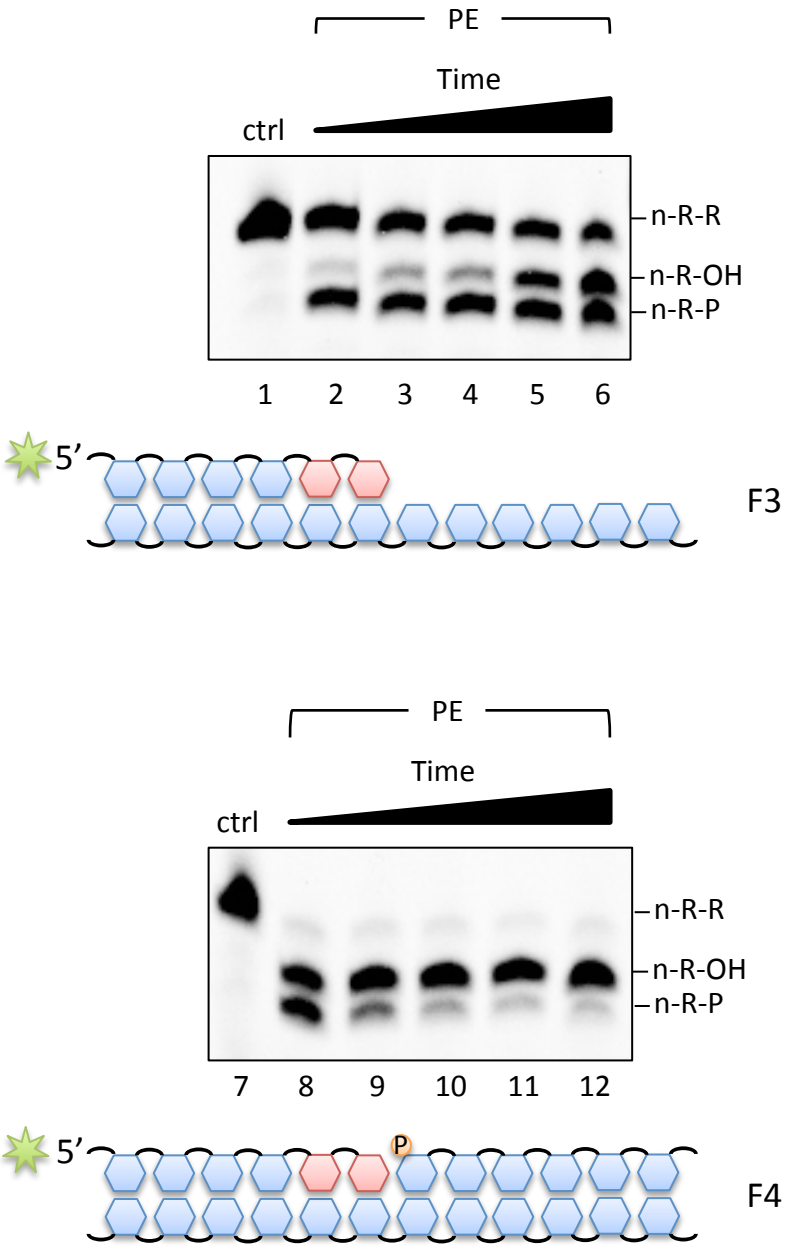


Figure 4.6 Mpa PE preferentially removes ribonucleosides from primers annealed to substrates with downstream DNA.

Schematics of the DNA substrates used for phosphoesterase reactions are shown below the scan each gel. Phosphoesterase reactions contained 30 nM 5’-fluorescein labelled DNA substrates F3 (primer 16-mer, template 35-mer) or F4 (primer 16-mer, template 35-mer, downstream 19-mer) as indicated. Reactions also contained 300 nM Mpa PE and 5 mM Mn and were incubated for 15, 30, 45, 60 and 75 minutes at 37 °C.

These results are not sufficient to conclude which elements of the downstream DNA (D-strand) that the PE recognises or interacts with. However, it is quite likely that this substrate preference has been the thwarting factor in other structural group's attempts to crystallise a PE bound to a DNA substrate (Smith et al., 2011; Natarajan et al., 2012). Following the identification of this preferential substrate of the PE, attempts to obtain X-ray crystallographic structural information were commenced, and they will be detailed in the following chapter. Despite this considerable increase in PE activity in the presence of a D-strand, the reasons for resecting a nicked RNA-DNA substrate that is prepared for ligation is questionable. Furthermore, in a situation of an annealed break with a single nucleotide gap the PE would not have a role to play, since a single incorporation would not offer an eligible substrate for the PE.

4.5.1 Mpa PE forms a complex with DNA-RNA substrates when a D-strand is present

To further determine whether a DNA-RNA substrate with a D-strand is the preferred substrate for PE, they were tested together and observed for stable complex formation via EMSA. It seemed unlikely that Mpa PE would bind stably to the substrate under native conditions if it were catalytically competent; therefore the co-factor manganese was removed from the buffer. Four different substrates were tested; an all DNA recessed end, an all DNA nick, a DNA-RNA recessed end, a RNA-DNA nick. This would assess whether Mpa PE could bind to a DNA-RNA recessed end, or if a D-strand on an all DNA substrate would also help localise the protein. The results showed that Mpa PE only formed a stable complex with DNA-RNA with a D-strand (Figure 4.7, lanes 10-12). This supports the conclusions drawn from the previous experiment, and validates the substrate as a crystallographic target. Since no complex was observed with PE and a DNA-RNA recessed end, it seemed increasingly likely that the PE does not interact with DNA ends during NHEJ but rather with annealed breaks. These assays describe an emerging role for PE in NHEJ for the first time, or at least the coordination of when the activities might occur. Further assays begin to impact on why the PE is drawn to resecting RNA at annealed breaks.

4.5.2 Mpa PE does not have endonucleolytic activity with a RNA-DNA hybrid

Following the finding that Mpa PE is most efficient in the presence of a D-strand when resecting RNA, it was questioned whether Mpa PE might also be able to interact with the same substrate once it was sealed by a ligase. To assess this a DNA-RNA hybrid primer was designed without a nick. Following incubation with Mpa PE over a time course no endonucleolytic cleavage of the substrate was observed (Figure 4.8). This suggests that although the PE can

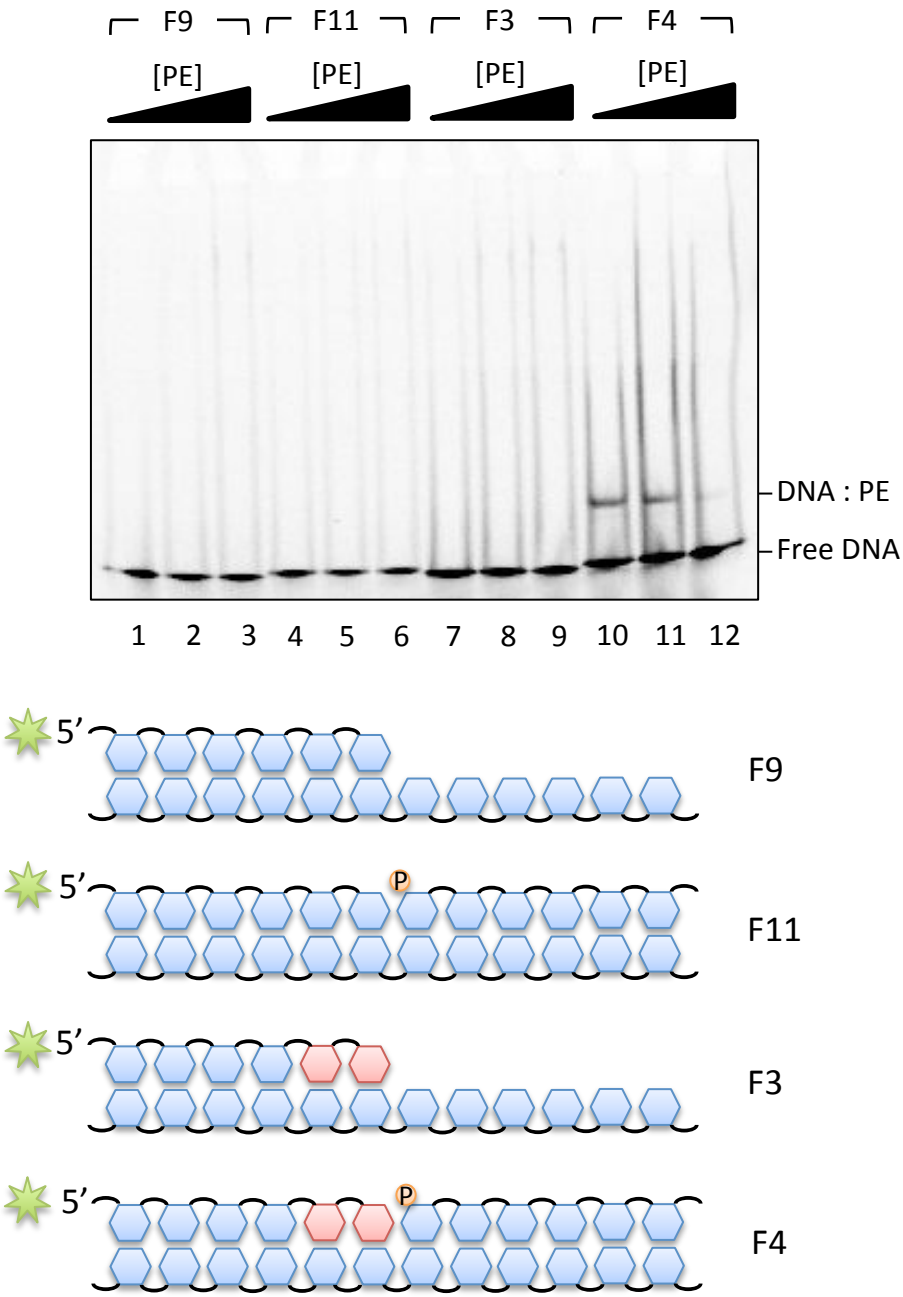


Figure 4.7 Mpa PE forms a complex with DNA-RNA when downstream DNA is present.

Schematics of the DNA substrates used for the DNA binding EMSAs are shown below the scan of each gel. EMSA reactions contained 60 nM 5’-fluorescein labelled DNA substrate F9, F3 (both with primer 16-mer, template 35-mer), F11 or F4 (both with primer 16-mer, template 35-mer, downstream 19-mer) and 200, 400 and 800 nM Mpa PE with no metal ion.

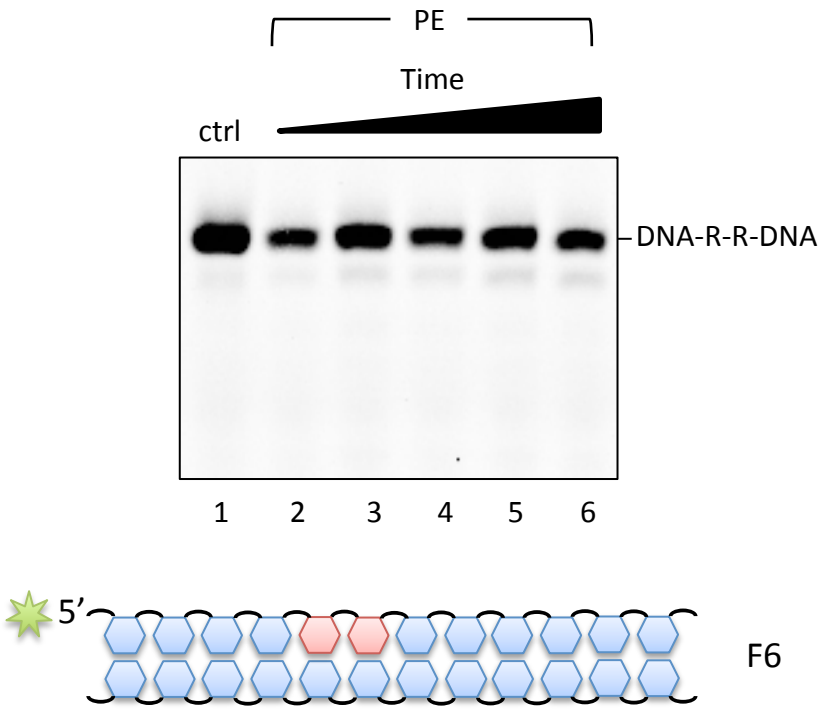


Figure 4.8 Mpa PE does not have endonucleolytic activity with a RNA-DNA hybrid. A schematic of the DNA substrate used for endonuclease reactions is shown below the scan of the gel. Reactions contained 30 nM 5’-fluorescein labelled DNA substrates F6 (primer 35-mer, template 35-mer) and 300 nM Mpa PE and 5 mM Mn and were incubated for 15, 30, 45, 60 and 75 minutes at 37 °C.

recognise a RNA-DNA nick, it does not have the requisite amino acids to catalyse cleavage of an intact phosphodiester backbone, even if RNA is present.

4.5.3 Presence of a D-strand does not significantly impact on the 3'-phosphatase activity of Mpa PE

Before progressing onto other aspects of NHEJ regulation, the implications of PE operating at annealed breaks was considered in context of its phosphatase ability. Figure 4.5 demonstrates that the PE can remove 3'-P from a gap at an annealed break, but it was unclear whether it also benefits from increased efficiency in the presence of a D-strand. PE was assayed with a DNA substrate with a 3'-P and also a D-strand that created a nick conformation. PE was able to remove the phosphate from the primer, and a comparison is shown with 3'-P removal from a primer with no D-strand from Figure 3.6 B (Figure 4.9). PE is slightly more efficient at 3'-P removal in the absence of a D-strand. This is an interesting finding, since the phosphodiesterase activity is greatly increased by the presence of a D-strand. The result suggests that the effect the D-strand has on phosphodiesterase activity does not alter the efficiency of the phosphomonoesterase activity. This separation of substrate specificities may reflect that the phosphatase activity might be necessary at damaged ends, prior to break annealing, whilst the ribonuclease activity is not. This supports the data found in Figure 4.5 that the PE could repair damaged DNA with 3'-phosphatase activity that is unrelated to the ribonuclease activity.

4.6 Mpa Pol can displace a D-strand whilst extending DNA primers

Although the NHEJ polymerase is the most thoroughly characterised and understood of the NHEJ enzymes, Mpa Pol had been observed to produce an unusual and unexpected result whilst gap filling. In the experiment shown in Figure 3.13 Mpa Pol was incubated with a single nucleotide gap substrate, and each NTP individually. This allowed the fidelity of the enzyme to be tested, to see if it incorporated any bases incorrectly. A NTP mix was unnecessary, since the primer was not expected to extend beyond one. However, in lane 2 of Figure 3.13, a +2 addition was observed, suggesting incorporation of two successive CTPs. Although the next downstream template base was dG, it was already base-paired with the D-strand and should not have been accessible to the polymerase. This result was initially considered an artifact, however a repeat of the experiment, this time with an NTP mix was run. Again using a 1-nt gap substrate, Mpa Pol was incubated with NTPs over a time course, and additions of +2 and +3

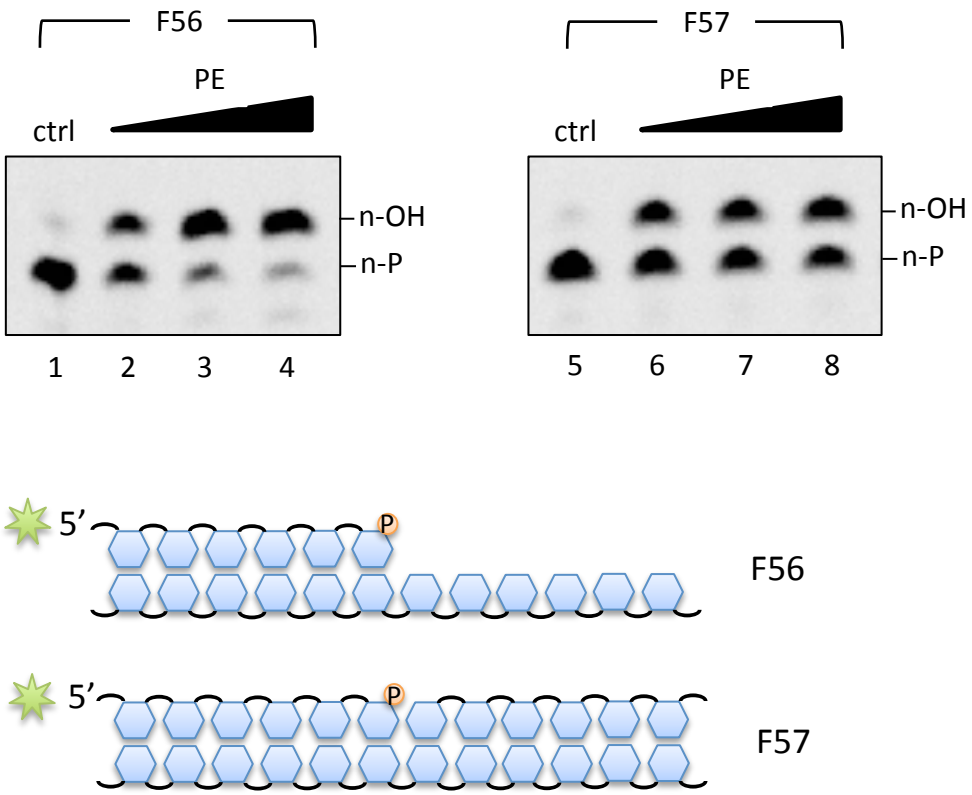


Figure 4.9 Presence of downstream DNA does not make a significant impact on the 3’-phosphatase activity of Mpa PE.

Schematics of the DNA substrates used for phosphomonoesterase reactions are shown below the scan of each gel. Phosphomonoesterase reactions contained 30 nM 5’-fluorescein labelled DNA substrates F56 (primer 16-mer, template 35-mer) or F57 (primer 16-mer, template 35-mer, downstream 19-mer) as indicated. Reactions also contained 250, 500 and 1000 nM Mpa PE and 5 mM Mn.

were observed (Figure 4.10, lanes 1-4). These data suggested that the NHEJ Pol was indeed extending the primer and displacing the D-strand in order to continue. Whilst strand displacement is not a novel activity for a polymerase, it has not been attributed to any of the NHEJ polymerases characterised so far. Further experiments were performed to see under what circumstances Mpa Pol caused strand displacement.

Mpa Pol was challenged with a DNA nick substrate, a substrate that offers no templating base at all, to see how robust the strand displacement activity was. Mpa Pol was able to incorporate the same amount of NTPs as with the 1-nt gap substrate, although the activity was less efficient (Figure 4.10, lanes 5-8). These results strongly corroborated the implication that Mpa Pol retains an innate ability to displace downstream DNA during primer extensions. The reason for Mpa Pol possessing such a capability was not immediately clear, since the obvious consequence of strand displacement was the production of overhang intermediates that the ligase could not seal (as noted before when Mpa Lig was assayed with an overhanging nick substrate in Figure 3.25, lanes 12-15).

Since Mpa Pol has template independent synthesis activities, it remained possible that the strand displacement activity was an addition to the primer end unrelated to the template DNA. The strand displacement first observed in Figure 3.13 appeared to be template dependent, but to stringently test this a DNA nick substrate was used (Figure 4.11). Mpa Pol again incorporated two CTP molecules where two dG templating bases could only be made available after interruption of base-pairing (Figure 4.11, lane 2). This confirms that the strand displacement activity is template dependent, and that Mpa Pol is indeed peeling back the D-strand in order to access the template DNA.

4.6.1 D-strand displacement is conserved in bacterial NHEJ polymerase

Having discovered a new, unexpected ability of the archaeal NHEJ Pol, the question arose as to whether this was a unique feature of this individual enzyme, or an ability possessed by the wider NHEJ Pol family. The polymerase domain of Mtu LigD was used as a comparison to Mpa Pol, and was tested with the 1-nt gap substrate and the DNA nick. Mtu PolDom was found to possess the same ability, apparently displacing the D-strand in order to access the DNA template for primer extension (Figure 4.12). This was true of both the 1-nt gap and DNA nick substrates. The implication of these results is that the NHEJ Pol does maintain an ability to displace downstream DNA, as primer extension is prioritised.

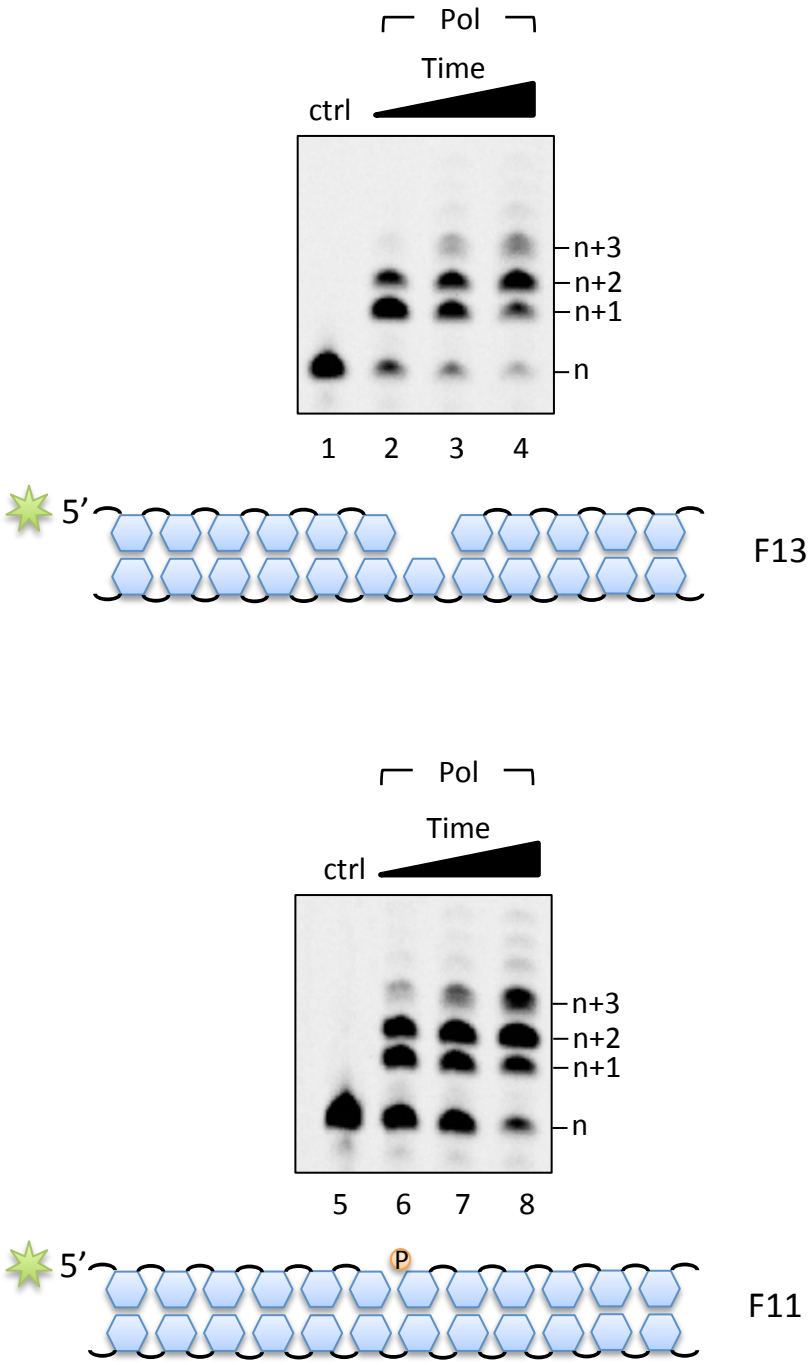


Figure 4.10 Mpa Pol can displace downstream DNA whilst extending DNA primers. Schematics of the DNA substrates used for downstream strand displacement are shown below the scans of each gel. Extension reactions contained 30 nM 5'-fluorescein labelled DNA substrate F13 (primer 16-mer, template 35-mer, downstream 18-mer) or F11 (primer 16-mer, template 35-mer, downstream 19-mer) and 300 nM Mpa Pol with 250 µM NTPs and 5 mM Mn. Reactions were incubated for 30, 60 and 90 minutes at 37 °C.

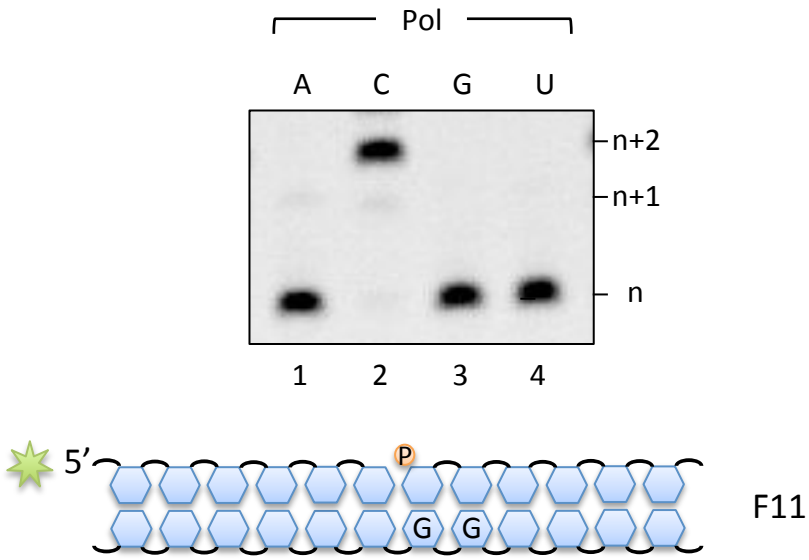


Figure 4.11 Mpa Pol mediated strand displacement at nick substrates is template dependent.

A schematic of the DNA substrate used for strand displacement is shown below the scan of the gel. Extension reactions contained 30 nM 5’-fluorescein labelled DNA substrate F11 (primer 16-mer, template 35-mer, downstream 19-mer) and 300 nM Mpa Pol with 62.5 µM of the indicated NTP and 5 mM Mn.

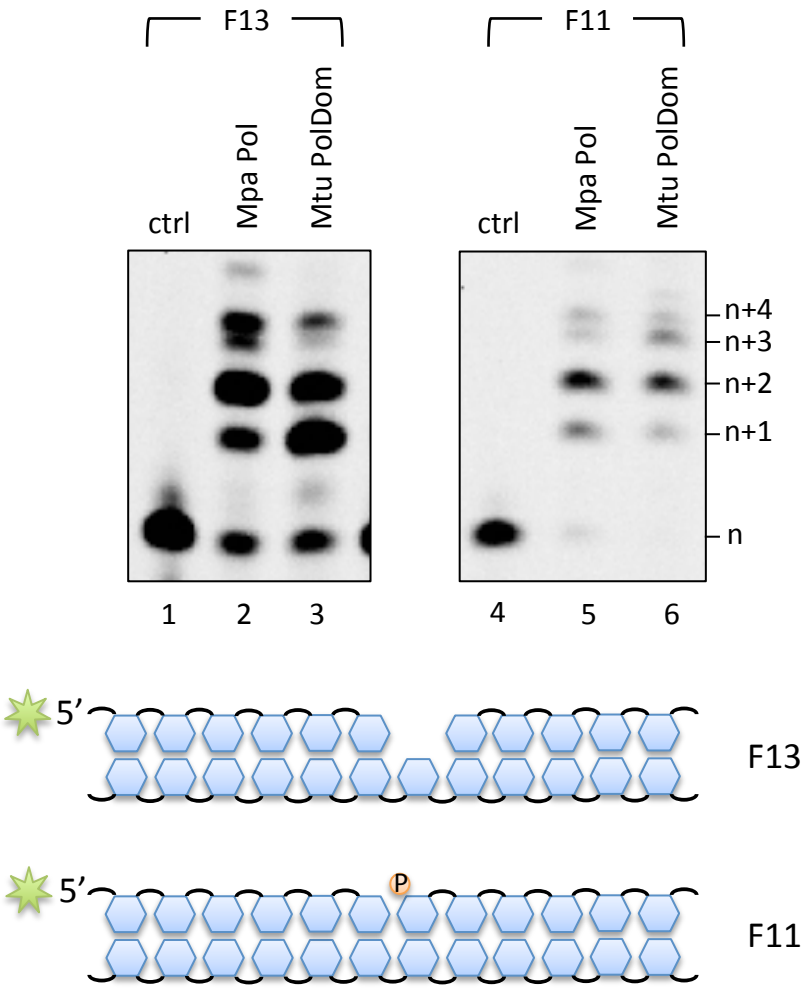


Figure 4.12 DNA strand displacement is conserved in bacterial NHEJ Pols.
Schematics of the DNA substrates used for downstream strand displacement are shown below the scans of each gel. Extension reactions contained 30 nM 5’-fluorescein labelled DNA substrate F13 (primer 16-mer, template 35-mer, downstream 18-mer) or F11 (primer 16-mer, template 35-mer, downstream 19-mer) and 300 nM Mpa Pol or 300 nM Mtu PolDmm with 250 µM NTPs and 5 mM Mn.

As noted above, the reason why microbial NHEJ polymerases might preserve such ability is unclear, not least because of the seemingly genotoxic product that it generates. If a D-strand was displaced by two or three bases, it would leave a 5'-flap of DNA. This would require either a 5'-DNA exonuclease or a 5'-flap endonuclease to resolve. Interestingly during BER in humans Pol β catalyses primer extension and displaces the downstream DNA, and this is followed by removal of the flap by FEN-1 (Imoto et al., 2008). A FEN-1 orthologue exists in both bacteria and archaea, although it is not known whether it plays a role during NHEJ. *M. tuberculosis* does not possess a discrete FEN-1 orthologue, nor do any of the plasmid repair junctions resemble what would be expected from this method of repair (Aniukwu et al., 2008). Given that strand displacement could cause a considerable problem during NHEJ, it seems likely that any enzyme that could solve this problem would be well conserved.

4.7 Mpa Pol and Lig can repair gapped DNA substrates

Although Mpa Pol's ability to fill gaps had been proven, it had not been demonstrated whether these were suitable for ligation by Mpa Lig. In order to test this, three different DNA gapped substrates were designed; 1, 2 and 3-nt gaps. Mpa Pol was able to efficiently fill all, and indeed catalyse D-strand displacement in all three cases (Figure 4.13) Mpa Lig was able to ligate all three substrates following Mpa Pol activity, however only the 1-nt gap substrate was robustly ligated (Figure 4.13, lanes 2, 4 and 6). It is interesting that these results contrast with those discovered for Mpa Lig in the previous chapter, when challenged with pre-made RNA tracts at the 3' side of the nick, in that Mpa Lig could efficiently ligate longer stretches of RNA as well as mono-ribo nicks. It may be that Mpa Pol is able to manipulate the D-strand of the 2 and 3-nt gap substrates more easily than with a 1-nt gap, and therefore reduce ligation efficiency. This would follow the trend of nicked DNA being the least efficient substrate for Mpa Pol to displace as observed in Figure 4.10. It is possible that more room at the point of primer extension might allow the polymerase a greater ability to control the position of the D-strand.

Another unanticipated result was observed during this experiment; the generation of two different ligation products (Figure 4.13, lanes 2, 4, and 6). The additional ligation product was observed using all three gap substrates, and a repeat of the assay with a size marker indicated that it was larger than the expected product (data not shown). The full-length ligation product was expected at 35 bases, and the 35 base size marker migrated next to the smaller of the two ligation products. This raised a number of possibilities, one of which was that the polymerase

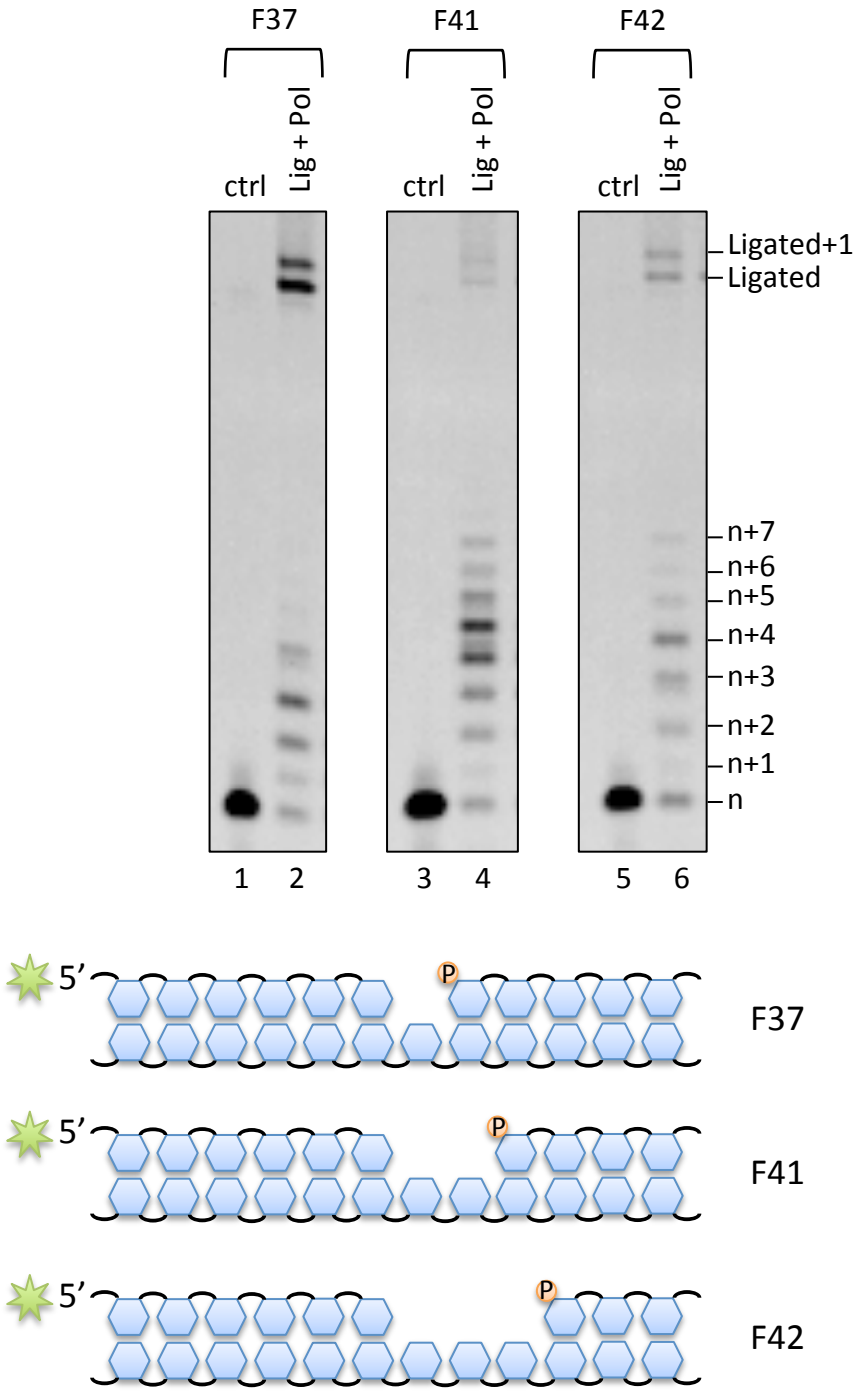


Figure 4.13 Mpa Pol fills gaps and promotes ligation.

Schematics of the DNA substrates used for gap filling and ligation are shown below the scans of the gels. Gap filling and ligation reactions contained 30 nM 5'-fluorescein labelled DNA substrate F37 (all primer 16-mer, template 35-mer, downstream 18-mer), F41 (all primer 16-mer, template 35-mer, downstream 17-mer) or F42 (all primer 16-mer, template 35-mer, downstream 16-mer) as indicated. Reactions also contained 300 nM Mpa Lig and 300 nM Mpa Pol with 250 µM NTPs and 5 mM Mn.

was enabling the ligase to seal the displaced D-strand to the nascent DNA-RNA primer. However another more likely possibility occurred, that Mpa Pol could be catalysing +1 extension at the 3' end of the substrate via template independent synthesis. This would not be observed in a normal gap-filling assay, since the D-strand is not joined to the primer, and any 3' extension would go unnoticed. However when the gap is filled and ligated, the D-strand becomes part of the fluorescently labelled primer, and any 3' additions would increase the size of the strand. In order to discover if this was the case a new 1-nt gap substrate was designed with a 3'-dideoxynucleotide at the unlabelled 3' end. Polymerases required a 3'-OH for extension, and in their absence, they cannot incorporate nucleotides. The assay produced the expected result, the 3'-dideoxy ablated the template independent synthesis, and only one ligation product was observed across a time course (Figure 4.14). D-strand displacement remained unchanged.

4.8 Mpa PE resects RNA from substrates with a displaced D-strand

In light of the recent discovery that Mpa PE prefers to resect ribonucleosides from annealed breaks, it became a potential candidate for being able to resolve the strand displacement issues of the NHEJ polymerase. A variety of DNA-RNA substrates were designed to assess the limits of Mpa PE's resecting activities, and to see whether it could resect RNA from an oligomer with a displaced D-strand. The substrates tested were; a +2 overhang (D-strand displaced), a +1 overhang, a RNA-DNA nick, a 1-nt gap, and a 2-nt gap. Mpa PE was able to resect all five substrates to varying degrees. The PE was most efficient at cleaving ribo and phosphate from the DNA-RNA nick substrate (Figure 4.15). Remarkably, the substrate that PE was next most efficient at ribonuclease cleavage of was the +1 overhang, rather than the 1-nt gap (Figure 4.15, lanes 4-6 and 10-12). Both were similar in profile, but the +1 overhang had more ribo and phosphate removal in the same time point. The contrast between +2 overhang and 2-nt gap was even more distinct (Figure 4.15, lanes 1-3 and 13-15). The +2 overhang substrate had ribo cleavage and phosphate removal only slightly slower than the 1-nt gap, but the 2-nt gap had distinctly slow ribo removal and minimal phosphate scission.

These experiments revealed two key findings; firstly, that the distance between the D-strand and RNA tract is critical for PE activity and, secondly, that the PE is capable of resecting RNA from substrates in which the D-strand has been displaced. The earlier revelation that Mpa PE is most active when a D-strand is present (Figure 4.6) raised queries as to what feature of the new substrate it was that the PE preferred. The finding that Mpa PE is significantly less

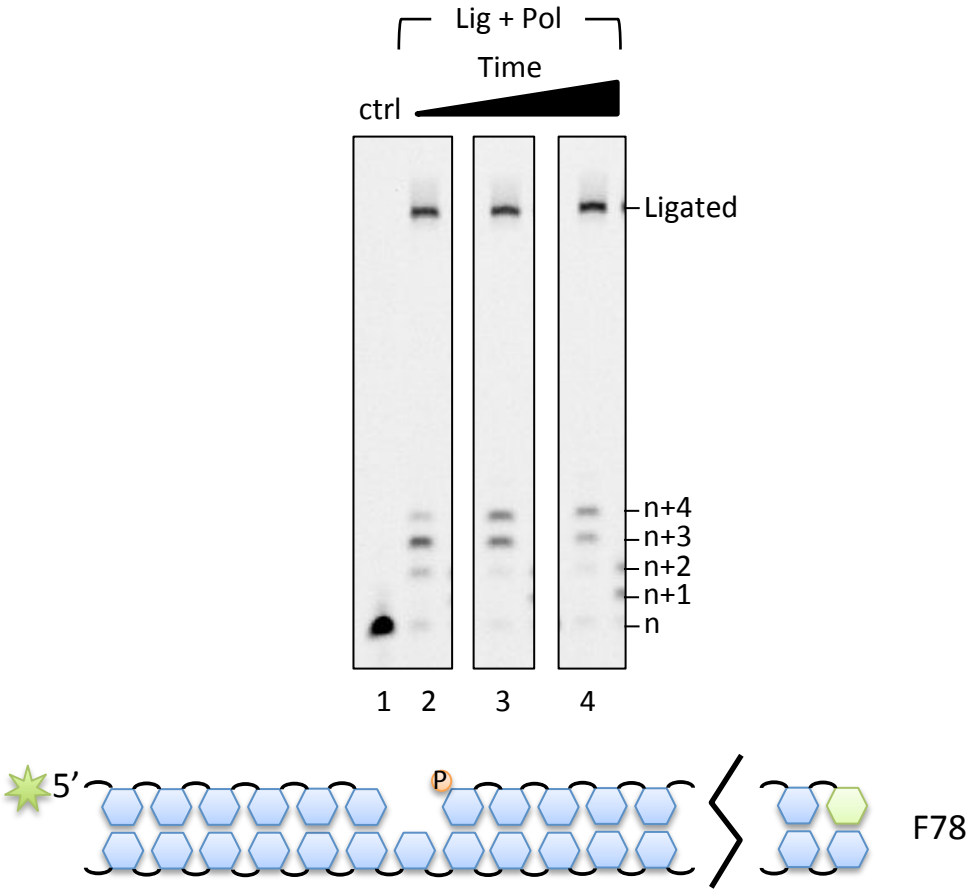


Figure 4.14 A 3'-dideoxynucleotide prevents Mpa Pol template independent synthesis.

A Schematic of the DNA substrate used for gap filling and ligation is shown below the scan of the gel. Gap filling and ligation reactions contained 30 nM 5'-fluorescein labelled DNA substrate F78 (all primer 16-mer, template 35-mer, downstream 18-mer) with a terminal 3' dideoxynucleotide on the D-strand (indicated in green), 300 nM Mpa Lig and 300 nM Mpa Pol with 250 µM NTPs and 5 mM Mn. Reactions were incubated for 30, 60 and 90 minutes at 37 °C.

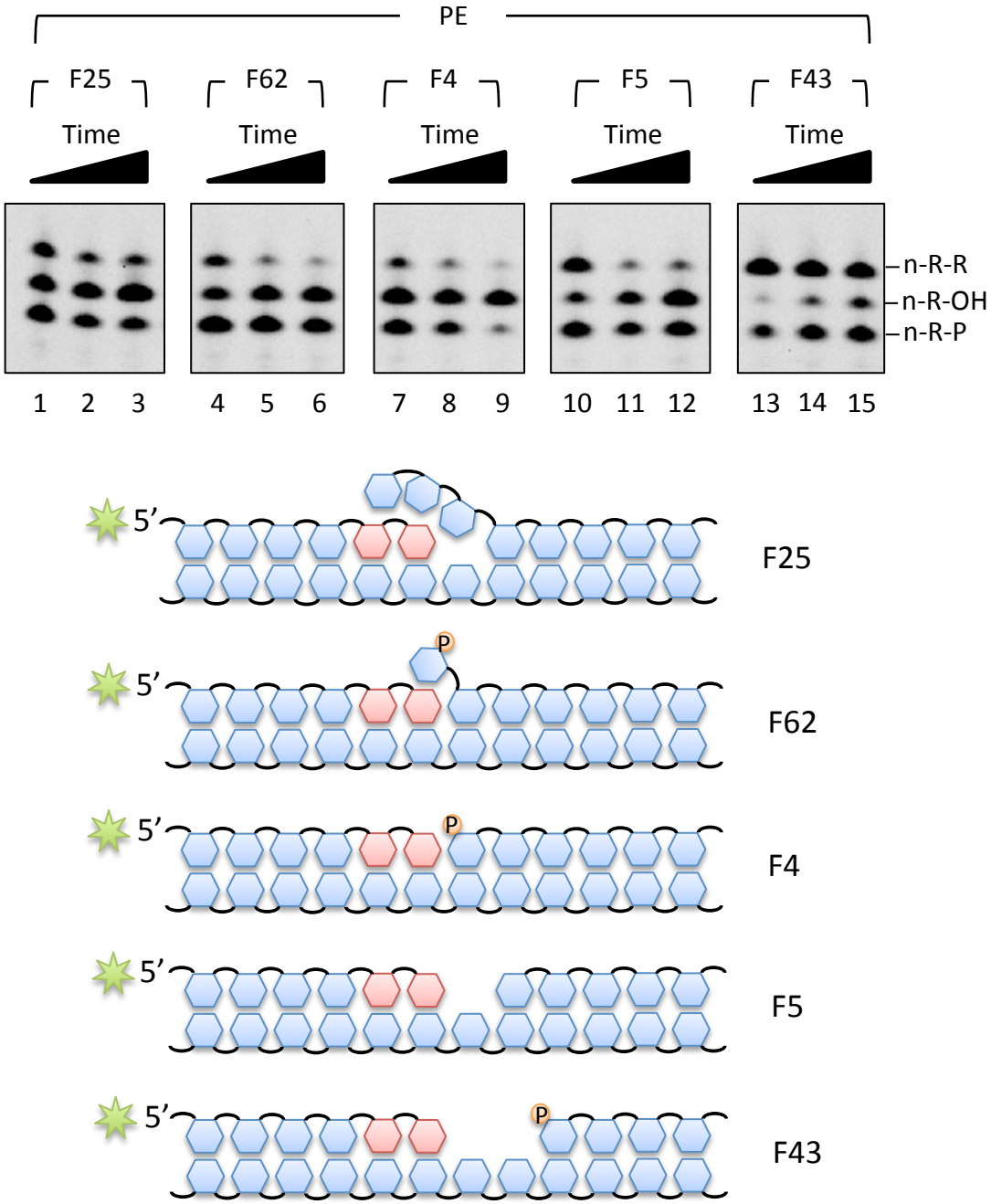


Figure 4.15 Mpa PE resects RNA from substrates with displaced downstream DNA. Schematics of the DNA substrates used for phosphoesterase reactions are shown below the scans of the gels. Phosphoesterase reactions contained 30 nM 5'-fluorescein labelled DNA substrate F25 (primer 16-mer, template 35-mer, downstream 21-mer), F62 (primer 16-mer, template 35-mer, downstream 20-mer), F4 (primer 16-mer, template 35-mer, downstream 19-mer), F5 (primer 16-mer, template 35-mer, downstream 18-mer) or F43 (primer 16-mer, template 35-mer, downstream 17-mer). Reactions also contained 300 nM Mpa PE and 5 mM Mn. Reactions were incubated for 20, 40 and 60 minutes at 37 °C.

efficient at ribo removal from a 2-nt gap substrate suggests that the proximity of the 5' end of the D-strand to the RNA is highly important in terms of PE activity. Mpa PE may physically interact with the D-strand that allows it to become properly positioned with respect to the RNA at the 3' side of the nick/gap. Whilst the presence of the D-strand is not essential, the improvement of catalysis that it stimulates would ensure that the PE would be active at annealed breaks, rather than interacting with the broken ends immediately. A curious addition to PE's affinity for the D-strand is that even when the strand is displaced and rendered as a flap over the RNA, PE can still efficiently remove ribonucleosides and phosphates. This result strongly implicated the PE as the nuclease that could resolve displaced D-strands to enable ligation.

4.8.1 Mpa PE can resect RNA from a displaced D-strand substrate to enable ligation by Mpa Lig

The initial step of investigating the role of Mpa PE in resolving strand displacement was to reconstitute a substrate with a displaced D-strand with a +1 overhang. The substrate was designed to replicate a filled 1-nt gap, with an additional ribonucleoside displacing the first base of the D-strand. Mpa Lig could not ligate this substrate alone (Figure 4.16), repeating the result observed in Figure 3.25 with a near identical substrate except with a 3'-monoribonucleoside. However upon addition of Mpa PE to the reaction mix, the 3' terminal ribonucleoside and phosphate were removed and ligated product was observed (Figure 4.16, lane 3). This was the expected result, and confirms that Mpa PE and Mpa Lig can cooperate to repair a substrate with a displaced D-strand. Combined with the previous results, Mpa PE emerged as a viable regulator of Mpa Pol mediated strand displacement.

4.9 Mpa PE regulates NTP incorporation by Mpa Pol

After exploring the ability of Mpa PE to resect RNA that is displacing D-strand DNA, the obvious course of action was to measure if Mpa PE could directly regulate the generation of D-strand intermediates. By incubating reactions with a set concentration of Mpa PE against an increasing titration of Mpa Pol, the effects of Mpa PE on RNA synthesis could be easily identified. This assay was run with two DNA substrates; a primer without D-strand, and a 1-nt gap. In the assay using the substrate without a D-strand, Mpa PE demonstrated negligible effects on Mpa Pol's RNA synthesis (Figure 4.17, A). When reaching a saturating concentration of Mpa Pol that was over four times the concentration of Mpa PE (Figure 4.17, A, lane 7), an

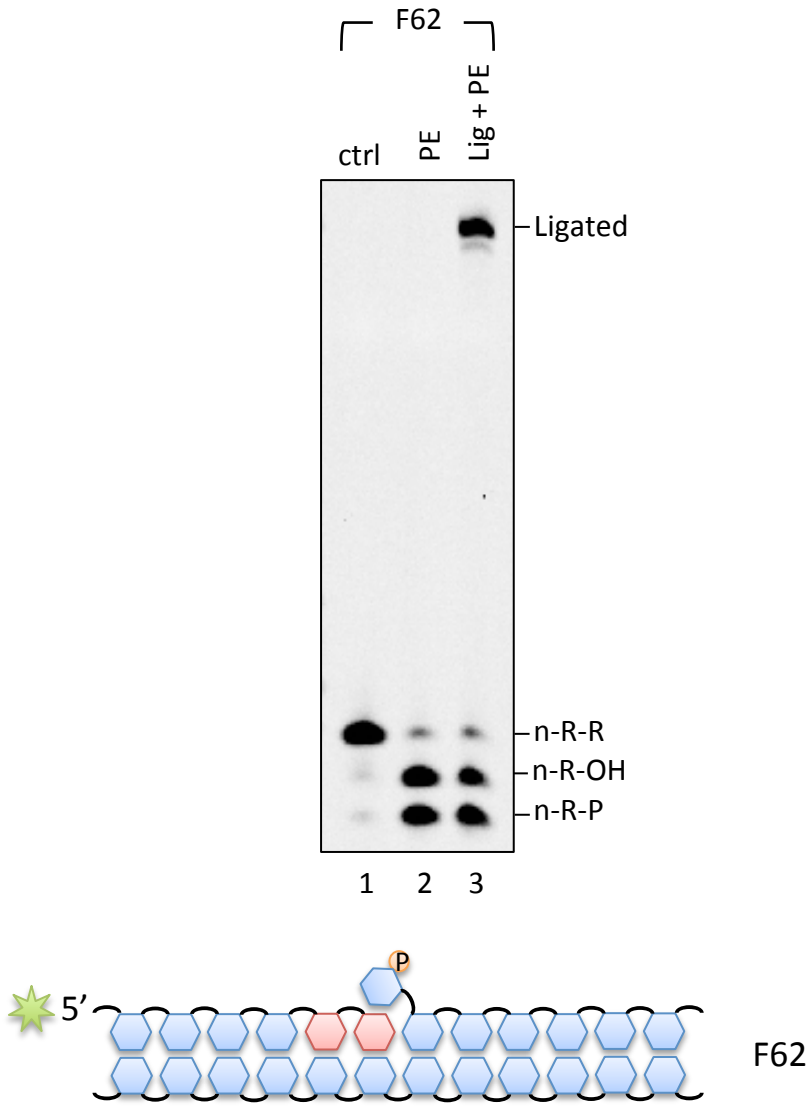


Figure 4.16 Mpa PE can resect RNA that is displacing downstream DNA to enable ligation.

A schematic of the DNA substrate used for phosphoesterase and ligation reactions is shown below the scan of the gel. Ligation reactions contained 30 nM 5’-fluorescein labelled DNA substrate F62 (primer 16-mer, template 35-mer, downstream 20-mer), 300 nM Mpa Lig, and 300 nM Mpa PE (where indicated) with 5 mM Mn.

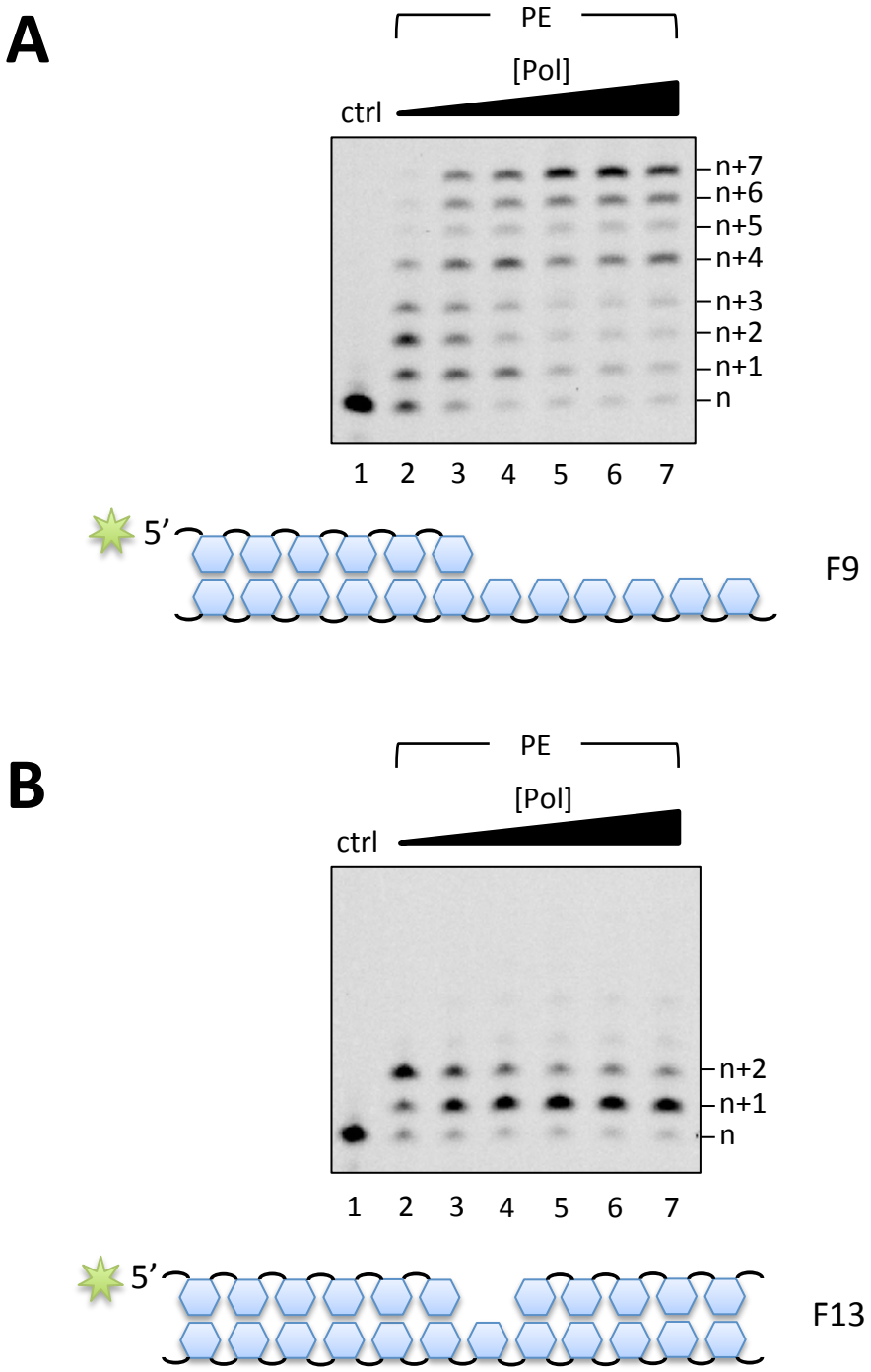


Figure 4.17 Mpa PE regulates NTP incorporation by Mpa Pol at annealed breaks. Schematics of the DNA substrates used for primer extension and phosphoesterase are shown below the scans of each gel. (**A, B**) Reactions contained 30 nM 5’-fluorescein labelled DNA substrate F9 (primer 16-mer, template 35-mer) or F13 (primer 16-mer, template 35-mer, downstream 18-mer). Reactions also contained 300 nM Mpa PE and 100, 200, 400, 800, 1000 and 1400 nM Mpa Pol with 250 µM NTPs and 5 mM Mn.

increasing DNA ladder was observed, with a typical maximum RNA tract of seven bases. Only at the lowest concentration of Mpa Pol was Mpa PE able to influence the length of RNA tract, with a maximum incorporation of 3 – 4 ribonucleotides (Figure 4.17, A, lane 2). When the enzymes reached equimolar concentrations, Mpa Pol was able to extend much of the template beyond four bases (Figure 4.17, A, lanes 4-7).

When both enzymes were incubated with a 1-nt gap substrate, the contrast between results was striking. Mpa PE was able to regulate Mpa Pol activity preventing extension beyond two ribonucleotides (Figure 4.17, B). Mpa PE reduced the activity of Mpa Pol, with the majority product remaining at +1 extension, even when Mpa Pol was at saturating concentrations (Figure 4.17, B, lanes 6 and 7). From equimolar concentrations of PE and Pol to an excess of Pol, the products of the reaction did not change. Mpa PE encouraged a maximum extension of +1, which filled the gap and generated a nick that Mpa Lig would be able to seal. It is notable that Mpa Pol extension on a substrate with a D-strand is reduced compared to a free template, and that Mpa PE is most efficient in the presence of downstream DNA. It is most likely these factors that combine to allow the enzymes to construct a suitable product for ligation. These data clearly demonstrate that Mpa PE has a strong regulatory effect on the activity of Mpa Pol at an annealed break, and convincingly suggest that the PE could prevent strand displacement intermediates and promote ligation of breaks.

4.10 Mpa PE optimises Mpa Pol products for ligation

In order to validate the theories on the role of Mpa PE at annealed breaks, Mpa Lig had to be included along with the PE and Pol. A 1-nt gap substrate was used (with a 3'-dideoxynucleoside to prevent any template independent addition by Mpa Pol) and pre-incubated with Mpa Pol for varying amounts of time to generate strand displacements. Then Mpa PE and Mpa Lig or Mpa Lig alone were added to the reactions, and variations in ligation efficiencies were compared with each other and a sample that was not pre-incubated with Mpa Pol. The Lig + PE sample compared to Lig only produced the same pair of extension profiles across the time course, regardless of the pre-incubation period with Mpa Pol (Figure 4.18). In the absence of Mpa PE a +4 extension was observed, in the presence of Mpa PE extension was limited to +2. This suggests two things; firstly that Mpa Pol is able to displace a D-strand by at least 3 bases under suitable conditions, and secondly that Mpa PE can efficiently reduce a +4 extension with a displaced D-strand to a +1 extension that can be ligated.

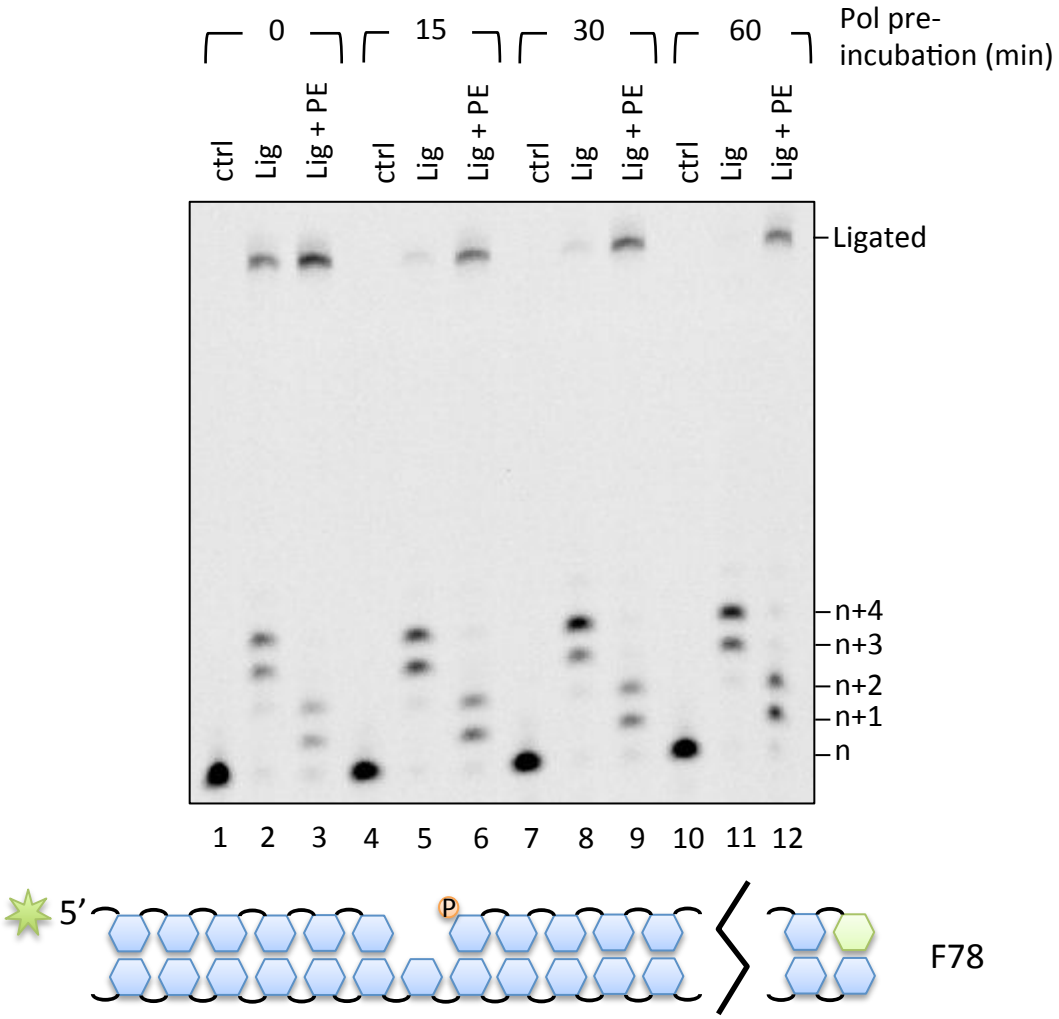


Figure 4.18 Mpa PE optimises Mpa Pol repair intermediates for ligation.
A schematic of the DNA substrate used for gap filling, phosphoesterase and ligation is shown below the scan of the gel. Ligation reactions contained 30 nM 5'-fluorescein labelled DNA substrate F78 (primer 16-mer, template 35-mer, downstream 18-mer) and 300 nM Mpa Lig, 300 nM Mpa Pol and 300 nM Mpa PE (where indicated) with 250 µM NTPs and 5 mM Mn.

Ligation was observed in both samples that were not pre-incubated with Mpa Pol but, critically, the sample with Mpa PE had an increased amount of ligated product compared to the sample without Mpa PE (Figure 4.18, lanes 2 and 3). This indicates that presence of Mpa PE confers an advantage for efficient ligation of annealed breaks. More over, when only Mpa Pol and Mpa Lig are available to repair annealed breaks, roughly 50 % of the starting substrate is converted to +3 or +4 extensions with a displaced D-strand, which is a repair intermediate that cannot be ligated. Whilst these assays use a large quantity of DNA in order to visualise results, it must be remembered that in a cell even a single unrepaired DSB could be lethal. Without Mpa PE, Mpa Pol is prone to generating repair intermediates that are potentially genotoxic and would require other specialised repair enzyme to process. This finding is even more salient when the samples pre-incubated with Mpa Pol are concerned. Following a 15 minute pre-incubation with Mpa Pol, almost no substrate remained that was eligible for ligation by Mpa Lig, and from 30 minutes onwards no ligation is observed (Figure 4.18, lanes 5, 8 and 11). Yet when Mpa PE is introduced to these samples ligation is rescued, even after an hour's worth of pre-incubation with Mpa Pol (Figure 4.18, lanes 6, 9 and 12).

These data robustly support the model of Mpa PE as the key agent in preventing and repairing Mpa Pol generated strand-displaced intermediates during annealed break repair. Mpa PE can prevent the occurrence of genotoxic repair intermediates, and is therefore potentially important during all DSB repair, since even a single break that cannot be repaired might be fatal to a cell. A schematic describing the activity of Mpa PE observed during these assays is shown in Figure 4.19.

4.10.1 Mpa PE can resect Mtu PolDom products and stimulate ligation

A repeat of the previous assay was conducted and duplicated, except with one set of reactions utilising Mtu PolDom. This would allow assessment of whether Mtu PolDom performed a similar role to that of Mpa Pol, and whether the repair profiles would be similar in the absence and presence of Mpa PE. Ideally this assay would encompass the individually truncated components of Mtu LigD, however only Mtu PolDom was readily available, and Mtu NucDom (PE) had previously shown resistance to purification and solubility. Even so, the assay could suggest if the bacterial LigD domains followed a similar repair pathway to the archaeal NHEJ enzymes, and whether the domains/proteins might be interchangeable. The extension profiles of Mtu PolDom were similar to those of Mpa Pol, except with slightly longer primer extensions observed, with some products up to +5 (Figure 4.20). As the Mtu PolDom pre-incubation

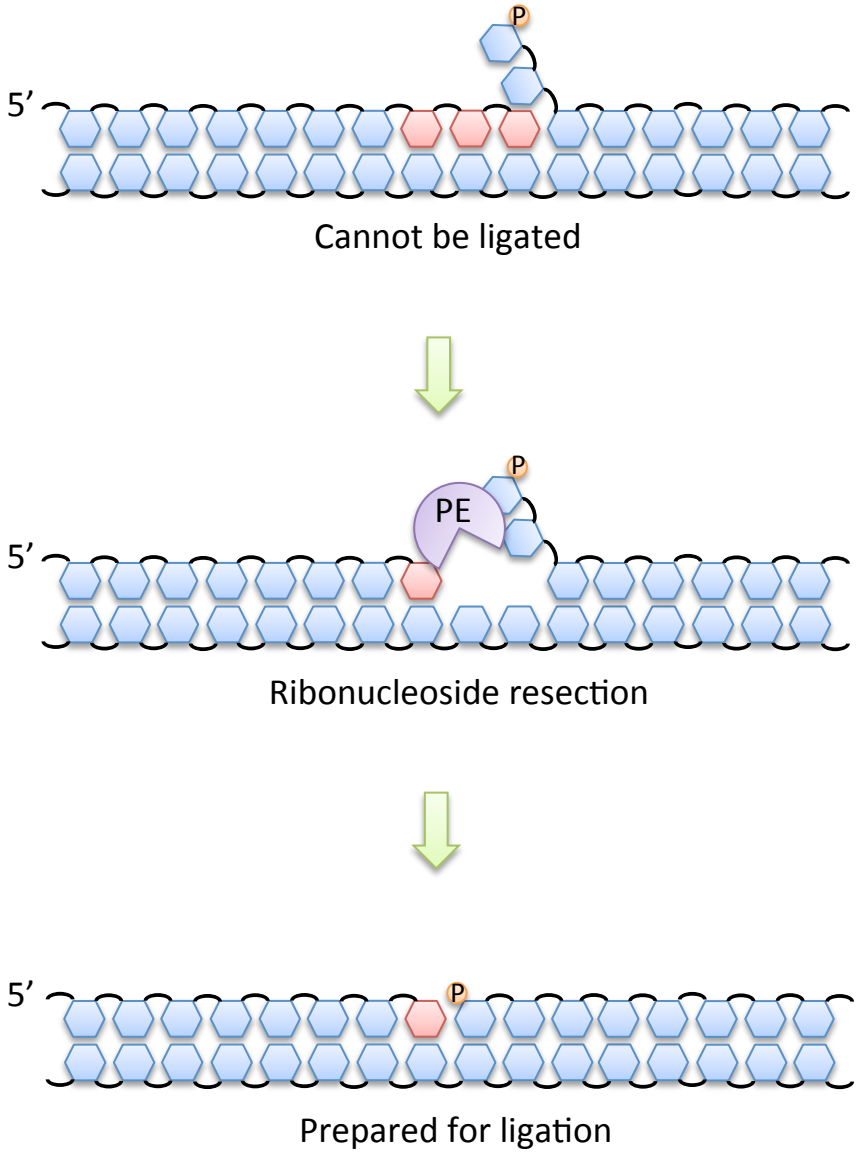


Figure 4.19 A schematic of Mpa PE optimising Mpa Pol generated repair intermediates for ligation.

Downstream DNA is displaced by Mpa Pol, creating a repair intermediate that cannot be ligated. Mpa PE is able to resect the RNA tract, by sequentially removing ribonucleosides and phosphates until a 3’-monoribonucleoside remains. The displaced DNA can then reanneal with the template, and is prepared for ligation by Mpa Lig.

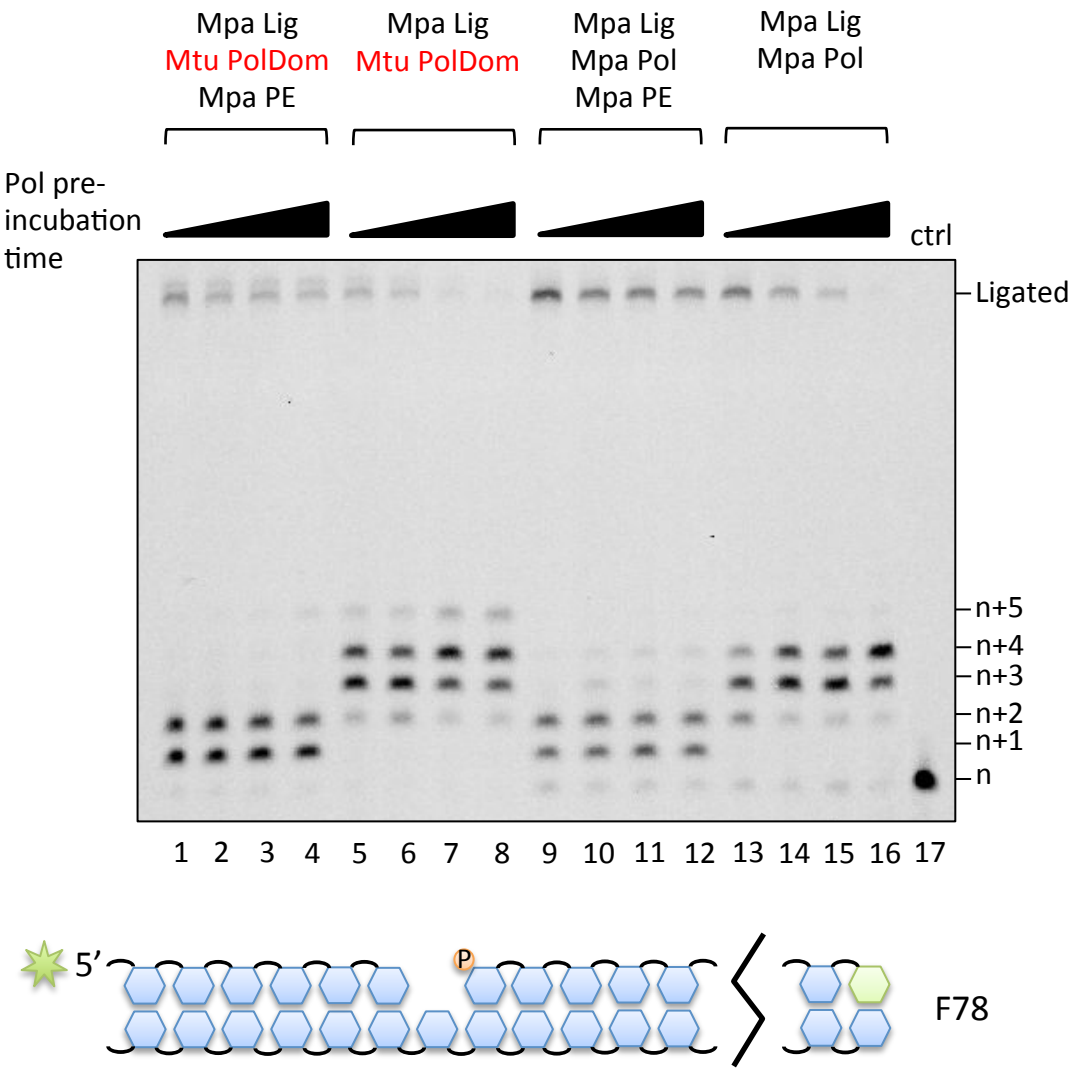


Figure 4.20 Mpa PE can resect Mtu PolDom repair intermediates to stimulate ligation.

A schematic of the DNA substrate used for gap filling, phosphoesterase and ligation is shown below the scan of the gel. Ligation reactions contained 30 nM 5'-fluorescein labelled DNA substrate F78 (primer 16-mer, template 35-mer, downstream 18-mer) and 300 nM Mpa Lig, 300 nM Mpa Pol or 300 nM Mtu PolDom and 300 nM Mpa PE (where indicated) with 250 µM NTPs and 5 mM Mn. The relevant polymerase was pre-incubated for 0, 15, 30 and 60 minutes.

period increased, the amount of ligated product reduced to zero (Figure 4.20, lanes 5-8), and following inclusion of Mpa PE extension is reduced mostly to +1 or +2 and ligation is observed across all time points (Figure 4.20, lanes 1-4). These data equate well with those detected when Mpa Pol is utilised (Figure 4.20, lanes 9-16), and demonstrate that Mpa PE can resect strand displacement generated by Mtu PolDom. These data also indicate that the bacterial NHEJ pathway probably operates in a similar way to that observed with the *M. paludicola* enzymes, since Mtu PolDom is also prone to generating strand displacement intermediates and Mtu LigD presumably maintains the PE domain to counteract this tendancy.

4.10.2 Mpa PE resection of NHEJ intermediates

Whilst the previous assays undoubtedly demonstrate that Mpa PE can play a significant role regulating the strand displacement activity of Mpa Pol, and that this is most likely the previously hidden role of the PE class of repair enzymes, confirmation of whether this role was also active at genuine broken ends awaited. An assay was designed using 3' overhanging end substrates and included Mpa Ku along with Mpa Lig, Pol and PE with hopes to replicate the effects seen using an annealed break substrate above. When using a substrate with a 5 base 3' overhang, of which four bases could create a microhomology with another substrate molecule, primer extension and ligation was observed in all samples; with and without PE, and with and without pol pre-incubation (Figure 4.21, lanes 1-5). It is interesting to note that the strand displacement activity of Mpa Pol is still evident, although to a lesser degree (Figure 4.21, lanes 2 and 4). The results from a similar assay without Mpa PE from Figure 4.4 concur that Mpa Pol struggles to displace the D-strand with this substrate. This is mostly likely because Mpa Pol needs to wait for the 3' primer strand before becoming active. Mtu PolDom has been shown to form a pre-ternary complex with DNA ends with a 3' overhang and a recessed 5'-P, and Mpa Pol likely shares this NHEJ Pol ability (Brissett et al., 2011). This allows the Pol to incorporate a NTP at unstable ends, but precludes it from extensive strand displacement. In this case the role of Mpa PE is less evident.

However, when using a substrate with a longer microhomology region, a 3' overhang of 11 bases, Mpa Pol mediated strand displacement was observed (Figure 4.21, lanes 7 and 9). The longer microhomology increases the stability of the annealed break, and becomes analogous to the gapped substrates used previously. Mpa Pol can be seen to generate a majority of strand-displaced intermediates following a 30' pre-incubation period (Figure 4.21, lane 7), even some products at +3 extension. Without pre-incubation, strand displacement is still

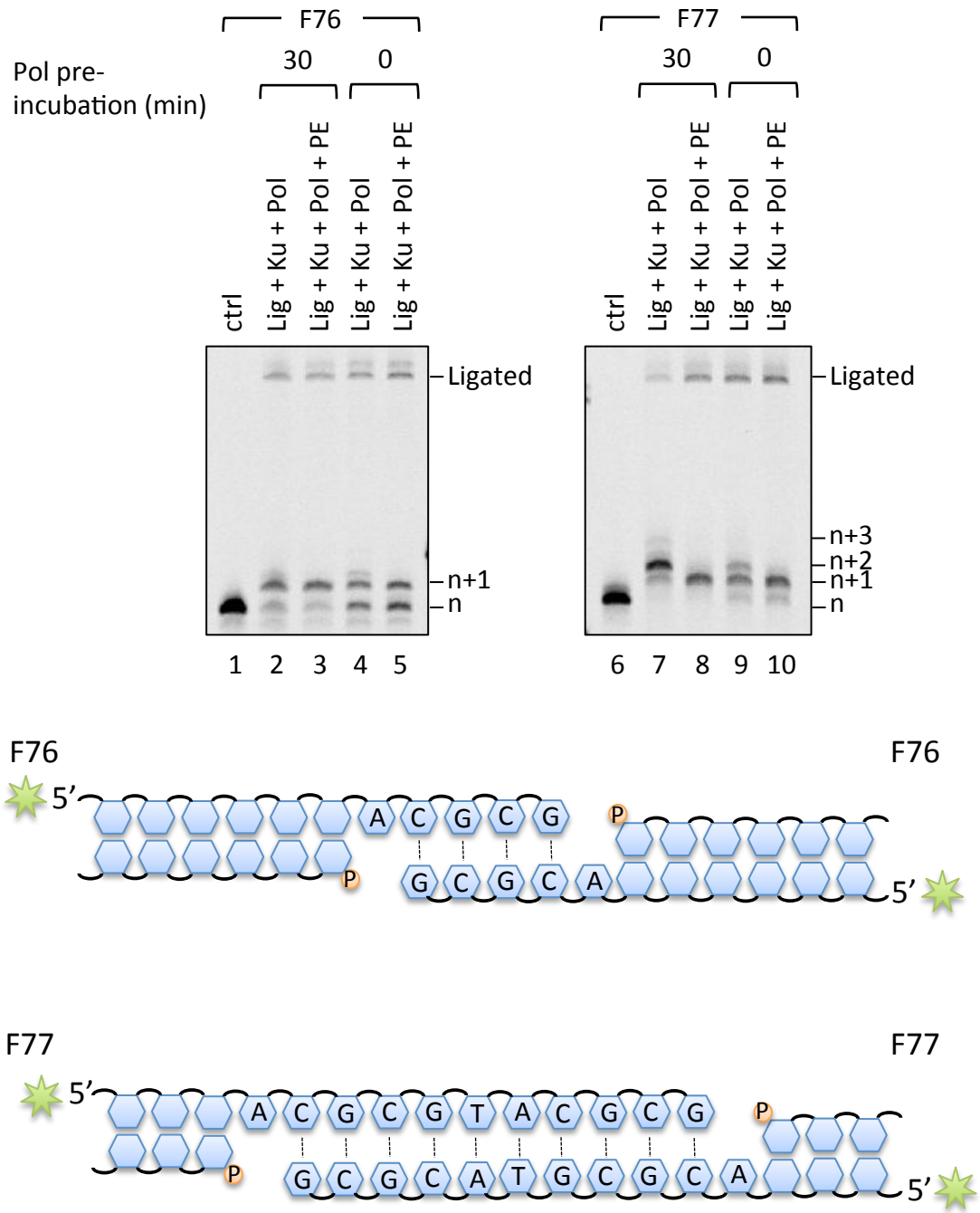


Figure 4.21 Mpa PE can optimise Mpa Pol repair intermediates to stimulate ligation in an end joining reaction.

Schematics of the DNA substrate used for end joining reactions are shown below the scans of the gels. End joining reactions contained 30 nM 5'-fluorescein labelled DNA substrate F76 (primer 42-mer, template 37-mer) or F77 (primer 48-mer, template 37-mer). Reactions also contained 300 nM Mpa Lig, 400 nM Mpa Ku 300 nM Mpa Pol and 300 nM Mpa PE (where indicated) with 250 µM NTPs and 5 mM Mn.

visibile, and Mpa PE can reverse this, although the difference in ligated products is difficult to determine (Figure 4.21, lanes 9 and 10). However addition of Mpa PE after the 30' Mpa Pol pre-incubation period reduces the RNA tract to +1, and causes an increase in ligated product (Figure 4.21, lane 8). This result consolidates the role of Mpa PE in regulating Mpa Pol activity; Mpa PE can repair strand-displaced intermediates that are generated during NHEJ, and stimulate ligation. The assays described in this chapter have revealed the role of PE in NHEJ in LigD based repair pathways that has remained unknown for a decade since the discovery of NHEJ in bacteria.

4.11 A model for Archaeo-Prokaryotic (AP)-NHEJ repair mechanisms

The new information garnered on the role of Mpa PE allows us to construct a model of NHEJ in the archaeal systems that possess LigD-like enzymes, and even further to apply these realisations to the orthologous bacterial NHEJ systems. This model is shown in Figure 4.22 and described in detail below.

Following a DSB the Ku homodimer binds the DNA ends. This protects the DNA ends from exonucleolytic degradation and also helps to tether them, preventing the termini from becoming distant. The NHEJ polymerase also plays a role in keeping the DNA termini proximal, by making use of unique DNA strand manipulation abilities that stem from the structural region known as Loop 1. The NHEJ Pol can foster DNA end-synapsis by utilising short regions of microhomology in the DNA ends, should the damaged ends have sufficient overhangs. Once the broken ends have been reintroduced and annealed, the NHEJ Pol can begin to fill any gaps that are present at either end of the microhomology region. The NHEJ Pol preferentially fills these gaps with ribonucleotides and generates a short stretch of RNA. In some situations, the NHEJ Pol can interrupt the base pairing between D-strand and the template, and the downstream DNA can become displaced by several bases. If the NHEJ Pol does generate any strand-displaced intermediates they require additional processing before the broken DNA can be successfully ligated. The phosphoesterase that is associated with almost all LigD enzymes possesses two distinct catalytic abilities; ribonuclease and phosphatase. This allows the PE to sequentially remove ribonucleosides and phosphates from the 3' ends of the tract generated by the NHEJ Pol. The PE is capable of resecting RNA that is displacing DNA, and therefore allows the reannealing of the downstream DNA to the template strand. A key feature of the PE ribonuclease activity is that it cannot remove a terminal ribonucleoside from a DNA strand, it requires a 2'-OH group on the penultimate base to allow for cleavage of the terminal base that

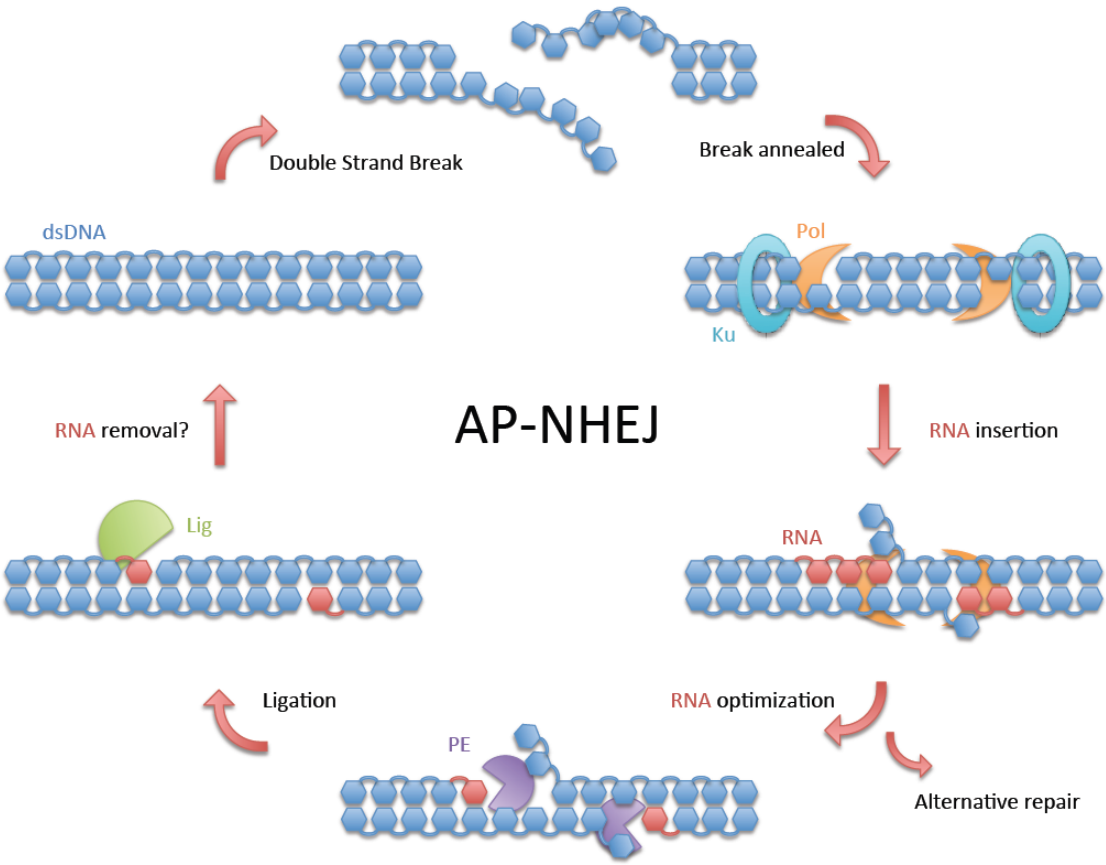


Figure 4.22 A model of Archaeo-prokayrotic (AP) Nonhomologous end joining.
A schematic of AP-NHEJ. Following a double strand break, the Ku homodimers bind the broken ends and bring them into proximity. Ku and Pol promote microhomology-mediated synapsis of overhanging termini, if present, to anneal the break. Pol fills in any resulting gaps at either end of the annealed break by template-dependent RNA synthesis. Pol may displace several bases of downstream DNA using its distinct ability to mobilise and direct DNA strands, offering a great flexibility to repair a variety of configurations of damaged DNA. Strand-displacement RNA synthesis by Pol is regulated by PE, which can remove unnecessary NMPs and allow the displaced DNA to realign with the template. PE always leaves a single terminal ribonucleoside in place in resected breaks, presenting Lig with an optimised nicked substrate for ligation. Although the broken DNA is reconnected, repair may not be fully completed, because several RNA bases still reside in the duplex. RNA removal may be performed by additional enzymes, such as RNase HII, before other repair Pols fill in the gaps with DNA.

it has engaged. This results in at least one ribonucleoside remaining present at the break following processing. Once the RNA resection is complete any displaced DNA can reanneal and the LigD ligase can complete the repair. LigD has a unique preference amongst ligases to seal nicked DNA with a 3'-monoribonucleoside, which dovetails neatly with the restriction of the PE not being able to remove the last ribonucleoside from a DNA strand. LigD seals the nick, and the DSB break is repaired. The complete DNA duplex now contains short tracts of RNA that presumably require processing and removal by other enzymes. The strict requirement of each of the AP-NHEJ enzymes for ribonucleotide or ribonucleoside usage is considered in the discussion section below.

The proposed model for AP-NHEJ considers a hypothetical damaged end configuration, but of course the system must be robust enough to deal with a large variety of different ends that can be generated by the extreme damage at a DSB. Data gathered over the previous two chapters compiled with previously published *in vitro* and *in vivo* work on the bacterial systems offers some insights of how these enzymes might cope with a staggering variety of damaged ends.

4.11.1 Repair of blunt ends

At first glance, blunt ends might seem the simplest types of DSBs to repair, in that they should require no processing and can simply be ligated end to end. Whilst this is true of the eukaryotic NHEJ pathway, the bacterial and archaeal enzymes are not capable of doing so. Plasmid repair assays in *M. smegmatis* using blunt ended substrates produced either +1 insertions or extensive deletions during repair (Aniukwu et al., 2008). Given our understanding of the strict requirements of the LigD ligase, it is not surprising that a + 1 insertion is the only detected outcome of AP-NHEJ with blunt ends. The necessity for a ribonucleoside at the 3' end of the ligation reaction dictates that an NTP must be incorporated, and the NHEJ Pol has a template independent activity to facilitate this. The schematic in Figure 4.23A shows how this repair likely occurs. These results have been repeated *in vitro* using the *M. paludicola* NHEJ enzymes, showing blunt end joining with a +1 incorporation (Figure 4.24), and inability to ligate DNA only blunt ends without NTP incorporation (data not shown). The NHEJ Pol is capable of inserting any nucleotide onto the end of a blunt DSB, but in order for the ribonucleosides to base pair, it is possible that one end is extended first, and the ends are connected, allowing the remaining gap to be filled in a template dependent manner.

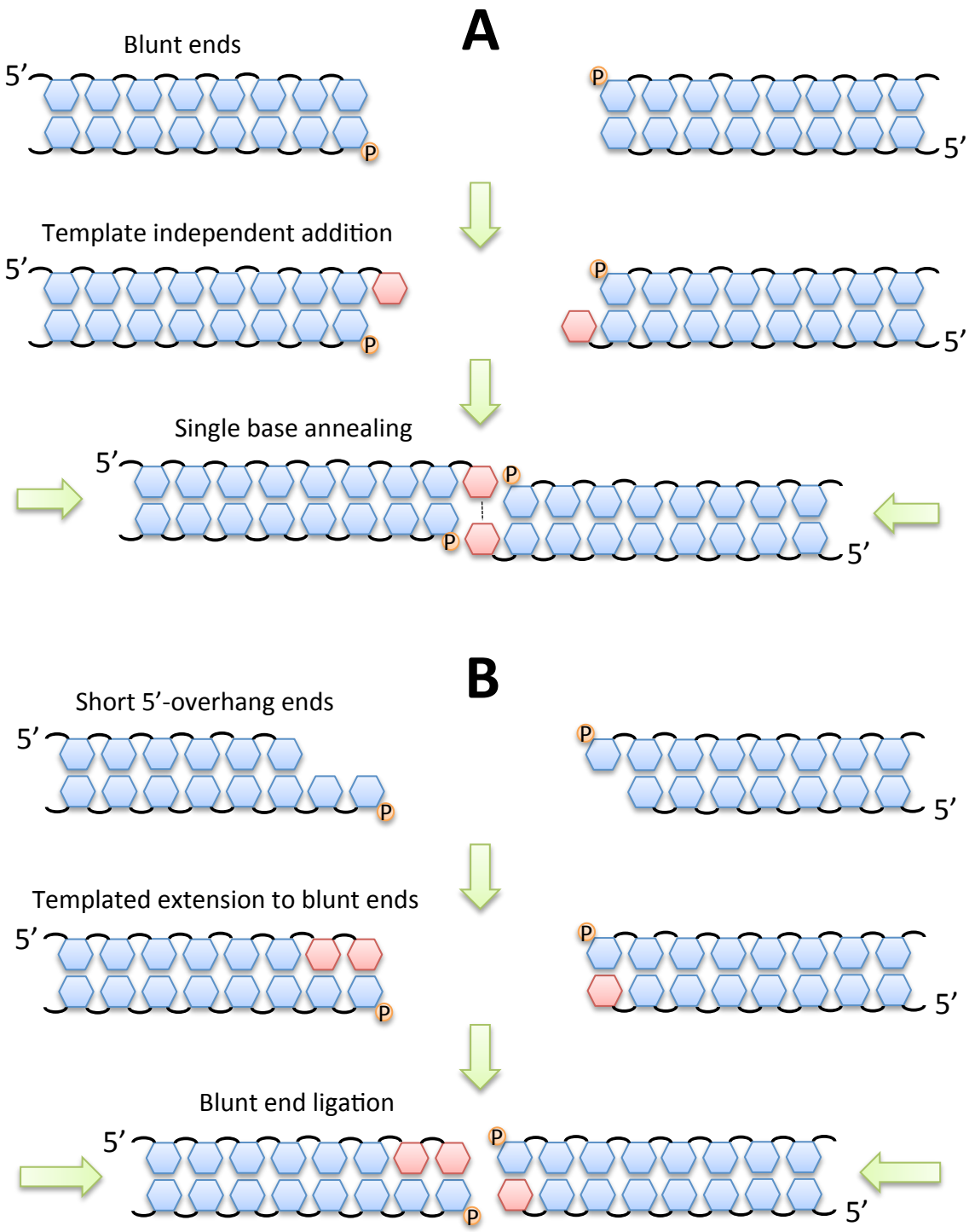


Figure 4.23 Schematics of how AP-NHEJ resolve blunt or short 5′ overhang ends. (**A**) A model of how AP-NHEJ might resolve blunt ends. The NHEJ Pol can incorporate a single ribonucleotide to blunt ends, creating a single base 3′ overhang. It is not known how the system ensures that the bases pair correctly to enable joining of the ends, but it may be that one end is extended first, and then offers a template to the second extension. (**B**) A model of how AP-NHEJ might resolve short 5′ overhang ends. The NHEJ Pol can fill the 5′ overhangs to blunt ends that can be directly ligated by the NHEJ Ligase. The Pol can also catalyse template independent extension of a single base 3′ overhang, which can then be sealed in a similar manner to blunt ends.

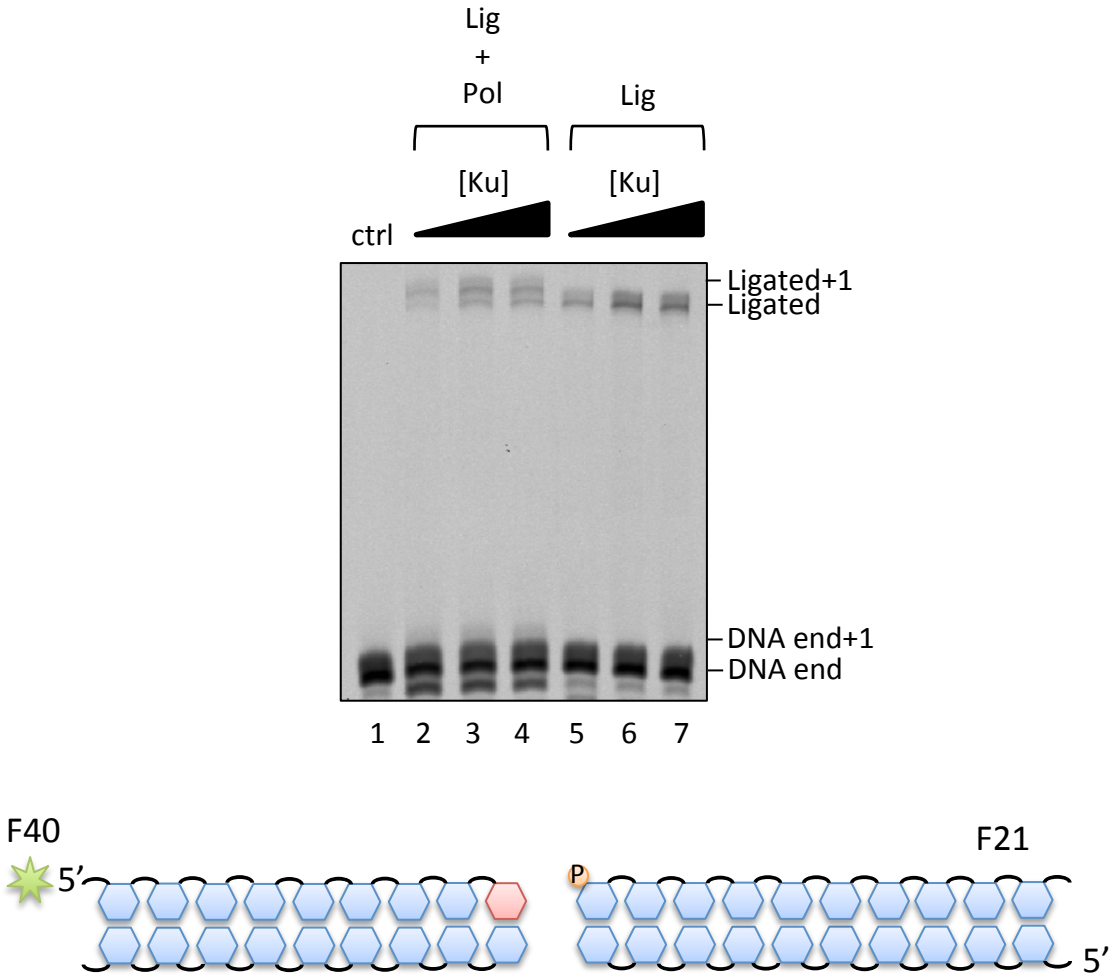


Figure 4.24 Mpa Pol can extend blunt ends to generate a +1 insertion that can be ligated by Mpa Ku and Mpa Lig.

A schematic of the DNA substrate used for extension and end joining is shown below the scan of the gel. Extension and end-joining reactions contained 30 nM 5’-fluorescein labelled DNA substrate F40 (primer 36-mer, template 36-mer) and unlabelled F21 (primer 36-mer, downstream 36-mer) and 300 nM Mpa Lig, 300 nM Mpa Pol (where indicated) and 200, 400 and 800 nM Mpa Ku with 250 µM NTPs and 5 mM Mn.

4.11.2 Repair of short 5'-overhang ends

Short 5'-overhang ends offer a relatively simple repair process for AP-NHEJ enzymes. The NHEJ Pol can efficiently fill 5' overhangs with ribonucleotides until the ends are blunt. This produces blunt dsDNA ends with 3'-ribonucleosides that the ligase can seal. A schematic of this method of repair is show in Figure 4.23, B. Experimental examples of this have been demonstrated in plasmid repair assays and *in vitro* with the *M. paludicola* NHEJ enzymes (Figure 4.25) (Aniukwu et al., 2008). Plasmid repair also showed that in a very small minority of cases, the NHEJ Pol will continue to extend the blunt end to a single base 3' overhang by template independent synthesis. With both short 5' overhangs and blunt ends it is unlikely that the PE plays a significant role in removing ribos, and is more likely useful to ensure that 3' ends are free of phosphates.

4.11.3 Repair of long 5'-overhang ends

Repair of 'long' 5'-overhang ends is potentially a more complex affair than with short ends, since two competing mechanisms may be operating at the same time. The NHEJ Pol can begin to fill the ends in a template dependent manner, whilst the ends are also directed towards a search for microhomology. Figure 4.26 demonstrates how repair of two long 5' overhangs might happen, although it is entirely possible that the NHEJ Pol has already begun filling the ends whilst microhomology searching is underway. Plasmid repair assays in *M. smegmatis* did not reveal any usage of microhomology for 4 base overhanging 5' ends, however plasmids repaired *in vitro* by Mtu LigD and Mtu Ku with the same length overhangs did detect utilisation of microhomology and gap-filling for repair (Aniukwu et al., 2008; Della et al., 2004). The Mpa NHEJ enzymes were able to join long 5'-overhang ends, and did so by creating a microhomology, although the assay was designed to encourage this method of repair (Figure 4.27). It is possible that the preference for repair method depends on the sequences of the ends, and how quickly a microhomology-mediated synapse is created. In the event of break annealing and subsequent gap filling, the PE plays a pivotal role in ensuring that the gaps are filled correctly and that any strand displacement is reversed.

4.11.4 Repair of short 3'-overhang ends

Repair of short 3'-overhang ends might employ a unique method of repair, by formation of a pre-ternary complex where the NHEJ Pol binds to the 5'-P and accommodates an incoming NTP in its active site that is correctly base paired with the template. The pre-ternary complex then awaits arrival of the 3' overhang end as primer to extend from (Figure 4.28).

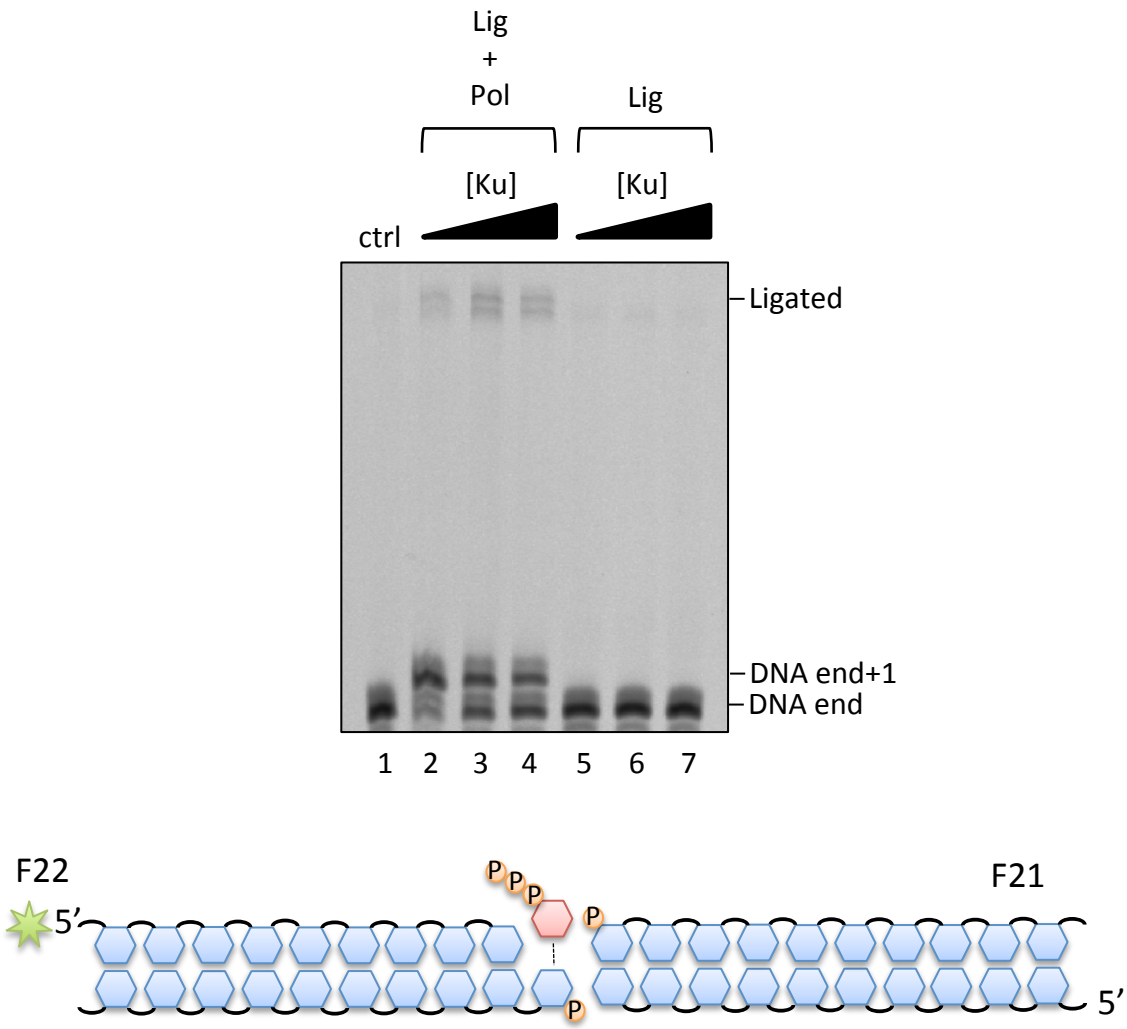


Figure 4.25 Mpa Pol fills in 5’-overhang ends to enable ligation by Mpa Ku and Mpa Lig.

A schematic of the DNA substrate used for extension and end joining is shown below the scan of the gel. Extension and end-joining reactions contained 30 nM 5’-fluorescein labelled DNA substrate F22 (primer 36-mer, template 37-mer) and unlabelled F21 (primer 36-mer, downstream 36-mer) and 300 nM Mpa Lig, 300 nM Mpa Pol (where indicated) and 200, 400 and 800 nM Mpa Ku with 250 µM NTPs and 5 mM Mn.

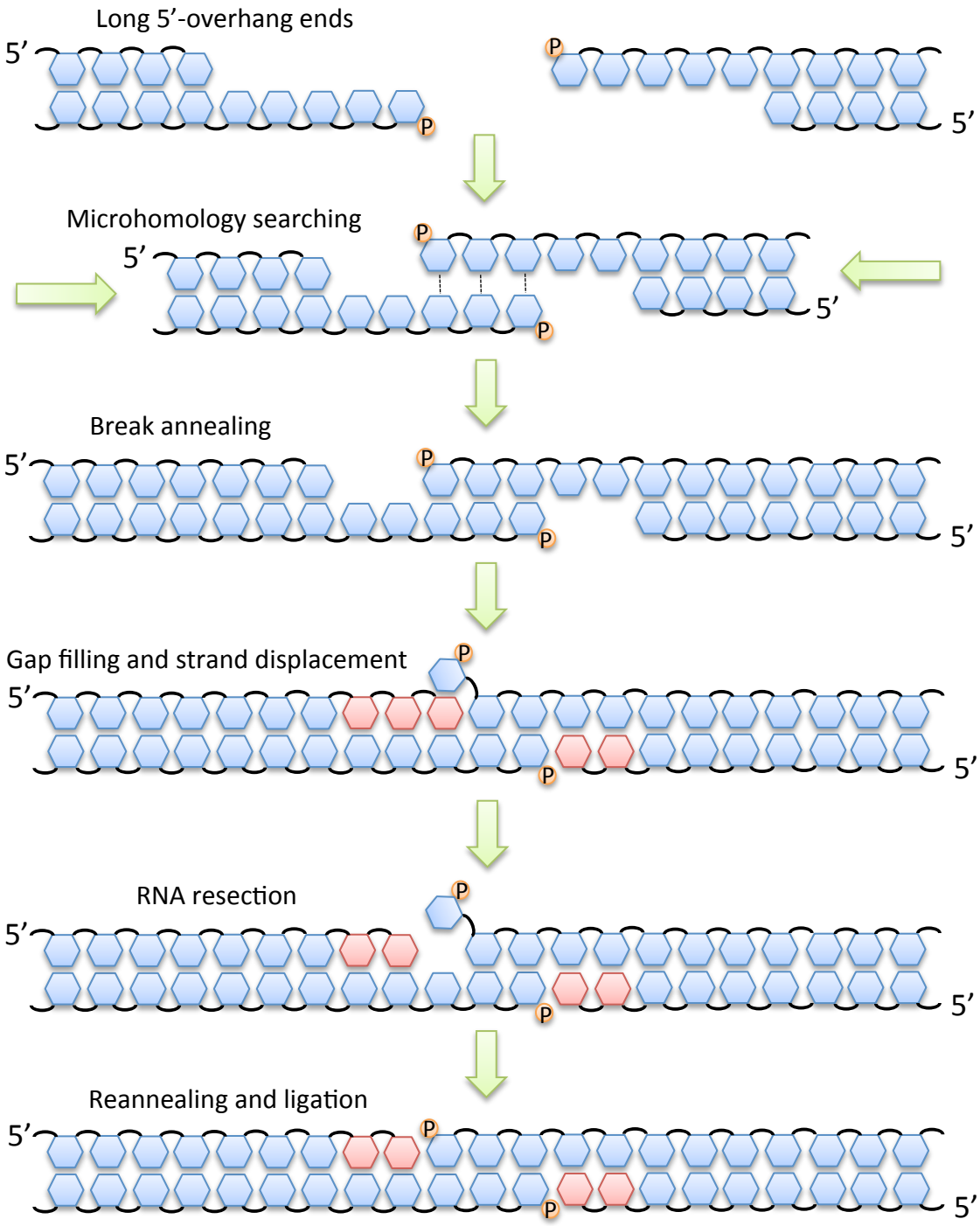


Figure 4.26 Schematics of how AP-NHEJ resolves long 5' overhang ends.
A model of how AP-NHEJ might resolve a DSB with long 5' overhang ends. The NHEJ Pol can utilise microhomologies found in 5' overhang ends to anneal the break. This ability mightcompete with the primer extension activity of Pol, since it can begin to fill the ends as soon as it engages them. In the event of a sufficiently long 5' end with a strong microhomology the break can be annealed, and the NHEJ Pol then gap-fills the remaining holes. If D-strand displacement occurs the PE can repair this by RNA resection, and allow the D-strand to reanneal. The repair is completed by Ligase D which seals the nicks.

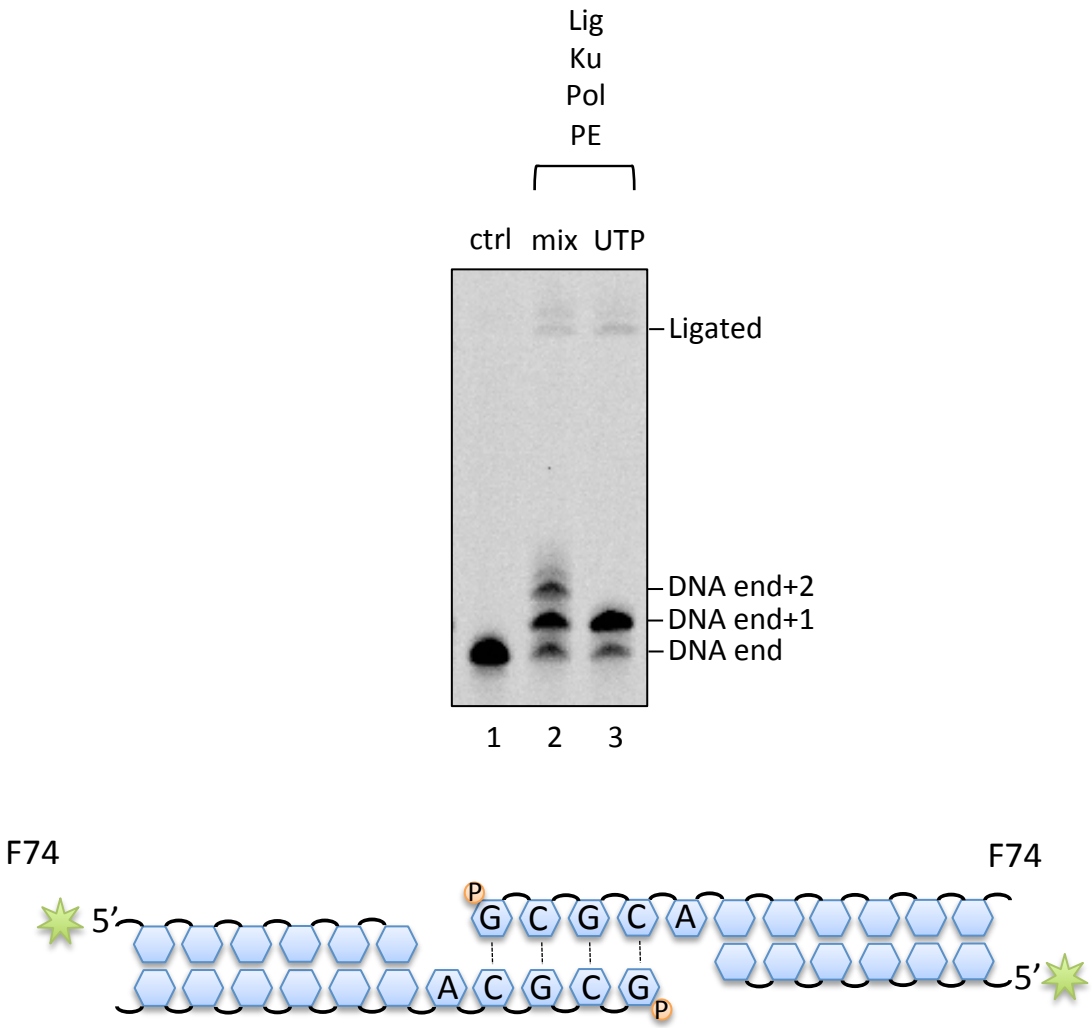


Figure 4.27 Mpa Lig, Ku, Pol and PE can join long 5’-overhang ends via microhomology.

A schematic of the DNA substrate used for end joining is shown below the scan of the gel. End-joining reactions contained 30 nM 5’-fluorescein labelled DNA substrate F74 (primer 37-mer, template 42-mer) which forms a microhomology of four bases. End-joining reactions also contained 300 nM Mpa Lig, 300 nM Mpa Pol, 300 nM Mpa PE and 400 nM Mpa Ku with 250 µM NTP mix or UTP 62.5 µM and 5 mM Mn.

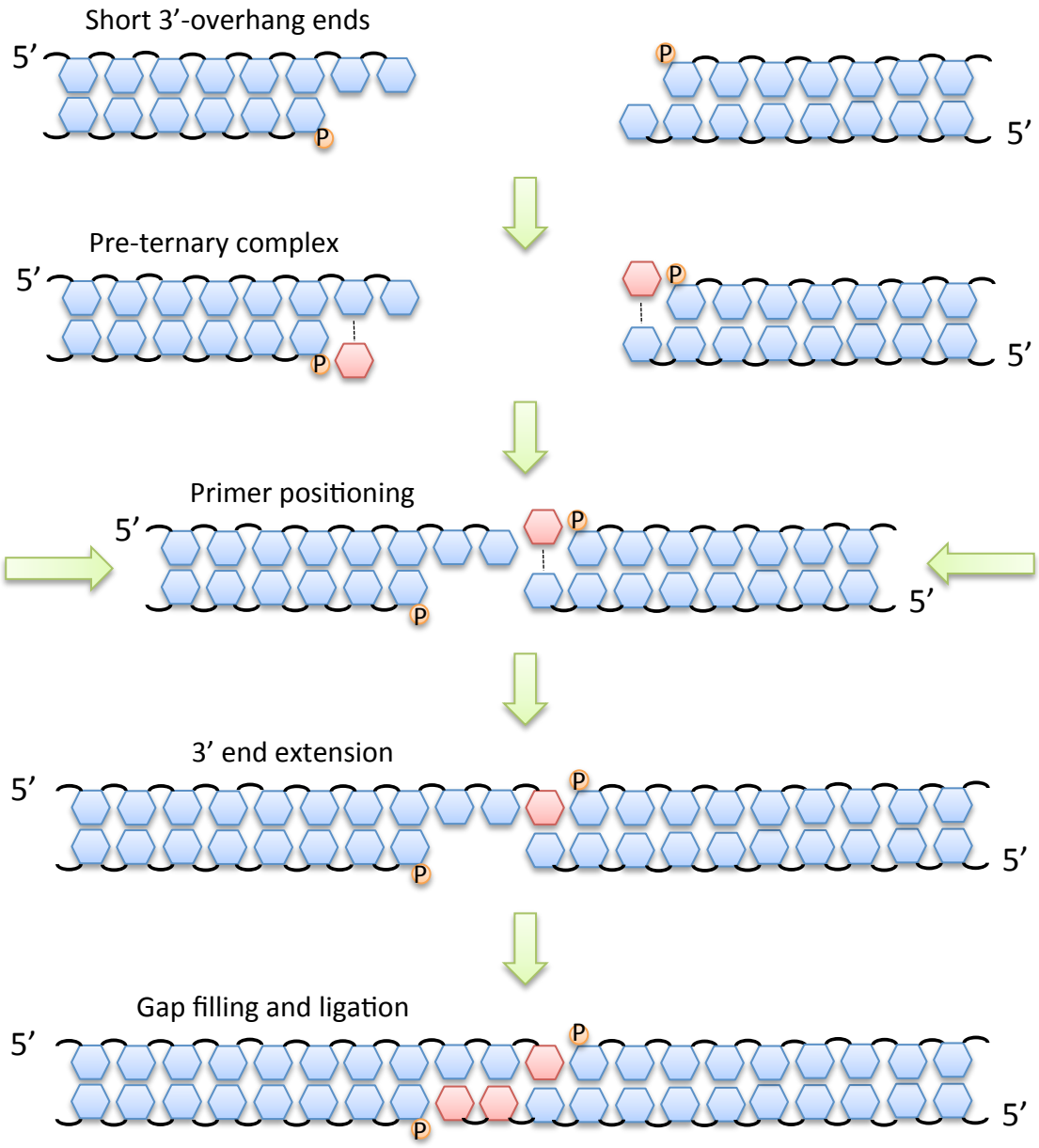


Figure 4.28 Schematics of how AP-NHEJ resolves short 3' overhang ends.
A model of how AP-NHEJ might resolve a DSB with short 3' overhang ends. The NHEJ Pol can bind strongly to a 5'-phosphate, and can even accept an incoming ribonucleotide into its active site in the absence of a primer. This pre-ternary complex has been observed in Figure 3.15 and Brissett et al., (2011). The 3' overhang end can then be positioned in the active site by the NHEJ Pol for extension in a template dependent manner. It is unknown whether the NHEJ Ligase can seal this nick if a gap remains on the other strand. Even so, the added stability could then allow for gap filling if the other strand could not be repaired by pre-ternary complex. Any remaining nicks could be sealed by the NHEJ Ligase.

Crystallographic and biochemical data support this method of short 3' overhang repair, and plasmid repair assays in *M. smegmatis* demonstrated at least one case of this mechanism of 3'-overhang end repair (Brissett et al., 2011; Aniukwu et al., 2008). Without a series of further experiments it is difficult to suggest if this mechanism is always used to repair short 3' overhangs, and therefore remains somewhat speculative. Repair via formation of the pre-ternary complex offers AP-NHEJ an alternative method to increase the overall flexibility of the system. It is possible that the PE is important in this pathway, since gap filling after the extension and joining of one end might cause strand displacement.

4.11.5 Repair of long 3'-overhang ends

Repair of long 3'-overhangs is potentially very similar to that of long 5'-overhang ends, except instead of templated fill in of ends, the AP-NHEJ system might construct pre-ternary complexes. However pre-ternary complex formation is less likely to interfere with microhomology formation, and should a gap of more than one base remain after a synapse is generated, any pre-ternary complex would dissolve and the NHEJ Pol could then fill the gap. This method of repair was demonstrated *in vitro* with *M. paludicola* (Figure 4.21) and was also detailed in repair of plasmids by Mtu LigD and Mtu Ku (Della et al., 2004). It was through exploration of repair of long 3'-overhang ends that the role of the PE was understood. Following break annealing the PE regulates the extension of the NHEJ Pol to enable LigD ligase to seal the remaining nick. Essentially the PE enables the NHEJ Pol to maintain an incredible flexibility, and the ability to manipulate DNA strands, and should the repair efforts become excessive, the PE can undo them and generate LigD's ideal ligation substrate (Figure 4.29). It is interesting to note that the role of the PE was effectively invisible during plasmid repair assays, but this is almost certainly because the strand displacement activity of the NHEJ Pol is also not visible in successfully ligated plasmids (Della et al., 2004; Aniukwu et al., 2008). Removal of the PE would condemn any strand-displaced plasmid repair to failed ligation or excessive exonucleolytic degradation. Indeed the plasmid repair assays did suffer a decreased efficiency of repair, especially with deletion of the catalytic histidine residue for 3'-overhang repair (Aniukwu et al., 2008). Increased deletions were also noted for repair of all ends with a catalytic dead PE domain. These changes in efficiency and fidelity were not significant enough to make clear the role of the PE, however since the unraveling of its ability to resect strand-displaced repair intermediates, the results begin to make sense.

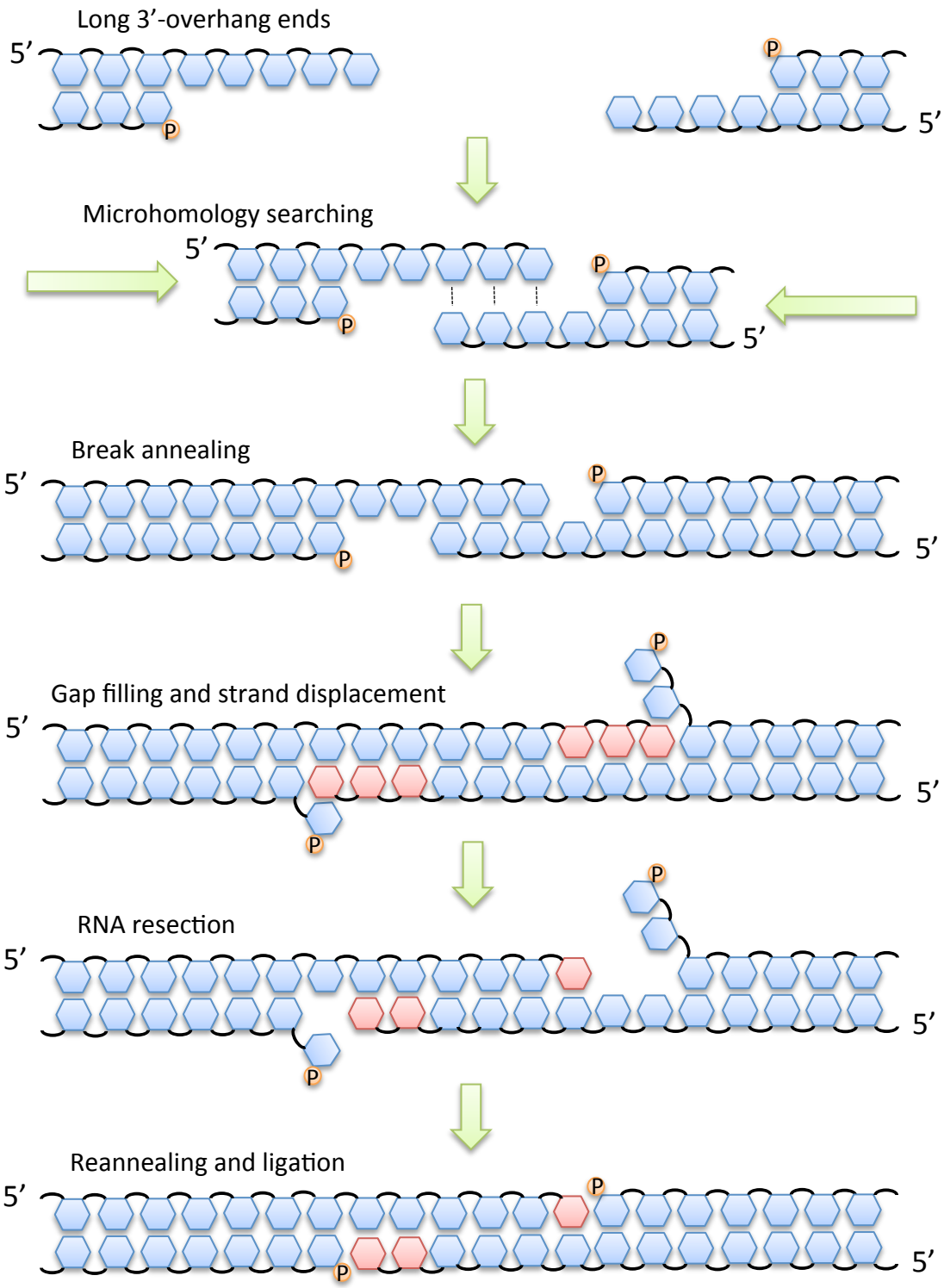


Figure 4.29 Schematics of how AP-NHEJ resolves long 3' overhang ends.
A model of how AP-NHEJ could resolve a DSB with long 3' overhang ends. Long enough ends would ensure that a small region of microhomology could be found as seen in Figure 4.21 and Brissett et al., (2007). After the break is annealed the NHEJ Pol can fill the remaining gaps, at which point it may displace the D-strand. If so, the PE can resect the displacing RNA and allow the D-strand to reanneal. The ligase can then complete the repair by sealing the nicks.

4.11.6 Insights into AP-NHEJ from *Mycobacterium tuberculosis* and *Archaeoglobus fulgidus*

Isolated proteins from the NHEJ systems from *M. tuberculosis* and *A. fulgidus* have also been used as comparisons to the *M. paludicola* NHEJ proteins. A gap-filling assay comparing wt Mtu PolDom to a 'quad' mutant Mtu PolDom with three residues on Loop 1 and a single residue of Loop 2 substituted with alanine, showed that the quad mutant was incapable of strand displacement (Figure 4.30). This result concurs with previously published information on the substitution of arginine 220 with alanine; that the enzyme is less active, however it also demonstrates a loss of strand displacement activity. Further investigation of this result is required to decipher whether the arginine substitution causes this loss of function, or if the lower activity of the mutated enzyme affects strand displacement activity indirectly.

Dr. Nigel Brissett and Natalie Day kindly donated purified Afu ligase and Afu Ku proteins in order to discern if these two components were capable of joining discontinuous DNA ends. The results of the experiment showed that Afu Lig and Ku were sufficient to join DNA ends with a four base microhomology with a ribonucleoside at the 3' end (Figure 4.31). As such, *A. fulgidus* can be considered NHEJ proficient. Despite the absence of an NHEJ Pol, Afu Lig displayed a preference for ligating nicks with a 3'-ribonucleoside and, unlike Mpa Lig, was capable of sealing a limited amount of all-DNA nicks (Figure 4.32). It is possible that Afu Lig might have to maintain additional substrate flexibility given the lack of associated domains or proteins.

4.12 Discussion

The targets of the research discussed in this chapter were to reconstitute a complete end-joining reaction *in vitro* to firmly establish the presence of NHEJ in archaea; to systematically test the role of the PE in NHEJ and discover why it is conserved with the other LigD enzymes; and finally to explore the capabilities of the other NHEJ components and find the limits of their capabilities. The *M. paludicola* NHEJ proteins proved capable of joining discontinuous DNA ends *in vitro* in a Ku dependent manner, although efficiency of the reactions was less than expected when compared to the annealed break substrates used prior to the end-joining experiments. Even so, the assays established that a fully functional NHEJ system operates in archaea and that it is highly analogous to the bacterial system. As such, any bacterial or archaeal system that possesses the trifecta of LigD Ligase, Pol and PE are likely to repair DSBs in a very similar way and can be classified as members of the AP-NHEJ pathway.

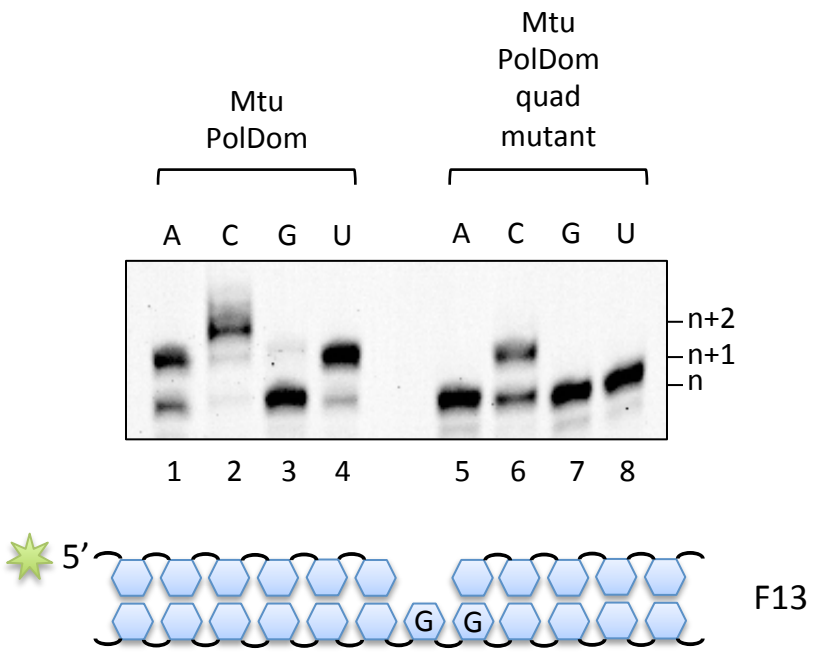


Figure 4.30 Mutations in Mtu PolDom Loop 1 and Loop 2 prevent strand displacement.

A schematic of the DNA substrate used for gap filling is shown below the scan of the gel. Gap filling reactions reactions contained 30 nM 5'-fluorescein labelled DNA substrate F13 (primer 16-mer, template 35-mer, downstream 18-mer) and 300 nM wt Mtu PolDom or 300 nM Mtu PolDom quad mutant (Loop 1 H83A, R84A, S85A, Loop 2 R220A) with 62.5 µM of ATP, CTP, GTP or UTP and 5 mM Mn.

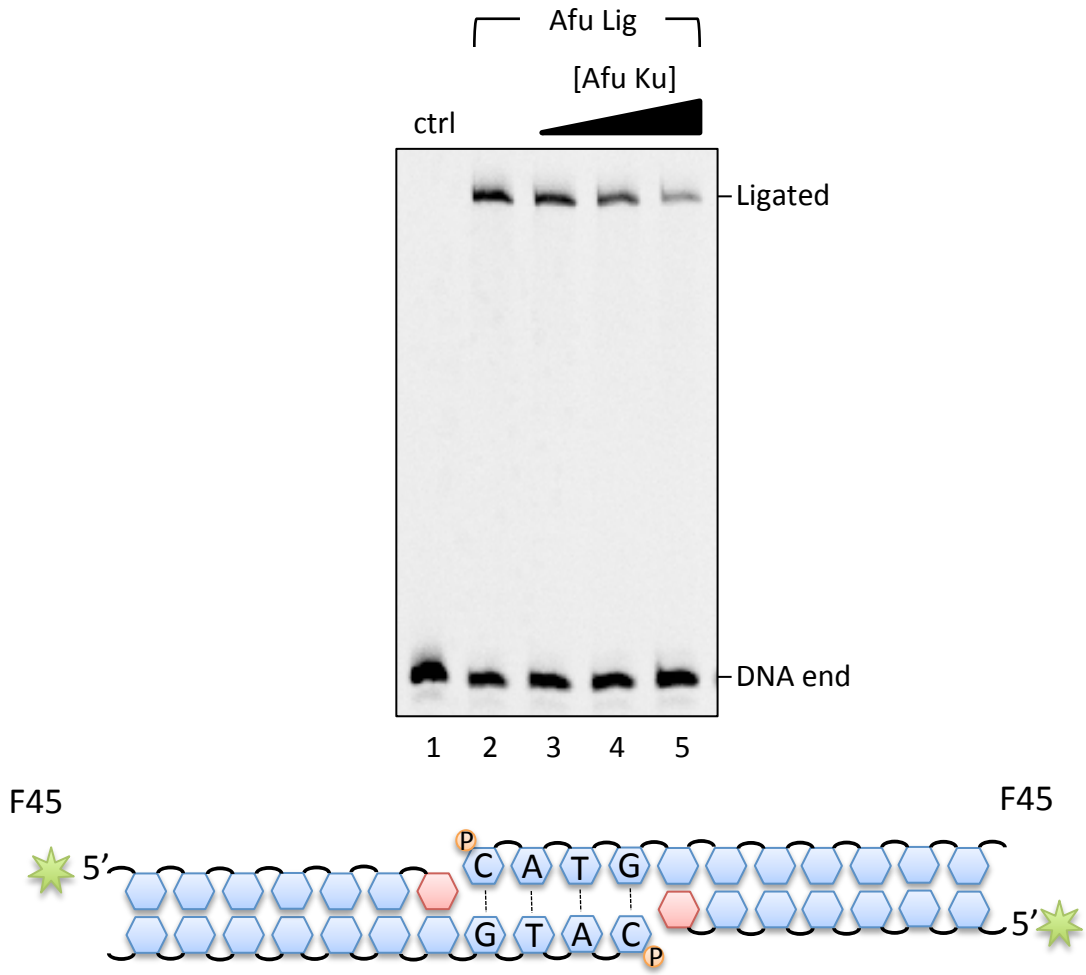


Figure 4.31 Afu Lig and Afu Ku can ligate DNA ends.
A schematic of the DNA substrate used for the end-joining reaction is shown below the scan of the gel. End-joining reactions contained 30 nM 5’-fluorescein labelled DNA substrate F45 (primer 36-mer, template 39-mer) which has four self-complementary bases that can form a microhomology. End-joining reactions also contained 300 nM Afu Lig and 200, 400 and 800 nM Afu Ku with 5 mM Mn.

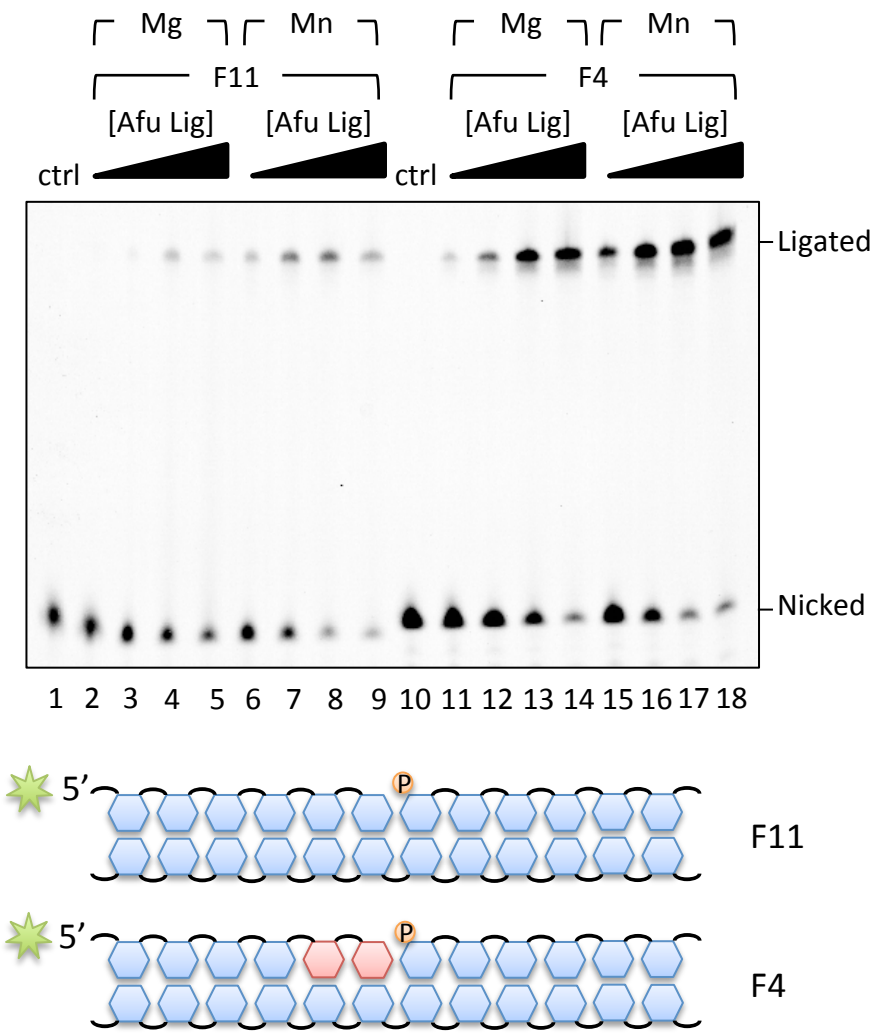


Figure 4.32 Afu Lig is most efficient at ligating RNA-DNA nicks, but can also seal all-DNA nicks.

Schematics of the DNA substrates used for ligation reactions are shown below the scan of the gel. Ligation reactions contained 30 nM 5'-fluorescein labelled DNA substrate F11 or F4 (both primer 16-mer, template 35-mer, downstream 19-mer) where indicated. Ligation reactions also contained 200, 400 and 800 nM Afu Lig and either 5 mM Mg or Mn.

The *M. paludicola* PE was active and performed similar tasks to the bacterial PE as described in the previous chapter. Examination of a variety of substrates revealed that the PE actually preferentially resects RNA in the presence of a D-strand. Further experiments with Mpa Pol revealed that it could displace D-strands and extend primers beyond the size of whichever gap it was filling. This led to experimentation with Mpa PE, to see if it might be able to regulate this activity and prevent the accumulation of potentially genotoxic repair intermediates. Mpa PE was capable of cleaving ribonucleosides that were displacing the D-strand, and was able to stringently regulate Mpa Pol, and prevent it from over extending DNA primers. The results revealed that the role of the PE was to remove strand-displaced repair intermediates and to stimulate ligation by doing so. Mpa PE was able to improve ligation efficiency both with an annealed break substrate and with disparate DNA ends.

One of the large, unanswered questions of AP-NHEJ is why the enzymes have co-evolved a specific requirement for ribonucleotides and ribonucleosides. Whilst it is true that the NHEJ Pol has a limited flexibility in dNTP incorporation, neither Mpa PE nor Mpa Lig can process all DNA substrates. The ratio of NTPs to dNTPs during senescence or growth is highly in favour of NTP usage, so the NHEJ components may have evolved to make use of the most abundant resources that are available (Pitcher et al., 2005). Fascinatingly, in the case of a precise break, the necessity for a ribonucleoside for ligation ensures that any successful repair will be mutagenic, which surely cannot be beneficial for a cell. It may be that a precise break, either blunt or with perfectly matching ends, is a rare occurrence, and it might also be favourable to rapidly repair a DSB in whichever manner possible. An unrepaired break might prove fatal for a cell, therefore mutagenic repair is preferable to none at all. A fascinating possibility of an RNA centric NHEJ mechanism is that once the repair has begun, the DNA ends are protected from any further loss. For example, once the NHEJ Pol has incorporated a ribonucleotide, the PE can never remove the terminal ribonucleoside. In this way the DNA beyond the ribonucleoside is protected from exonucleolytic degradation by the unique requirements of the PE for a 2'-OH on the penultimate base of a 3' strand. Essentially, the RNA is demarcating the boundaries of the repair, and the NHEJ complex is only able to process and remodel the RNA intermediates that it creates itself.

A secondary consideration of AP-NHEJ employing RNA as a repair intermediate is that at least one ribonucleoside remains in the repaired duplex, and possibly more depending specifics of the break. It is unlikely that the RNA would then remain in the DNA for a long time, given that even a single RNA base constitutes a form of DNA damage, in that it increases the likelihood of

a replication error by a DNA polymerase (Nick McElhinny et al., 2010). There are enzymes that perform the specific role of removing RNA from DNA, such as RNase HII, that are thought to be maintained to prevent an accumulation of such sites following replication (Rydberg & Game, 2002). RNase HII is demonstrated to remove single ribonucleosides, or short tracts from DNA, and is thought to be present in every organism (Rydberg & Game, 2002). RNase HII has been shown to be essential in murine systems (Reijns et al., 2012). *M. paludicola* does encode an orthologue of RNase HII, and it is likely responsible for removal of RNA post AP-NHEJ repair (Appendix A). A gathering body of evidence is building that argues that AP-NHEJ is not alone in utilisation of RNA during NHEJ. Studies on human repair polymerases have discovered that they also preferentially insert NTPs over dNTPs (Nick McElhinny & Ramsden, 2003; Martin et al., 2013). It is unclear what effect the NTPs have on the efficiency of repair, and as yet no ribonucleases have been identified that operate with the eukaryotic NHEJ enzymes. Lig IV also has a distinct ability to ligate all DNA ends, so the necessity for any ribonucleotide incorporation is unclear. It is likely that the eukaryotic NHEJ system also seeks to make use of the high availability of NTPs whilst it is most active.

Other unanswered queries remain about AP-NHEJ, the most intriguing of which is why the NHEJ Pol retains the ability to displace DNA that is downstream of the primer that it is extending. This ability has apparently generated the need for a separate enzyme that can undo the strand displacement, the PE. A simpler option would be to do without the strand displacement and the PE altogether, and then gaps could be efficiently filled and ligated. A reasonable response to this is that the PE probably plays a more sophisticated role than only deleting excessive extensions, and that the strand displacement ability of the NHEJ Pol is integral for other aspects of NHEJ. It is possible that the strand displacement ability of the polymerase is linked to the ability of Loop 1 to manipulate DNA ends or the phenylalanine pair that occupy a position on the edge of the active site (Figures 1.14 and 1.16). The phenylalanines have been shown to possess the ability to splay DNA strands and distort the geometry of the helix; it is possible that they play a role in separating the D-strand from the template strand.

Recent publications in the field of bacterial NHEJ have characterised the *B. subtilis* NHEJ enzymes (de Vega, 2013). As noted in Figure 1.7, *B. subtilis* is different from all other studied bacterial NHEJ systems because it does not have a complete LigD-like operon. The Bsu NHEJ operon is composed of a continuous LigD ligase and an NHEJ Pol gene, and is adjacent to a Ku gene. The polymerase domain does have a preference for NTP incorporation, but is able to

incorporate dNTPs in a similar manner to Mtu, Pae and Mpa NHEJ Pols. Interestingly the LigD domain was able to catalyse ligation of a gap filled entirely by dNTPs, and although is not as efficient in doing so as with an RNA filled gap, it is still a striking departure from the AP-NHEJ system discussed so far. The dNTP flexibility might reflect the absence of the PE domain. It is also unclear whether the Bsu PolDom has strand displacement ability; it would be interesting to know how a system without a PE could repair the damage done in such a situation. One gap-filling gel result appears to show a +2 addition to a 1-nt gap, but the author makes no comment (de Vega, 2013). A fascinating feature of the Bsu NHEJ system is that Bsu LigDom has an incredibly relaxed requirement for nick ligation, and is able to seal a variety of mismatched ends and even gaps. Mpa Lig is unable to process the substrates that Bsu LigDom is shown to ligate, although no experiments using mismatched or displaced 5' ends are shown, so it unknown whether Bsu LigDom could actually seal a displaced 5' strand. Such an experiment would be highly interesting, and potentially informative about this AP-NHEJ outlier.

Other more minimal NHEJ systems have been detected; the species that has been longest known for possessing LigD-like ligase and Ku genes is *Archaeglobus fulgidus*. Whether the Afu proteins are capable of end joining has long been speculated, and even doubted since no NHEJ Pol or PE genes are present within the genome. Results described above demonstrate that the Afu ligase and Ku operate like other orthologous AP-NHEJ proteins, despite the lack of an NHEJ Pol or PE. The genomes of three other members in the Archaeoglobae family have recently been sequenced (*A. profundis, A. veneficus* and *A. sulfaticallidus*), and the latter two possess a LigD-like ligase gene adjacent to a Ku gene (*A. profundis* appears to have a LigB-like gene, but no Ku). Interestingly, all three of the newly sequenced species possess a PE gene that is highly similar to the Mtu and Mpa PE, whilst none had an NHEJ Pol (Appendix B). In *A. profundis*, the PE domain is found adjacent to a DnaG type primase. There is no information on whether these two genes produce proteins that interact, but the prospect of a PE gene next to a different kind of primase gene is provocative. Could a DnaG primase fulfill the role that an evolved NHEJ Pol satisfies in other species? There is increasing information about primases in archaea having evolved unusual activities that are not strictly associated with that of a DNA/RNA primase. Evidence has emerged from research in the hyperthermophilic archaeon *Sulfolobus sulfataricus* (Sso) that the PriSL complex is capable of extending a primer across two discontinuous ends, described as polymerisation across discontinuous templates (PADT) (Hu et al., 2012). This activity is distinct from that discerned in the AP-NHEJ systems, since the Sso PriSL complex synthesises DNA in a processive manner, and can create incredibly long DNA strands that are capable of joining two long single stranded DNA ends. The NHEJ Pol, in

comparison, is distributive and can only extend primers by short stretches, and is more suited to gap filling. However, Huang and colleagues believe that Sso PriSL may play a role in DSB repair (Hu et al., 2012). Combined with our knowledge on the NHEJ Pol that has evolved from AEP origins, there is a precedent for a functionally analogous set of primase-polymerases that play a role in the repair in the DSBs. As yet there is no evidence for DnaG primases sharing this capability, but one must wonder if the arrangement of DnaG and PE in *A. profundis* is mere chance. Although none of the Archaeglobae PE have been characterised, and *A. fulgidus* survives without one, the role of the protein seems adrift without an associated polymerase.

The major remaining questions on AP-NHEJ are: why is the *in vitro* end-joining function inefficient relative to the gap filling and nick sealing? Secondly, what other components also play a regular role in AP-NHEJ? The first question may lead onto the second, in that missing components may play a vital role in stabilising the end-joining reaction. However the first query can also be taken at face value. The abilities of all three Mpa NHEJ enzymes are effective when using annealed break substrates, but when using discontinuous ends the efficiency drops remarkably. One key difference is the inclusion of Mpa Ku, which is essential for all of the end-joining reactions. However, Mpa Ku was included with all types of annealed break assays and did not reduce the catalytic functions of any of the proteins, until saturating conditions of Ku were reached and the DNA became coated in the ring-shaped dimer (data not shown). More over, we would not expect Ku to have a negative effect on ligation efficiency. Given that Mpa Pol and PE both seem effective during end-joining assays, and the major difference is the slow ligation, it is likely that Mpa Lig is the rate-limiting step of end joining. Evidence from Mycobacterium and Agrobacterium Ligase C and Ligase D, along with eukaryotic viral DNA ligases, has shown that these ligases have a tendency to adenylate the 5'-P and then dissociate from the substrate (Gong et al., 2004; Shuman, 1995; Zhu & Shuman, 2007; Sriskanda & Shuman, 1998). If an excess of ATP is present in the reaction, the dissociated ligase immediately forms a ligase-AMP complex and cannot rebind the 5'-adenylated phosphate. This might be a factor in the inefficiency of Mpa Lig in joining ends, especially given the necessity to include a NTP mix for RNA polymerisation. However in assays designed to require only the addition of a single ribonucleoside that isn't ATP, ligation was not noticeably more efficient than in reactions that contained a mix of NTPs (Figure 4.27).

When considering more efficient end-joining reactions using the eukaryotic DNA LigIV, it is noticeable that the enzyme is always accompanied by XRCC4 (Gu, Lu, Tippin, et al., 2007a). XRCC4 is thought to form a scaffold supporting LigIV and other repair enzymes, and may play a

role on stabilising the broken ends. No bacterial or archaeal orthologue is known to exist for this enzyme, although it is not inconceivable that there are structural factors that help to stabilise DNA ends and/or proteins during end joining. Bacteria and archaea are known to have DNA packaging proteins equivalent to histones, and these elements may also contribute to the stability and positioning of DNA during repair processes (White & Bell, 2002; Drlica & Rouviere-Yaniv, 1987). Another component of AP-NHEJ that remains almost completely unknown is the regulation of the pathway. In eukaryotes the signal transduction pathway from break recognition down to phosphor-regulation of the repair enzymes is well studied. Although serine/threonine protein kinases are known to exist in bacteria and archaea, it is not known if they play a role in DSB repair (Leonard et al., 1998). Whether or not kinases are involved in a kind of signal transduction to induce AP-NHEJ is speculation, however there must be an enzyme which functions in 5'-phosphoryl transfer of polynucleotides. In eukaryotes this role is performed by PNK, an enzyme that harbours both 3'-phosphatase and 5'-kinase functions. The PE in AP-NHEJ most likely fulfills the role of the 3'-phosphatase, but has no kinase domain. Addition of a phosphate to a 5'-OH is as essential a function as scission of a 3'-phosphate to reveal an OH, and yet, curiously, no 5'-kinase is found to be associated along with NHEJ operons in bacteria or archaea. It may be that these enzymes are valuable for a number of DNA repair pathways, like PNK, and are elsewhere in the genome. It is possible that these enzymes have a unique structural conformation, like PE, and thus avoid easy identification by either sequence homology or threading searches.

In conclusion to the results of to these two chapters concerning the pivotal enzymes of AP-NHEJ, we have an increased understanding of the capabilities of the NHEJ Pol, and a new concept of what the PE does during the NHEJ repair of DSBs. We can begin to piece together potential mechanisms of repairing different kinds of DSBs with relative detail. Tasks that remain are to discover other enzymes that influence the process of AP-NHEJ, and to capture structures of the canonical NHEJ enzymes engaging with their preferred substrates. The following chapter covers the structural studies in *M. paludicola* that sought to reveal more information on AP-NHEJ.

Chapter 5

Structural studies of the components of the *Methanocella paludicola* NHEJ repair complex

5.1 Introduction

The efforts to obtain atomic structures of the *M. paludicola* NHEJ complex began alongside the biochemical work. The key initial key target was to crystallise the PE since, at the time, the structure of the protein was unknown. Both bacterial and archaeal structures of a PE were published by another research lab prior to the completion of the Mpa PE structure (Nair et al., 2010; Smith et al., 2011). After determining the preferred substrate of Mpa PE (detailed in Chapter 4), attempts were made to crystallise the PE in the presence of several short DNA-RNA substrates. A structure of substrate bound to PE has proven elusive for other laboratories, whom have moved onto NMR solution structures. However, it was hoped that knowledge of the preferred substrate specificity in the presence of the D-strand would allow more progress with X-ray crystallography (Natarajan et al., 2012).

The NHEJ DNA polymerase domain has been the subject of a large number of structural studies; alone, in complex with nucleotides, and with several DNA substrates that illustrate annealed breaks (Pitcher, Brissett, Picher, et al., 2007b; Brissett et al., 2007; Brissett et al., 2011; Akey et al., 2006; Zhu et al., 2006). One aim of the research with the Mpa NHEJ proteins was to crystallise a complex of them together, to demonstrate how they interact. This work was to begin with Mpa Pol and Mpa Ku, bound to DNA. This offered a two-fold incentive, since no structure of bacterial or archaeal Ku yet exists, and no AP NHEJ proteins have been crystallised in complex together.

The final target of structural studies in *M. paludicola* was the NHEJ DNA Ligase. Whilst an apo structure could prove interesting to compare to the *M. tuberculosis* LigDom of LigD, the key target was to capture the enzyme engaging the DNA-RNA nick substrate (Akey et al., 2006). Structures exist of other DNA ligases bound to their relevant substrates, and it was hoped that a structure of the NHEJ ligase would explain its specific necessity for RNA at the 3' side of the nick (Nandakumar et al., 2007; Pascal et al., 2004). This chapter explores our attempts to generate novel atomic structures of the AP-NHEJ components, using a variety of DNA substrates described in Figure 5.1.

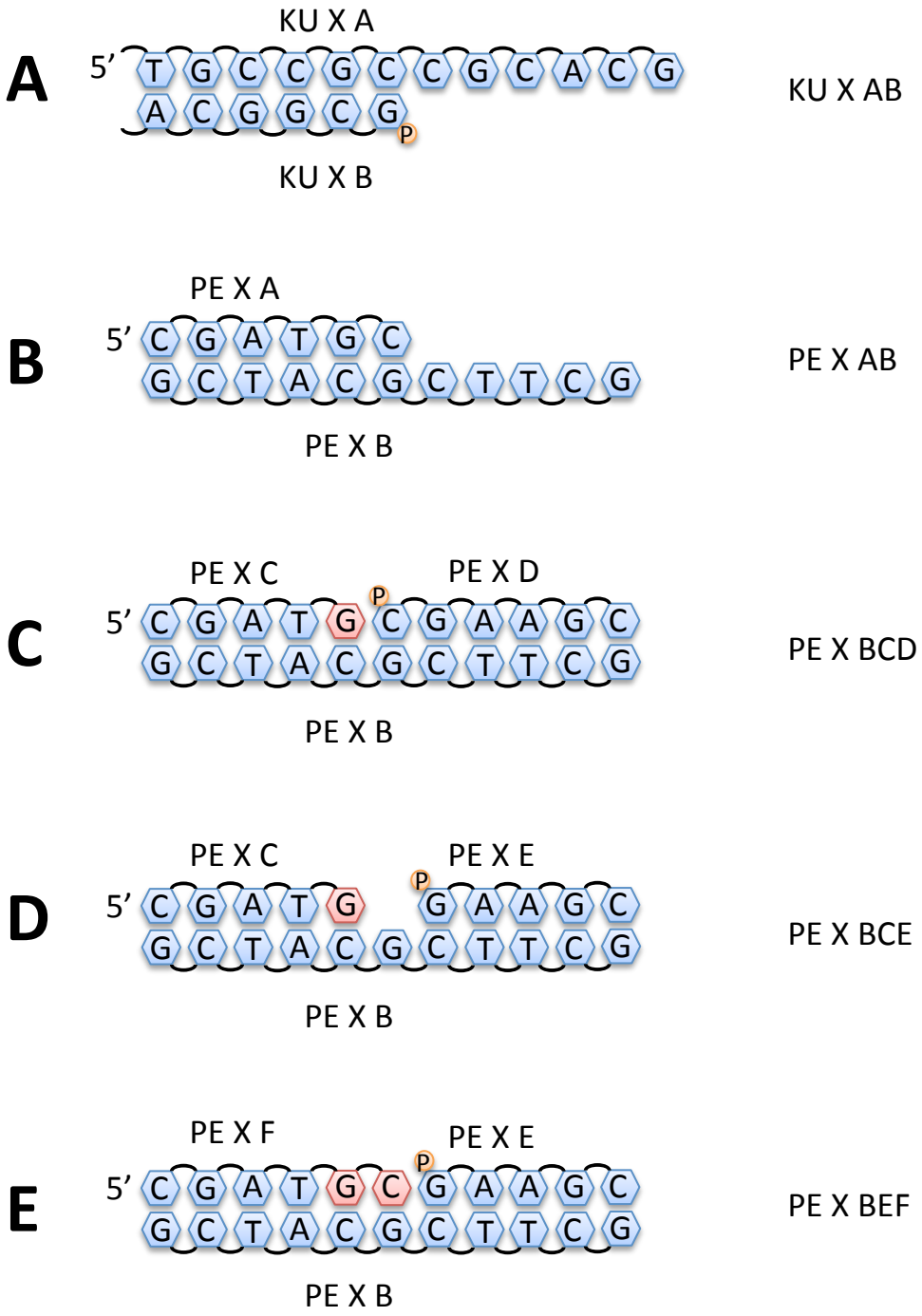


Figure 5.1 DNA substrates used for crystallography.
Schematics of DNA substrates used for crystal screening with Mpa Lig, Pol, Ku and PE. (**A**) KU X AB is composed of a 37-mer (KU X A) and a 31-mer (KU X B) with a 5’-phosphate, creating a recessed 5’ end of 6 bases. (**B**) PE X AB is composed of a 6-mer (PE X A) and an 11-mer (PE X B) generating a 3’-overhang. (**C**) PE X BCD is composed of a 5-mer (PE X C) with a 3’-ribonucleoside, a 6-mer (PE X D) with a 5’-P, and an 11-mer (PE X B). These oligos generate a nick substrate which is amenable to ligation by Mpa Lig. (**D**) PE X BCE is composed of a 5-mer (PE X C) with a 3’-ribonucleoside, a 5-mer (PE X E) with a 5’-P, and an 11-mer (PE X B). These oligos generate a 1-nt gap substrate. (**E**) PE X BEF is composed of a 6-mer (PE X F) with two 3’-ribonucleosides, a 5-mer (PE X E) with a 5’-P, and an 11-mer (PE X B). These oligos generate a substrate which can be ligated by Mpa Lig or cleaved by Mpa PE.

5.2 Crystallisation screening of Mpa Ku

5.2.1 Crystal screening studies of the Mpa Ku protein

Prior to crystallisation screening, all protein solutions were examined with the Pre-Crystallisation Test (Hampton Research) to discern a reasonable starting concentration of protein for crystallisation trials. All crystal screens were set as sitting drops and incubated at 4 °C unless otherwise stated. Initially, Mpa Ku alone was tested, and appeared to be in an acceptable concentration range at 4.7 mg/mL. An initial screen at a 1:1 protein to reservoir ratio with 0.7 µL protein using the Qiagen PEGs suite produced heavy precipitation in almost all drops, suggesting that the protein was quickly precipitating. In order to avoid this rapid precipitation the protein concentration was lowered. Several conditions of the PEGs suite were repeated with Mpa Ku at half the previous concentration, now 2.35 mg/mL. These drops produced much less precipitation, and some remained clear, indicating that Mpa Ku might be more amenable to crystallisation at relatively low protein concentration levels.

Further screening was conducted at 2.35 mg/mL, and a single hit was detected in the Hampton Research Crystal Screen. Very small crystals were observed that were prone to dissolving at room temperature (Figure 5.2, A). Whilst the crystals reappeared after further incubation at 4 °C, they did not continue growing. The crystals observed only grew to ~ 10 µm, and although the crystal growth conditions could be repeated, they were all unstable and too small for data collection. A grid was designed around the condition that produced the hit; 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodoylate trihydrate pH 6.5, 30% (v/v) 2-Methyl-2,4-pentanediol (MPD). The grid varied the concentration of Magnesium acetate tetrahydrate from 0.15-0.25M, and the concentration of MPD from 10-40 % (Figure 5.2, B). The customised screen did replicate the small crystals produced before, all in the range of 30 % (v/v) MPD, but at varying concentrations of magnesium acetate tetrahydrate. The crystals did not grow any further, and were not suitable for data collection.

5.2.2 Mpa Ku and DNA crystal screening with Mpa Pol

The DNA used for crystallographic experiments with Mpa Ku was designed as a ds 31bp duplex with a 6 base 3’ overhang and a 5’ phosphate on the recessed end, labelled KU X (Figure 5.1, A). The only structure of any Ku dimer bound to DNA utilised a clover-leaf like formation of DNA hairpins at one end in order to trap Ku (Walker et al., 2001). The design of the KU X AB substrate intended for the NHEJ Pol to bind to the recessed 5’-P on the KU X B strand. The

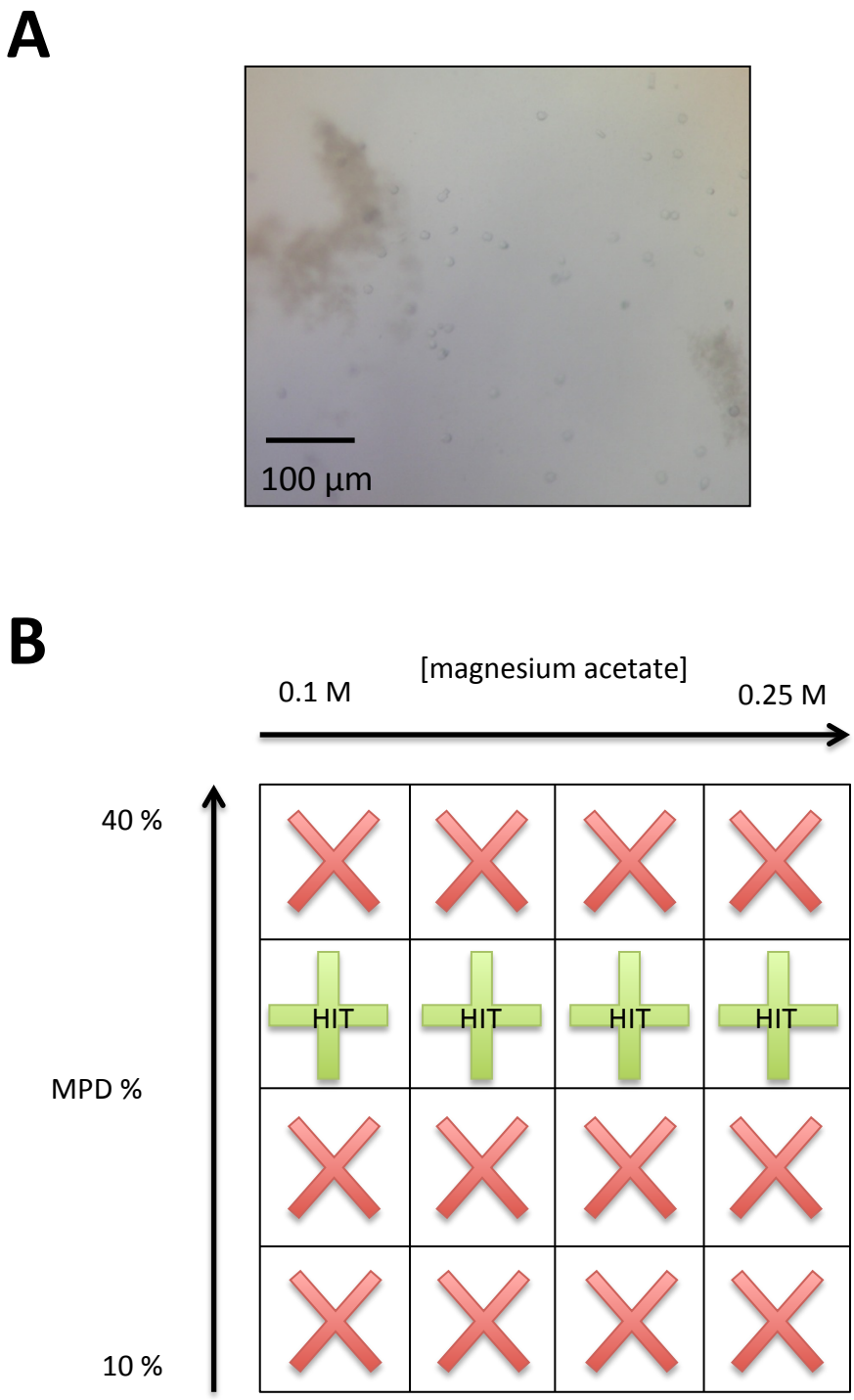


Figure 5.2 Attempts to crystallise Mpa Ku.
Crystals produced during screening with Mpa Ku, and a grid used to optimise crystal size. (**A**) Mpa Ku crystals grown using Crystal Screen (Hampton Research). 0.7 µL protein: Mpa Ku 2.35 mg/mL, 0.7 µL reservoir: 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate ph 6.5, and 30 % (v/v) MPD. Crystals grew after a week, but did not grow visibly beyond the initial period during nucleation. (**B**) A grid design to explore variations around the condition which produced the initial Mpa Ku crystal hit. Magnesium acetate was varied from 0.1 to 0.25 M and MPD from 10 – 40 %. Similar small crystals were observed in all conditions at 30 % (v/v) MPD, although all were very small, and none grew beyond the size initially observed.

NHEJ Pol could then act as a tether at one end (like the clover-leaf DNA in the Walker paper), preventing the Ku dimer from translocating and sliding off of the duplex. The DNA molecule was designed to encompass a potential 14 bases that traversed the centre of the *H. sapiens* Ku structure, and another 16 bases to encompass the NHEJ Pol and engage the phosphate-binding pocket of the NHEJ Pol.

The initial concentrations of Mpa Ku, Mpa Pol and KU X AB for crystal screening were tested using the PCT, with a ratio of 1:1:0.5 for Ku:DNA:Pol at 125 µM Ku. Screening was performed using Qiagen Nucleix, PEGs and MPD suites. Most drops produced mild precipitation, but no crystals. Any significant increase of complex concentration yielded heavy precipitation in most drops. Extended screening did not produce any detectable crystals.

A small sample of Mpa Ku and Mpa Pol without DNA was screened using the PEGs suite, at concentrations of 171 µM (5.75 mg/mL) Ku and 85 µM (2.85 mg/mL) Pol. The screen had more positive results than the same screen with DNA, and did not produce excessive precipitation. Several crystals grew, and reached sizes of ~50 µm, however the drops featured high nucleation with many small crystals. Two techniques were used to attempt to reproduce the crystals, except with a lower nucleation rate. First, the conditions that produced crystals were repeated [0.2 M sodium sulphate with 20 % (w/v) polyethylene glycol [PEG] 3350, 0.2 M ammonium sulphate with 20 % (w/v) PEG 3350] except using lower protein concentrations. Second, the initial drops were repeated as before but were streak seeded. Streak seeding is a method of collecting tiny fragments or 'seeds' of an existing crystal and transferring them into a new protein drop. The seeds are nucleation sites for crystal growth before the rest of the protein reaches the limit of solubility and starts to produce additional nucleation. A cat whisker was used for streaking, since they have suitable surface properties to remove small fragments of a crystal. Streaking involves simply wiping the whisker over the crystals, and then wafting the whisker into the fresh protein drop. The crystal streaking was successful, with new crystals that grew overnight to ~200 x 25 x 25 µm (Figure 5.3, A). The conditions that the crystals were grown in were very similar to those which produced the original Mtu PolDom crystals, so there was a concern that the crystals contained Mpa Pol only, and not a stable interaction of Ku and Pol (Pitcher, Brissett, Picher, et al., 2007b). Several crystals were gathered for data collection, and the space group and initial unit cell estimations $P2_1$, 44.42/60.55/59.41 [a, b, c (Å)], and 90.00/101.02/90.00 [α, β, γ (°)] concurred with the suggestion that the crystals contained Mpa Pol only. The data was processed, but this will be discussed in the following section concerning the NHEJ polymerase only.

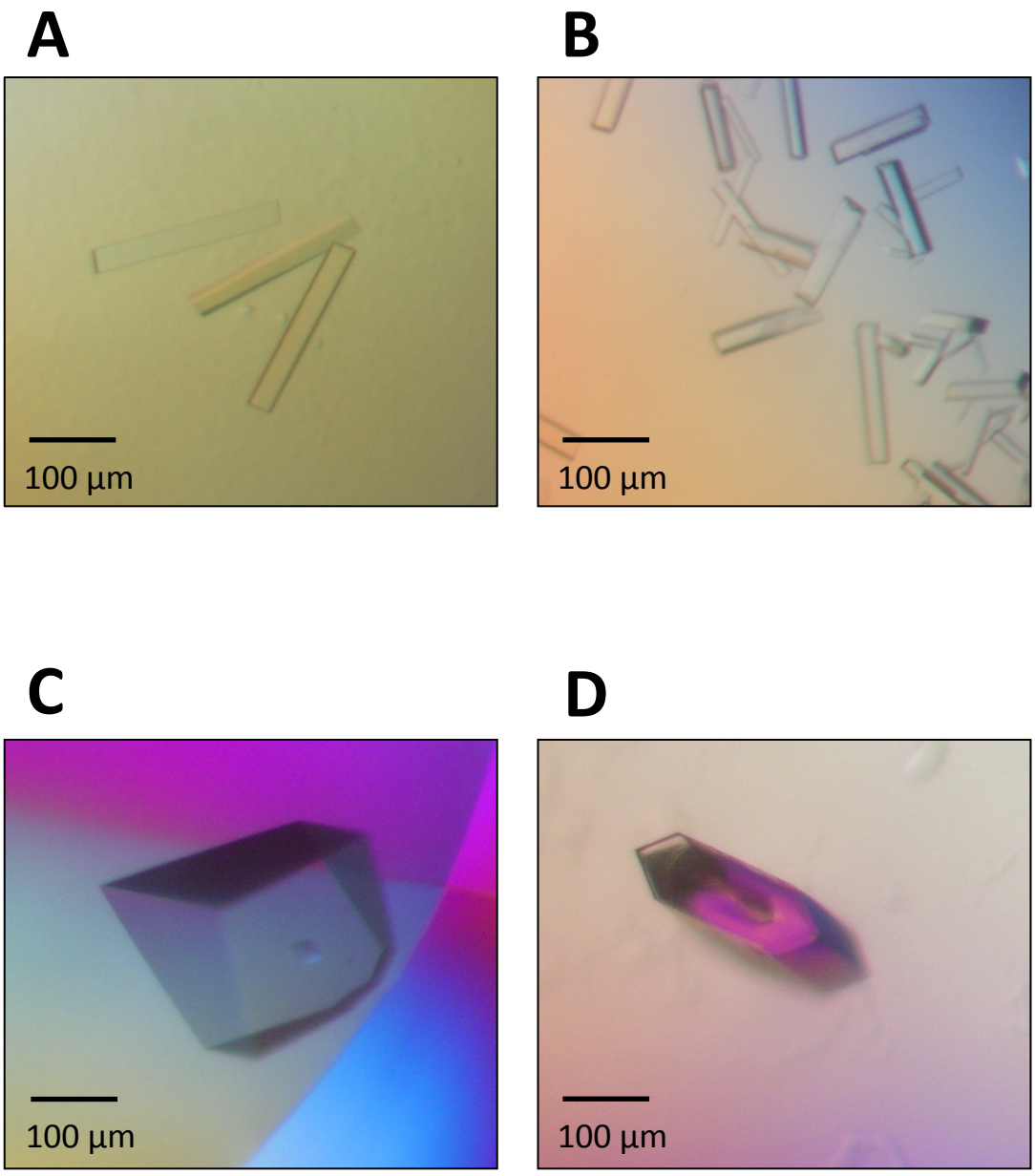


Figure 5.3 Mpa Pol crystals.
Crystals produced during screening for Mpa Pol using the PEGs suite (Qiagen). (**A**) 0.7 µL protein: Mpa Pol 2.75 mg/mL, Mpa Ku 4.7 mg/mL. 0.7 µL reservoir: 0.2 M Ammonium sulphate, 20 % (w/v) PEG 3350. These crystals were ultimately used to resolve the structure of Mpa Pol, but no electron density was observed for Mpa Ku. (**B**) 0.7 µL protein: Mpa Pol 3.825 mg/mL. 0.7 µL reservoir: 0.2 M Ammounium sulphate, 20 % (w/v) PEG 3350. (**C**) 0.7 µL protein: Mpa Pol 7.65 mg/mL. 0.7 µL reservoir: 0.2 M Ammounium sulphate, 20 % (w/v) PEG 3350. (**D**) 0.7 µL protein: Mpa Pol 7.65 mg/mL, KU X DNA 250 µM. 0.7 µL reservoir: 0.1 M Citric acid pH 4.0, 40 % (v/v) MPD.

5.3 Crystallisation of Mpa NHEJ Polymerase

Concurrent to the screening with Mpa Ku, Mpa Pol and DNA, the NHEJ Pol was also screened alone. A PCT of Mpa Pol protein solution suggested a screening concentration of 5 mg/mL. The PEGs suite was chosen as the first screen, and using a 96 well plate with two sitting drop wells two protein concentrations were screened; 3.82 mg/mL and 7.65 mg/mL. Rod shaped crystals grew overnight (100 µm in length, but only 20 µm width and depth) at the lower concentration in the condition of 0.2 M ammonium sulphate and 20 % (w/v) PEG (Figure 5.3, B). After two days, a single very large crystal (~300 x 200 x 200 µm) grew in the same condition with the higher protein concentration (Figure 5.3, C). The larger crystals were repeated, and several large Mpa Pol protein crystals were grown. They were suitable for data collection, and several data sets were collected, although ultimately the data gathered from the protein crystal grown in the solution with Mpa Pol and Mpa Ku was used to generate the atomic structure of Mpa Pol.

5.3.1 Mpa Pol and DNA crystal screening

Two screens of Mpa Pol and KU X AB were generated, using the PEGs and MPD crystal screening suites. The screens were used as a comparison to those including Mpa Ku, which seemed to be causing heavy precipitation even at low protein concentrations. Mpa Pol was set at 7.65 mg/mL at 1:1.2 to KU X AB (208 / 250 µM). No excessive precipitation was observed in the drops, suggesting that Mpa Ku, and not the DNA, was the cause of the precipitation observed in prior screens. No crystals were observed in the weeks following set-up of the trays, however after several months a single crystal was observed in the MPD screen (Figure 5.3, D). The crystal had dimensions of ~200 x 75 x 300 µm and was grown in 0.1 M citric acid pH 4.0 with 40% (v/v) MPD. This condition requires future screening to attempt to replicate the crystal growth before data collection.

5.3.2 Data collection, processing and refinement of Mpa Pol

The data for Mpa Pol were collected in house from the crystals described in section 5.2.2 using a Rigaku MicroMax 007-HF. The diffraction patterns showed spots out to the edge of the detector, up to a resolution limit of 1.94 Å (Figure 5.4, A, and B). The diffraction images were indexed using iMOSFLM, and the space group was chosen based on the selection with the highest symmetry and the lowest penalty score (Battye et al., 2011). Cell refinement and

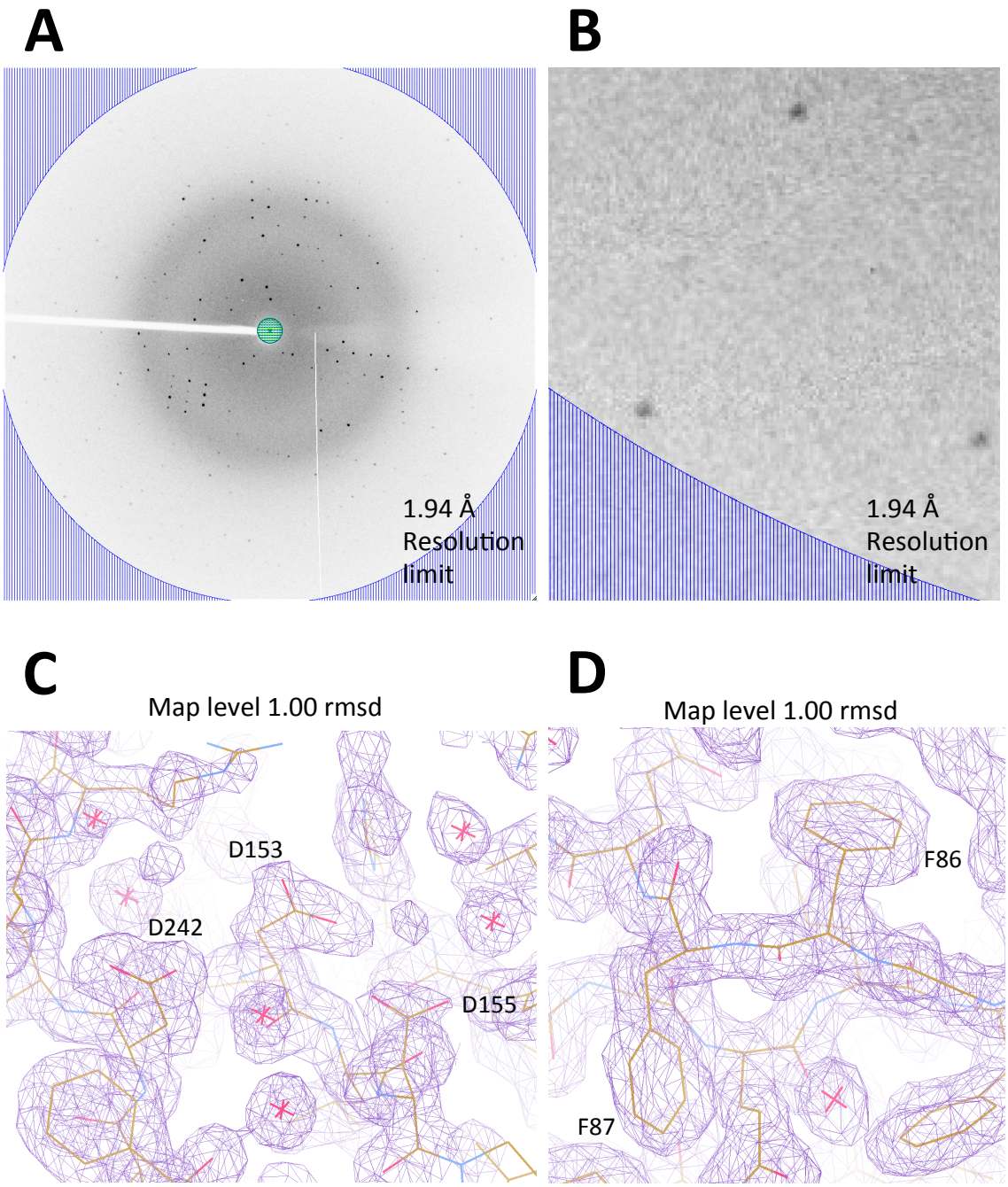


Figure 5.4 Diffraction patterns and electron density map of Mpa Pol.
(**A, B**) Diffraction patterns from a Mpa Pol crystal collected on the in-house Rigaku MicroMax HF-007. (**A**) A complete image showing the diffraction pattern of Mpa Pol with the outer blue edge indicating the 1.94 Å resolution limit. The small blue and green circles indicate the 17.00 Å lower resolution limit, and the back stop cut off limit. (**B**) A magnified view of diffraction spots out to the edge of the resolution limit. Images of diffraction pattern were made using iMOSFLM (Battye et al., 2011). (**C, D**) Electron density maps (purple mesh) over the structure of Mpa Pol displayed as sticks (yellow). Water molecules can be seen as red crosses in small spheres of purple mesh. (**C**) A view of the active site of Mpa Pol, with the catalytic aspartate triad (D153, D155 and D242) shown with the electron density map. (**D**) A view of the active site of Mpa Pol, with the twin phenylalanine residues (D153, D155 and D242) shown with the electron density map. Images of density maps were made using Coot from the CCP4 suite (Emsley et al., 2004).

integration was also completed using iMOSFLM. The remainder of the processing was completed using the programs from the CCP4 suite (E. Potterton et al., 2002; L. Potterton et al., 2004; Collaborative Computational Project, Number 4, 1994). The Laue group symmetry was additionally confirmed using the program Pointless. The average reflection intensities for each image were produced by Scala and the key statistics (with the values from the highest resolution shell shown in brackets) were R_{sym} (%): 0.71 (0.20), Redundancy: 3.5 (2.6), and Overall I/(σI): 25.1 (4.9) (Evans, 2006; Evans, 2011). Crystals of Mpa Pol contained one protein molecule in the asymmetric unit, giving a V_M of 2.34 $Å^3$/Da with 47.45 % solvent content.

Molecular replacement was used to determine the phases for solving the structure of Mpa Pol, using the program PHASER with the Mtu PolDom PDB 2IRU as the search model (McCoy et al., 2007; Pitcher, Brissett, Picher, et al., 2007b). Of multiple potential solutions, one stood out based on the statistical parameters output by the program. The MR phases and solution model were used as a starting point for automated building, using ARP/wARP (Cohen et al., 2008). The output of ARP/wARP underwent 3 rounds of model building and refinement prior to the final structural refinements. After each round of manual remodelling using the program Coot, final structural refinements were performed with REFMAC5 (Vagin et al., 2004; Murshudov et al., 1997; Murshudov et al., 2011; Emsley & Cowtan, 2004). The Mpa Pol structure and electron density map (with a map level of 1.00 rmsd) showing a reliable fit around several key residues can be seen in Figure 5.4, C and D.

A final refined model at 1.95 Å resolution, with an R_{cryst} of 18.95% and R_{free} of 22.87%, was obtained. Over 98% of residues in the structure are in the favoured region of the Ramachandran plot. The data collection and refinement statistics for Mpa Pol are shown in Table 5.1.

5.4 The structure of the NHEJ Mpa Polymerase

The structure of Mpa Pol was solved to 1.95 Å resolution and is shown in Figures 5.5 and 5.6. Mpa Pol shares a very strong structural resemblance to both polymerase domains from *M. tuberculosis* and *P. aeruginosa*. Based on the structural similarities between Mpa Pol and Mtu and Pae PolDoms, we can assume that the key structural elements (Loop 1, Loop 2, 5'-P binding pocket, catalytic residues) are the same. Mpa Pol is primarily composed of eight α-helices surrounding two triple-stranded anti-parallel β-sheets (β3, β4, β7 and β8, β9, β10) that encompass the open active site (Figure 5.5, A and B). There are two additional β-hairpins, each

Data collection	
Source	Rigaku MicroMax 007-HF
Space group	P21
Unit cell dimensions	
a, b, c (Å)	44.42/60.55/59.41
α, β, γ	90.00/101.02/90.00
Wavelength (Å)	1.5418
Resolution (Å)	13.934-1.949
Total number of observations	77013
Number of unique reflections	22140
Overall I/(σI)[a]	25.1 (4.9)
Overall completeness (%)[a]	91.1 (84.0)
R_{sym} (%)[a,b]	0.71 (0.20)
Redundancy[a]	3.5 (2.6)
Refinement	
Resolution(Å)	13.934-1.949
Number of reflections	20996
R_{factor} / R_{free} [c,d]	0.1895/0.2287
Contents of asymmetric unit	1 protein molecule
Number of atoms	
Protein	2506
Water molecules	154
Mean B value (Å^2)	14.379
Rmsds	
Bonds (Å)	0.02
Angles (°)	1.98
Ramachandran statistics	
Favoured regions (%)	98.59
Allowed regions (%)	1.41
Disallowed regions (%)	0

Table 5.1. Data collection and refinement statistics of Mpa Pol.
[a] Values for highest resolution shell (2.05–1.95 A˚) is in parentheses. [b] *R*sym = Σ|*I* – <*I*>| / Σ <*I*>, where *I* is the observed intensity. [c] *R*factor = Σ||*F*o – |*F*c|| / Σ|*F*o|, where *F*o and *F*c are the observed and calculated structure factor, respectively. [d] *R*free is equal to R factor for a randomly selected 5% subset of reflections not used in the refinement.

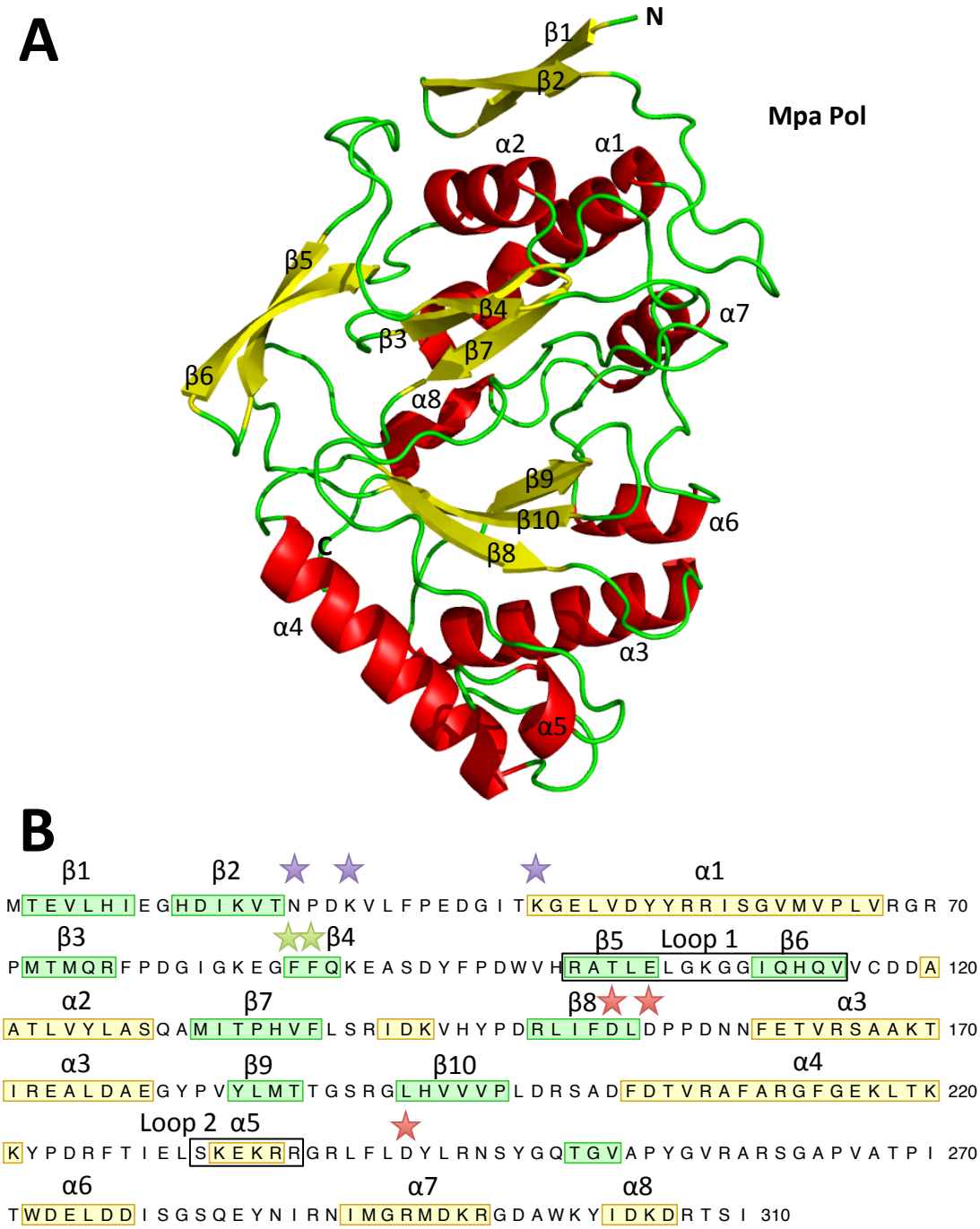


Figure 5.5 Structure of Mpa Pol.

(**A**) The structure of Mpa Pol at 1.95 Å resolution. The secondary structure elements are labelled in order from N terminus to C terminus. Mpa Pol is composed of ten distinct β-strands and 8 α-helices. The β-strands form two antiparallel sheets, which cradle the active site (β8, 9, 10 and β3, 4, 7), and two β-hairpins. Loop 1 comprises β5 and β6 and protrudes away from the DNA binding surface. β1 and β2 form the outside edge of the phosphate binding pocket along with α1 and α2. Loop 2 is a flexible loop region with a helical element, α5, which points towards the active site. (**B**) An amino acid sequence of Mpa Pol with secondary structure elements highlighted. β-strands are shown in green, α-helices are yellow. Purple stars indicate conserved phosphate binding residues, green stars indicate the DNA template twisting phenylalanine residues, and the red stars indicate the catalytic aspartate triad. The amino acid sequence with secondary structure labels was generated using UCSF Chimera (Pettersen et al., 2004).

comprising two antiparallel β-strands that sit on the periphery of the enzyme. The first two β-hairpins (β1 and β2) compose the outside edge of the 5'-P binding pocket, which is completed by two α-helices. The second periphery β-hairpin (β5 and β6) protrudes from the surface of the enzyme on the edge of the active site, and composes Loop 1. Loop 2 is composed of a short α-helical region (α5) on a loop that is in contact with the active site.

The surface representation of Mpa Pol reveals an obvious active site groove in between Loop 1 and Loop 2, and an indentation for the phosphate binding pocket that is divided from the active site by a 'wedge' structure (Figure 5.6, A). A ribbon diagram of Mpa Pol is shown beneath the surface model in the same orientation, in order to indicate which structural elements form key elements of the surface. The active site of Mpa Pol is situated in the open cleft, and the catalytic residues are positioned on the β-sheet shown in Figure 5.7, A and B. The conserved AEP aspartate triad and histidine are shown coloured blue. The positioning of these residues is virtually identical to those seen in Mtu and Pae PolDom. The manganese ions from the bacterial structures can be superimposed for an excellent fit with the Mpa Pol active site, along with the incoming nucleotide (Zhu et al., 2006; Pitcher, Brissett, Picher, et al., 2007b).

Opposite the β-sheet that contains the catalytic residues is another three-stranded anti-parallel β-sheet, and the outside edge of this sheet forms a ridge that separates the active site from the 5'-P binding region. On the cusp of this ridge two phenylalanine residues are positioned, protruding in opposite directions (coloured blue, Figure 5.7, A and 5.8, A). These residues are strictly conserved in the NHEJ polymerases (shown in Figure 1.16). These residues form a wedge, the role of which is to position the template DNA strand for base pairing with incoming nucleotides. The first phenylalanine, F86 in Mpa, causes a kink in the DNA backbone and together with F87 forces the templating base to face towards to the nucleotide-binding pocket in the active site. The aromatic hydrocarbon ring of the phenylalanine base stacks with the DNA, which allows for the manipulation of the strand. Figure 5.8, B shows the surface of Mpa Pol with both phenylalanine residues coloured blue, and with DNA superposed from the Mtu PolDom pre-ternary structure in green. The phenylalanine residues can clearly be seen to form a wedge that positions the DNA template strand, with bases pointing towards the active site and towards the phosphate binding pocket.

Two key surface loops are present on Mpa Pol, Loop 1 and Loop 2. A comparison of these structural features in Mpa Pol and Mtu PolDom is shown in Figure 5.9. These regions have been implicated in template and primer positioning, and regulation of the active site (Brissett

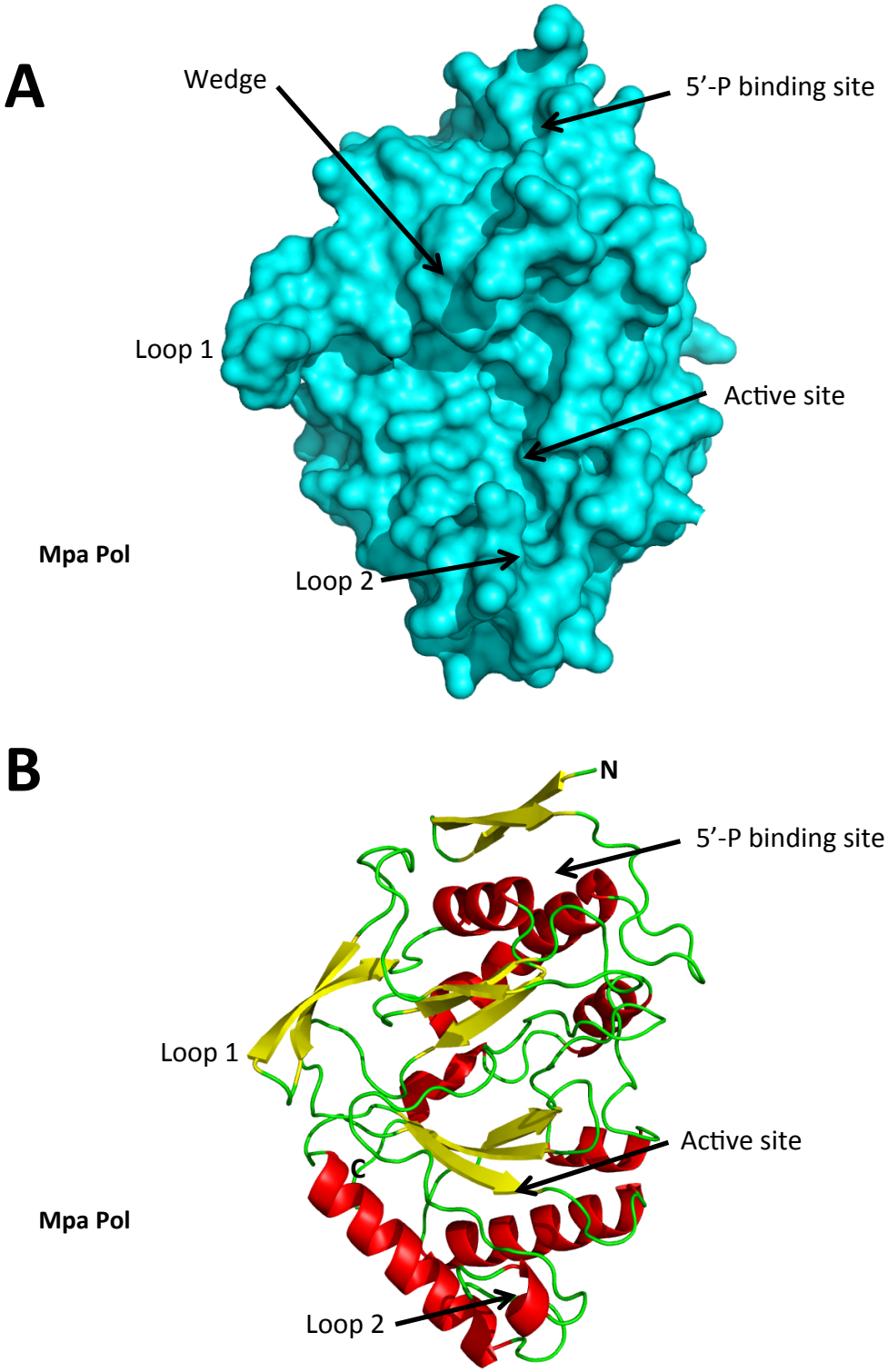


Figure 5.6 Structure of Mpa Pol.
The 1.95 Å resolution structure of Mpa Pol. The structure of Mpa Pol shown as a surface representation (**A**) and in ribbon format (**B**). The postions of Loop 1 and Loop 2 are indicated on the periphery of the protein. The active site is shown as a distinct cavity in the surface representation, and is shown cradled between two anti-parallel β-sheets in the ribbon representation. The phosphate binding pocket is indicated, composed of two α-helices and a β-hairpin. The surface representation clearly shows how the active site and phosphate binding pocket are separated by a structural wedge. This structural element is discussed in more detail later in this chapter.

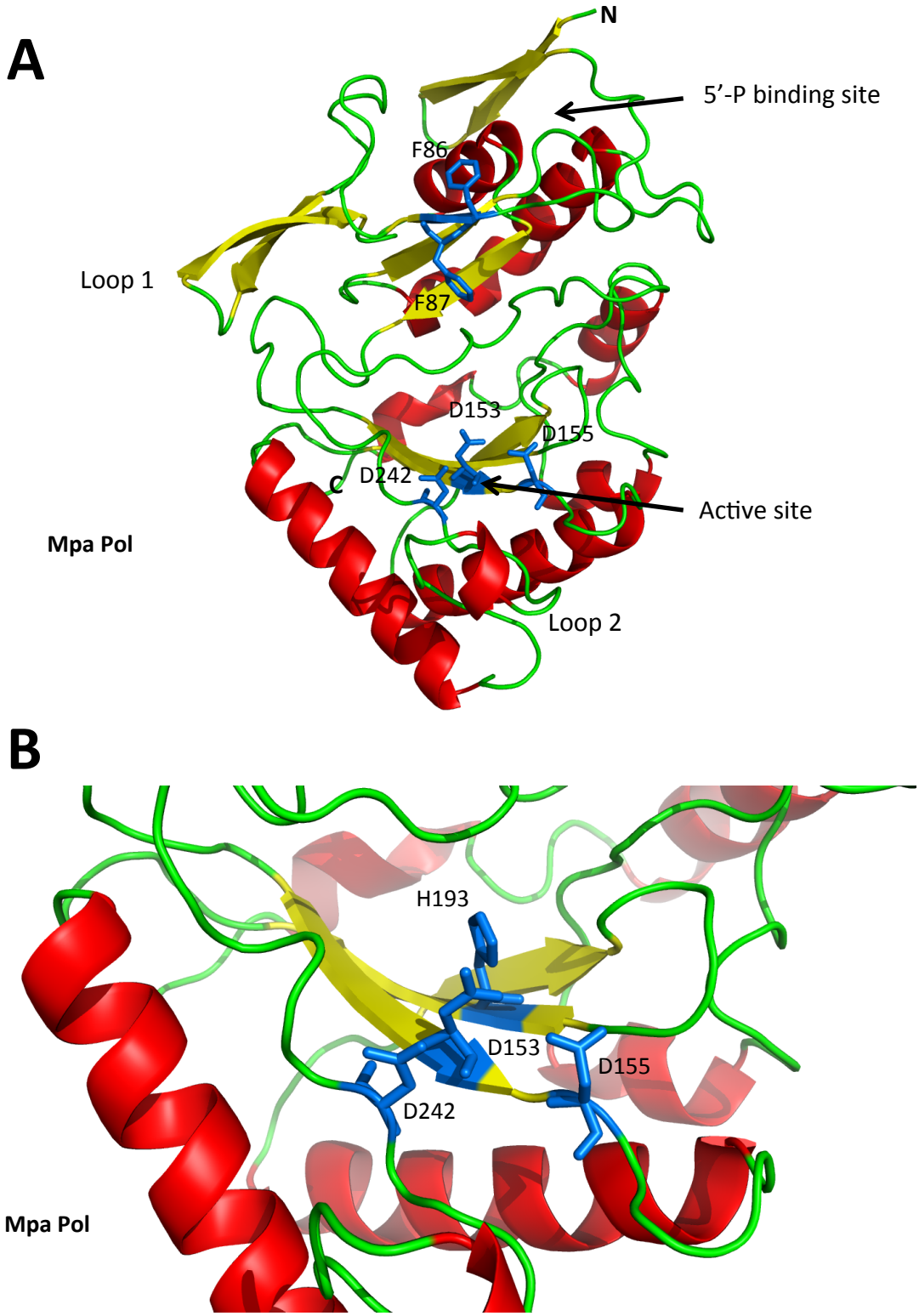


Figure 5.7 Ribbon diagrams and active site detail of Mpa Pol.
(**A**) A ribbon diagram of Mpa Pol with active site detail (D153, D155 and D242) and the two phenylalanine residues which form part of the structure wedge shown in 5.4 A. (**B**) A magnified view of the Mpa Pol active site. The aspartate triad (D153, D155 and D242) which coordinate the metal ion are shown in blue. H193 is also shown in blue, and is a highly conserved residue in NHEJ Pols and is shown to contact the β phosphate of an incoming nucleotide.

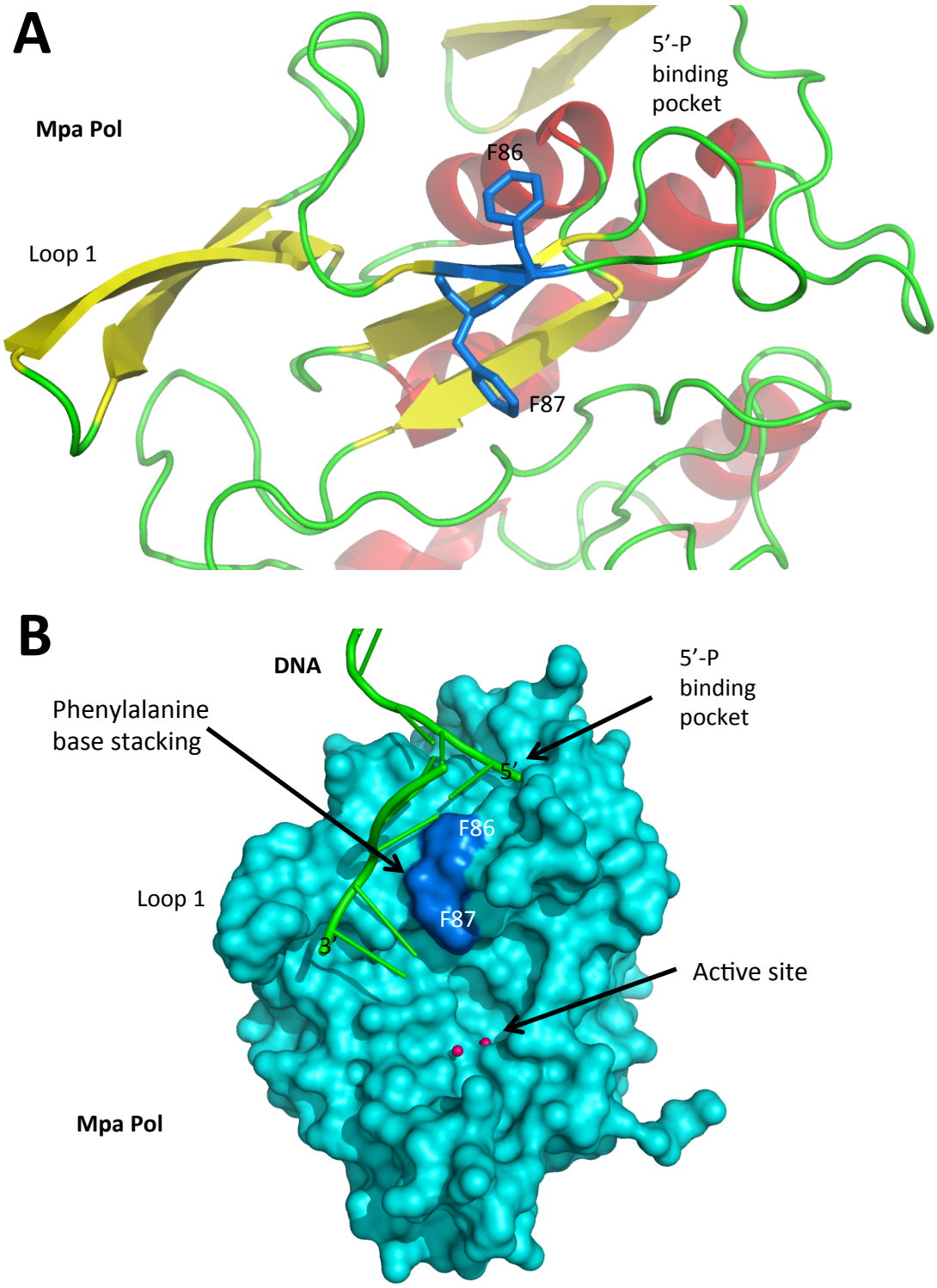


Figure 5.8 Conserved phenylalanine base stacking residues of Mpa Pol.
(**A**) A ribbon diagram of Mpa Pol that shows the twin phenylalanine residues, F86 and F87, sit either side of a ridge that divides the active site from the phosphate binding pocket. (**B**) A surface representation of Mpa Pol with DNA (green) superimposed from the Mtu PolDom pre-ternary structure (Brissett *et al* 2011). The phenylalanine residues are highlighted in blue, and are shown to be in position to base stack with the DNA template, which could splay the DNA template in the same manner as Mtu PolDom. In this representation Loop 1 cradles the DNA template strand, ensuring that the templating base sits opposite the active site (indicated by the manganese ions modelled by superposition from Brissett et al., 2011 Mtu PolDom structure PDB 3PKY).

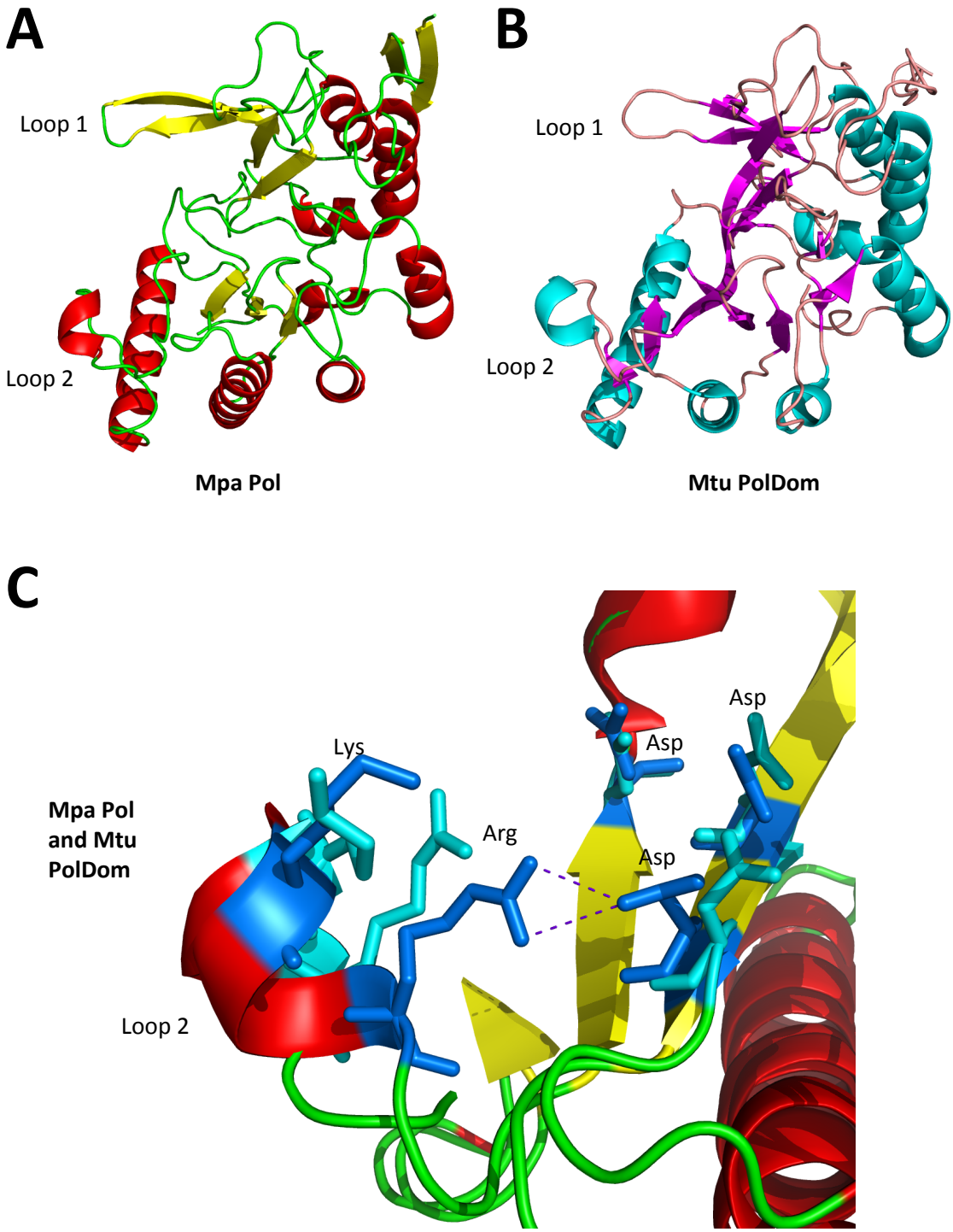


Figure 5.9 Comparing Loop 1 and Loop 2 positions in NHEJ Pol structures.
A comparison of Loop positioning between Mpa Pol and Mtu PolDom. (**A**, **B**) A view of Loop 1 and Loop 2 from Mpa Pol and Mtu PolDom respectively. The loops occupy very similar positions, despite differing in amino acid composition, suggesting that the function of the loops is conserved. (**C**) An overlay of Mpa Pol and Mtu PolDom Loop 2, Mpa residues are coloured cyan, and Mtu residues are coloured blue. Mtu Loop 2 Arg (R220) forms a salt bridge with a catalytic Asp (D139), but this is not seen in the Mpa Pol structure, with the Loop 2 residues adopting a slightly distant conformation. The Mtu PolDom figures are created from the Pitcher (et al., 2007) structure, PDB 2IRU.

et al., 2007; Brissett et al., 2011). Whilst the amino acid residues that compose both Loop 1 and Loop 2 in Mpa Pol are not strictly conserved with those found in Mtu or Pae PolDoms (Figure 3.10), the overall secondary structure and positions are conserved. The Loop 2 residues that face towards the active site, K217 and R220 in Mtu PolDom, are conserved in Mpa Pol, although the lysine is not conserved in Pae PolDom. Only the arginine is seen to directly contact an aspartate of the catalytic residues, by formation of a salt bridge which is subsequently disrupted by the binding of metal ions (Brissett et al., 2011). Interestingly, when the structure of Mpa Pol is overlaid on Mtu PolDom, the Loop 2 arginine (R235 in Mpa Pol) does not form a salt bridge with the aspartate (D155 in Mpa Pol). The salt bridge seen in the Mtu PolDom structure was thought to prevent the proper coordination of the active site until a correct incoming nucleotide was selected by the NHEJ polymerase and Loop 2 was considered a regulator of this function (Brissett et al., 2011). Since the Mpa Pol structure does not contain the relevant metal ions for coordination of the active site, it is difficult to speculate whether this new positioning of Loop 2 is significant.

The phosphate-binding pocket is a conserved structural feature of all observed NHEJ Pols (Figure 3.10 and Figure 5.5, B). The region consists of a small groove on the surface that can accommodate a 5'-P of a DNA strand, a D-strand in the case of an annealed break. The residues that compose the pocket are positively charged, as demonstrated in Figure 5.10, with the DNA from the Mtu PolDom pre-ternary structure positioned into the groove. Both electrostatic surface representations show the phenylalanine base-stacking region dividing the positively charged DNA binding regions around the active site and at the phosphate binding pocket.

5.5 Structural insights on the activities of the Mpa NHEJ Polymerase

5.5.1 5'-phosphate binding by Mpa Pol

5'-P binding is a conserved activity of NHEJ polymerases, and the structural feature that they use to achieve this is a positively charged grouping of amino acid residues that can accommodate a phosphate (Figure 5.11). Although Mpa Pol also possesses the 5'-P binding ability, it binds the substrate with a significantly lower affinity than Mtu PolDom. Figure 5.12, A shows an EMSA with Mtu PolDom significantly retarding the migration of a substrate with a 5'-P compared with Mpa Pol samples. The his-tag of Mpa Pol was not removed prior to this point, but it was possible that it could be interfering with the DNA binding. The 5'-P binding pocket is

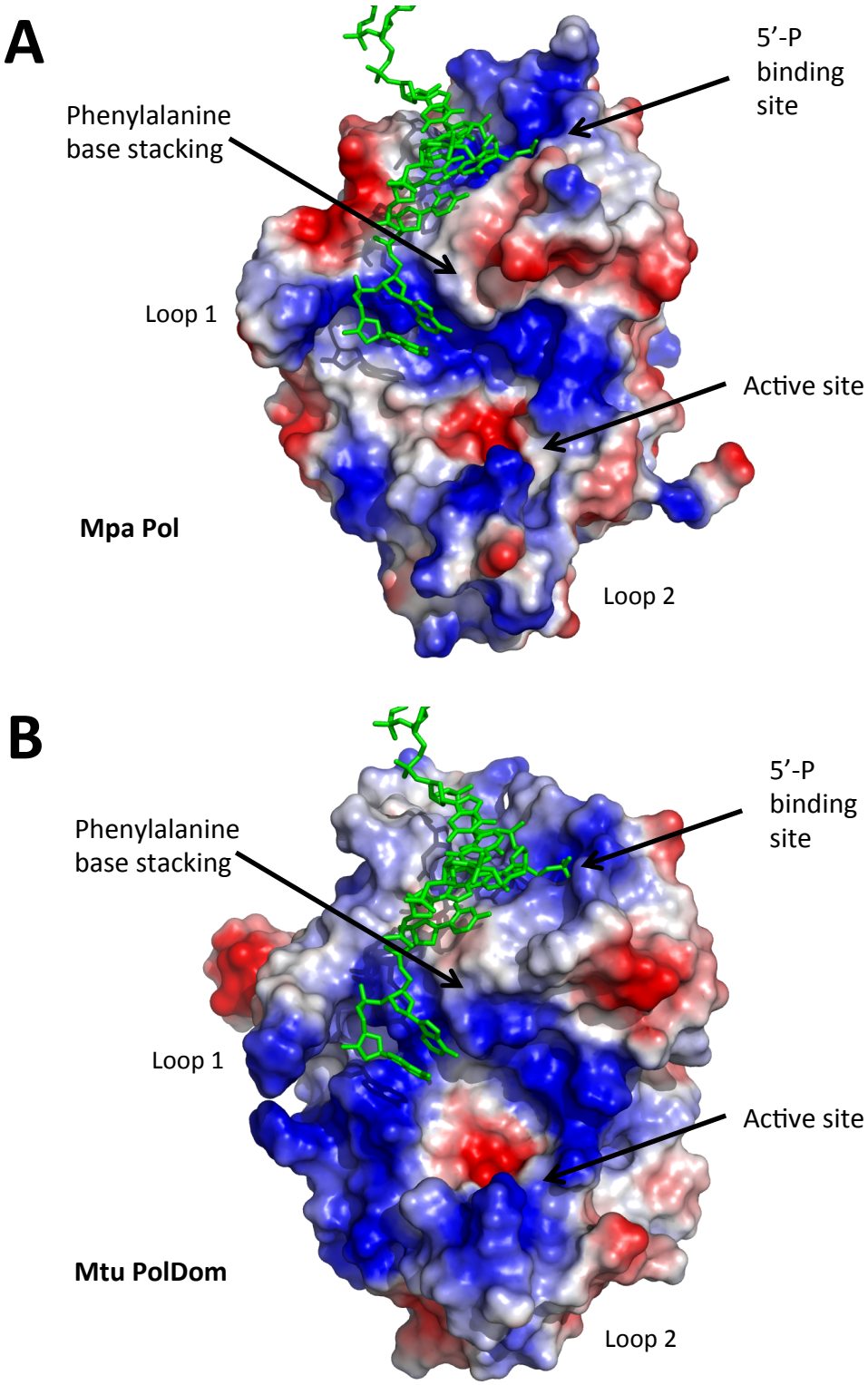


Figure 5.10 Electrostatic surface representations of Mpa Pol and Mtu PolDom. (**A**) An electrostatic surface representation of Mpa Pol with DNA superposed from the Mtu PolDom pre-ternary structure (Brissett et al., 2011). The 5’-P fits into the binding pocket, which has a positive charge. (**B**) An electrostatic surface representation of Mtu PolDom. The electro positive regions are largely in the same place as Mpa Pol, although Mtu PolDom has a more defined ‘ring’ of positive charge around the active site. The figures were made using the structure from Brissett et al., 2011, PDB 3PKY.

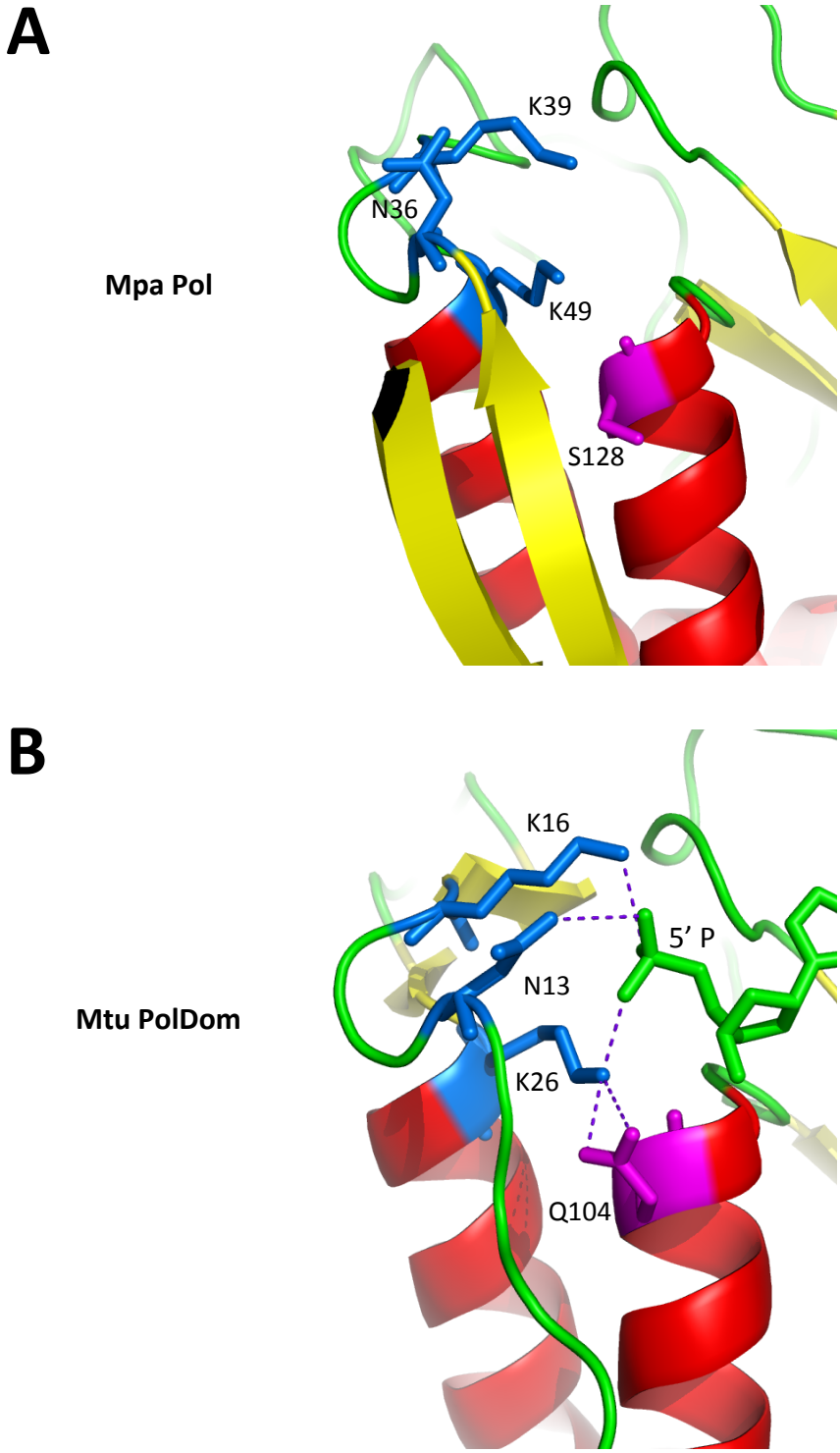


Figure 5.11 A comparison of the 5’-phosphate binding site of NHEJ polymerases. (**A**) A ribbon diagram of Mpa Pol, showing the phosphate binding pocket, composed of α1, α2, β1, and β2. The conserved 5’-P interacting residues are shown in blue (N36, K39, K49). S128 is in place of a conserved glutamine residue found in Mtu, Msm and Pae PolDoms. (**B**) A ribbon diagram of Mtu PolDom showing an incoming 5’-P interacting with conserved residues in blue (N13, K16, K26). The conserved glutamine (Q104) is shown in magenta, and is stabilising K26 by hydrogen bonding. The Mtu PolDom figure was made using the structure from Brissett et al., 2011, PDB 3PKY.

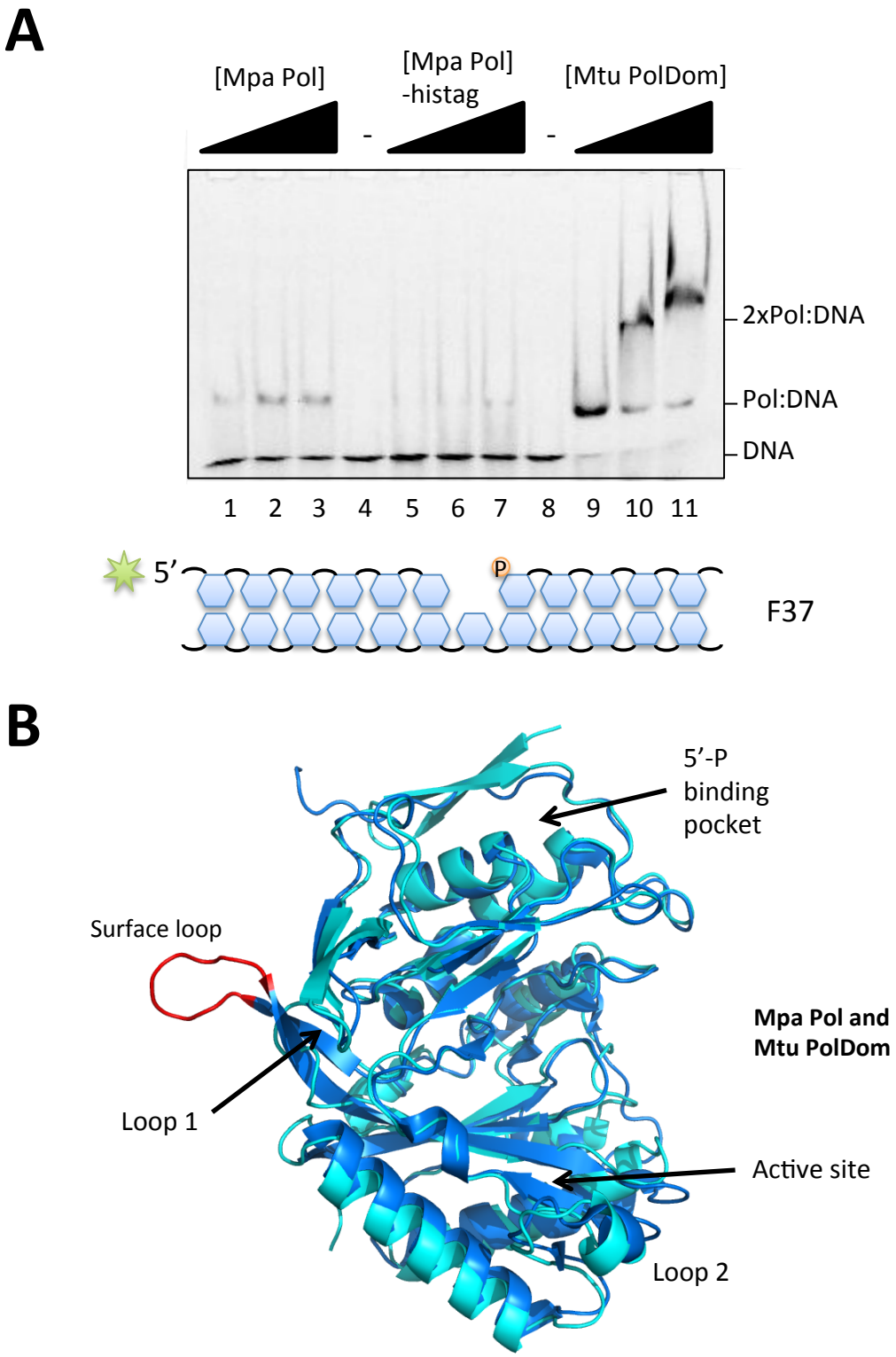


Figure 5.12 Comparison of the structures of Mpa Pol and Mtu PolDom.
(**A**) A schematic of the DNA substrate used in the EMSA is shown below the scan of the gel. EMSA reactions contained 200, 400 and 800 nM Mpa Pol, Mpa Pol without histag or Mtu PolDom where indicated, and 60 nM 5'-fluorescein labelled substrate F37 (primer 16-mer, template 35-mer, downstream 18-mer). Reaction mixtures also contained 5 mM Mn. (**B**) A ribbon diagram of Mpa Pol superposed over Mtu PolDom. The extra surface loop on Mtu PolDom is coloured in red. The figure was made using the structure from Pitcher et al., 2007, PDB 2IRU.

very close to the N-terminus and is largely composed of the first 30 amino acids of Mpa Pol. The his-tag was cleaved with thrombin, and Mpa Pol was isolated from the protease before inclusion in assays. The removal of the his-tag did not positively impact on the 5'-P binding ability of Mpa Pol, suggesting that the deficiency was caused by another factor. A comparison of Mpa Pol and Mtu PolDom reveal remarkably similar 5'-P binding pockets, with two α-helices and two loops positioned near identically (Figure 5.11, A and B). The two lysine and single asparagine residues which contact the terminal phosphate are shown in blue for Mpa Pol (Figure 5.11, A) and Mtu PolDom along with the 5'-P of the D-strand from the pre-ternary structure (Figure 5.11, B) (Brissett et al., 2011). The Mpa Pol structure included density for an extended area of the N-terminus, which comprises a β-hairpin (Figure 5.11, A). It may be that this region affects the DNA binding of Mpa Pol, although it is also possible that Mtu PolDom possess a similar structural motif, but it was not discernable from the electron density when the structure was solved. It is more likely that Mpa Pol has a slightly more flexible 5'-P binding pocket because of a loss of an otherwise conserved glutamine residue. Mtu PolDom has a glutamine on the α-helix that helps to indirectly stabilise the phosphate binding with hydrogen bonding (Figure 5.11, B). Instead, Mpa Pol has a serine residue that does not make contacts with the key lysine residue (5.11, A). Although Mpa Pol is a comparable activity to Mtu PolDom with a variety of substrates, it does not appear to form a stable complex with a DNA substrate with a 5'-P under native conditions. This may be one of the reasons why the prior crystallographic screening with Mpa Ku, DNA and Mpa Pol were unsuccessful.

5.5.2 Mpa Pol is missing a surface loop that is present on Mtu PolDom

The most noticeable aspect of a comparison between Mpa Pol and Mtu PolDom structures is how similar they are. There is only one large structural region that is conspicuously absent on Mpa Pol, a surface loop that sits adjacent to Loop 1 on Mtu PolDom (Figure 5.12, B). On Mtu PolDom this element comprises a β-hairpin with an extended apical loop region (Pitcher, Brissett, Picher, et al., 2007b; Brissett et al., 2011). A superimposition of Mtu PolDom over Mpa Pol highlights this surface loop as the only structural motif that is not conserved, shown in red in Figure 5.12, B. function of this loop is unknown; it does not appear to add any direct differences to the behaviour of Mtu PolDom, since Mpa Pol can replicate the same catalytic activities under the same conditions. It may be that the loop is used for interacting with other proteins, since the loop extends away from the active site, and is not close enough to engage with bound DNA. Furthermore, the loop represents an insertion in the amino acid sequence

that is not present in the closely related Msm PolDom, or Pae PolDom (Figure 3.10). It is probably that the function is unique to *M. tuberculosis*.

5.5.3 Structural similarities between NHEJ polymerases and archaeal replicative primases

Mpa Pol, like the bacterial NHEJ polymerases, is an archaeo-eukaryotic primase (AEP). The NHEJ fold bears significant similarities with several archaeal primases that reflects their evolutionary history. Figure 5.13 shows a ribbon structure of Mpa Pol juxtaposed with a structure of *Pyrococcus horikoshii* (Pho) (Ito et al., 2003). Both enzymes have an active site cleft that is cradled by two β-sheets composed of anti-parallel strands, with surrounding α-helices. The Mpa Pol has function specific regions highlighted in red; Loops 1, 2 and the phosphate-binding pocket. Pho primase has additional helical regions highlighted in orange and a zinc-binding site in blue that allows the enzyme to bind ssDNA (Ito et al., 2003). It is interesting to consider that these two enzymes have evolved to perform completely different roles in the cell, whilst maintaining a similar core. The nucleotidyltransferase activity remains present in both enzymes, and additional components have allowed the enzymes to specify different substrates. When viewed in this context, it is perhaps not surprising that the NHEJ Pols prefer to insert small RNA sequences when interacting with broken DNA, since their roles probably overlapped with those of the primases at some point in their history.

5.5.4 NHEJ polymerase mediated strand-displacement

Studies described in Chapter 4 revealed that the NHEJ polymerases have an intrinsic ability to displace downstream DNA whilst they are extending primer strands. It is unknown exactly how the polymerases manage to disrupt the base pairing and mobilise the D-strand to achieve strand displacement. A recent structure of two Mtu PolDom molecules annealing two 3'-overhang ends can be used as a basis to explore how strand displacement might occur (Brissett et al., 2013, under review). Figure 5.14, A shows a ribbon diagram of the complex, with the 3'-OH termini of each strand occupying a catalytically active position in the active site of each NHEJ Pol. The 5'-P on the D-strand of each broken end is engaging the 5'-P binding pocket, leaving one base of the template strand unpaired. This DNA configuration is broken ends annealed to create a single nucleotide gap. The DNA from the Mtu PolDom structure is overlaid onto a surface model Mpa Pol, and edited to fit a single molecule (Figure 5.14, B). The single un-paired base is positioned towards the active site, and the 3'-OH is positioned to the side, both awaiting the next nucleotide. The DNA template is 7 bases long, with a 3-mer primer and a 3-mer D-strand. Figure 5.15, A shows the arrival of an incoming ribonucleotide (UTP),

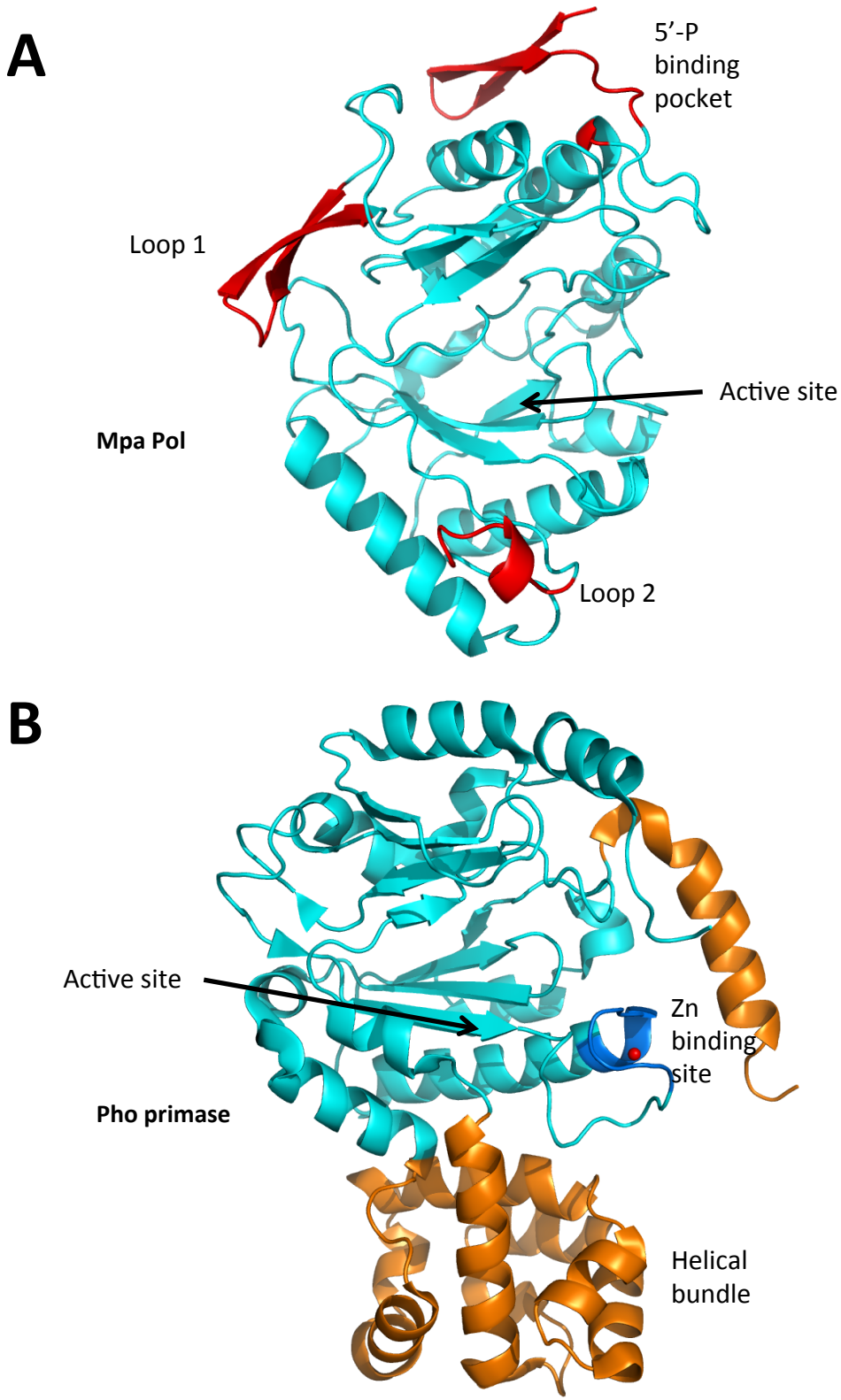


Figure 5.13 Comparing the structures of Mpa Pol and Pho primase.
(**A**) A ribbon diagram of Mpa Pol, with Loops 1 and 2 and the phosphate binding pocket coloured red. The active site area is cradled by two β-sheets similar to those in an archaeal primase. (**B**) A ribbon diagram of the structure of *Pyrococcus horikoshii* (Pho) primase, with additional α-helices coloured orange, and a zinc binding motif in blue with the ion coloured red. The diagram of Pho primase was made using the structure from Ito et al., 2004, PDB 1V34. Adapted from Zhu et al., (2006) and Pitcher et al., (2007).

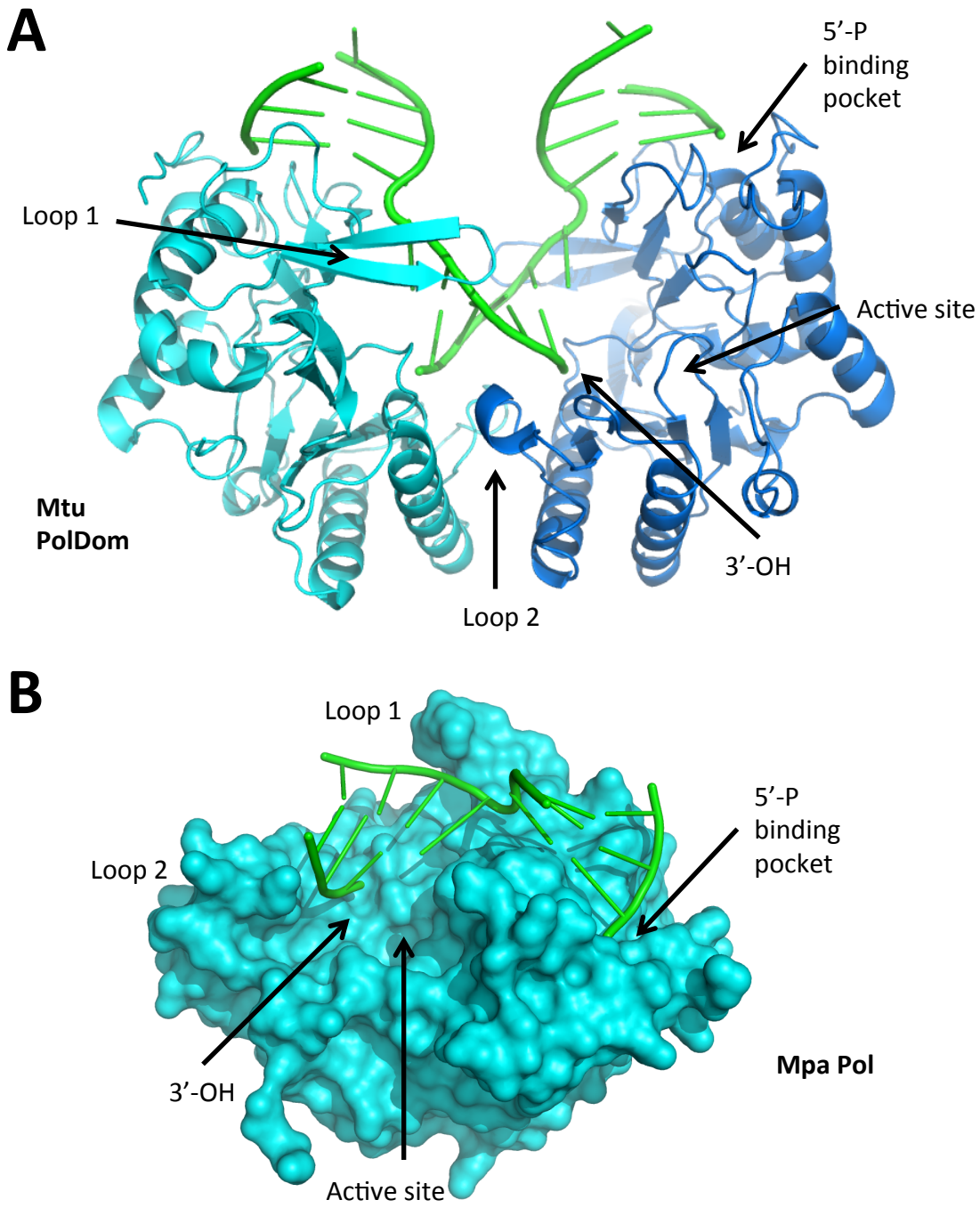


Figure 5.14 Two Mtu PolDom molecules anneal 3'-overhangs and position the 3'-OH into the active site of the opposite polymerase.

(**A**) Two Mtu PolDom molecules annealing a break of two 3' overhangs and positioning the primers in the active site of the opposite enzyme. PolDom uses Loop 1 and Loop 2 to direct the 3' strand into the active site of the in trans PolDom. Both DNA molecules are also engaged at the 5'-P binding pocket of PolDom. The figure was made using the structure from Brissett et al., 2013 (under review). (**B**) A surface representation of Mpa Pol with the DNA from the above Mtu PolDom structure superposed. The DNA has been edited to highlight the single nucleotide gap. This representation shows the primer in a catalytically competent position in the active site. The wedge that separates the template bases and the 5'-P binding pocket might be one of the structural features that cause D-strand displacement. The figures were made using the structure from Brissett et al., 2013 (under review).

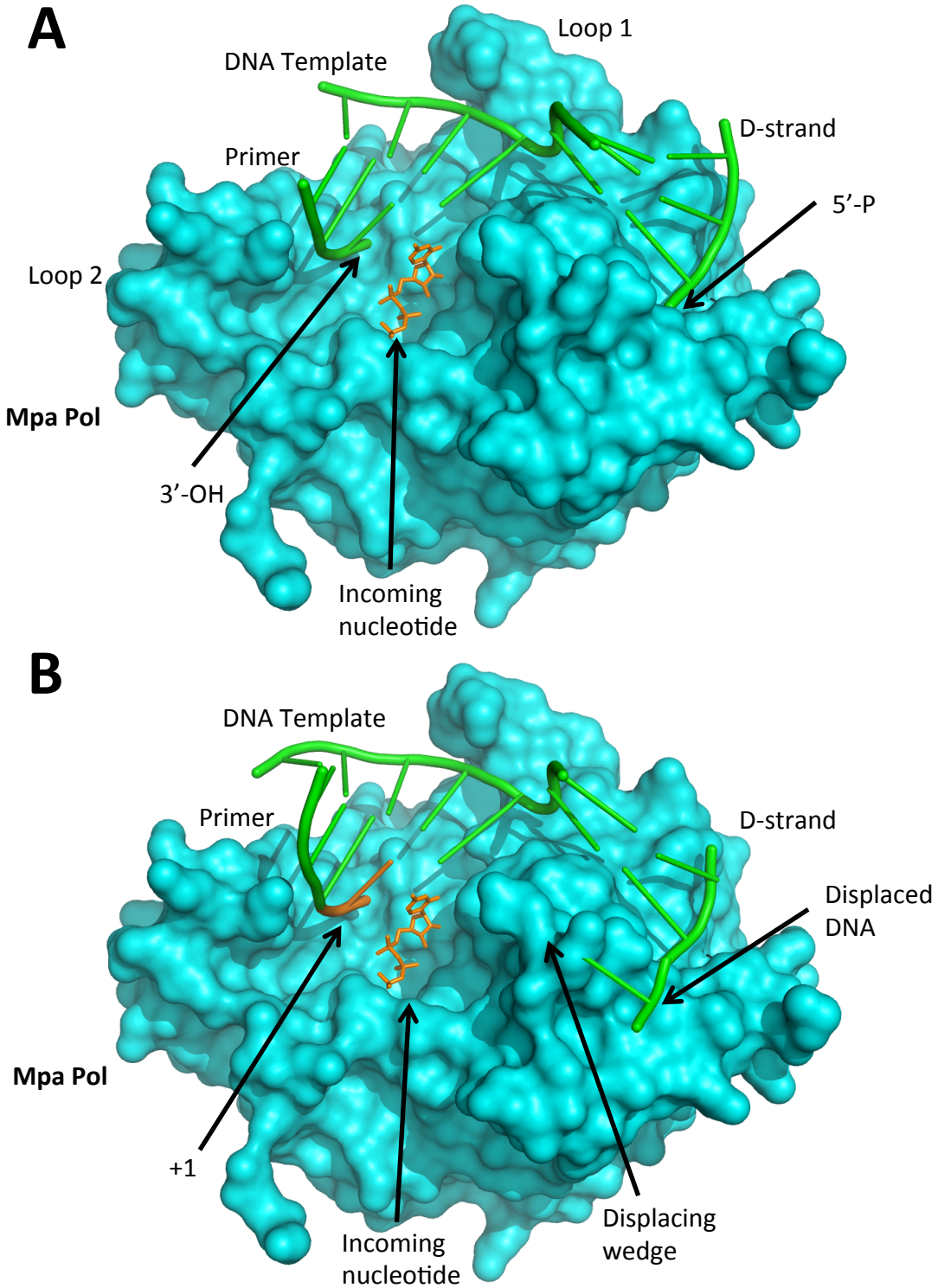


Figure 5.15 Structural insights into strand displacement with Mpa Pol.
(**A**) A surface representation of Mpa Pol with a single nucleotide gapped DNA substrate modelled in. The primer, template and D-strand of the DNA are labelled. An incoming nucleotide is shown in orange, in the active site adjacent to the 3'-OH of the primer strand. (**B**) The same representation of Mpa Pol shown after a +1 extension of the primer strand. The +1 addition is shown in orange on the primer strand, and a second incoming nucleotide is shown in the active site. The polymerase has moved along the template strand to position the next templating base opposite the active site. This base had previously been paired with the D-strand, but the polymerase activity has caused the D-strand to become displaced. The figures were made using the structure from Brissett et al., 2013 (under review).

and suggests that the DNA positioning is correct for catalysis (Brissett et al., 2011). It is the next step, following the +1 extension, which would force mobilisation of the D-strand. Figure 5.15, B shows an extended primer of 4 bases, and the arrival of a second ribonucleoside. In order to accommodate the next nucleotide in the active site the polymerase must move along the template. Previously, at this point we would have expected the polymerase to stop and dissociate. However, if the polymerase is to continue extending the primer, the first base of the D-strand must become un-paired with the template DNA, and moved from the 5'-P binding site. The D-strand in Figure 5.15, B has been moved out of the pocket and shown moving over the surface of the polymerase. The DNA positioning in this figure is speculative, and it is unknown whether there are any amino acids that direct the un-paired DNA following D-strand displacement. It is possible that the displacement is caused by the structural area labelled as the 'displacement wedge' in Figure 5.15, B. The twin phenylalaine residues splay the DNA template that allows the surface structure to protrude between the template and D-strand. As the polymerase moves along the template, it would presumably force the wedge into the base-paired duplex and displace the D-strand.

5.6 Crystallisation screening of the Mpa NHEJ Ligase

An apo crystal structure of the LigD ligase domain was solved early on in the research of bacterial NHEJ, from *M. tuberculosis* (Akey et al., 2006). This paper remains the only insight into the structure of an AP-NHEJ ligase, and therefore a target of this research was to capture of structure of Mpa Lig, ideally engaging a DNA-RNA nick. The ligase domain had previously been resistant to crystallographic studies, likely owing to the highly flexible nature of the protein, which is composed of two globular domains (OB and NTase) that are connected by a hinge region (shown in Figure 1.9). It was hoped that inclusion of a suitable substrate during the crystallisation process could stabilise the enzyme and allow for crystal growth.

5.6.1 Attempts to discern a potential DNA substrate for crystallising Mpa Ligase

Following the confirmation that Mpa Lig preferentially sealed DNA-RNA nicks in the presence of manganese, a small DNA-RNA nick substrate was designed for crystallography (Figure 5.1, C). This substrate, PE X BCD, offered potential for screening with both Mpa Lig and Mpa PE (discussed in 5.8.2). A primary issue with supplying an enzyme with its preferred substrate for crystallography is that the enzyme tends to catalyse the intended reaction and then dissociate from the substrate. Previous structures of DNA ligases overcame this problem with some

clever experiment design. The *H. sapiens* Ligase I structure allowed the ligase to adenylate the 5'-P of the DNA nick, but prevented the final ligation step by usage of a 3'-dideoxycytidine monophosphate on the opposite side of the nick (Pascal et al., 2004). The *E. coli* Lig A structure was achieved by a different approach, pre-adenylating the ligase, and then quenching any further adenyltransfer by addition of 10 mM EDTA (Nandakumar et al., 2007). The EDTA was intended to chelate the metal ions necessary for adenylylation of the 5'-P, allowing Lig A to interact with the substrate, but not to catalyse the second step of the ligation reaction. However, the structure revealed adenylated DNA, but Lig A had been unable to seal the nick in the absence of metal ions. The latter of these two techniques was chosen for generating a Mpa Lig – DNA structure.

The PCT for crystallographic screening of Mpa Lig and DNA suggested that a screening concentration of 9 mg/mL (230 µM) Mpa Lig, 460 µM PE X BCD and 50 µM EDTA would be suitable, however the first 96 condition suite, Nucleix, produced only clear drops over night. The concentration was increased to 14 mg/mL Mpa Lig (and the DNA kept to 2:1 ratio) and the screen repeated, and this time the drops produced some granular precipitation. Additional screens of Ammonium Sulphate, PEGs and MPD were also set using an increased concentration. At present no significant hits have been observed, so extended screening is required. It may be necessary to vary the concentration of DNA to protein, perhaps initially lowering the substrate to 1.5 or 1.2 to that of the ligase. It is possible that Mpa Lig has difficulty properly engaging the DNA substrate in the absence of a metal ion. If this is the case, and should screening require addition of manganese, a different approach must be taken to stall the ligase on the substrate. A new substrate with a 3'-deoxyribonucleoside might be the only way to capture Mpa Lig bound to a substrate.

5.7 Crystallisation of the Mpa NHEJ Phosphoesterase

The primary crystal screening of Mpa PE was aimed at capturing an apo structure since, at the time no atomic structure had been published. Prior to the completion of the Mpa PE structure the Shuman laboratory published a structure of the truncated PE domain from the multi-domain LigD from *P. aeruginosa*, followed by structures of discrete archaeal PEs from *M. barkeri* and *C. korachaeum cryptofilum* (Nair et al., 2010; Smith et al., 2011). The work on the Mpa PE structure was underway, and could still offer an interesting comparision to the other PE structures, which will be discussed later in this chapter. The initial PCT for the Mpa PE screening suggested that a concentration of between 4.8 and 9.3 mg/mL would be favourable.

The first screen with the PEGs suite at 7 mg/mL produced almost entirely clear drops with little precipitation, and a repeat at 9.3 mg/mL did not improve the results appreciably. A third repeat, with Mpa PE at 18.6 mg/mL produced several, small crystals in a single condition; magnesium sulphate 0.2 M with 20 % (w/v) PEG 3350. Given that only a handful of small crystals grew, the screen was repeated with a much higher concentration of protein to see if additional protein would increase the crystal size rather than the nucleation amount. A screen at 30 mg/mL produced a single, large crystal in the same condition as those seen before, ~150 x 150 x 300 µm (Figure 5.16, A). The crystal looked suitable for data collection, although it took several weeks to appear in the drop, so several repeats of the single condition were used to attempt to reproduce the large crystal. The condition proved reliable, and a number of crystals over 100 x 100 x 100 µm were grown for data collection (Figure 5.16, B-C). The data collected from these crystals was used to solve the structure of Mpa PE, as described in section 5.7.2, below.

5.7.1 Mpa PE crystal screening with $NaVO_4$ and short ssDNA oligomers

The screening of Mpa PE generated a generous number of large crystals that could be used for soaking in additional components to improve the value of the Mpa PE structure. The Doherty lab possessed a number of short ssDNA oligomers that had been used in previous crystallographic projects that could be used to soak into Mpa PE crystals. In order to ensure that the oligomer would bind with the enzyme, sodium vanadate was chosen as a co-factor. Sodium vanadate ($NaVO_4$) has been used previously by other research groups to solve structures of phosphoryltransfer enzmes as it is an analogue for phosphate, for example the structure of TDP1 bound to a phospho-tyrosyl analogue composed of peptide fragment, DNA and $NaVO_4$ (Davies et al., 2002; Davies et al., 2003; Davies & Hol, 2004). $NaVO_4$ should enable the capture of a transition state analogue with the PE and DNA, since it adopts a highly similar geometry to a phosphate, but cannot be hydrolysed. The existing Mpa PE crystals were soaked overnight with 6 mM DNA oligomers (5'-P-CGCG, 5'-CTCGG, and 5'-P-GCGGC) and 10 mM $NaVO_4$. The crystals assumed a yellow hue (because of oxidation of the vanadate) and some became fragile, and began to disintegrate (Figure 5.12, D). Several crystals remained suitable for data collection and diffraction patterns were collected using a synchrotron light source (Diamond). The data did not reveal any discernable electron density in difference maps for the DNA oligos, with only the vanadate ion present in the active site. The structure of Mpa PE vanadate was still solved however, and generated a higher resolution structure than the Mpa PE structure that was solved using the home source equipment. Given that the crystal soaking

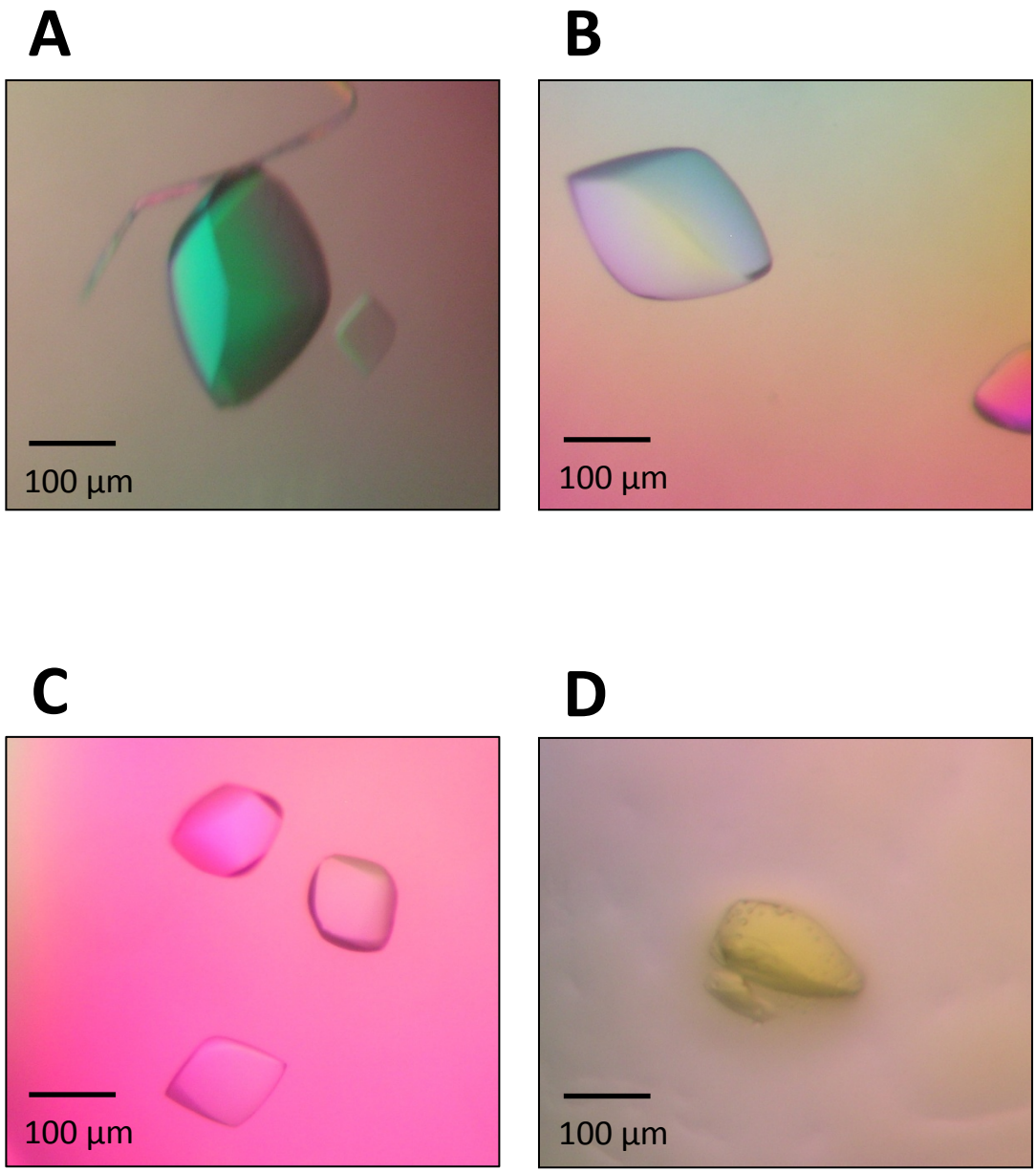


Figure 5.16 Mpa PE crystals.
(**A-C**) Mpa PE crystals grown using the PEGs suite (Qiagen). 0.7 µL protein: Mpa PE 30 mg/mL. 0.7 µL reservoir: 0.2 M Ammounium sulphate, 20 % (w/v) PEG 3350. These crystals were used to determine the structure of Mpa PE. (**D**) The crystal in this drop was grown under the same conditions as those in A-C, but was subsequently soaked with 6 mM 5'-P-GCGGC DNA and 10 mM Sodium Vanadate. No density was observed for the DNA fragment in the data collected from these crystals.

process affected the crystal stacking, co-crystallisation was the next step for generating Mpa PE + substrate crystals and is discussed in section 5.8.1.

5.7.2 Data collection, processing and refinement of Mpa PE

Dr. Nigel Brissett processed the data and refined the apo and vanadate structures of Mpa PE. The data for Mpa PE was collected using a Rigaku MicroMax 007-HF, and was processed using the CCP4 suite as described in the materials and methods. Crystals of Mpa PE contained one protein molecule in the asymmetric unit, giving a V_M of 2.32 $Å^3$/Da with 47 % solvent content. Phases were solved by molecular replacement, using the Pae PE as the search model (Nair et al., 2010). A final refined model at 2.51 Å resolution, with an R_{cryst} of 22.80 % and R_{free} of 31.14 %, was obtained. Over 95 % of residues in the structure are in the allowed region of the Ramachandran plot. The data collection and refinement statistics of apo Mpa PE are shown in Table 5.2.

5.7.3 Data collection, processing and refinement of Mpa PE vanadate

Dr. Nigel Brissett processed the data and refined the apo and vanadate structures of Mpa PE. The data for Mpa PE vanadate was collected at a synchrotron light source, Station IO3, Diamond, and was processed using the CCP4 suite as described in the materials and methods. Crystals of Mpa PE vanadate contained one protein molecule and one metal ion in the asymmetric unit, giving a V_M of 2.32 $Å^3$/Da with 47 % solvent content. Phases were solved by molecular replacement, using the Pae PE as the search model (Nair et al., 2010). A final refined model at 1.79 Å resolution, with an R_{cryst} of 18.02 % and R_{free} of 23.80 %, was obtained. Over 97 % of residues in the structure are in the favoured region of the Ramachandran plot. The data collection and refinement statistics of Mpa PE vanadate are shown in Table 5.3.

5.8 The structure of the Mpa NHEJ Phosphoesterase

The structures of Mpa PE and Mpa PE + vanadate [herein referred to as Mpa PE (Van)] were solved to resolutions of 2.51 and 1.79 Å respectively, and are shown in Figures 5.17, 5.18 and 5.19. Ribbon diagrams of Mpa PE and Mpa PE (Van) are shown in Figures 5.17 and 5.18, showing the hydrophobic core formed of a β-barrel super secondary structure, composed of 9 anti-parallel strands. In the first Mpa PE structure the β-barrel is surrounded by two α-helices and a 3_{10} helix, the same as observed in the Pae PE structure (Nair et al., 2010). In the Mpa PE (Van) structure, there are two additional short α-helices, α1 and α4, which are also noted in

Data collection	
Source	Rigaku MicroMax 007-HF
Space group	$P3_121$
Unit cell dimensions	
a, b, c (Å)	56.65/56.65/104.58
α, β, γ	90.00/90.00/120.00
Wavelength (Å)	1.5418
Resolution (Å)	23.07-2.51
Total number of observations	100231
Number of unique reflections	6958
Overall I/(σI)[a]	47.5 (5.6)
Overall completeness (%)[a]	94.9 (98.1)
R_{sym} (%)[a,b]	0.042 (0.456)
Redundancy[a]	15.8 (13.1)
Refinement	
Resolution(Å)	23.07-2.51
Number of reflections	6614
R_{factor} / R_{free} [c,d]	0.2280/0.3114
Contents of asymmetric unit	1 protein molecule
Number of atoms	
Protein	1263
Water molecules	58
Mean B value (Å^2)	28.022
Rmsds	
Bonds (Å)	0.016
Angles (°)	1.963
Ramachandran statistics	
Favoured regions (%)	89.03
Allowed regions (%)	6.45
Disallowed regions (%)	4.52

Table 5.2. Data collection and refinement statistics of Mpa PE.
[a] Values for highest resolution shell (2.65–2.51A˚) is in parentheses. [b] *R*sym = Σ|*I* – <*I*>| / Σ <*I*>, where *I* is the observed intensity. [c] *R*factor = Σ||*F*o – |*F*c|| / Σ|*F*o|, where *F*o and *F*c are the observed and calculated structure factor, respectively. [d] *R*free is equal to R factor for a randomly selected 5% subset of reflections not used in the refinement.

Data collection	
Source	Station I03, Diamond
Space group	$P3_121$
Unit cell dimensions	
a, b, c (Å)	56.89/56.89/104.69
α, β, γ	90.00/90.00/120.00
Wavelength (Å)	0.9787
Resolution (Å)	49.27-1.79
Total number of observations	119283
Number of unique reflections	18922
Overall I/(σI)[a]	33.8 (3.2)
Overall completeness (%)[a]	87.1 (99.8)
R_{sym} (%)[a,b]	0.030 (0.509)
Redundancy[a]	5.6 (6.4)
Refinement	
Resolution(Å)	49.27-1.79
Number of reflections	17915
R_{factor} / R_{free} [c,d]	0.1802/0.2380
Contents of asymmetric unit	1 protein molecule, 1 metal
Number of atoms	
Protein	1338
Water molecules	118
Mean B value (Å^2)	34.645
Rmsds	
Bonds (Å)	0.018
Angles (°)	1.969
Ramachandran statistics	
Favoured regions (%)	97.01
Allowed regions (%)	2.40
Disallowed regions (%)	0.60

Table 5.3. Data collection and refinement statistics of Mpa PE vanadate.
[a] Values for highest resolution shell (1.89–1.79 A˚) is in parentheses. [b] *R*sym = Σ|*I* – <*I*>| / Σ <*I*>, where *I* is the observed intensity. [c] *R*factor = Σ||*F*o – |*F*c|| / Σ|*F*o|, where *F*o and *F*c are the observed and calculated structure factor, respectively. [d] *R*free is equal to R factor for a randomly selected 5% subset of reflections not used in the refinement.

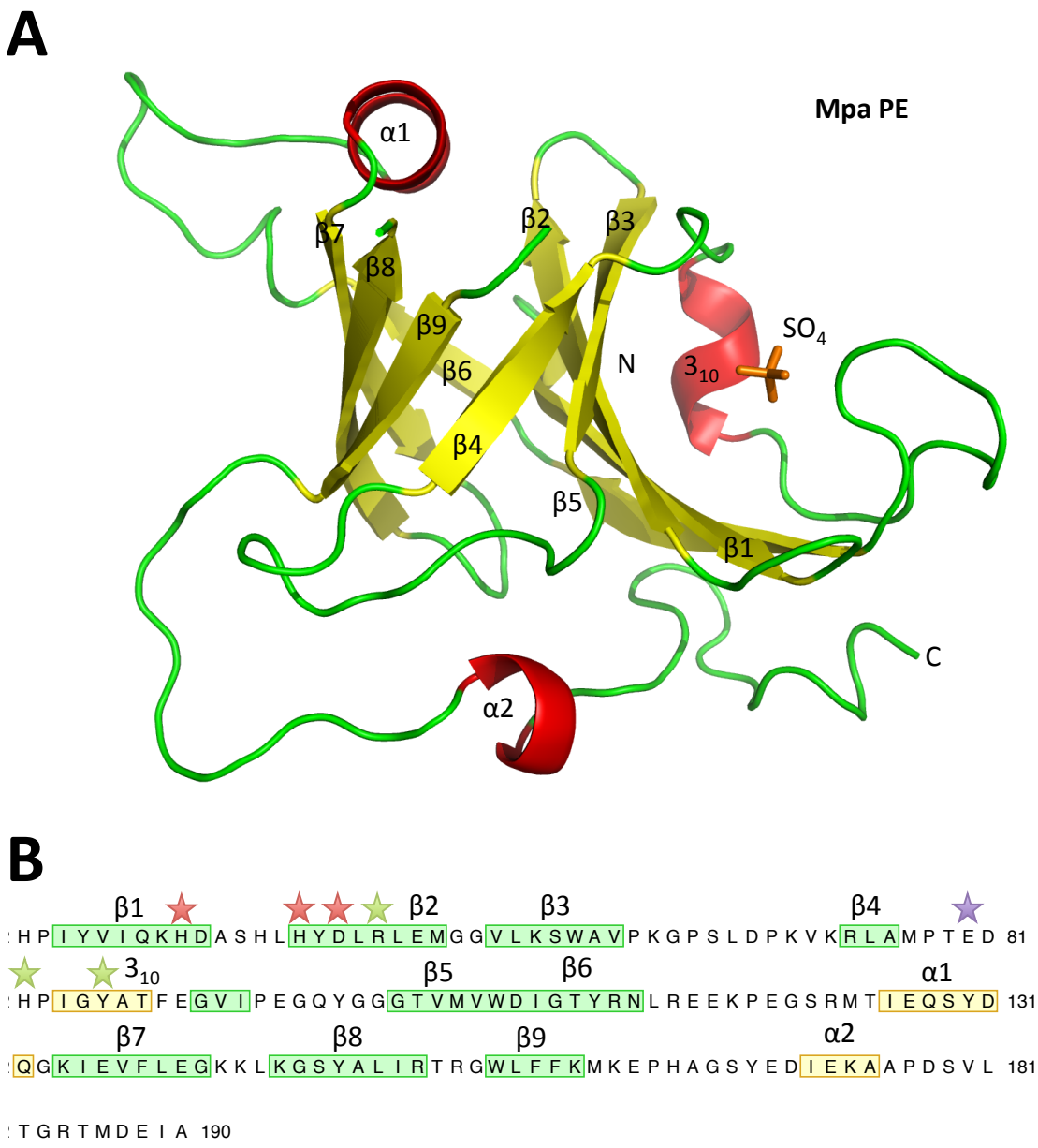


Figure 5.17 Structure of Mpa PE.
(**A**) A ribbon diagram of the 2.51 Å resolution structure of Mpa PE. The hydrophobic core of the enzyme is comprised of a β-barrel containing 9 anti-parallel β-strands. The active site position is indicated by the sulphate ion (orange) and is adjacent to the 3^{10} helix. (**B**) An amino acid sequence of Mpa PE with secondary structure elements highlighted. β-strands are shown in green, α-helices are yellow. The red stars indicate conserved catalytic site residues which coordinate the metal ion, the green stars indicate the catalytic site residues which position the scissile phosphate, and the purple star indicates the conserved glutamate residue which is essential for phosphatase activity. The amino acid sequence with secondary structure labels was generated using UCSF Chimera (Pettersen et al., 2004).

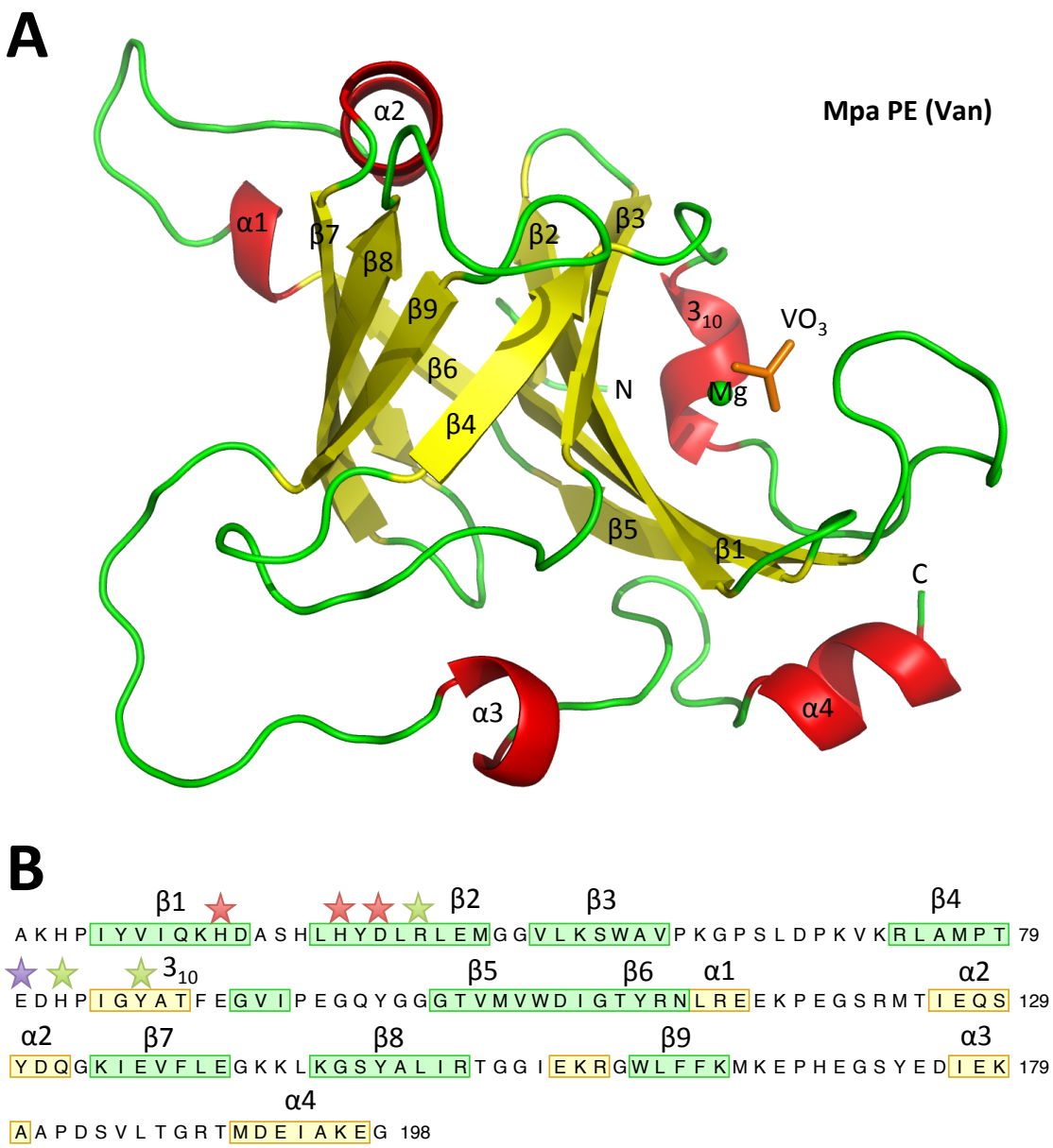


Figure 5.18 Structure of Mpa PE (Van).
(**A**) A ribbon diagram of the 1.79 Å resolution structure of Mpa PE (Van). The hydrophobic core of the enzyme is comprised of a β-barrel containing 9 anti-parallel β-strands. The active site position is indicated by the vanadate ion (orange) and is adjacent to the 3^{10} helix. (**B**) An amino acid sequence of Mpa PE with secondary structure elements highlighted. β-strands are shown in green, α-helices are yellow. The red stars indicate conserved catalytic site residues which coordinate the metal ion, the green stars indicate the catalytic site residues which position the scissile phosphate, and the purple star indicates the conserved glutamate residue which is essential for phosphatase activity. The amino acid sequence with secondary structure labels was generated using UCSF Chimera (Pettersen et al., 2004).

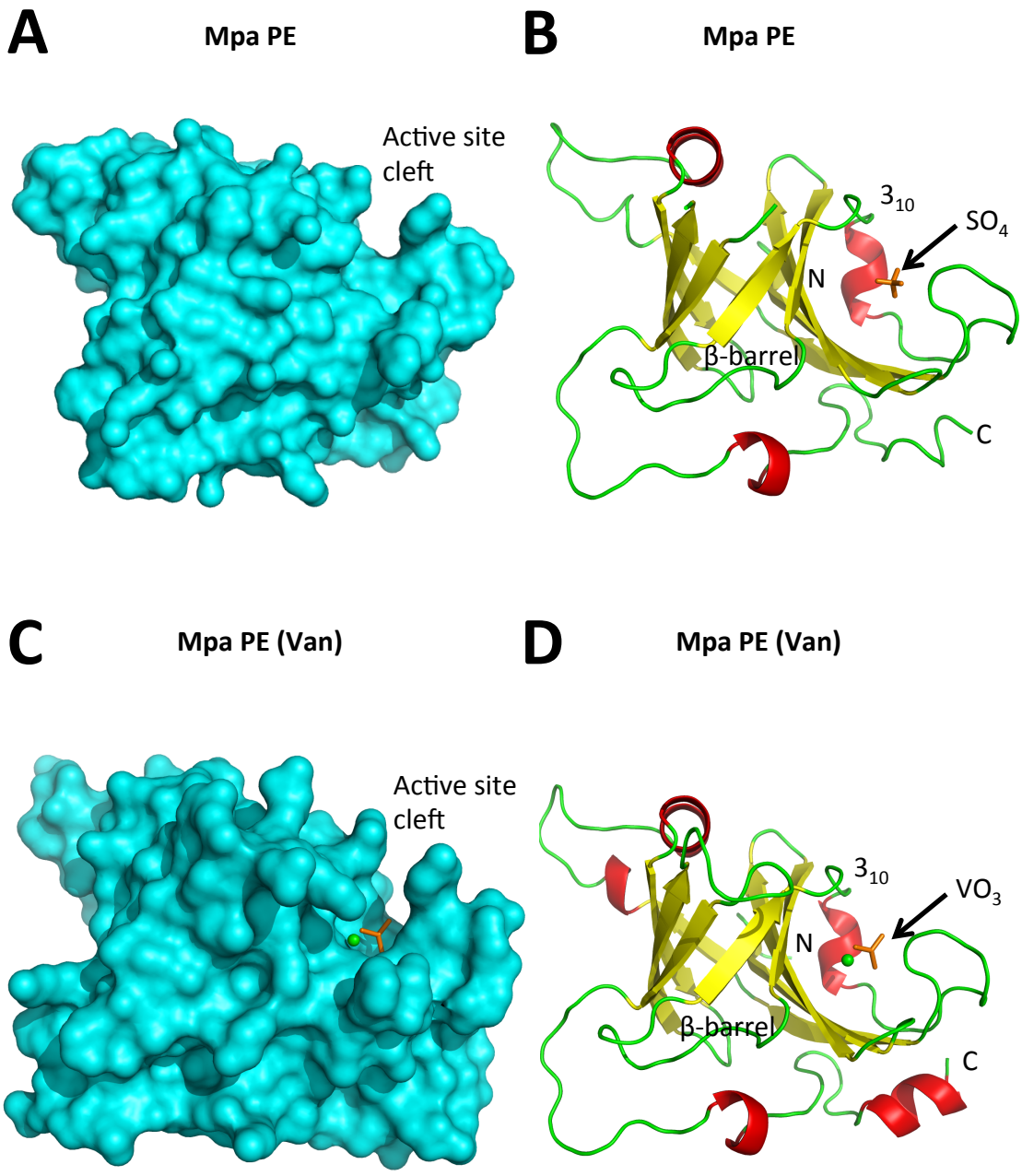


Figure 5.19 A comparison of Mpa PE and Mpa PE (Van) surface and ribbon representations.

Surface representations and ribbon diagrams of the 2.51 and 1.79 Å resolution structures of Mpa PE (**A, B**) and Mpa PE (Van) (**C, D**). (**A, C**) Mpa PE is a small globular enzyme with an active site cleft that sits on the outside edge of the core. (**B, D**) The ribbon diagrams show that the active site cleft is enclosed by a flexible loop region. It is likely that in solution this loop region 'opens up' to allow DNA to access the active site, before 'closing' on the substrate.

the amino acid sequence in 5.18, B. A comparison between the two ribbon diagrams in Figures 5.17 A and 5.18 A, suggests that these two α-helical regions are present in both structures but have not been assigned in the Mpa PE structure without vanadate. The outside of the β-barrel provides a base for the active site, which is enclosed by the 3_{10} helix and a surface loop (Figures 5.17, A and 5.18, A). Mpa PE is a small globular protein; a surface representation of the enzyme is shown in Figure 5.19, A and C. The active site is situated in a groove on the surface of the molecule, suggesting a channel in which a DNA substrate could be bound. Figure 5.19 offers a direct comparison between the surface and ribbon representations of Mpa PE and Mpa PE (Van), with the key structural elements labelled. The scissile phosphate is predicted to sit in the active site, positioned in between the six key catalytic residues; H40, H46, D48, R50, H82 and Y86. Figure 5.20, A and B shows the position of the six active site residues (in blue) and the phosphate mimetic ions (orange); sulphate and vanadate. Both occupy a similar position, although the Mpa PE (Van) structure also has a metal ion (magnesium, in green) coordinated by H40, H46 and D48. The overall fold of the PE has previously been found to be unique to this clade of NHEJ-related phosphoesterases, and no other 3' ribonucleases/phosphatases have a catalytic site organisation that is the same (Nair et al., 2010).

5.8.1 Comparison to bacterial and archaeal PE structures

The overall fold of the PE appears highly conserved between bacteria and archaea, a comparison of the ribbon structures of all of the solved PE crystal structures is shown in Figure 5.21 (Nair et al., 2010; Smith et al., 2011). All four proteins have a β-barrel core and a 3_{10} helix forming one side of the active site. All three archaeal proteins have a surface loop opposite to the β-barrel which forms the active site cleft, and it seems likely that the Pae PE also has this loop, although no electron density for this region was observed in the Pae structure, suggesting that it is highly flexible (Nair et al., 2010). The architecture of the structures implicate that the surface loop is flexible, and encloses the DNA once the proper substrate is engaged for catalysis. This speculation has since been supported by NMR solution structures that imply that the crystal structures represent a 'closed' active site formation, and that the loop moves away from the active site in the absence of a substrate (Natarajan et al., 2012). The studies showed that upon addition of substrate, spectral pertubations were observed that indicated movement of the loop region, and formation of a 'closed' active site from an 'open' one.

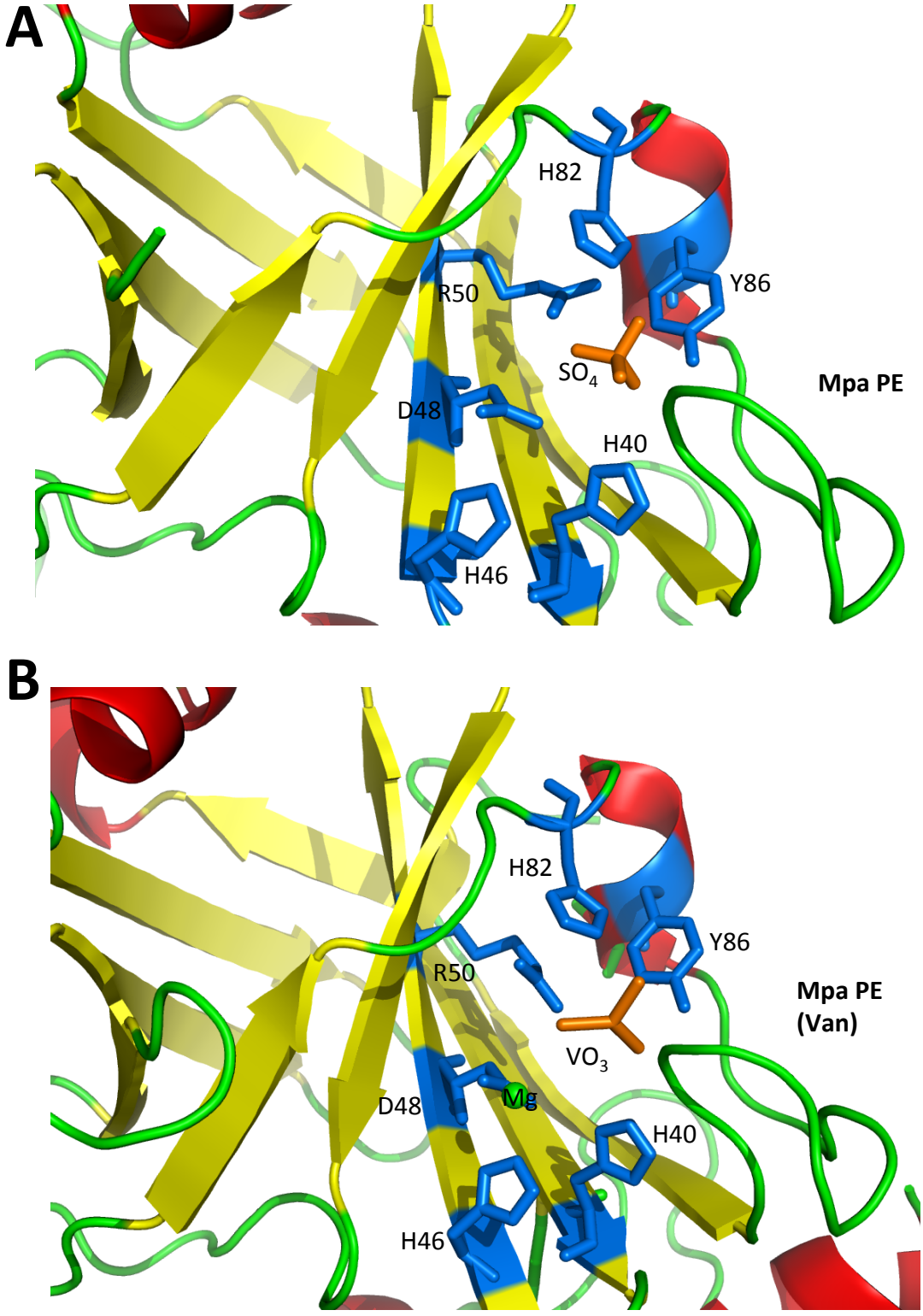


Figure 5.20 Ribbon diagrams of Mpa PE with active site detail.
Ribbon diagrams with a close view of the active site of Mpa PE (**A**) and Mpa PE (Van) (**B**). The six key catalytic residues are coloured blue, H40, H46 and D48 coordinate the metal ion. Magnesium is shown in the Mpa PE (Van) structure, although the ion is not in an optimal position to coordinate the potential phosphate properly. The residues R50, H82 and Y86 coordinate the sulphate and vanadate anions by a network of water molecules.

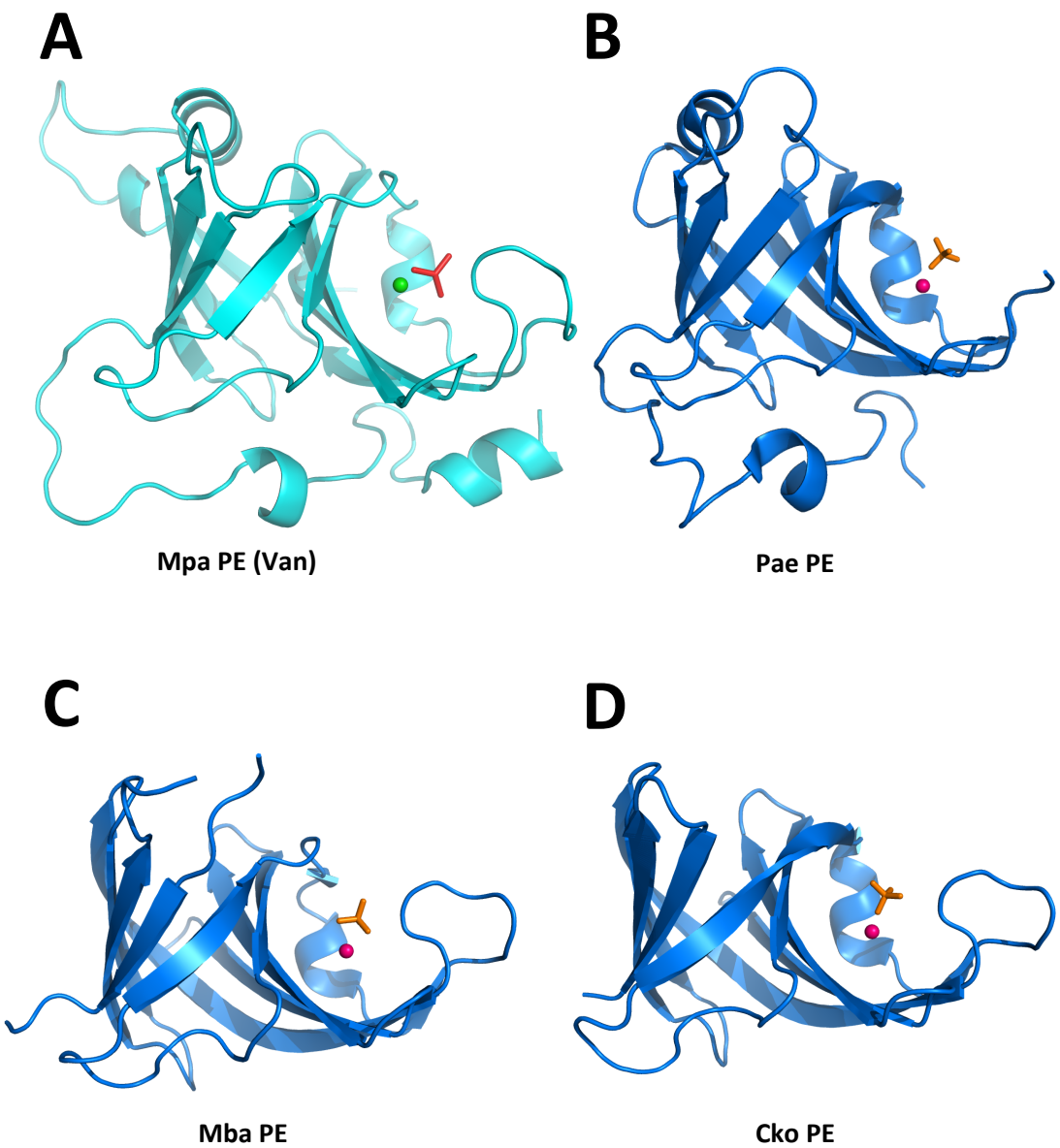


Figure 5.21 Contrasting the structures of Mpa, Pae, Mba and Cko PE.
(**A**-**D**) Ribbon diagrams showing the structures of Mpa, Pae, Mba and Cko PE. Mpa PE (Van) is coloured cyan, Pae, Mba and Cko are coloured blue. The sulphate ions are coloured orange, whilst vanadate is coloured red. The metal ions, magnesium and manganese, are coloured green and pink respectively. All four PEs have a β-barrel core and a 3_{10} helix adjacent to the active site. Pae PE has two additional short α-helices, and Mpa PE has four, which are not present in Mba or Cko PE. The surface loops which comprise the outer edge of the active site are visible in the Mpa, Mba and Cko structures. The diagrams were made using the structures from Nair et al., (2010), PDB 3N9D , and Smith et al., (2011), PDB 3P43 and PDB 3P4H.

Aside from the mobile surface loop, Mpa PE most closely resemble the Pae PE with two additional α-helices that are not seen in the other two archaeal structures, and a further two α-helices that are not in the bacterial PE. The Mpa PE and Pae PE represent 'full length' PE enzymes, which have both phosphodiesterase and phosphomonoesterase activity, whilst Mba PE has a slightly truncated N-terminus and Cko PE is missing the region entirely (Smith et al., 2011). The phosphomonoesterase activity has been ascribed to the N-terminus of the PE, following site directed mutagenesis of the Pae PE and the weak activity of the Cko PE (Zhu et al., 2005; Zhu & Shuman, 2006). Sadly, this region is not discernable in electron density maps in any of the PEs that have the N-terminal sequence, suggesting that it is a highly flexible region which undergoes a conformational change to enable catalysis. The N-terminal sequence emerges from 'behind' the 3_{10} helix, indicated in Figures 5.17, and 5.18, A. If the residues that are 'essential' for phosphatase activity are directly involved in catalysis, the N-terminus must wrap around the 3_{10} helix to access the active site. It seems more likely that the N-terminus helps to stabilise the binding of the DNA substrate. If so, it is unclear why these residues would not also be essential for the ribonuclease activity. The presence of a D-strand does not greatly affect the phosphomonoesterase activity, unlike the phosphodiesterase activity, so it is possible that the N-terminus compensates in some way for this.

The active site of Mpa PE is highly similar to the bacterial equivalent, and superposition of the catalytic residues shows a near identical conformation (Figure 5.22, A). The Pae PE structure features a manganese ion in the active site, which can be modelled into the Mpa PE structure to coordinate with the two histidines, aspartate and sulphate anion. The remaining arginine, histidine and tyrosine coordinate water molecules that also aids positioning of the phosphate mimetic. A figure of the Pae PE active site hydrogen bonding network including water molecules is shown in Figure 5.22, B. These residues are found to be essential for both phosphomonoesterase and phosphodiesterase activities (Zhu et al., 2005).

5.9 Structural insights on the activities of the Mpa NHEJ phosphoesterase

5.9.1 Phosphoesterase selection for ribonucleoside substrates

Only two residues in the PE have been identified as essential for ribonuclease activity, and not the phosphatase activity; an arginine and lysine (Zhu et al., 2005). Viewing these two residues, which were determined to be essential for ribonuclease activity by alanine substitution, in the structure of PE reveals that they are a considerable distance (19.3Å) from the predicted scissile

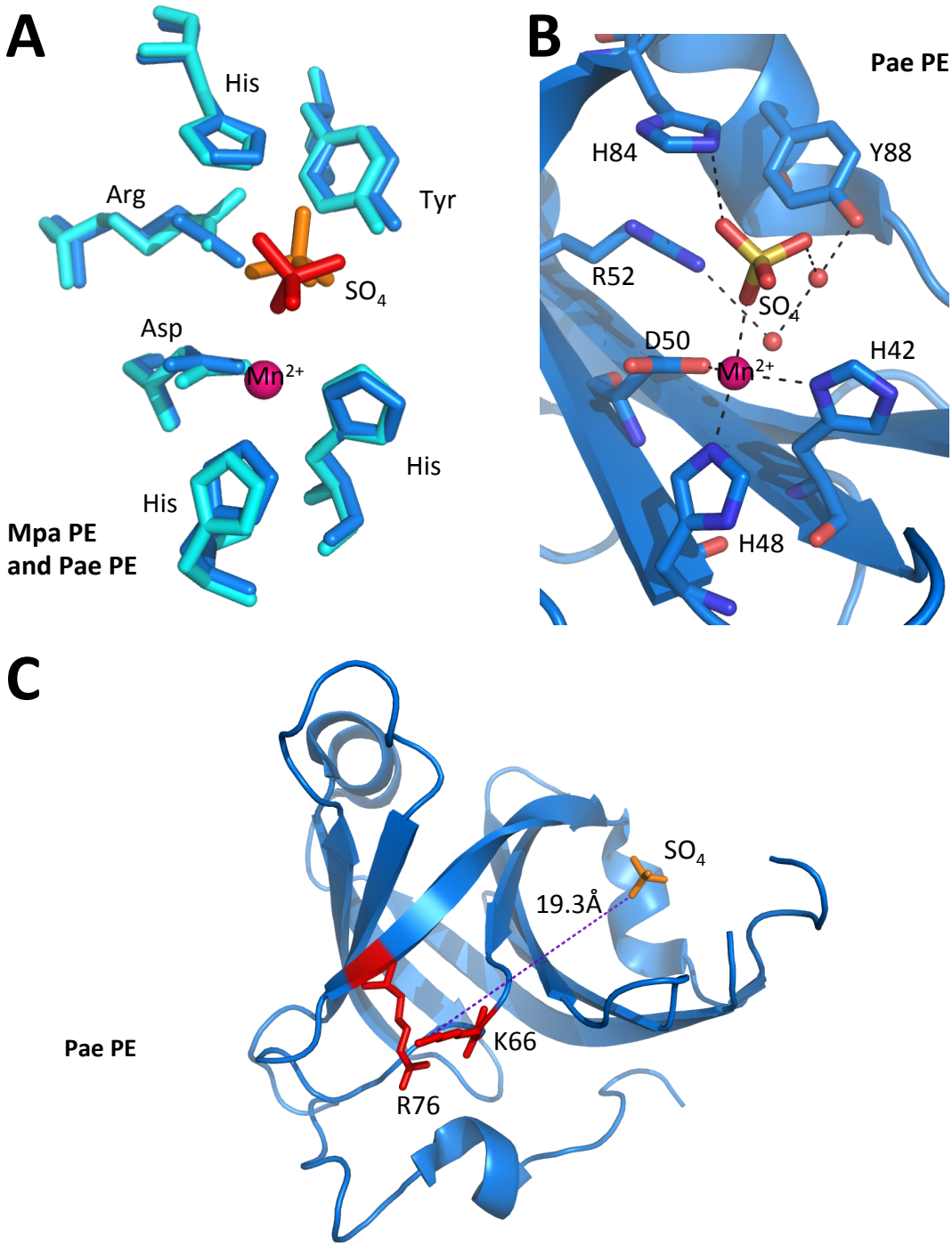


Figure 5.22 The conserved active site architecture of the PE, and the distance to ribonuclease essential residues from the active site.

(**A**) An overlay of the catalytic residues of Mpa PE (cyan) and Pae PE (blue). The sulphate anion from Mpa is shown in orange whilst the sulphate from Pae is shown in red. The manganese ion (pink) is from the Pae structure and fits with both sets of residues. (**B**) A ribbon diagram of Pae PE with the six key active site residues shown with a hydrogen bonding network with sulphate and manganese ions and water molecules. (**C**) A ribbon diagram of Pae PE with ribonuclease specific residues K66 and R76 coloured red. The distance from the active site to the residues is shown. The distance is too great for either residue to be directly involved in 2'-OH recognition of the penultimate ribonucleoside at the active site. The diagrams of Pae PE were made using the structures from Nair et al., (2010), PDB 3N9D.

phosphate in the active site (Figure 5.22, C). Prior to the structure of PE being solved, a reasonable prediction might have been that these residues contacted the 2'-OH of the penultimate ribonucleoside, since this would offer them an activity specific to ribonuclease activity but not 3'-phosphate removal. However, they are far from the phosphate binding residues, it would be impossible for them to contact the hydroxyl of the penultimate ribonucleoside whilst the ultimate base was engaged in the active site. A maximum distance of ~6 Å is feasible from the position of the phosphate in the active site that would allow access to the 2'-OH of the ribose of the next base. It is perhaps more likely that these residues offer structural stability, or interact with the upstream duplex region, although it is unclear how they might do so in a way that would only benefit one of the two phosphoryl transfer reactions. Many of the unresolved questions about the importance of individual residues of the PE will remain unanswered until a structure in complex with a substrate emerges.

5.9.2 Orientation of the DNA-RNA substrate in the PE active site

None of the structures of the NHEJ PE have revealed precisely how the enzyme binds DNA, and in which orientation the strands access the active site. However, there seems a likely channel in which the substrate docks (Figure 5.23, A). The figure shows an electrostatic surface representation, with the vanadate molecule opposite an electro-positive region, suggesting the area of the channel that binds the DNA substrate. Given that the scissile phosphate almost certainly adopts the position taken by the sulphate and vanadate anions seen in the PE structures, it is likely that the phosphate backbone of the primer and D-strands lays in the channel indicated in Figure 5.23, A. Another view of this representation is shown in Figure 5.23, B, in which the loop region occludes the active site. This view allows for the superposition of an idealised DNA duplex. It suggests that when the DNA-RNA is bound by the PE in a closed formation that the ribonucleoside to be removed is buried inside the cleft. Whilst we cannot be sure of the geometry of the DNA on either side of the active site, since it may be distorted to allow easier access to the scissile phosphate, we can speculate on which side the primer and D-strand occupy. Given the that the H82 residue is essential for the removal of annealed ribonucleosides in Mpa PE, one might surmise that it plays a role in either interrupting the base-pairing of the terminal ribonucleoside, or stabilising the position of it on one side of the active site. When considering the position of this histidine residue in Figure 5.20, relative to the presumed phosphate position, it suggests that the terminal ribonucleoside must on the arginine, histidine and tyroside 'side' of the active site. The penultimate ribonucleoside, therefore, must be the other side of the active site in the position indicated by the orange base

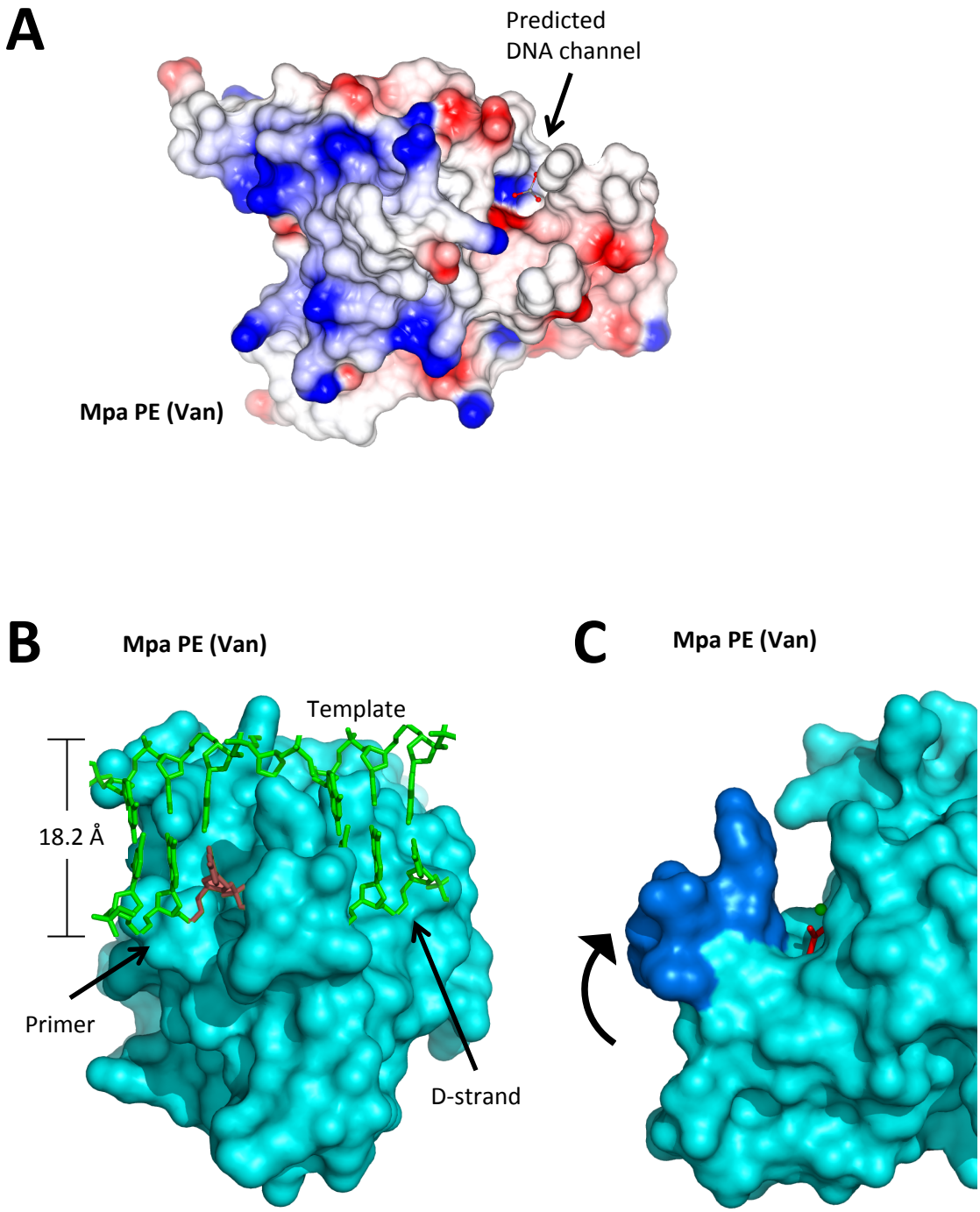


Figure 5.23 Structural insights into Mpa PE substrate binding.
(**A**) An electrostatic surface representation of Mpa PE (Van) with vanadate shown in red and grey. The predicted channel for DNA to access the active site is indicated by an arrow. (**B**) A surface representation of Mpa PE with an overlay of a DNA-RNA substrate. The position of the DNA is purely hypothetical, although the width of the helix is accurate to the Mpa PE structure. The direction of the primer and D-strand is estimated by the position of the scissile phosphate mimetic in the active site, and the position of the H82 residue, which may help to orient the terminal ribonculeoside for cleavage. (**C**) A surface representation of Mpa PE showing the active site cleft which binds the DNA, and the mobile loop region (shown in blue) which encloses around the substrate.

in Figure 5.23, B. It also seems likely that the D-strand should occupy the other side of the cleft, as shown in Figure 5.23, C. This region of the PE is more open and could accommodate a displaced D-strand. It is also worth considering that the N-terminal 30 residues that are highly flexible could interact with the substrate in this region. There is no electron density observed for these residues in any of the existing structures, but they have been shown to affect phosphatase activity only, in alanine substitution mutants. Figure 5.23, C also shows the flexible loop region highlighted in blue, with an indication of its movement in adopting the closed conformation.

5.10 Crystallisation screening of Mpa PE with DNA-RNA substrate

5.10.1 Early attempts to co-crystallise Mpa PE with DNA substrates

The DNA fragments that were soaked into Mpa PE crystals along with vanadate did not produce any PE-DNA data. The DNA soaked crystals also became fragile, and it seemed that co-crystallisation might provide more crystals. At this point, the importance of the D-strand in PE activity discussed in chapter 4 had not yet been discovered, and early substrate choice revolved around small oligomers and even single nucleotides. A co-crystallisation screen was set up with Mpa PE and AMP, using the same PEGs suite condition that produced the apo crystals. At 5 mg/mL PE and 1.25 mM AMP, a large number of small crystals grew (Figure 5.24, A), and eventually reached 100 x 100 x 100 µm size. Data was collected from several of these crystals, but no discernable electron density for the AMP could be detected. The structures were not fully refined, since they did not offer any additional information that wasn't already revealed by the prior Mpa PE and Mpa PE (Van) structures.

5.10.2 Mpa PE with DNA-RNA, $NaVO_4$, manganese and cytidine

The studies described in chapter 4 discovered a preferred substrate for the phosphoesterase; an annealed break with RNA at the 3' side of the nick. New substrates for crystallography were designed to match these findings, with a view to constructing multiple different substrates using the same template DNA. The primary target was to crystallise Mpa PE with a DNA-RNA nick like substrate. The ideal substrate would be a di-ribonucleoside at the 3' side of the nick with a D-strand on a small DNA template, however Mpa PE would cleave the di-ribo substrate and then dissociate. Earlier experiments described in chapter 3 established several metal ions that could 'poison' the reaction and prevent Mpa PE from cleaving the di-ribonucleoside, however it was not known how these ions might affect coordination of the active site residues.

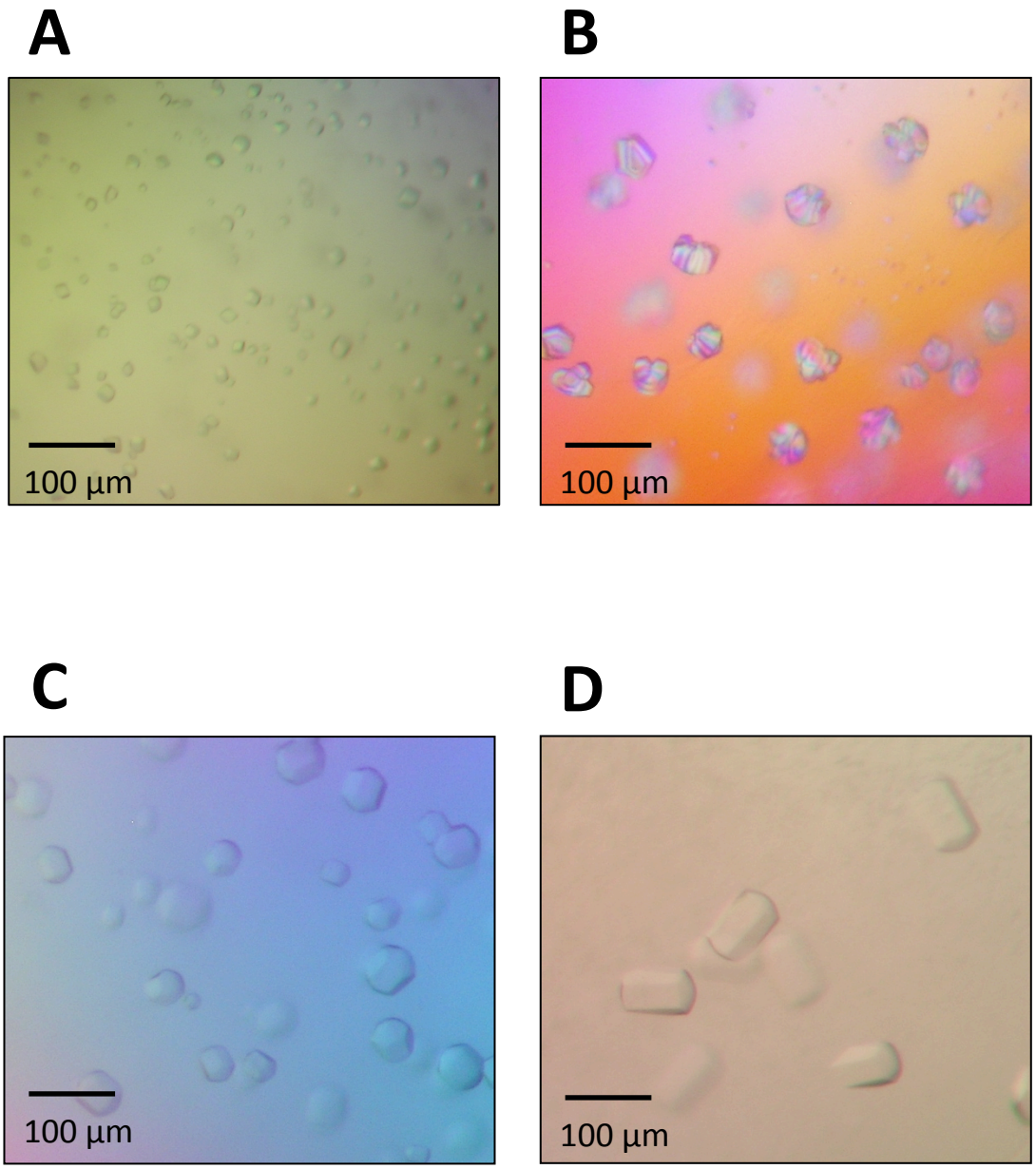


Figure 5.24 Attempts to crystallise Mpa PE with DNA-RNA substrate.
(**A**) Mpa PE crystals grown using the PEGs suite (Qiagen). 0.7 µL protein: Mpa PE 5 mg/mL, AMP 1.25 mM. 0.7 µL reservoir: 0.2 M Ammonium sulphate, 20 % (w/v) PEG 3350. (**B**) Mpa PE crystals grown using the Nucleix suite (Qiagen). 0.7 µL protein: Mpa PE 9 mg/mL, PE X BCE 800 µM, $NaVO_4$ 2 mM, Cytidine 2mM, and $MnCl_2$ 2 mM. 0.7 µL reservoir: 0.05 M Tris pH 8.5, 30 % (w/v) PEG 4000, 0.2 M NH_4Cl, 0.01 CaCl. (**C**) Mpa PE crystals grown using the Ammonium sulphate suite (Qiagen). 0.7 µL protein: Mpa PE 9 mg/mL, PE X BCE 800 µM, $NaVO_4$ 2 mM, Cytidine 2mM, and $MnCl_2$ 2 mM. 0.7 µL reservoir: 0.1 M MES pH 6.0, 2.4 M NH_4SO_4. (**D**) Mpa PE crystals grown using the Ammonium sulphate suite (Qiagen). 0.7 µL protein: Mpa PE 9 mg/mL, PE X BCE 800 µM, $NaVO_4$ 2 mM, Cytidine 2mM, and $MnCl_2$ 2 mM. 0.7 µL reservoir: 0.2 M di-sodium tartrate, 2.2 M NH_4SO_4.

It was decided to use manganese as the co-factor to ensure proper coordination, and instead use a substrate that would hold the enzyme in place. A 1-nt gap substrate with an incoming ribonucleoside could utilise vanadate as a phospho-mimetic, creating a phosphodiester bond that the PE could not hydrolyse (Figure 5.25, A). Theoretically this should capture the PE engaging the terminal ribonucleoside, with the vanadate in the active site, and the penultimate ribonucleoside 2'-OH contacts being made elsewhere. Using vanadate to solve crystal structures of phosphoryltransfer enzymes was discussed briefly above, two examples of structure solved using vanadate are shown in Figure 5.26. The figures show two substrate fragments linked by vanadate, that is then engaged by the enzymes; a TDP1 histidine coordinating vanadate that is joining a TOP1 peptide fragment and a short DNA fragment (Figure 5.26, A), and Hairpin-ribozyme bases positioning a vanadate that is bonded to two RNA fragments (Figure 5.26, B) (Davies et al., 2003; Rupert et al., 2002). Vanadate might also be used to capture Mpa PE engaging with a 3'-phosphate-like substrate. A DNA substrate was designed with a recessed 3' end that could bind PE via vanadate (Figure 5.25, C).

Crystal screening proceeded with 9 mg/mL Mpa PE, 800 µM PE X BCE, 2 mM $NaVO_4$ 2 mM cytidine and 2 mM manganese, and produced crystals in the initial screen using the Nucleix suite (Qiagen), shown in Figure 5.24, B. The crystals were small, ~60 x 40 x 40 µm, but a promising early hit. A grid was created around the condition that produced the hit; 0.05 M Tris pH 8.5, 0.2 M NH_4Cl_2, 0.01 M CaCl, and 30 % PEG 4000. The grid varied the PEG and ammonium chloride concentrations, and a sister grid was created using ammonium sulphate in place of ammonium chloride. The grids were not successful in generating any new crystals, and screening continued using the Ammonium Sulphate suite (Qiagen). Crystals grew in two drops in the NH_4SO_4 screen, shown in Figure 5.24, C and D. Screening continued with both conditions, whilst varying the substrate-to-protein ration from 2:1 to 1.5:1. Crystals eventually grew to 75 x 100 x 75 µM using both ratios with the condition: 0.2 M di-sodium tartrate, 2.2 M NH_4SO_4. These crystals were used for data collection, using the Rigaku MicroMax 007-HF. The crystals were prone to dissolution during cryoprotection with ethylene glycol, and paratone-N was eventually used to stabilise the crystals to gather data. Diffraction patterns suggested that the crystals were largely composed of salt, although some closely spaced spots resembling protein diffraction were observed. The crystals were most likely not composed of Mpa PE : DNA complexes, and fresh screening with the complex is required.

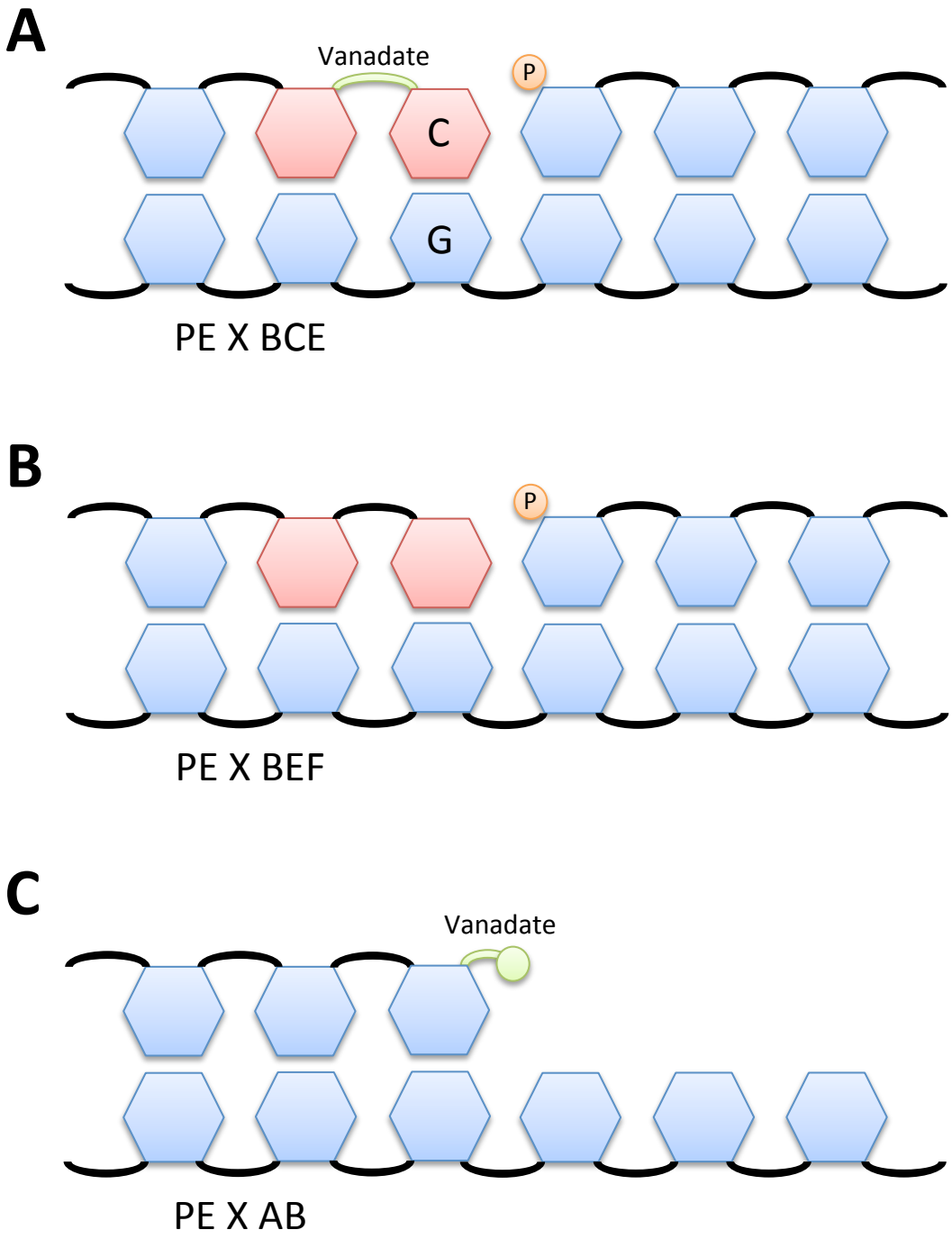


Figure 5.25 Schematics of potential crystal structure compositions.
Schematics of DNA substrates designed to form a stable complex with Mpa PE or Mpa PE H82A. (**A**) PE X BCE DNA, which has a 1-nt gap. The gap is filled using cytidine, and $NaVO_4$ is added as a non-cleavable phosphate mimetic, which should allow Mpa PE to bind as if it was engaging a DNA-RNA nick. Manganese could also be added to allow proper positioning of the vanadate in the active site. (**B**) PE X BEF DNA is the hydrolysable equivalent of the complex shown in A, and should make a suitable substrate for Mpa PE H82A. Although any potential structure would be missing the catalytic histidine, hopefully the enzyme could still bind to the substrate and deliver valuable information on DNA positioning.(**C**) PE X AB DNA offers a 3' recessed end that could bind Mpa PE in the presence of $NaVO_4$, potentially capturing the enzyme engaging in phosphoesterase activity.

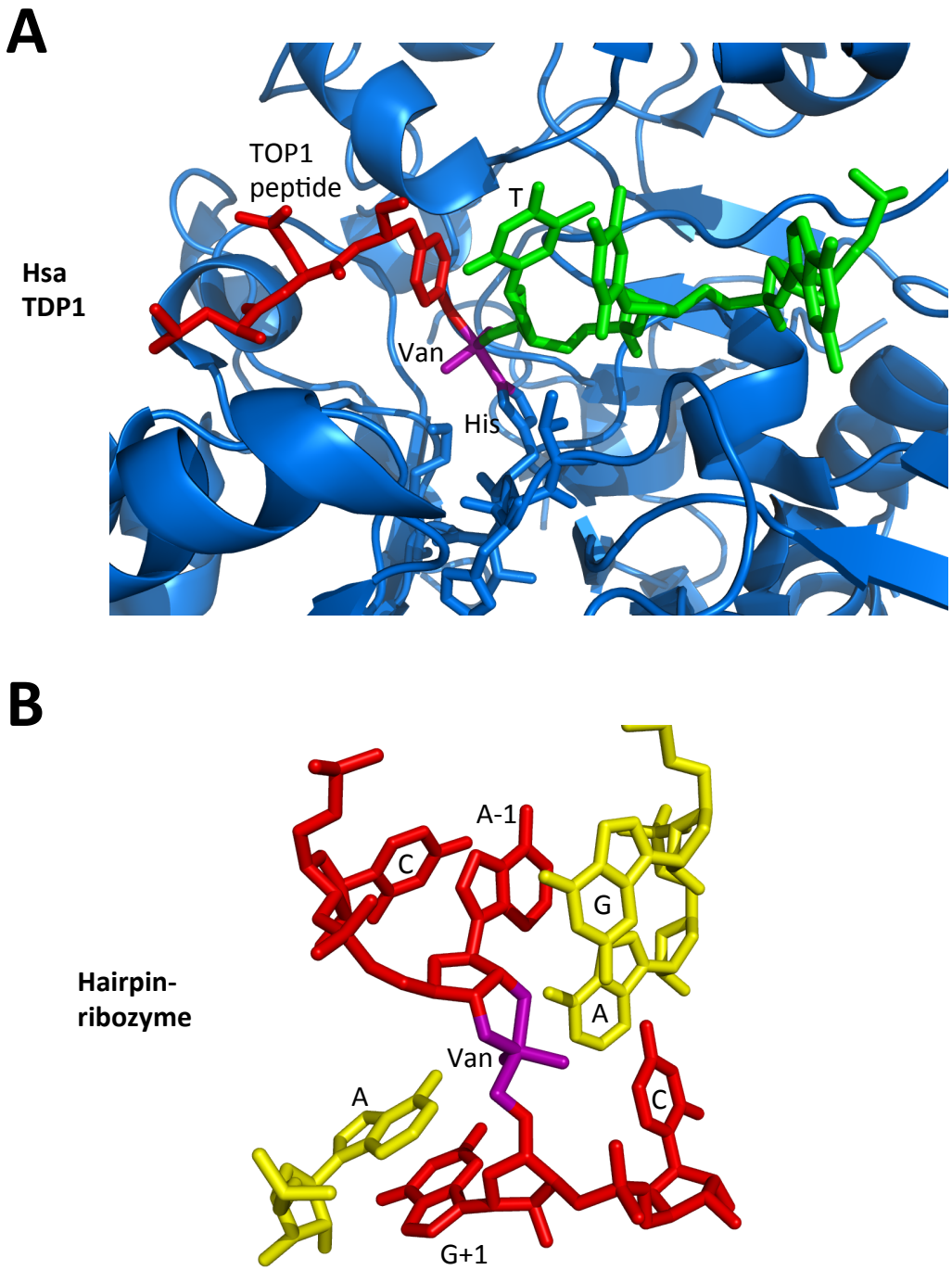


Figure 5.26 Crystal structures using sodium vanadate as a phospho-mimetic.
(**A**) A crystal structure of TDP1 bound to a vanadate-TOP1 peptide-DNA fragment substrate complex. TDP1 is shown in blue, with a histidine residue contacting the vanadate (purple). The vanadate molecule is mimicking the phosphotyrosyl linkage to DNA that is formed following abortive TOP1 activity, with the TOP1 peptide in red, and the DNA fragment in green. The figure of TDP1 was made using the structure from Davies et al., 2003, PDB 1NOP. (**B**) A figure of Hairpin-ribozyme engaging two RNA fragments that are joined by vanadate. The bases of Hairpin-ribozyme are shown in yellow, whilst vanadate is purple, and the RNA fragments are red. The figure of Hairpin-ribozyme was made using the structure from Rupert et al., 2002, PDB 1M5O. Figures adapted from Davies et al., 2004.

5.10.3 Towards crystallising Mpa PE with a phosphatase substrate

It was thought possible to capture crystals of Mpa PE engaging with a DNA substrate during phosphatase activity if vanadate was used as a mimetic for the scissile phosphate. A substrate was designed using the same template as those for PE DNA-RNA nicks, shown in Figure 5.25, C. Only primary screening has commenced so far with 9 mg/mL Mpa PE, 800 µM PE X AB, 2 mM $NaVO_4$ and 2 mM manganese. No crystals have been observed thus far, but extensive screening would be worthwhile, given that the role and structure of the N-terminus is elusive, and it may stabilise during phosphatase activity.

5.10.4 Towards crystallising Mpa PE H82A with DNA-RNA and manganese

Another method of generating crystals of Mpa PE bound to a DNA-RNA nick was to use the active site mutant H82A, which cannot cleave an annealed di-ribonucleoside. A single screen of Mpa PE H82A only was set up using the same condition that produced Mpa PE wt crystals; 0.2 M Ammonium sulphate, 20 % (w/v) PEG 3350. Crystals grew more slowly than wt, but continued growth over several months until sizes of over 400 x 400 x 400 µm were reached (Figure 5.27, A and B). These crystals were not used for data collection, but as an indicator of whether Mpa PE H82A could crystallise in a similar manner to the wt. Given the promising nature of the Mpa PE H82A crystals, a similar substrate to PE X BCE was made, except with a di-ribonucleoside (Figure 5.25, B). Screening with Mpa PE H82A could include manganese, since the enzyme cannot cleave the substrate, and vanadate was also not necessary. Primary screening with 9 mg/mL Mpa PE, 1.2 mM PE X BEF, and 3 mM manganese produced two hits, shown in Figure 5.27, C and D, which require further screening around.

5.11 Discussion

The aims of the studies described in this chapter were to expand our understanding of the function of the Mpa NHEJ complex, and to determine structural information that could inform on additional aspects of the preferences of the enzymes. One of the primary targets was to finally solve the structure of the phosphoesterase since, when this research began, there was no structural information published on any NHEJ PE. We have learned that both bacterial and archaeal enzymes consist of a conserved hydrophobic core with an active site that sits on the outside, with a flexible loop region that can enclose around an incoming substrate. The NHEJ PE has a unique fold, and has an active site configuration that is not typical of any 3' polynucleotide exonuclease or phosphatase (Nair et al., 2010). It seems that the PE has

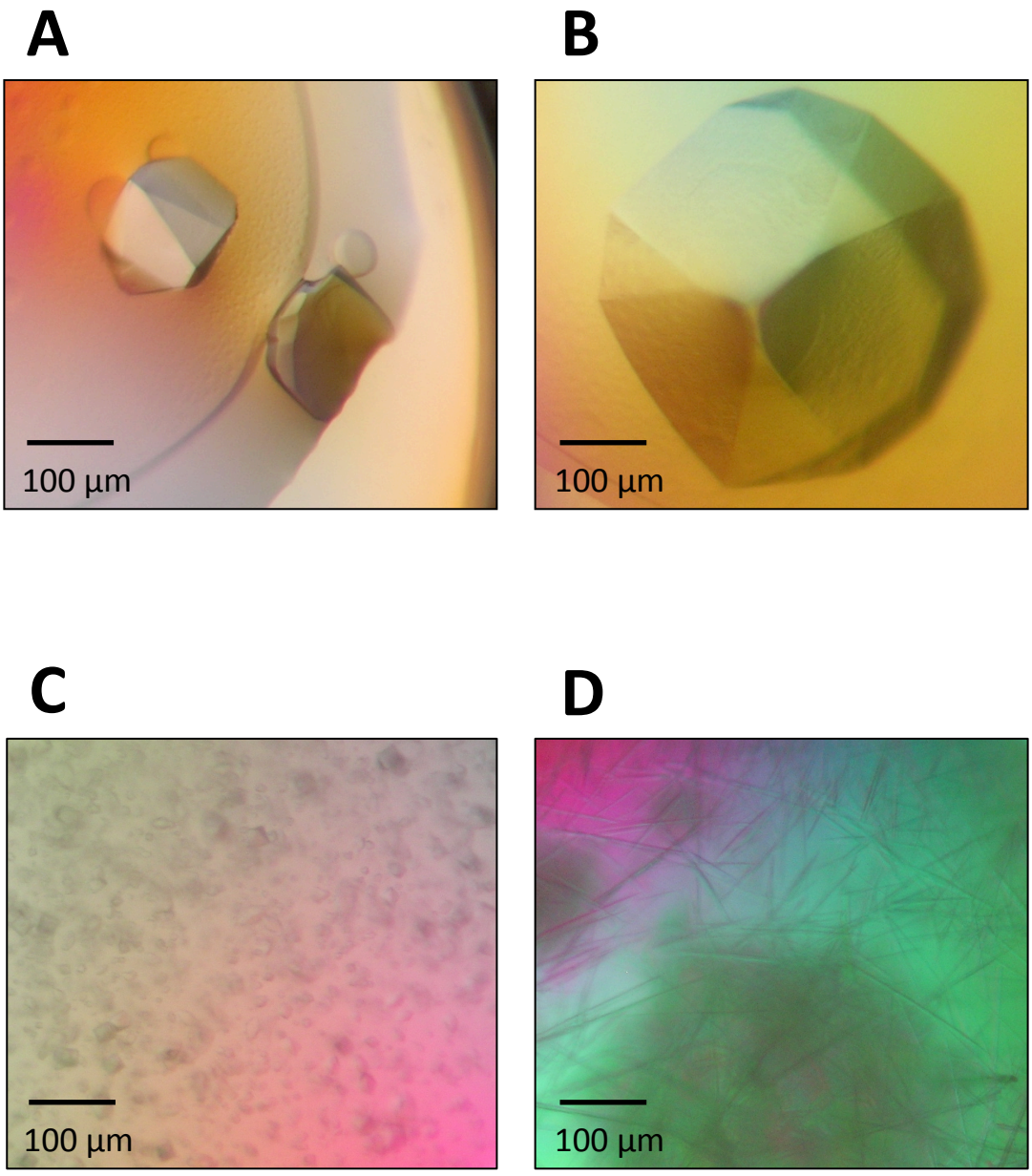


Figure 5.27 Mpa PE H82A apo crystals.
(**A-B**) Mpa PE H82A crystals grown using the PEGs suite (Qiagen). 0.7 µL protein: Mpa PE H82A 14.95 mg/mL (**A**), 57.2 mg/mL (**B**), 0.7 µL reservoir: 0.2 M Ammonium sulphate, 20 % (w/v) PEG 3350. Crystals grew slowly over a three month period. (**C**) Mpa H82A DNA crystals grown with the Ammonium sulphate suite. Mpa PE H82A 15 mg/mL, PE X BEF 1.2 mM, manganese 3 mM. 0.7 µL reservoir: 3 M Ammonium sulphate, 1 % MPD. (**D**) Mpa H82A DNA crystals grown with the PEGs suite. Mpa PE H82A 15 mg/mL, PE X BEF 1.2 mM, manganese 3 mM. 0.7 µL reservoir: 0.1 M sodium acetate ph 4.6, 15 % (w/v) PEG 20000.

evolved to fill a specific niche that is unlike any other previously characterised. We recently determined that the PE regulates NHEJ Pol activity when it is synthesising RNA at annealed breaks, and that it specialises in optimising NHEJ repair intermediates for ligation. In light of this, it should not be surprising that the NHEJ PE has a novel structural configuration. Whilst the detailed structures of the PE allow us to speculate on the likely orientation of an incoming substrate, and estimate the whereabouts of the amino acid residues that confer ribonucleoside specifity, only a structure of PE bound to DNA-RNA will fully disclose this information.

Experimental work on determining a crystal structure of the PE bound to a substrate is still underway, and there is still scope for substantial crystal screening using the DNA substrates detailed in Figure 5.1. The availability of different substrates which either consist of a di-ribonucleoside, or can form a di-ribonucleoside with vanadate, along with a variety of metal ions that alter the activities of Mpa PE, offer a solid platform to continue structural work. The H82A PE mutant further expands the opportunities to crystallise the enzyme bound to a relevant substrate. If no new crystals are forthcoming after future screening, the DNA substrate could be altered. The PE is a small enzyme, and long DNA substrates may affect crystal packing. Although the 11-mer currently in use is quite short, it may be possible to reduce this even further. Another possibility is to use very small components that could be soaked into the existing Mpa PE wt crystals. When vanadate and short oligomers were soaked into PE crystals, they became fragile. However, it may be possible to soak in vanadate with a ribonucleoside and a ribonucleoside monophosphate. Although this substrate would not offer the full advantages of a ds duplex with a D-strand, it may be stable enough to allow the bases to interact properly with the active site. Much has been learned about the nature and role of the PE in NHEJ, but it remains the most mysterious of the four key NHEJ proteins, and the one that could potentially reveal the most exciting information in the near future.

One of the objectives for the archaeal NHEJ Pol was to use it as an anchor for Mpa Ku on DNA. The screening so far has not produced any positive results, although progress is still possible with modified experimental design. It may be preferable to redesign the DNA substrate to utilise a cloverleaf set of hairpins at one terminus of the DNA to trap the Ku dimer, as seen in the *H. sapiens* Ku structure (Walker et al., 2001). Then a short 5' recessed end with a 5'-P should attract the polymerase, and the two enzymes might form a stable complex with the DNA. This is perhaps the next most pressing target that exists, structurally, for AP-NHEJ research. It should inform on how the two proteins directly interact, and whether Ku remains

in contact once the Pol has engaged the break. The orientation of the complex would also be very interesting to learn.

The recent discovery of the ability of the NHEJ Pol to displace D-strand DNA in order to extend a 3' terminus has also raised an intriguing new structural project. It would be most interesting to learn which residues engender the ability to displace the downstream DNA strand. Whilst previous efforts to solve a structure of the NHEJ Pol bound to a gap substrate have met with frustration (Brissett and Doherty unpublished observations), a recent structure of two Mtu PolDom molecules annealing a break allows for speculation (Brissett et al., 2013, under review). Figures 5.14 and 5.15 speculate on how the DNA might be configured in gap-filling repair, and suggest how the polymerase might separate the paired downstream duplex, and where the displaced DNA might traverse to. It may be possible to crystallise an NHEJ Pol in the process of displacing a D-strand, and it would be vital to note whether there was any electron density for the position of the D-strand. If there are structural features on the NHEJ Pol that facilitate the ongoing displacement, then this may be possible. It is also possible that the DNA is simply free to move over the surface of the Pol, and is not attached, in which case it is unlikely that it could be visualised in a crystal structure. It is pertinent to consider the gap-filling assay shown in Figure 4.30, where a quadruple alanine mutant of Mtu PolDom could still gap fill, but did not have any discernable strand displacement ability. The mutated residues were on Loop 1 and Loop 2, which both seem structurally distant to the strand displacement area. It may be that these features are necessary for efficient function of the NHEJ Pol, and that without them usual activity is somewhat impaired. The structural information from Mtu PolDom and Mpa Pol implicate the 'wedge' in the disruption of the downstream duplex, and it would be very surprising if future experiments prove that this is not the case.

Attempts to produce a structure of Mpa Lig bound to a DNA-RNA have not yet been successful, although continued screening may yet yield crystals. If not, the substrate may need redesigning, possibly making use of a smaller oligo, or perhaps changing the mechanism of trapping the ligase. If the terminal 3' ribonucleoside of the nick was altered to remove the 3'-OH group instead of the 2'-OH, the nick sealing step could not occur, and the ligase should remain bound to the adenylylated 5'-P. Such a modification might be very costly to produce at the yield necessary for crystallography, therefore other methods could be used, such as replacing the metal ion with one which would poison the reaction. Zinc or nickel might allow the ligase to bind the substrate, but be unable to catalyse adenyltransfer.

The crystal structures of the NHEJ Pol and PE confirm that the archaeal NHEJ components are highly similar to the bacterial equivalents. The archaeal PE has proven a good choice for biochemical and structural research, given that it is easier to purify at high yields than the bacterial orthologue, and remains soluble even at high concentrations. Whilst the research described here did not produce DNA bound NHEJ enzymes, the foundations for continued structural studies have been laid, and they remain realistic targets for the future.

Chapter 6

Conclusion

This thesis details the existence of an NHEJ repair pathway in an archaeon, *Methanocella paludicola*, which is closely related to an orthologous repair mechanism present in many bacteria. Previously, the only NHEJ-like genes identified in an archaeal genome were Ligase D and Ku in *Archaeoglobus fulgidus*. The NHEJ operon discovered in *M. paludicola* has a 'complete' set of repair proteins that is highly similar to those studied in Mycobacteria, including a NHEJ polymerase and phosphoesterase along with a LigD and Ku. The NHEJ complex of *M. paludicola* was used to establish that an archaea can repair DSBs using similar proteins to bacteria, and also to answer outstanding questions that the bacterial orthologues had not. An overview of AP-NHEJ is shown in Figure 6.1, adapted from (Lieber & Wilson, 2010).

6.1 The Nonhomologous end joining pathway can repair a variety of different double strand breaks

6.1.1 AP-NHEJ can repair many configurations of DNA termini

The studies described in this thesis, and previous research in the bacterial NHEJ pathway, have demonstrated that the minimal AP-NHEJ complex is capable of repairing a vast variety of different DSBs (Aniukwu et al., 2008; Della et al., 2004; Pitcher et al., 2005; Brissett et al., 2007; Brissett et al., 2011; Zhu & Shuman, 2005b; Zhu & Shuman, 2008). The AP-NHEJ pathway can repair blunt-ended breaks and DNA termini with different configurations of 5' and 3'-overhanging ends. It is conceivable that almost any combination of long or short overhanging ends could be repaired by the combined activities of the NHEJ Pol, PE and Ligase D. It is unknown if the NHEJ Pol is capable of synthesising RNA opposite damaged bases to allow for repair of DNA termini with compound damage. Recent research in the Doherty lab has characterised a eukaryotic replicative primase that is a member of the AEP superfamily, and is capable of bypassing or extending primers opposite bulky DNA lesions caused by UV irradiation (Bianchi et al., 2013, manuscript submitted). Preliminary experiments with the Mpa NHEJ Pol suggest that it does not share the same abilities of DNA damage bypass. However, prior experiments with the *M. tuberculosis* NHEJ Pol established that it can accurately replicate opposite an 8-oxo-G lesion (Pitcher, Brissett, Picher, et al., 2007b). This is certainly an area of interest for future research, and it is necessary to explore the limits of which varieties of DNA damage the NHEJ Pols can tolerate whilst repairing DNA termini. The AP-NHEJ repair pathway is certainly comparable to that of the eukaryotic NHEJ system, and can restore a similar scope of DSBs (Aniukwu et al., 2008, Della et al, 2004, Pitcher et al., 2005).

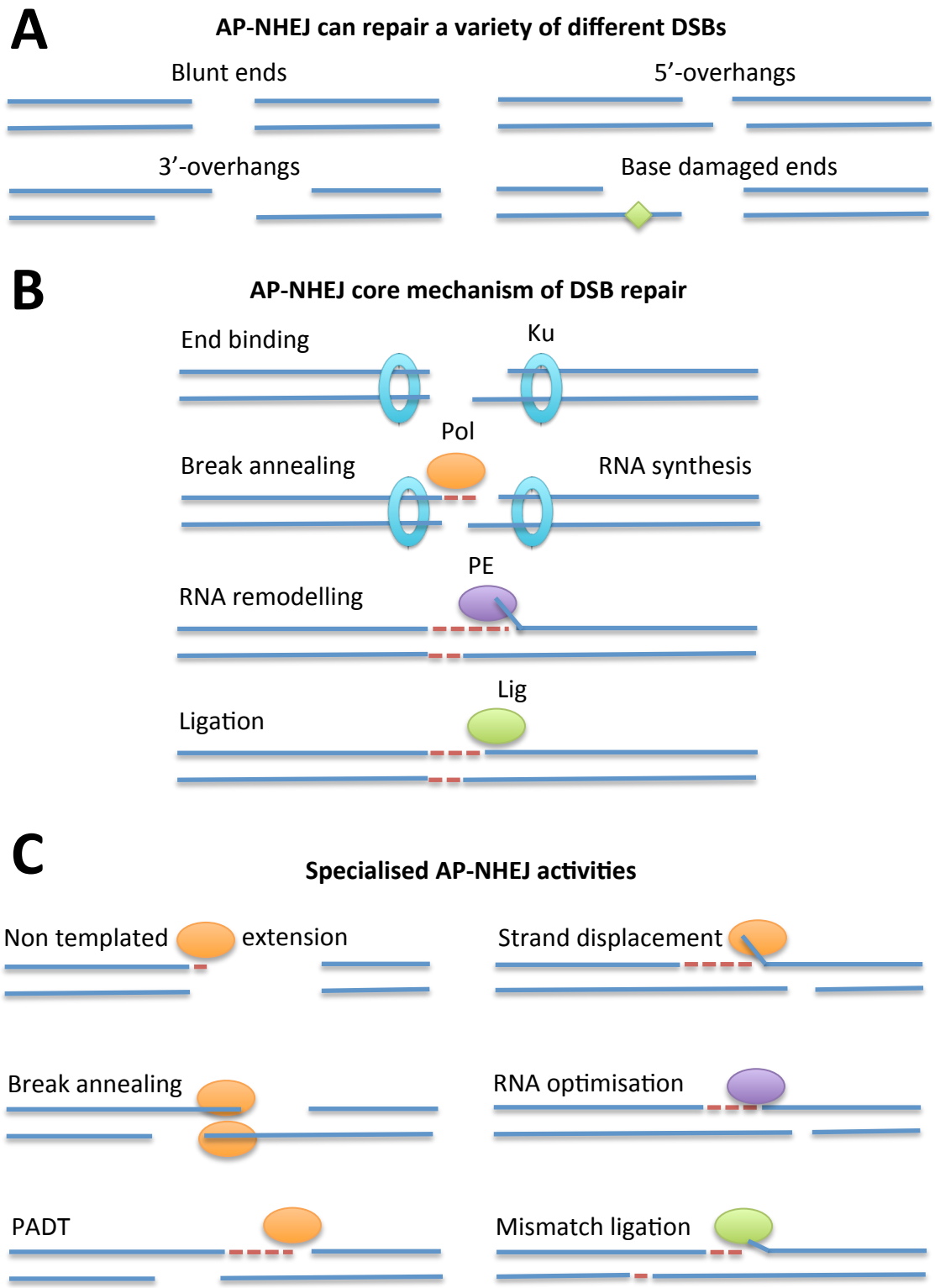


Figure 6.1 An overview of AP-NHEJ.
(**A**) AP-NHEJ can repair a variety of different DSBs; blunt, 3’-overhang, 5’-overhang, damaged bases, or a mixture of ends. (**B**) A schematic of the core AP-NHEJ mechanism. Ku binds ends, before the Pol can anneal the breaks and synthesise RNA to begin the repair. The PE aids the break processing and optimises the RNA tract for ligation. The Lig then seals the RNA-DNA nick. (**C**) The specialised AP-NHEJ activities; the Pol can extend from blunt ends, anneal breaks, extend across broken ends and displace D-strand DNA. The PE can repair D-strand intermediates and the Lig can seal mismatched nicks. Adapted from Lieber and Wilson (2010).

6.1.2 AP-NHEJ utilizes a range of mechanisms for repairing DSBs

One of the incredible aspects of all NHEJ repair pathways is the range of mechanisms that can be used to fix broken DNA ends. It is likely that the highly flexible nature of the NHEJ repair pathway allows the complex to cope with such a wide array of different DNA damage configurations. One of the key features of NHEJ is the ability to harness the instrinsic base-pairing capability of DNA to create microhomology regions that tether broken ends. Some broken DNA ends may have sequences that are amenable for generating a microhomology, whilst the eukaryotic system also makes use of exonucleases that can expose suitable sequences further upstream or downstream (Mahaney et al., 2009; Lieber, 2010). Whilst there is no clear candidate for the 3'-exonuclease in AP-NHEJ, resection is observed during *in vivo* DSB plasmid repair assays in mycobacteria (Aniukwu et al., 2008). An interesting addition to end-joining abilities is that the AP-NHEJ Pols can play an active role in enabling break annealing and creating a microhomology (Brissett et al., 2007; Brissett et al., 2013, under review). It is unknown whether any of the eukaryotic family X polymerases can replicate this ability, or whether other components of the NHEJ complex fulfill this role.

One of the distinct variations between AP-NHEJ and eukaryotic NHEJ is that the archaeo-prokaryotic enzymes almost exclusively prefer to repair with RNA. One of the repercussions of this mechanism is that the AP-NHEJ enzymes cannot repair a blunt-ended DSB without inserting a ribonucleotide (Aniukwu et al., 2008). A characteristic of all types of NHEJ repair is the variety of repair outcomes that occur even when the exact same sequence of broken ends is fixed. Whilst AP-NHEJ does not offer 'accurate' repair of precisely broken blunt ends, it can repair them with a single nucleotide addition. Neither AP nor eukaryotic NHEJ complexes create lengthy insertions during plasmid repair, and this is most likely because of the distributive nature of the related NHEJ polymerases. It is more common to observe extensive deletions, although this may follow failed attempts to ligate DSBs. It is notable that the NHEJ repair systems opt to reconfigure the broken DNA termini with small additions, if possible. Presumably this prevents significant alteration of the genetic sequence following a break. The AP-NHEJ system uses the phosphoesterase to regulate the RNA synthesis and strand displacement of the NHEJ Pol, but it never resects beyond the first ribonucleoside. This continues with the theme of offering considerable flexibility of repair options whilst altering as little of the original sequence as possible.

6.2 Specialised AP-NHEJ abilities

The most striking facet of AP-NHEJ is the preference for RNA based repair of DSBs. Future research should investigate how each of the enzymes select for RNA usage, and then why they evolved to do this. It is possible that the Ligase and PE developed stringency for RNA based repair to encompass the preferences of a polymerase that evolved from an RNA primase. However, recent research has disclosed that some eukaryotic polymerases also preferentially insert ribonucleotides, so there may be other functional benefits to repairing breaks in this manner (Martin et al., 2013).

6.2.1 NHEJ Polymerase

The NHEJ Pol is a member of the AEP superfamily, and is closely related to the PriS sub-unit of the PriSL component of the replication machinery found in archaea and eukaraya (Iyer et al., 2005). The canonical role of PriS is to generate short RNA primers (between 2-12 nt), however some archaeal PriS enzymes have been shown to generate DNA primers of a much greater length (Frick & Richardson, 2001; Bocquier et al., 2001). Recently an archaeal PriS has been shown to polymerise DNA across two discontinuous ends (polymerisation across discontinuous templates; PADT), and it was suggested that PriS could support NHEJ mediated DSB repair (Hu et al., 2012). It is possible that PriS and NHEJ Pol could have some overlapping functions. It would be interesting if this was the case, and may indicate how the NHEJ Pol evolved into an end-joining protein from a primase. Future experiments should seek to discover if PriS enzymes, that are very similar to NHEJ Pol, could fill DNA gaps and even displace D-strands.

The nucleotidyltransferase activities of the NHEJ Pols are extensive; distributive template-dependent and non-templated extension of DNA primers, filling of single-strand gaps in DNA, and limited PADT (Lieber, 2010; Mahaney et al., 2009). AP-NHEJ Pols are capable of the same, and can also perform DNA overhang end synapsis, 5’-phosphate binding and pre-ternary complex formation, and the ability to displace downstream DNA strands (Brissett et al., 2007; Brissett et al., 2011). Whilst the benefits of encouraging a microhomology and forming a pre-ternary complex are clear, the ability to displace downstream DNA is not. Yet, the ability for a DNA polymerase to mobilise DNA around itself is not a novel concept. Replicative polymerases have a proofreading ability which can excise a single base in the 3’-5’ direction and *H. sapiens* Pol β can unwind DNA in the 5’-3’ direction whilst extending a DNA strand (Garg & Burgers, 2005; Chary et al., 2012). Archaeal B-family polymerases are found to possess an extended proofreading activity when they encounter a uracil base on the templating strand (Richardson

et al., 2013). The *Pyrococcus furiosus* (Pfu) polymerase can unwind and cleave the DNA duplex in the 3'-5' direction by up to four bases, leaving the uracil templating base at the +4 position. Structural studies show that the upstream strand displacement by the Pfu Pol is caused by arginine and tyrosine residues (Richardson et al., 2013). The disruption created by these residues is similar to that of arginine (R53) and proline (P55) residues seen in Mtu PolDom, suggesting that they may play a similar role (Figure 6.2, A). An interesting consideration as to why the AP-NHEJ Pol may have this ability is that it could potentially unravel short hairpins or DNA snap-backs at broken ends. A 5'-overhanging end should offer a template to replicate, but the DNA could snap-back and self anneal given the necessary sequence, creating a hairpin end. The hairpin-ended DNA would essentially be similar to a gap or nick DNA substrate, depending on the DNA sequence (Figure 6.2, B). The AP-NHEJ Pol could potentially unravel such a structure and allow for the ends to be repaired. This function would be somewhat analogous to one of the roles of Artemis, the DNA exo/endonuclease in higher eukaryotes. Future experiments could see if DNA substrates designed with a snap-back sequence on an overhanging end could be unraveled by an NHEJ Pol, and whether discontinuous ends that could not be joined as hairpins, could be repaired in the presence of a NHEJ Pol.

6.2.2 NHEJ Phosphoesterase

One of the most pressing unanswered questions still concerning the AP-NHEJ system was the role of the phosphoesterase. The preliminary predictions for the domain was a 3'-exonuclease that could perform a similar role to those found in the eukaryotic system (Della et al., 2004; Pitcher et al., 2005). Later studies elucidated the catalytic abilities of the enzyme as a 3'-ribonuclease and 3'-phosphatase (Zhu & Shuman, 2005b). The reasons why these activities might be valuable during NHEJ, and why the PE was so well conserved along with the Ligase and Pol, remained unclear. A role for the PE emerged once it was challenged with substrates analogous to annealed breaks, rather than DNA termini. The PE was also shown to be capable of resecting RNA synthesised by the Pol during strand displacement. Further experiments, which included the NHEJ ligase, showed that the PE could regulate NHEJ Pol activity of gap filling. The results suggested the PE could optimise gap filling to allow efficient ligation and that, effectively, the PE was capable of repairing potentially genotoxic strand displaced intermediates. Both the strand displacement ability of the NHEJ Pol and the displacement repair of the PE were also seen during ligation of discontinuous DNA ends, suggesting that these activities also occur during NHEJ *in vivo*. These data have expanded the abilities of an already multifaceted NHEJ polymerase, and finally implicated why the PE is required during

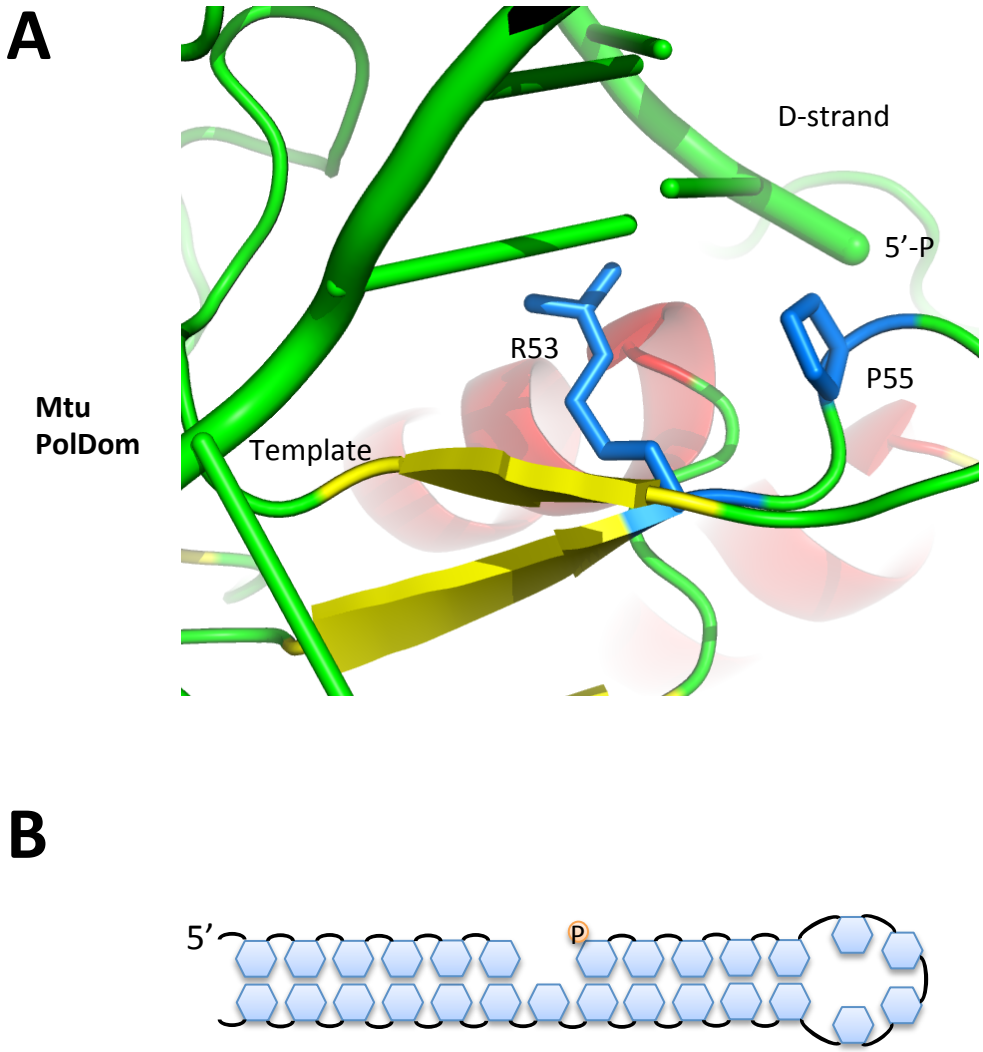


Figure 6.2 Potential residues involved in NHEJ strand displacement, and a schematic of a future strand displacement assay substrate.
(**A**) A view of two residues which have been identified as potentially encouraging disruption of base pairing. R53 and P55 in Mtu PolDom sit inbetween the phosphate binding pocket, and the twin phenylalanine residues which splay the DNA template strand. The figure was made using the Mtu PolDom structure from Brissett et al., 2011 PDB 3PKY. (**B**) A schematic of a potential strand displacement substrate that could demonstrate if the NHEJ Pol can unravel hairpins or snap-back DNA ends.

NHEJ. At present there are no known eukaryotic equivalents of the NHEJ PE, although PNK is able to remove 3'-phosphates that block extension or ligation in the same way as the PE (Karimi-Busheri et al., 1998; Bernstein et al., 2005).

6.2.3 NHEJ Ligase

The NHEJ Ligase shares the RNA preferences of the other members of its operon, and preferentially seals nicks with a 3'-monoribonucleoside, shown here and in bacteria (Zhu & Shuman, 2008). The necessity for the 3'-ribo is unknown, both structurally and in terms of what advantage it offers for NHEJ. It may be that a nick with a ribonucleotide on one side is a signal from the processing enzymes that the repair is complete, and ready for finalising. This could potentially reduce erroneous ligations of ends prior to Pol and PE interaction. Studies from this thesis also show that the NHEJ Lig has a limited mismatch sealing ability, and is able to ligate nicks that have a single base mismatch on the 5' side. It is notable that the 3'-ribo side of the nick cannot be ligated if it is not base-paired. This reflects the rigid requirement for the ribonucleoside for nick sealing, and suggests that the Ligase has evolved some key structural elements which reconigise the 2'-OH of the ribose, and the position of the base. The Ku dimer is shown to stimulate the mismatch ligation ability of LigD, although only if the mismatch is on the 5' side.

6.3 Future questions about AP-NHEJ

6.3.1 Other potential components of AP-NHEJ

The further expansion of AP-NHEJ studies should also concern the identification of other proteins that play supporting roles in NHEJ. The most obvious absent partners are a 5'-polynucleotide kinase, exonucleases and a structural protein akin to XRCC4 or XLF. A 5'-polynucleotide kinase can repair 5'-OH ends without the need for 5'-exonucleolytic resection, or endonucleolytic removal of the end without the necessary phosphate for ligation. A 5'-OH or a 3'-P that goes unprocessed could significantly slow repair, and would likely require nucleolytic resection incurring loss of genetic material. A single enzyme PNK, provides both of these functions in mammals (Karimi-Busheri et al., 1998; Bernstein et al., 2005). In AP-NHEJ only the 3'-phosphatase activity is associated with the canonical repair enzymes. It is most likely that the 5'-polynucleotide kinase has roles in other DNA repair pathways, and is probably found elsewhere in the genome than NHEJ operons. Identification of which molecule is responsible for this important task still remains a significant target.

Studies into exonucleases that affect the outcomes of AP-NHEJ would be highly interesting. As yet, it is unclear if either any 3' or 5' exonucleases play a significant role in NHEJ. Plasmid repair assays in mycobacteria did reveal large deletions in some cases of repair, but it is unknown if these are attributable to NHEJ (Aniukwu et al., 2008). Given that AP-NHEJ has developed a system that appears to protect the remaining original DNA sequence around a break, it would be confusing if a 3' or 5' exonuclease was common component of the system and could resect lengthy sequences from the DNA ends. Most bacteria and archaea do have the 5'-3' flap structure-specific endonuclease 1 (FEN-1) enzyme (present as a discrete enzyme in *M. paludicola*, but not in *M. tuberculosis*; Appendix C), which is shown to play a role in excising Okazaki fragments and long patch base-excision repair (BER) (Friedrich-Heineken & Hübscher, 2004; Balakrishnan & Bambara, 2013). Whilst FEN-1 could potentially help resolve some of the strand displacement problems caused by the NHEJ Pol, it is unknown if it plays a specific role in AP-NHEJ. Studies in yeast have suggested that FEN-1 does play a role in processing NHEJ involving 5'-flap intermediates (Wu et al., 1999). The Artemis enzyme is a central exonuclease/endonuclease in mammalian NHEJ repair and V(D)J recombination (Ma et al., 2002; Ma et al., 2005; Goodarzi et al., 2006). One of the main roles of Artemis is to remove loops and hairpins at DNA ends, and currently there is no known equivalent for this function in AP-NHEJ. The NHEJ Pol strand displacement ability might be able to perform a similar function, but this activity requires stringent testing before this can be asserted.

There is currently no research on any potential structural proteins that could stimulate the repair of DSBs by AP-NHEJ. In yeast and mammals, XRCC4 and XLF are thought to form nucleoprotein filaments which stabilise DNA ends and recruit LigIV, forming the LXX complex (Ahnesorg et al., 2006; Ropars et al., 2011). It would be interesting to discover if there are proteins that perform a similar task in AP-NHEJ, and if they would increase the efficiency of *in vitro* discontinuous-end ligation. It seems likely that DNA packaging factors, whether nucleoprotein filaments, or histone-like proteins, would stabilise AP-NHEJ reactions. Very little is known in this area in bacteria or archaea, in terms of the regulating the access to damaged DNA for repair.

6.3.2 Continued structural studies in AP-NHEJ

Apo structures of the archaeal NHEJ Pol and PE were solved, and at the outset of this research there were no published structures of any NHEJ PE, archaeal or bacterial. Both structures serve

to demonstrate that the archaeal NHEJ components are remarkably similar to those of their bacterial counterparts, and that any organism which possesses conserved LigD, NHEJ Pol, PE and Ku repair genes probably repairs DSBs in the same way as the AP-NHEJ model detailed in this thesis.

One significant target of the research described in this thesis was to gain new structural information on one or more of the enzymes interacting with a DNA substrate. The major aims were to solve structures of Ku and Pol bound to one substrate, and of the Ligase and PE bound to their individual preferred substrates. For the latter two, it was hoped that the structures could reveal which specific residues or motifs of the enzymes that generate the necessity for RNA interactions. Although none of these structures were solved, good progress has been made and the foundations for future work have been set. Hopefully continued screening with the current complexes will generate some new crystals, and if not small alterations could be made in terms of metal ion choice, or soaking vanadate and a ribonucleoside into existing crystals.

6.3.3 Future biochemical work on the canonical AP-NHEJ components

Future biochemical studies on AP-NHEJ could continue from this thesis, making use of the *in vitro* fluorescently labelled DNA end joining system. This could be used to expand our understanding of the NHEJ Pol's lesion bypass abilities, and whether the strand displacement could be used to unravel hairpins or loops. Further plasmid repair assays would be useful, perhaps with a comparison of the canonical AP-NHEJ complex *in vitro* and whole cells *in vivo*, which might indicate missing components for the *in vitro* system. It would also be useful to compare relative repair efficiencies and fidelities.

Future structural studies could allow for experimentation with the specificities of the enzyme. For example if the potential 2'-OH binding elements in the PE were identified, it might be possible to substitute the residue(s) and create a gain-of-function mutant which could cleave 3'-DNA. In the same manner, it might be possible to remove the strand displacement ability of the NHEJ Pol with alanine substitutions of key residues. These experiments could confirm predicted functions of structural elements.

6.4 Conclusion

This thesis presents the first *in vitro* characterisation of an archaeal NHEJ system, and informs on the work of the established bacterial pathway. The role of the PE has finally been clarified, and additional activities have been discovered for the NHEJ Pol. This research has allowed for the proposition of a model of AP-NHEJ, which encompasses the repair of a vast array of distinctive double strand breaks. The research in AP-NHEJ has begun to offer explanations of how specific configurations of broken ends might be repaired, describing in molecular detail the mechanics of NHEJ that have broad implications for double strand break repair.

References

Ahel I, Rass U, El-Khamisy SF, Katyal S, Clements PM, McKinnon PJ, Caldecott KW and West SC (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*. 443 (7112), 713–716.

Ahnesorg P, Smith P and Jackson SP (2006) XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell*. 124 (2), 301–313.

Akey D, Martins A, Aniukwu J, Glickman MS, Shuman S and Berger JM (2006) Crystal structure and nonhomologous end-joining function of the ligase component of Mycobacterium DNA ligase D. *The Journal of biological chemistry*. 281 (19), 13412–13423.

Aniukwu J, Glickman MS and Shuman S (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes & development*. 22 (4), 512–527.

Aravind L and Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome research*. 11 (8), 1365–1374.

Aravind L, Leipe DD and Koonin EV (1998) Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*. 26 (18), 4205–4213.

Bakkenist CJ and Kastan MB (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*. 421 (6922), 499–506.

Balakrishnan L and Bambara RA (2013) Flap endonuclease 1. *Annual review of biochemistry*. 82, 119–138.

Batra VK, Shock DD, Beard WA, McKenna CE and Wilson SH (2012) Binary complex crystal structure of DNA polymerase β reveals multiple conformations of the templating 8-oxoguanine lesion. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (1), 113–118.

Battye TGG, Kontogiannis L, Johnson O, Powell HR and Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography*. 67 (Pt 4), 271–281.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW (2010) GenBank. *Nucleic Acids Research*. 38 (Database issue), D46–51.

Bernstein NK, Williams RS, Rakovszky ML, Cui D, Green R, Karimi-Busheri F, Mani RS, Galicia S, Koch CA, Cass CE, Durocher D, Weinfeld M and Glover JNM (2005) The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Molecular cell*. 17 (5), 657–670.

Bianchi J, Rudd S, Jozwiakowski SK, Bailey LJ, Soura V, Taylor E, Stevanovic I, Green AJ, Stracker T, Lindsay HD and Doherty AJ (2013) Eukaryotic PrimPol bypasses UV photoproducts during chromosomal DNA replication. *Molecular cell.* **Manuscript submitted**.

Bocquier AA, Liu L, Cann IK, Komori K, Kohda D and Ishino Y (2001) Archaeal primase: bridging the gap between RNA and DNA polymerases. *Current biology : CB*. 11 (6), 452–456.

Boulton SJ and Jackson SP (1998) Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *The EMBO journal*. 17 (6), 1819–1828.

Brissett NC, Martin MJ, Bartlett EJ, Bianchi J, Blanco L and Doherty AJ (2013) Molecular basis for a DNA double-strand break annealing and primer extension by a NHEJ DNA polymerase. *Cell Reports*. **Under review**.

Brissett NC, Martin MJ, Pitcher RS, Bianchi J, Juarez R, Green AJ, Fox GC, Blanco L and Doherty AJ (2011) Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. *Molecular cell*. 41 (2), 221–231.

Brissett NC, Pitcher RS, Juarez R, Picher AJ, Green AJ, Dafforn TR, Fox GC, Blanco L and Doherty AJ (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science*. 318 (5849), 456–459.

Bunting SF, Callén E, Wong N, Chen H-T, Polato F, Gunn A, Bothmer A, Feldhahn N, Fernandez-Capetillo O, Cao L, Xu X, Deng C-X, Finkel T, Nussenzweig M, Stark JM and Nussenzweig A (2010) 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell*. 141 (2), 243–254.

Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, Lawrence CW, Nakanishi M, Ohmori H, Prakash L, Prakash S, Reynaud CA, Sugino A, Todo T, Wang Z, Weill JC and Woodgate R (2001) Eukaryotic DNA polymerases: proposal for a revised nomenclature. *The Journal of biological chemistry*. 276 (47), 43487–43490.

Busch CR and DiRuggiero J (2010) MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon Halobacterium salinarum NRC-1. *PLoS ONE*. 5 (2), e9045.

Caldecott KW (2008) Single-strand break repair and genetic disease. *Nature reviews. Genetics*. 9 (8), 619–631.

Chapman JR, Taylor MRG and Boulton SJ (2012) Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*. 47 (4), 497–510.

Chappell C, Hanakahi LA, Karimi-Busheri F, Weinfeld M and West SC (2002) Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *The EMBO journal*. 21 (11), 2827–2832.

Chary P, Beard WA, Wilson SH and Lloyd RS (2012) DNA polymerase β gap-filling translesion DNA synthesis. *Chemical research in toxicology*. 25 (12), 2744–2754.

Chayot R, Montagne B, Mazel D and Ricchetti M (2010) An end-joining repair mechanism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (5), 2141–2146.

Cheng C and Shuman S (1997) Characterization of an ATP-dependent DNA ligase encoded by Haemophilus influenzae. *Nucleic Acids Research*. 25 (7), 1369–1374.

Cheng KC, Cahill DS, Kasai H, Nishimura S and Loeb LA (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *The Journal of biological chemistry*. 267 (1), 166–172.

Chung CT, Niemela SL and Miller RH (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America*. 86 (7), 2172–2175.

Ciccarelli FD, Doerks T, Mering von C, Creevey CJ, Snel B and Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science*. 311 (5765), 1283–1287.

Clements PM, Breslin C, Deeks ED, Byrd PJ, Ju L, Bieganowski P, Brenner C, Moreira M-C, Taylor AMR and Caldecott KW (2004) The ataxia-oculomotor apraxia 1 gene product has a role distinct from ATM and interacts with the DNA strand break repair proteins XRCC1 and XRCC4. *DNA Repair*. 3 (11), 1493–1502.

Cobb JA and Lees-Miller SP (2010) DNA double strand breaks: not all foci are created equal. *Cell Cycle*. 9 (3), 442–443.

Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, Sixma TK, Lamzin VS, Murshudov GN and Perrakis A (2008) ARP/wARP and molecular replacement: the next generation. *Acta crystallographica. Section D, Biological crystallography*. 64 (Pt 1), 49–60.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*. 50 (Pt 5), 760–763.

Cooke MS, Evans MD, Dizdaroglu M and Lunec J (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal*. 17 (10), 1195–1214.

Coquelle N, Havali-Shahriari Z, Bernstein N, Green R and Glover JNM (2011) Structural basis for the phosphatase activity of polynucleotide kinase/phosphatase on single- and double-stranded DNA substrates. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (52), 21022–21027.

Critchlow SE, Bowater RP and Jackson SP (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current biology : CB*. 7 (8), 588–598.

Crook EM, Mathias AP and Rabin BR (1960) Spectrophotometric assay of bovine pancreatic ribonuclease by the use of cytidine 2‘:3’-phosphate. *The Biochemical journal*. 74, 234–238.

Daley JM, Palmbos PL, Wu D and Wilson TE (2005) Nonhomologous end joining in yeast. *Annual review of genetics*. 39, 431–451.

Davies DR and Hol WGJ (2004) The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS LETTERS*. 577 (3), 315–321.

Davies DR, Interthal H, Champoux JJ and Hol WGJ (2003) Crystal structure of a transition state mimic for Tdp1 assembled from vanadate, DNA, and a topoisomerase I-derived peptide. *Chemistry & biology*. 10 (2), 139–147.

Davies DR, Interthal H, Champoux JJ and Hol WGJ (2002) The crystal structure of human tyrosyl-DNA phosphodiesterase, Tdp1. *Structure*. 10 (2), 237–248.

Davis BJ, Havener JM and Ramsden DA (2008) End-bridging is required for pol mu to efficiently promote repair of noncomplementary ends by nonhomologous end joining. *Nucleic Acids Research*. 36 (9), 3085–3094.

De Bont R and Larebeke N (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*. 19 (3), 169–185.

de Vega M (2013) The Minimal Bacillus subtilis Nonhomologous End Joining Repair Machinery. *PLoS ONE*. 8 (5), e64232.

Della M, Palmbos PL, Tseng H-M, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE and Doherty AJ (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science (New York, N.Y.)*. 306 (5696), 683–685.

Difilippantonio MJ, Zhu J, Chen HT, Meffre E, Nussenzweig MC, Max EE, Ried T and Nussenzweig A (2000) DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature*. 404 (6777), 510–514.

Dobbs TA, Tainer JA and Lees-Miller SP (2010) A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair*. 9 (12), 1307–1314.

Doherty AJ and Jackson SP (2001) DNA repair: how Ku makes ends meet. *Current biology : CB*. 11 (22), R920–4.

Doherty AJ and Suh SW (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research*. 28 (21), 4051–4058.

Doherty AJ, Jackson SP and Weller GR (2001) Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS LETTERS*. 500 (3), 186–188.

Drlica K and Rouviere-Yaniv J (1987) Histonelike proteins of bacteria. *Microbiological reviews*. 51 (3), 301–319.

Dutta K, Natarajan A, Nair PA, Shuman S and Ghose R (2011) Sequence-specific 1H, 13C and 15N assignments of the phosphoesterase (PE) domain of Pseudomonas aeruginosa DNA ligase D (LigD). *Biomolecular NMR assignments*. 5 (2), 151–155.

Dynan WS and Yoo S (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Research*. 26 (7), 1551–1559.

Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*. 60 (Pt 12 Pt 1), 2126–2132.

Evans P (2006) Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*. 62 (Pt 1), 72–82.

Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica. Section D, Biological crystallography*. 67 (Pt 4), 282–292.

Fedoroff OY, Salazar M and Reid BR (1993) Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *Journal of molecular biology*. 233 (3), 509–523.

Feng L and Chen J (2012) The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. *Nature Structural & Molecular Biology*, 19 (2) 201-206.

Frank EG and Woodgate R (2007) Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. *The Journal of biological chemistry*. 282 (34), 24689–24696.

Franklin RE and Gosling RG (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*. 172 (4369), 156–157.

Frick DN and Richardson CC (2001) DNA primases. *Annual review of biochemistry*. 70, 39–80.

Friedrich-Heineken E and Hübscher U (2004) The Fen1 extrahelical 3'-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Research*. 32 (8), 2520–2528.

Fuss JO and Cooper PK (2006) DNA repair: dynamic defenders against cancer and aging. *PLoS biology*. 4 (6), e203.

Garces F, Pearl LH and Oliver AW (2011) The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Molecular cell*. 44 (3), 385–396.

Garcia V, Phelps SEL, Gray S and Neale MJ (2011) Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*. 479 (7372), 241–244.

Garg P and Burgers PMJ (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical reviews in biochemistry and molecular biology*. 40 (2), 115–128.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. 31 (13), 3784–3788.

Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annual review of biochemistry*. 71, 101–132.

Goedecke W, Pfeiffer P and Vielmetter W (1994) Nonhomologous DNA end joining in Schizosaccharomyces pombe efficiently eliminates DNA double-strand-breaks from haploid sequences. *Nucleic Acids Research*. 22 (11), 2094–2101.

Gong C, Bongiorno P, Martins A, Stephanou NC, Zhu H, Shuman S and Glickman MS (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nature Structural & Molecular Biology*. 12 (4), 304–312.

Gong C, Martins A, Bongiorno P, Glickman M and Shuman S (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *The Journal of biological chemistry*. 279 (20), 20594–20606.

Goodarzi AA and Jeggo PA (2012) The heterochromatic barrier to DNA double strand break repair: how to get the entry visa. *International journal of molecular sciences*. 13 (9), 11844–11860.

Goodarzi AA, Yu Y, Riballo E, Douglas P, Walker SA, Ye R, Härer C, Marchetti C, Morrice N, Jeggo PA and Lees-Miller SP (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *The EMBO journal*. 25 (16), 3880–3889.

Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M and Lieber MR (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature*. 388 (6641), 492–495.

Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF and Lieber MR (2007a) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *The EMBO journal*. 26 (4), 1010–1023.

Gu J, Lu H, Tsai AG, Schwarz K and Lieber MR (2007b) Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Research*. 35 (17), 5755–5762.

Hammel M, Yu Y, Fang S, Lees-Miller SP and Tainer JA (2010a) XLF regulates filament architecture of the XRCC4·ligase IV complex. *Structure*. 18 (11), 1431–1442.

Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S, Abolfath RM, Chen DJ, Lees-Miller SP and Tainer JA (2010b) Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *The Journal of biological chemistry*. 285 (2), 1414–1423.

Hentges P, Ahnesorg P, Pitcher RS, Bruce CK, Kysela B, Green AJ, Bianchi J, Wilson TE, Jackson SP and Doherty AJ (2006) Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *The Journal of biological chemistry*. 281 (49), 37517–37526.

Hsieh P and Yamane K (2008) DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mechanisms of ageing and development*. 129 (7-8), 391–407.

Hsu GW, Huang X, Luneva NP, Geacintov NE and Beese LS (2005) Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication. *The Journal of biological chemistry*. 280 (5), 3764–3770.

Hu J, Guo L, Wu K, Liu B, Lang S and Huang L (2012) Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon Sulfolobus solfataricus. *Nucleic Acids Research*. 40 (8), 3470–3483.

Iles N, Rulten S, El-Khamisy SF and Caldecott KW (2007) APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Molecular and cellular biology*. 27 (10), 3793–3803.

Imoto S, Bransfield LA, Croteau DL, Van Houten B and Greenberg MM (2008) DNA tandem lesion repair by strand displacement synthesis and nucleotide excision repair. *Biochemistry*. 47 (14), 4306–4316.

Ito N, Nureki O, Shirouzu M, Yokoyama S and Hanaoka F (2003) Crystal structure of the Pyrococcus horikoshii DNA primase-UTP complex: implications for the mechanism of primer synthesis. *Genes to cells : devoted to molecular & cellular mechanisms*. 8 (12), 913–923.

Iyer LM, Koonin EV, Leipe DD and Aravind L (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Research*. 33 (12), 3875–3896.

Iyer RR, Pluciennik A, Burdett V and Modrich PL (2006) DNA mismatch repair: functions and mechanisms. *Chemical reviews*. 106 (2), 302–323.

Jackson SP (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 23 (5), 687–696.

Jackson SP and Bartek J (2009) The DNA-damage response in human biology and disease. *Nature*. 461 (7267), 1071–1078.

Jahnen-Dechent W and Ketteler M (2012) Magnesium basics. *Clinical Kidney Journal*. 5 (Suppl 1), i3–i14.

Joyce CM (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proceedings of the National Academy of Sciences of the United States of America*. 94 (5), 1619–1622.

Jozwiakowski SK and Connolly BA (2011) A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *Chembiochem : a European journal of chemical biology*. 12 (1), 35–37.

Kanno S-I, Kuzuoka H, Sasao S, Hong Z, Lan L, Nakajima S and Yasui A (2007) A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. *The EMBO journal*. 26 (8), 2094–2103.

Karimi-Busheri F, Lee J, Tomkinson AE and Weinfeld M (1998) Repair of DNA strand gaps and nicks containing 3‘-phosphate and 5’-hydroxyl termini by purified mammalian enzymes. *Nucleic Acids Research*. 26 (19), 4395–4400.

Kastan MB and Lim DS (2000) The many substrates and functions of ATM. *Nature reviews. Molecular cell biology*. 1 (3), 179–186.

Keck JL, Roche DD, Lynch AS and Berger JM (2000) Structure of the RNA polymerase domain of E. coli primase. *Science*. 287 (5462), 2482–2486.

Kelman Z and White MF (2005) Archaeal DNA replication and repair. *Current Opinion in Microbiology*. 8 (6), 669–676.

Keppetipola N and Shuman S (2007) Characterization of the 2‘,3’ cyclic phosphodiesterase activities of Clostridium thermocellum polynucleotide kinase-phosphatase and bacteriophage lambda phosphatase. *Nucleic Acids Research*. 35 (22), 7721–7732.

Khanna KK and Jackson SP (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nature genetics*. 27 (3), 247–254.

Kisker C, Kuper J and Van Houten B (2013) Prokaryotic nucleotide excision repair. *Cold Spring Harbor perspectives in biology*. 5 (3), a012591.

Komori K, Miyata T, DiRuggiero J, Holley-Shanks R, Hayashi I, Cann IK, Mayanagi K, Shinagawa H and Ishino Y (2000) Both RadA and RadB are involved in homologous recombination in

Pyrococcus furiosus. *The Journal of biological chemistry*. 275 (43), 33782–33790.

Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD and Rehrauer WM (1994) Biochemistry of homologous recombination in Escherichia coli. *Microbiological reviews*. 58 (3), 401–465.

Kramer KM, Brock JA, Bloom K, Moore JK and Haber JE (1994) Two different types of double-strand breaks in Saccharomyces cerevisiae are repaired by similar RAD52-independent, nonhomologous recombination events. *Molecular and cellular biology*. 14 (2), 1293–1301.

Krishna SS and Aravind L (2010) The bridge-region of the Ku superfamily is an atypical zinc ribbon domain. *Journal of structural biology*. 172 (3), 294–299.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227 (5259), 680–685.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. 23 (21), 2947–2948.

Larrea AA, Lujan SA and Kunkel TA (2010) SnapShot: DNA mismatch repair. *Cell*. 141 (4), 730.e1.

Lavin MF and Kozlov S (2007) ATM activation and DNA damage response. *Cell Cycle*. 6 (8), 931–942.

Lee JH, Hwang GS and Choi BS (1999) Solution structure of a DNA decamer duplex containing the stable 3‘ T.G base pair of the pyrimidine(6-4)pyrimidone photoproduct [(6-4) adduct]: implications for the highly specific 3’ T --> C transition of the (6-4) adduct. *Proceedings of the National Academy of Sciences of the United States of America*. 96 (12), 6632–6636.

Lee JW, Blanco L, Zhou T, Garcia-Diaz M, Bebenek K, Kunkel TA, Wang Z and Povirk LF (2004) Implication of DNA polymerase lambda in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts. *The Journal of biological chemistry*. 279 (1), 805–811.

Leonard CJ, Aravind L and Koonin EV (1998) Novel families of putative protein kinases in bacteria and archaea: evolution of the ‘eukaryotic’ protein kinase superfamily. *Genome research*. 8 (10), 1038–1047.

Letunic I and Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*. 39 (Web Server issue), W475–8.

Li X and Heyer W-D (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell research*. 18 (1), 99–113.

Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry*. 79, 181–211.

Lieber MR and Wilson TE (2010) SnapShot: Nonhomologous DNA end joining (NHEJ). *Cell*. 142 (3), 496–496.e1.

Lieber MR, Yu K and Raghavan SC (2006) Roles of nonhomologous DNA end joining, V(D)J

recombination, and class switch recombination in chromosomal translocations. *DNA Repair*. 5 (9-10), 1234–1245.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature*. 362 (6422), 709–715.

Liu Y, Prasad R, Beard WA, Kedar PS, Hou EW, Shock DD and Wilson SH (2007) Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. *The Journal of biological chemistry*. 282 (18), 13532–13541.

Loeb LA, Springgate CF and Battula N (1974) Errors in DNA replication as a basis of malignant changes. *Cancer research*. 34 (9), 2311–2321.

Lu D, Silhan J, MacDonald JT, Carpenter EP, Jensen K, Tang CM, Baldwin GS and Freemont PS (2012) Structural basis for the recognition and cleavage of abasic DNA in Neisseria meningitidis. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (42), 16852–16857.

Lukas J, Lukas C and Bartek J (2011) More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nature cell biology*. 13 (10), 1161–1169.

Lusetti SL and Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annual review of biochemistry*. 71, 71–100.

Ma Y, Pannicke U, Schwarz K and Lieber MR (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*. 108 (6), 781–794.

Ma Y, Schwarz K and Lieber MR (2005) The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair*. 4 (7), 845–851.

Mahaney BL, Meek K and Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *The Biochemical journal*. 417 (3), 639–650.

Mansour WY, Rhein T and Dahm-Daphi J (2010) The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. *Nucleic Acids Research*. 38 (18), 6065–6077.

Martin MJ, Garcia-Ortiz MV, Esteban V and Blanco L (2013) Ribonucleotides and manganese ions improve non-homologous end joining by human Polμ. *Nucleic Acids Research*. 41 (4), 2428–2436.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography*. 40 (Pt 4), 658–674.

McCoy AJ, Grosse-Kunstleve RW, Storoni LC and Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta crystallographica. Section D, Biological crystallography*. 61 (Pt 4), 458–464.

Moeller R, Stackebrandt E, Reitz G, Berger T, Rettberg P, Doherty AJ, Horneck G and Nicholson

WL (2007) Role of DNA repair by nonhomologous-end joining in Bacillus subtilis spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *Journal of bacteriology*. 189 (8), 3306–3311.

Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH and Pedersen LC (2007a) The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair*. 6 (12), 1709–1725.

Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA and Pedersen LC (2007b) Structural insight into the substrate specificity of DNA Polymerase mu. *Nature structural & molecular biology*. 14 (1), 45–53.

Moore JK and Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and cellular biology*. 16 (5), 2164–2173.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography*. 67 (Pt 4), 355–367.

Murshudov GN, Vagin AA and Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography*. 53 (Pt 3), 240–255.

Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *The EMBO journal*. 12 (3), 861–867.

Nair PA, Smith P and Shuman S (2010) Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (29), 12822–12827.

Nandakumar J, Nair PA and Shuman S (2007) Last stop on the road to repair: structure of E. coli DNA ligase bound to nicked DNA-adenylate. *Molecular cell*. 26 (2), 257–271.

Natarajan A, Dutta K, Temel DB, Nair PA, Shuman S and Ghose R (2012) Solution structure and DNA-binding properties of the phosphoesterase domain of DNA ligase D. *Nucleic Acids Research*. 40 (5), 2076–2088.

Nick McElhinny SA and Ramsden DA (2003) Polymerase mu is a DNA-directed DNA/RNA polymerase. *Molecular and cellular biology*. 23 (7), 2309–2315.

Nick McElhinny SA, Havener JM, Garcia-Diaz M, Juarez R, Bebenek K, Kee BL, Blanco L, Kunkel TA and Ramsden DA (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Molecular cell*. 19 (3), 357–366.

Nick McElhinny SA, Snowden CM, McCarville J and Ramsden DA (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. *Molecular and cellular biology*. 20 (9), 2996–3003.

Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PMJ, Johansson E, Chabes A and Kunkel TA (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (11), 4949–4954.

Pascal JM, O'Brien PJ, Tomkinson AE and Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature*. 432 (7016), 473–478.

Pitcher RS, Brissett NC and Doherty AJ (2007a) Nonhomologous end-joining in bacteria: a microbial perspective. *Annual review of microbiology*. 61, 259–282.

Pitcher RS, Brissett NC, Picher AJ, Andrade P, Juarez R, Thompson D, Fox GC, Blanco L and Doherty AJ (2007b) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *Journal of molecular biology*. 366 (2), 391–405.

Pitcher RS, Green AJ, Brzostek A, Korycka-Machala M, Dziadek J and Doherty AJ (2007c) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair*. 6 (9), 1271–1276.

Pitcher RS, Tonkin LM, Green AJ and Doherty AJ (2005) Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. *Journal of molecular biology*. 351 (3), 531–544.

Postow L (2011) Destroying the ring: Freeing DNA from Ku with ubiquitin. *FEBS LETTERS*, 1–7.

Potterton E, McNicholas S, Krissinel E, Cowtan K and Noble M (2002) The CCP4 molecular-graphics project. *Acta crystallographica. Section D, Biological crystallography*. 58 (Pt 11), 1955–1957.

Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A and Noble M (2004) Developments in the CCP4 molecular-graphics project. *Acta crystallographica. Section D, Biological crystallography*. 60 (Pt 12 Pt 1), 2288–2294.

Rastogi RP, Richa, Kumar A, Tyagi MB and Sinha RP (2010) Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of nucleic acids*. 2010, 592980.

Reijns MAM, Rabe B, Rigby RE, Mill P, Astell KR, Lettice LA, Boyle S, Leitch A, Keighren M, Kilanowski F, Devenney PS, Sexton D, Grimes G, Holt IJ, Hill RE, Taylor MS, Lawson KA, Dorin JR and Jackson AP (2012) Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*. 149 (5), 1008–1022.

Riballo E, Woodbine L, Stiff T, Walker SA, Goodarzi AA and Jeggo PA (2008) XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Research*. 37 (2), 482–492.

Richardson TT, Wu X, Keith BJ, Heslop P, Jones AC and Connolly BA (2013) Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. *Nucleic Acids Research*. 41 (4), 2466–2478.

Roberts JA, Bell SD and White MF (2003) An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Molecular microbiology*. 48 (2), 361–371.

Roberts SA, Strande N, Burkhalter MD, Strom C, Havener JM, Hasty P and Ramsden DA (2010) Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature*. 464 (7292), 1214–1217.

Ropars V, Drevet P, Legrand P, Baconnais S, Amram J, Faure G, Márquez JA, Piétrement O, Guerois R, Callebaut I, Le Cam E, Revy P, de Villartay J-P and Charbonnier J-B (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex

involved in nonhomologous DNA end-joining. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (31), 12663–12668.

Roth DB and Wilson JH (1986) Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Molecular and cellular biology*. 6 (12), 4295–4304.

Rouillon C and White MF (2010) The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. *The Journal of biological chemistry*. 285 (14), 11013–11022.

Rupert PB, Massey AP, Sigurdsson ST and Ferré-D'Amaré AR (2002) Transition state stabilization by a catalytic RNA. *Science*. 298 (5597), 1421–1424.

Rydberg B and Game J (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of the National Academy of Sciences of the United States of America*. 99 (26), 16654–16659.

Sakai S, Imachi H, Hanada S, Ohashi A, Harada H and Kamagata Y (2008) Methanocella paludicola gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellales ord. nov. *International journal of systematic and evolutionary microbiology*. 58 (Pt 4), 929–936.

Sakai S, Takaki Y, Shimamura S, Sekine M, Tajima T, Kosugi H, Ichikawa N, Tasumi E, Hiraki AT, Shimizu A, Kato Y, Nishiko R, Mori K, Fujita N, Imachi H and Takai K (2011) Genome sequence of a mesophilic hydrogenotrophic methanogen Methanocella paludicola, the first cultivated representative of the order Methanocellales. *PLoS ONE*. 6 (7), e22898.

Sartori AA and Jiricny J (2003) Enzymology of base excision repair in the hyperthermophilic archaeon Pyrobaculum aerophilum. *The Journal of biological chemistry*. 278 (27), 24563–24576.

Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J, Baer R, Lukas J and Jackson SP (2007) Human CtIP promotes DNA end resection. *Nature*. 450 (7169), 509–514.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E and Ye J (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. 38 (Database issue), D5–16.

Schrodinger LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

Shrivastav M, De Haro LP and Nickoloff JA (2008) Regulation of DNA double-strand break repair pathway choice. *Cell research*. 18 (1), 134–147.

Shuman S (1995) Vaccinia virus DNA ligase: specificity, fidelity, and inhibition. *Biochemistry*. 34 (49), 16138–16147.

Shuman S and Lima CD (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Current opinion in structural biology*. 14 (6), 757–764.

Shuman S and Schwer B (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Molecular microbiology*. 17 (3), 405–410.

Shuman S, Liu Y and Schwer B (1994) Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (25), 12046–12050.

Smith P and Shuman S (2012) The adenyltransferase domain of bacterial Pnkp defines a unique RNA ligase family. *Proceedings of the National Academy of Sciences of the United States of America*. *109* (7), 2296–2301.

Smith P, Nair PA, Das U, Zhu H and Shuman S (2011) Structures and activities of archaeal members of the LigD 3'-phosphoesterase DNA repair enzyme superfamily. *Nucleic Acids Research*. 39 (8), 3310–3320.

Sriskanda V and Shuman S (1998) Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Research*. 26 (20), 4618–4625.

Sung P and Klein H (2006) Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews. Molecular cell biology*. 7 (10), 739–750.

Suvarnapunya AE, Lagassé HAD and Stein MA (2003) The role of DNA base excision repair in the pathogenesis of Salmonella enterica serovar Typhimurium. *Molecular microbiology*. 48 (2), 549–559.

Takahara PM, Rosenzweig AC, Frederick CA and Lippard SJ (1995) Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature*. 377 (6550), 649–652.

Tang Y, Liu Z, Ding S, Lin CH, Cai Y, Rodriguez FA, Sayer JM, Jerina DM, Amin S, Broyde S and Geacintov NE (2012) Nuclear magnetic resonance solution structure of an N(2)-guanine DNA adduct derived from the potent tumorigen dibenzo[a,l]pyrene: intercalation from the minor groove with ruptured Watson-Crick base pairing. *Biochemistry*. 51 (48), 9751–9762.

Tomkinson AE, Vijayakumar S, Pascal JM and Ellenberger T (2006) DNA ligases: structure, reaction mechanism, and function. *Chemical reviews*. 106 (2), 687–699.

Truong LN, Li Y, Shi LZ, Hwang PY-H, He J, Wang H, Razavian N, Berns MW and Wu X (2013) Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. *110* (19), 7720–7725.

Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F and Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta crystallographica. Section D, Biological crystallography*. 60 (Pt 12 Pt 1), 2184–2195.

Vincze T, Posfai J and Roberts RJ (2003) NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Research*. 31 (13), 3688–3691.

Walker JR, Corpina RA and Goldberg J (2001) Structure of the Ku heterodimer bound to DNA

and its implications for double-strand break repair. *Nature*. 412 (6847), 607–614.

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25 (9), 1189–1191.

Watson JD and H CF (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 171 (4356), 737–738.

Weinfeld M, Mani RS, Abdou I, Aceytuno RD and Glover JNM (2011) Tidying up loose ends: the role of polynucleotide kinase/phosphatase in DNA strand break repair. *Trends in biochemical sciences*. 36 (5), 262–271.

Weller GR and Doherty AJ (2001) A family of DNA repair ligases in bacteria? *FEBS LETTERS*. 505 (2), 340–342.

Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F, Devine KM, Bowater RP, Jeggo PA, Jackson SP and Doherty AJ (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science*. 297 (5587), 1686–1689.

White MF (2003) Archaeal DNA repair: paradigms and puzzles. *Biochemical Society transactions*. 31 (Pt 3), 690–693.

White MF and Bell SD (2002) Holding it together: chromatin in the Archaea. *Trends in genetics : TIG*. 18 (12), 621–626.

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD and Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology*. 112, 531–552.

Wilkinson A, Day J and Bowater R (2001) Bacterial DNA ligases. *Molecular microbiology*. 40 (6), 1241–1248.

Woese CR, Kandler O and Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*. 87 (12), 4576–4579.

Wu X, Wilson TE and Lieber MR (1999) A role for FEN-1 in nonhomologous DNA end joining: the order of strand annealing and nucleolytic processing events. *Proceedings of the National Academy of Sciences of the United States of America*. 96 (4), 1303–1308.

Wuttke DS (2004) Targeting the End: The Structure of the Ku80 C-Terminal Domain. *Structure*. 12 (3), 355-356.

Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ and Livingston DM (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Molecular cell*. 22 (6), 719–729.

Yang SW, Burgin AB, Huizenga BN, Robertson CA, Yao KC and Nash HA (1996) A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. *Proceedings of the National Academy of Sciences of the United States of America*. 93 (21), 11534–11539.

Yu K, Chedin F, Hsieh C-L, Wilson TE and Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature immunology*. 4 (5), 442–451.

Zhu H and Shuman S (2005a) A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). *The Journal of biological chemistry*. 280 (1), 418–427.

Zhu H and Shuman S (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *The Journal of biological chemistry*. 283 (13), 8331–8339.

Zhu H and Shuman S (2007) Characterization of Agrobacterium tumefaciens DNA ligases C and D. *Nucleic Acids Research*. 35 (11), 3631–3645.

Zhu H and Shuman S (2010) Gap filling activities of Pseudomonas DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *The Journal of Biological Chemistry*. 285 (7), 4815–4825.

Zhu H and Shuman S (2005b) Novel 3‘-Ribonuclease and 3’-Phosphatase Activities of the Bacterial Non-homologous End-joining Protein, DNA Ligase D. *Journal of Biological Chemistry*. 280 (28), 25973–25981.

Zhu H and Shuman S (2006) Substrate specificity and structure-function analysis of the 3'-phosphoesterase component of the bacterial NHEJ protein, DNA ligase D. *The Journal of biological chemistry*. 281 (20), 13873–13881.

Zhu H, Bhattarai H, Yan H-G, Shuman S and Glickman MS (2012) Characterization of Mycobacterium smegmatis PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. *Biochemistry*. 21 (51),10147-58.

Zhu H, Nandakumar J, Aniukwu J, Wang LK, Glickman MS, Lima CD and Shuman S (2006) Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (6), 1711–1716.

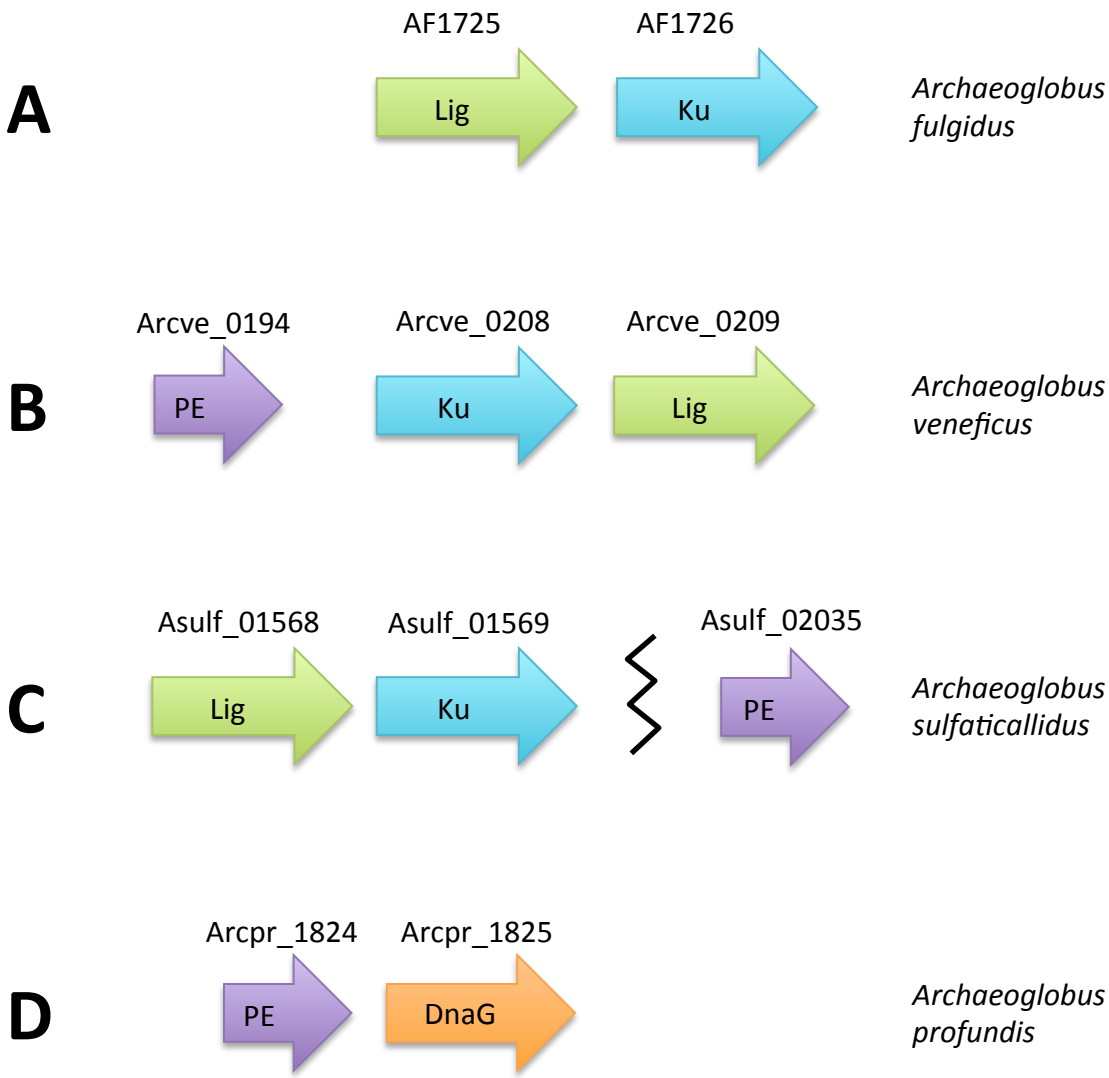
Zhu H, Wang LK and Shuman S (2005) Essential constituents of the 3'-phosphoesterase domain of bacterial DNA ligase D, a nonhomologous end-joining enzyme. *The Journal of biological chemistry*. 280 (40), 33707–33715.

Appendix

Description	Max score	Total score	Query cover	E value	Max ident	Accession
ribonuclease HII [Methanocella paludicola SANAE] >gi|502665746|ref|WP_012901703.1| ribonuclease HII [Methanocella paludicola] >gi|282157945|dbj|BAI63033.1| ribo	444	444	100%	8e-157	100%	YP_003358016.1
ribonuclease H, mammalian HI/archaeal HII subfamily [Methanocella conradii HZ254] >gb|AFC99255.1| ribonuclease H, mammalian HI/archaeal HII subfamily [Methanoce	333	333	99%	7e-113	70%	YP_005379774.1
ribonuclease H (RNase H) [Methanocella arvoryzae MRE50] >ref|WP_012036245.1| ribonuclease H (RNase H) [Methanocella arvoryzae] >sp|Q0W5R8.1|RNH2_UNCMA	300	300	99%	6e-100	65%	YP_685601.1
ribonuclease HII [Methanococcus maripaludis C6] >ref|WP_012194056.1| ribonuclease HII [Methanococcus maripaludis] >sp|A9A9T9.1|RNH2_METM6 RecName: Full=R	193	193	99%	7e-58	45%	YP_001549344.1
ribonuclease HII [Pyrococcus yayanosii CH1] >ref|WP_013904999.1| ribonuclease HII [Pyrococcus yayanosii] >gb|AEH23941.1| ribonuclease HII [Pyrococcus yayanosii C	190	190	93%	6e-57	49%	YP_004623213.1
ribonuclease HII [Methanococcus maripaludis C7] >ref|WP_011977015.1| ribonuclease HII [Methanococcus maripaludis] >sp|A6VGW0.1|RNH2_METM7 RecName: Full=I	191	191	99%	6e-57	44%	YP_001329837.1
ribonuclease HII [Methanococcus maripaludis C5] >ref|WP_011867978.1| ribonuclease HII [Methanococcus maripaludis] >sp|A4FWE5.1|RNH2_METM5 RecName: Full=F	190	190	99%	2e-56	44%	YP_001096735.1
Ribonuclease HII [Thermococcus sibiricus MM 739] >sp|C6A1I4.1|RNH2_THESM RecName: Full=Ribonuclease HII; Short=RNase HII >gb|ACS89479.1| Ribonuclease HI	187	187	93%	9e-56	48%	YP_002993828.1
ribonuclease HII [Thermococcus zilligii]	186	186	93%	2e-55	49%	WP_010478817.1
RNase HII [Thermoplasmatales archaeon SCGC AB-539-N05] >gb|EMR73812.1| RNase HII [Thermoplasmatales archaeon SCGC AB-539-N05]	185	185	99%	7e-55	44%	WP_008441849.1
ribonuclease HII [Pyrococcus furiosus DSM 3638] >ref|YP_006493152.1| ribonuclease HII [Pyrococcus furiosus COM1] >ref|WP_011012922.1| ribonuclease HII [Pyrococc	184	184	99%	1e-54	45%	NP_579510.1
ribonuclease HII [Aciduliprofundum boonei T469] >ref|WP_008084170.1| ribonuclease HII [Aciduliprofundum boonei] >gb|EDY35688.1| ribonuclease HII [Aciduliprofundu	183	183	99%	2e-54	47%	YP_003483327.1
ribonuclease HII [Aciduliprofundum boonei] >gb|EDY35782.1| ribonuclease HII [Aciduliprofundum boonei T469]	183	183	99%	3e-54	47%	WP_008084303.1
RNase HII [Methanoplanus limicola] >gb|EHQ35852.1| RNase HII [Methanoplanus limicola DSM 2279]	182	182	99%	4e-54	46%	WP_004077776.1
ribonuclease HII [Methanococcus maripaludis X1] >ref|WP_013999733.1| ribonuclease HII [Methanococcus maripaludis] >gb|AEK20394.1| ribonuclease HII [Methanococ	183	183	99%	8e-54	42%	YP_004743137.1
ribonuclease HII [Thermococcus sp. CL1] >gb|AFL96105.1| ribonuclease HII [Thermococcus sp. CL1]	182	182	93%	9e-54	48%	YP_006425899.1
ribonuclease HII [Thermococcus onnurineus NA1] >ref|WP_012571730.1| ribonuclease HII [Thermococcus onnurineus] >sp|B6YVT5.1|RNH2_THEON RecName: Full=RIt	182	182	93%	1e-53	49%	YP_002307155.1
ribonuclease HII [Archaeoglobus veneficus SNP6] >ref|WP_013683901.1| ribonuclease HII [Archaeoglobus veneficus] >gb|AEA47239.1| ribonuclease HII [Archaeoglobus	181	181	96%	1e-53	46%	YP_004341954.1
ribonuclease HII [Thermococcus barophilus MP] >ref|WP_013466946.1| ribonuclease HII [Thermococcus barophilus] >gb|ADT83648.1| ribonuclease HII [Thermococcus b	182	182	93%	1e-53	46%	YP_004070871.1
Chain A, Crystal Structure Of Tk-Rnase Hii(1-200)-C	181	181	94%	1e-53	48%	2DFE_A
ribonuclease HII [Methanococcus maripaludis S2] >ref|WP_011171318.1| ribonuclease HII [Methanococcus maripaludis] >sp|Q6LXH7.1|RNH2_METMP RecName: Full=F	182	182	99%	1e-53	42%	NP_988494.1
RNase HII [Methanosarcina barkeri str. Fusaro] >ref|WP_011307697.1| RNase HII [Methanosarcina barkeri] >sp|Q468N6.1|RNH2_METBF RecName: Full=Ribonuclease I	181	181	94%	1e-53	47%	YP_306236.1
Chain A, Crystal Structure Of Tk-Rnase Hii(1-204)-C	181	181	96%	2e-53	47%	2DFF_A
ribonuclease HII [Thermococcus gammatolerans EJ3] >gb|ACS33593.1| Ribonuclease HII (rnhB) [Thermococcus gammatolerans EJ3]	181	181	93%	2e-53	48%	YP_002959457.1
ribonuclease H, mammalian HI/archaeal HII subfamily [Methanomethylovorans hollandica DSM 15978] >gb|AGB49289.1| ribonuclease H, mammalian HI/archaeal HII sub	181	181	94%	2e-53	46%	YP_007312819.1
Ribonuclease HII [Thermococcus sp. AM4] >ref|WP_014121569.1| Ribonuclease HII [Thermococcus sp. AM4] >gb|EEB73038.1| Ribonuclease HII [Thermococcus sp. AM4	180	180	93%	8e-53	49%	YP_002581715.1
ribonuclease HII [Methanosalsum zhilinae DSM 4017] >ref|WP_013897755.1| ribonuclease HII [Methanosalsum zhilinae] >gb|AEH60316.1| ribonuclease HII [Methanosal	178	178	94%	2e-52	45%	YP_004615535.1
Chain A, Crystal Structure Of Tk-Rnase Hii(1-212)-C	178	178	93%	3e-52	48%	2DFH_A
Chain A, Crystal Structure Of Type 2 Ribonuclease H From Hyperthermophilic Archaeon, Thermococcus Kodakaraensis Kod1	177	177	93%	3e-52	48%	1IO2_A

Appendix A. RNase HII sequence homology search in archaea.

An amino acid sequence homology search using *M. paludicola* RNase HII as a query protein. The search produced results of many other archaeal RNase HII proteins. This search indicates that RNase HII is conserved in archaea as well as eukarya and bacteria. This search was performed using BLAST.



Appendix B. Operonic arrangement of NHEJ genes in Archaeoglobae.
The domain organisation of LigD, Ku and PE genes in archaeoglobae (**A**, **B**, **C** & **D**). The gene name is listed above each cartoon representation. The black jagged line between LigD and PE seen in the *Archaeoglobus sulfaticallidus* (**C**) cartoon represents an omitted genetic sequence, indicating that the LigD and PE are not operonic in this case. The PE in the Archaeoglobus veneficus (**B**) cartoon is probably not operonic with Ku and LigD, but is only a few genes distant. *Archaeoglobus profundis* (**D***)* does not seem to posses either LigD or Ku genes, yet has an NHEJ PE next to a DnaG primase. The other three Archaeoglobae all possess DnaG genes, but they are not operonic with any of the NHEJ genes.

flap structure-specific endonuclease [Methanocella paludicola SANAE]

Sequence ID: ref|YP_003357814.1| **Length:** 339 **Number of Matches:** 1

▸ See 2 more title(s)

Range 1: 1 to 339 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
367 bits(941)	8e-128	Compositional matrix adjust.	187/339(55%)	246/339(72%)	3/339(0%)

```
Query  1    MGADIGDLFEREEVELEYFSGKKIAVDAFNTLYQFISIIRQPDGTPLKDSQGRITSHLSG  60
            MG D+ DL    E  L   +GK IA+DAFNTLYQF+SIIRQ DGTPL D +G +TSHLSG
Sbjct  1    MGVDLTDLVPEHETTLVDLNGKVIAIDAFNTLYQFLSIIRQMDGTPLVDDKGEVTSHLSG  60

Query  61   ILYRVSNMVEVGIRPVFVFDGEPPEFKKAEIEERKKRRAEAEEMWIAALQAGDKDAKKYA  120
            I+YRV+N+VE GI+PVFV+DG+PP  K   I+ R++ R  A +M+ AA  AG  +A KYA
Sbjct  61   IIYRVTNLVEQGIKPVFVYDGKPPVLKAETIKARREVREAARQMYEAARAAGSAEAYKYA  120

Query  121  QAAGRVDEYIVDSAKTLLSYMGIPFVDAPSEGEAQAAYMAAKGDVEYTGSQDYDSLLFGS  180
            QA+  ++  I+  +K LL YMG+PF+ APSEGEAQAAYM  KG  ++ GSQDYDSLLFG+
Sbjct  121  QASTSINAQIIKDSKELLGYMGMPFLIAPSEGEAQAAYMVQKGAADFVGSQDYDSLLFGA  180

Query  181  PRLARNLAITGKRKLPGKNVYVDVKPEIIILESNLKRLGLTREQLIDIAILVGTDYNEGV  240
            PR+ RN+ ITG+RK+P + VYVDVKP+I+ L+  L+ L +TREQLID+ ILVGTD+N G+
Sbjct  181  PRMVRNVTITGRRKIPRRGVYVDVKPQIVELKEVLETLEVTREQLIDMGILVGTDFNPGI  240

Query  241  KGVGVKKALNYIKTYGDIFRALKALKVNIDHVEEIRNFFLNPPVTDDYRIEFREPDFEKA  300
              VG K AL  +K + D+   L  L  +I++ +EIR FFL+PP TDDY I++  P+ EK
Sbjct  241  YKVGPKTALKLVKKHPDMRAILDELGQDIENYQEIREFFLHPPTTDDYSIKWGRPEPEKI  300

Query  301  IEFLCEEHDFSRERVEKALEKL-KALKST--QATLERWF  336
              FLC+EH+FS ERV+K +E+L +A+   T  Q TL  WF
Sbjct  301  RHFLCDEHNFSPERVDKVIERLDRAVSETGKQKTLSAWF  339
```

Appendix C. A sequence homology search for FEN-1 in Methanocellae.

An amino acid sequence alignment following a BLAST search using the protein sequence of Afu FEN-1 in the subject Methanocellae. The alignment shows a significant similarity between Afu FEN-1 and putative FEN-1 in *M. paludicola*. It is likely that this protein performs similar tasks to all known FEN-1 enzymes.