DNA Repair Enzymes in

Prochlorococcus marinus strain MIT9312

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science (Research) in Molecular and Cellular Biology

at

The University of Waikato

by

Bronwyn Rachel Kirby

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2021

Abstract

Prochlorococcus marinus, a marine cyanobacterium, is one of the most abundant photosynthetic organisms in the ocean. *P. marinus* contributes towards approximately 8.5% of global ocean primary productivity. *P. marinus* is of interest because of the cell's simple and unique photosynthetic system, efficient carbon concentrating mechanism, and streamlined genome. *P. marinus* strain MIT9312 is a high-light adapted strain that thrives in the upper euphotic zone where there are minimal nutrients and high levels of UV exposure, which would typically put an organism at risk. Understanding the function of repair proteins is significant for being able to identify the DNA repair pathways used by *Prochlorococcus* to survive under increasingly more harmful conditions. My research, outlined in this thesis, aimed to characterize DNA repair enzymes in *Prochlorococcus marinus* strain MIT9312 to further our understanding of cyanobacterial growth, survival, and repair.

LigW is an ATP-dependent ligase that is unique to strain MIT9312. It was found to have low ligase activity on nicked DNA compared to other ligases, making its retention in the small *Prochlorococcus* genome unusual. It is suspected that LigW could achieve higher ligation rates by interacting with other proteins. LigW is in an operon with other proteins that are predicted to promote LigW activity or have their own DNA modifying activities. The three proteins of interest have been annotated as Pmar3, Pmar4, and Pmar5. The individual function and relationships between each protein were investigated in this thesis.

This thesis focused on recombinantly producing and biochemically characterizing three of the proteins adjacent to LigW and assessing their activities *in vitro*. Pmar3 was successfully expressed, purified, and characterized as a potential nuclease. Further mutagenesis work will confirm the proposed function of Pmar3. Attempts were made to characterize Pmar4, with further optimization of expression and purification protocols necessary. Pmar5 has been characterized as a pentameric Mg-dependent topoisomerase-like protein, with an inability to re-join DNA. It is likely that these proteins on the LigW operon work together with LigW in a currently unidentified repair pathway. Once the LigW DNA repair system is characterized, then this can be more easily identified and understood in other organisms where it may not be detected at present.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Adele Williamson. Your endless teaching, support, patience, and faith in my abilities from day one have been greatly appreciated. I have learned so much from you and am thankful for your guidance. I am so thankful to have had the opportunity to work on this project with you as both a summer and master's student, and to have you as a mentor.

Secondly, I would like to thank the staff of the Proteins and Microbes lab group. I am thankful to have had the opportunity to learn from you throughout my time in this group. Thank you for your input and challenging questions that have encouraged me to grow as a researcher. A special thank you to Judith, for always being caring and supportive on top of keeping the lab running smoothly.

Thank you to my newfound lab family. Each of you have made every day in the Proteins and Microbes lab more enjoyable. Trying to complete a master's during a global pandemic has been a challenge, but you all made it a lot more enjoyable. A special thank you to my undergraduate mentors Andrew and Emily, who have continued to mentor me during my masters.

I would like to thank the University of Waikato for supporting my research through the University of Waikato Research Masters Scholarship.

Thank you to my family for their unconditional love and support. You have always encouraged me in my studies, and it has meant more to me than I can put in words. A special thank you to my Mum and Dad for showing interest in my "bugs" and listening to my daily rants.

Lastly, I would like to thank Ian. You have been my rock and number one supporter. Thank you for always encouraging me to do my best, and congratulations on becoming an expert on *Prochlorococcus* too.

Finally, I would like to thank God for blessing me with this research opportunity.

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1

Introduction

Prochlorococcus marinus, a marine cyanobacterium, is one of the most abundant photosynthetic organisms in the ocean. *P. marinus* is thought to contribute towards approximately 8.5% of global ocean primary productivity (Laurenceau et al., 2020). Its significant contribution to global oxygen production is beneficial for helping other organisms to survive. *P. marinus* is an organism of interest because of the cell's simple and unique photosynthetic system, efficient carbon concentrating mechanism, and minimal genome. *Prochlorococcus* is a good model for naturally occurring ecological variation and evolutionary diversification (Prabha et al., 2014). *P. marinus* strain MIT9312 has adapted to thrive in the upper euphotic zone where there are minimal nutrients and high levels of UV exposure, which would typically put an organism at risk. Little research has focused on genes and proteins involved in DNA repair in *Prochlorococcus*. Understanding the function of repair proteins is significant for being able to identify the DNA repair pathways used by *Prochlorococcus*. Understanding how DNA repair enzymes among others within *Prochlorococcus* function could help us to utilize and engineer bacteria to survive under increasingly more harmful conditions, such as those caused by climate change. My research aimed to characterize DNA repair enzymes in *Prochlorococcus marinus* strain MIT9312 to assist in furthering our understanding of cyanobacterial growth, survival, and repair.

1.1 Classification of *Prochlorococcus marinus*

Prochlorococcus are commonly found in the euphotic zone between 45°N and 45°S, in the tropical and subtropical oligotrophic ocean at surface level and down to depths of 150-200m (Biller et al., 2015; Flombaum et al., 2013; Partensky et al., 1999). They have adapted to be small oligotrophs with minimal energy requirements, to aid their survival in these environments which lack macronutrients typically required for photosynthesis and cell growth. *Prochlorococcus* is the smallest known photosynthetic organism, with a diameter of 0.45-0.7μm (Partensky et al., 1999; Yan et al., 2020) and an estimated mean global abundance of $3x10^{27}$ cells (Flombaum et al., 2013; Partensky et al., 1999; Schattenhofer et al., 2009). *Prochlorococcus* has one of the smallest known

photoautotrophic non-symbiotic genomes, with a genome size of 1.6-2.4Mb (Coleman et al., 2006; Hess, 2011), GC content of 30.8-50.7% (Kettler et al., 2007), and it contains only ~1,700 genes (Rocap et al., 2003). It has a high surface-area:volume ratio, which promotes efficient light absorption (Biller et al., 2015), and allows it to dominate oligotrophic areas and outcompete other species for nutrients (Laurenceau et al., 2020). However, the small genome has been found to lack genes that are present in other species of cyanobacteria, including those involved in key processes such as DNA repair and photosynthesis (Dufresne et al., 2003).

1.1.1 Habitat and Morphology

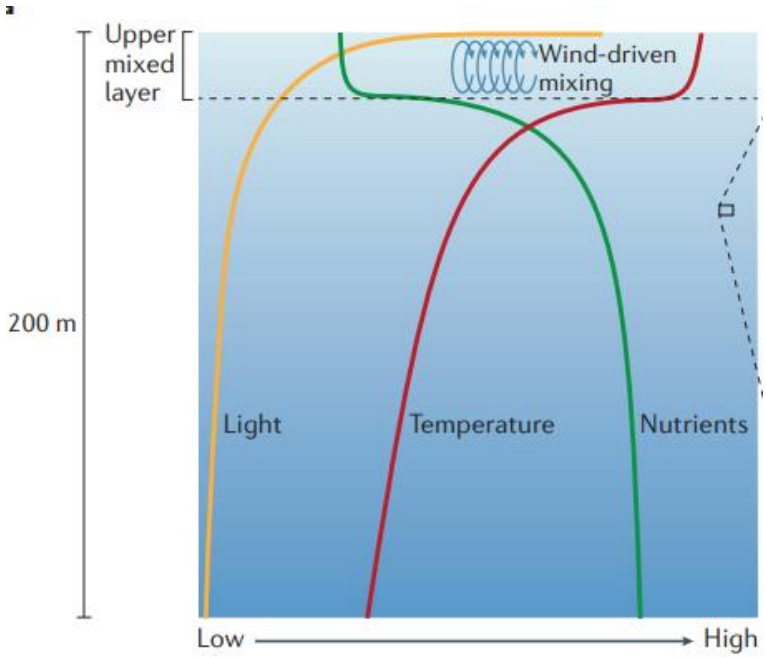


Figure 1-1: Diagram depicting the euphotic zone of the ocean, which *Prochlorococcus marinus* occupies. The euphotic zone water column is characterized by its light, temperature, and nutrient gradients in the lower layers. Different strains of *P. marinus* are located at different levels of the water column, depending on their light, temperature, and nutrient requirements (Biller et. al, 2015).

Prochlorococcus marinus is a micro-diverse organism that can occupy the entire euphotic zone. Subgroups of *Prochlorococcus* have adapted to survive in different levels of light, temperature, and nutrients, so are therefore capable of living at different levels of the three gradients. Figure 1-1, shows the light, temperature, and nutrient gradients at different depths of the euphotic zone water column. Nutrients and cells are distributed in the upper mixed layer by heat and wind turbulence. Without turbulence, both nutrient and

temperature gradients form in relation to ocean depth. As the depth increases, light levels and temperatures decrease, whereas nutrient concentrations increase (Biller et al., 2015).

Prochlorococcus is an example of a genetically diverse group of organisms that exhibit global biogeographical distribution due to its genetic composition and is influenced by external environmental factors. Phylogenetic variance has been observed between ocean regions, with lineages varying in relative frequencies across regions. Varying nitrate concentration and temperature cause most of the phylogenetic diversity across regions, while phosphate concentration, light availability, and coastal influence have been found to have less of an effect (Kent et al., 2016). Biogeographical patterns of gene content have been observed across regions and are similarly mainly influenced by nitrate concentrations and temperature. Variances in gene content and phylogenetics in *Prochlorococcus* populations are significantly correlated, indicating that phylogenetically similar regions have higher similarity in gene content. This suggests a model of vertical inheritance of gene content. Some non-core genes are environmentally influenced and have contributed to the existence of different *Prochlorococcus* lineages in different environments (Kent et al., 2016).

1.1.2 Nutrient Acquisition

Prochlorococcus requires nutrients such as nitrogen and phosphorous for growth and survival. Marine picocyanobacteria genomes have adapted to different niches and have acquired different methods to metabolise nitrogen depending on the niche they occupy and each niche's nitrogen availability. *P. marinus* strain MIT9312 is exposed to nanomolar concentrations of nitrogen in the surface layer of the oligotrophic (sub)tropical ocean (Scanlan et al., 2009). All marine picocyanobacteria are unable to fix dinitrogen as they lack the genes that encode nitrogenase. Most nitrogen is obtained by assimilating ammonium into organic nitrogen compounds, in addition to other inorganic and organic nitrogen compounds. Phosphorous acquisition genes in *Prochlorococcus* can vary between species and within a genus, which suggests that phosphorous acquisition strategies differ depending on the environment. Most of these genes are in variable genomic islands and it is suspected that *Prochlorococcus* have adapted over time to their environment and have acquired genes for nutrient assimilation via horizontal gene transfer (Coleman et al., 2006; Rocap et al., 2003). *Prochlorococcus* can thrive in areas with minimal nutrients, but when starved of light and nutrients for extended periods,

Prochlorococcus depends on co-occurring heterotrophic bacteria for survival (Roth-Rosenberg et al., 2020).

1.1.3 Photosynthesis

As the most abundant photosynthetic organism on Earth, *Prochlorococcus* is one of the main primary producers in low to mid-latitude oceans (Kettler et al., 2007), contributing 30-80% of total photosynthesis, and significantly contributing to climate regulation and global carbon cycling (Bryant, 2003; Paul et al., 2010). *Prochlorococcus* and *Synechococcus* are equipped with different photosynthetic systems which are well adapted to the different conditions that each species is exposed to. *Prochlorococcus* uses chlorophyll-binding proteins as a light harvesting antenna to harvest divinyl chlorophyll a/b pigments, whereas *Synechococcus* uses phycobilisomes on monovinyl chlorophyll pigments (Biller et al., 2015; Kettler et al., 2007; Scanlan et al., 2009; Yan et al., 2020). The Pcb (pigment binding) proteins used by *Prochlorococcus* allows the adaptation of different strains to low-light environments (Yan et al., 2020). Divinyl chlorophyll a/b pigmentation causes a red-shift in the absorption spectrum, which assists with the absorption of blue light, which is dominant in deeper, low-light environments. Different *Prochlorococcus* strains have adapted to varying light levels by adjusting their ratios of pigmentation to meet light absorption requirements (Scanlan et al., 2009; Yan et al., 2020).

1.1.4 *Prochlorococcus* Ecotypes

There is significant variation between *Prochlorococcus* strains, which has resulted in the classification of strains by ecotype. There are two ecotypes: high-light (HL) and low-light (LL), which are named as such due to the abundance of different strains at the surface and deeper levels of the ocean respectively (Prabha et al., 2014). This phenotypical classification has been explored further, leading to the discovery of at least 12 major clades (Biller et al., 2015; Yan et al., 2020).

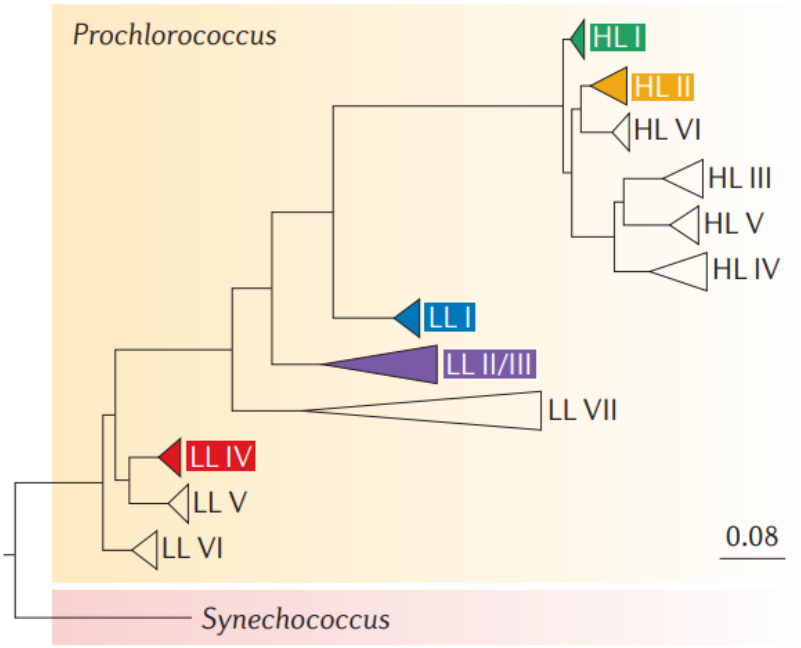


Figure 1-2: Phylogenetic tree of major *Prochlorococcus* clades, based on the rRNA internal transcribed spacer (ITS) sequences. The five clades highlighted have been successfully cultured, the remainder have been identified from environmental sequencing (Biller et. al, 2015).

The strains comprising these 12 clades have 16s rRNA sequences that have more than 97% sequence similarity. Despite the vast difference in traits held by HL and LL strains, *Prochlorococcus* remains as a monophyletic group that is closely related to *Synechococcus* (shown in Figure 1-2). The HL strains form a monophyletic group, consisting of six clades (HLI-HLVI), whereas the LL strains for a polyphyletic group, also consisting of six clades (HLI-HLVII) (Biller et al., 2015; Prabha et al., 2014; Yan et al., 2020).

P. marinus strain MIT9312 is a HLII strain. The HLI and HLII clades consist of the most abundant strains, contributing over 90% of the global *Prochlorococcus* population. The HLII strains are located in higher-temperature oceans and can grow at temperatures over 28^{o}C (Johnson et al., 2006). Physiological studies have confirmed that HL strains grow more quickly and efficiently in HL conditions than LL strains, and vice versa (Moore et al., 1998).

1.1.5 *Prochlorococcus* Genomics

The differentiation of ecotypes is based on the differences in the genomes of different strains. *Prochlorococcus* has a streamlined genome, which formed when the genome

decreased in size once the species diverged from *Synechococcus* (Scanlan et al., 2009). This led to the removal of genes that may not have provided a benefit to *Prochlorococcus* that outweighed the cost of retention (Sun & Blanchard, 2014). HL-adapted strains have a smaller genome of 1.60-1.80Mbp that encodes 1800-2100 genes, compared to LL-adapted strains, which are 1.75-1.95Mbp, and encodes 2000-2300 genes (Hess, 2011). The genome of the *Prochlorococcus* genus is described as being an open pan-genome of more than 13,000 genes, which consists of 'core' and 'flexible' sections. The core genome consists of essential housekeeping genes and comprises approximately 50% of the genome (Kettler et al., 2007). The flexible genes have been described as being part of hypervariable genomic islands which different strains draw from and contribute towards (Biller et al., 2015; Kettler et al., 2007; Tschoeke et al., 2020). Flexible genes that have been retained by different strains are likely to be beneficial for survival in the strain's particular environment and are subject to selection pressure (Biller et al., 2015). Horizontal gene transfer often occurs with genomic islands and is a source of new genes which allows *Prochlorococcus* to continuously adapt to its environment.

There is a notable difference in the number of enzymes present in low and high-light-adapted ecotypes of *Prochlorococcus*. High-light strains have been found to contain more genes related to DNA repair enzymes such as DNA ligases compared to low-light strains (Partensky & Garczarek, 2010). *P. marinus* MIT9312 is a high-light adapted strain that thrives in the upper euphotic zone (Zinser et al., 2006), and it contains more DNA modifying enzymes than other strains. The function of these additional enzymes is of interest because of their retention in the small, minimal *Prochlorococcus* genome over other 'essential' enzymes (Dufresne et al., 2003).

1.1.6 An Evolving Field of Research

The taxonomy of *Prochlorococcus marinus* is an active field of research that is rapidly evolving. Large-scale sequencing technology has led to the rapid identification of new species and genes which make up the *Prochlorococcus* collective, and the *Prochlorococcus* pan-genome (Tschoeke et al., 2020; Yan et al., 2020). For example, since the commencement of this project, 13 new isolates from the HLII clade were discovered and sequenced from the western Pacific Ocean and the South China Sea by the China University of Geoscience (NCBI accession number: PRJNA611498). In contrast to the widely adopted ecology-based classification of *Prochlorococcus*, a

genomic taxonomy approach has recently led to a proposal of the reclassification of *P. marinus* strain MIT9312 as part of a new genus: *Eurycolium*, and more specifically, *Eurycolium neptunius* (Tschoeke et al., 2020). This is an example of how our understanding of *Prochlorococcus* from ecological, taxonomic, and biochemical research is ever-growing.

1.2 DNA Repair Pathways in Prokaryotes

DNA repair is an essential process for all living organisms to allow DNA replication and gene expression to proceed correctly. As photoautotrophs, *Prochlorococcus* amongst other cyanobacteria are exposed to ultraviolet radiation which although necessary for photosynthesis, can be lethal to the organism (Cassier-Chauvat & Chauvat, 2015). DNA can easily be damaged by the external environment of organisms, hence the necessity of a well-functioning DNA repair system. Cyanobacteria can be severely harmed biochemically and physiologically by UV-B irradiation, where metabolic processes, growth, development, survival, morphology, pigmentation, and other essential life processes are severely affected (Pathak et al., 2019). In addition to UV damage, bacteria can develop DNA lesions in response to a range of endogenous and environmental factors. Figure 1-3 outlines the main DNA lesion types and the corresponding DNA repair pathways that are employed in response. In summary, base excision repair (BER) is carried out in response to alkylating agents and oxygen radicals modifying the bases and phosphates of DNA. BER is also used for damage caused by spontaneous DNA mutations. Nucleotide excision repair (NER) is used to repair DNA adducts and interstrand cross-links. Errors can arise during DNA replication, and these are repaired by mismatch repair (MMR). X-rays can induce double-stranded breaks in DNA, which are repaired by recombinational repair (Morita et al., 2010).

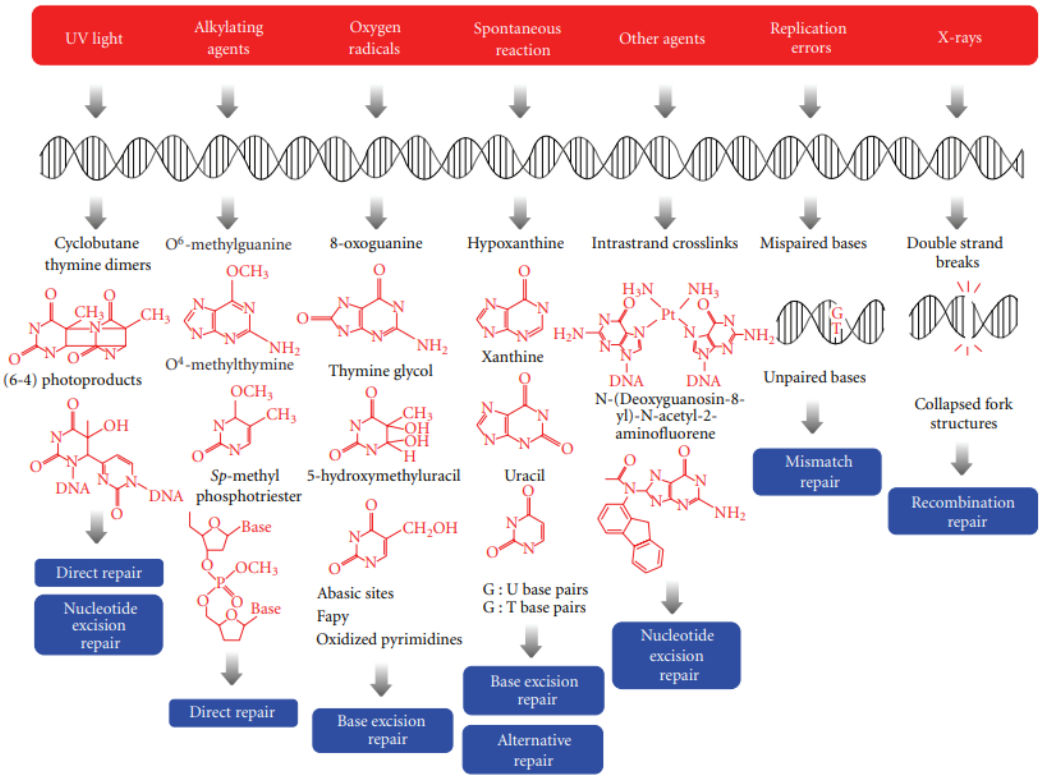


Figure 1-3: Different DNA repair systems for primary types of DNA lesions observed in prokaryotes (Morita et al., 2010). DNA repair in prokaryotes can be split into three main categories: direct reversal, single-stranded break repair, and double-stranded break repair.

1.2.1 Prokaryotic SOS Response

The SOS response in bacteria occurs in response to increased DNA damage, where DNA repair and mutagenesis are initiated. An inducer named RecA, and a repressor named LexA, are the two key proteins involved in this response. During normal bacterial growth, SOS gene expression is prevented as the LexA repressor binds to the SOS box in the promoter region of the SOS genes. When DNA damage increases, RecA filaments form at the DNA damage site, which activates self-cleavage of the LexA repressor, enabling SOS gene expression. The SOS response is driven by the accumulation of single-stranded DNA at replication forks (i.e., the damage site), as DNA polymerase is blocked from creating DNA duplexes. As LexA self-cleaves the repression of SOS genes decreases, and different DNA repair mechanisms are therefore activated based on the affinity of LexA for the SOS box (Kreuzer, 2013; Little & Mount, 1982; Maslowska et al., 2019; Michel, 2005).

1.2.2 Direct Reversal of DNA Damage

DNA damage that is induced by UV light and alkylating agents can sometimes be directly reversed without the need for the resynthesis of DNA. Schematic models of both direct reversal pathways are shown in Figure 1-4.

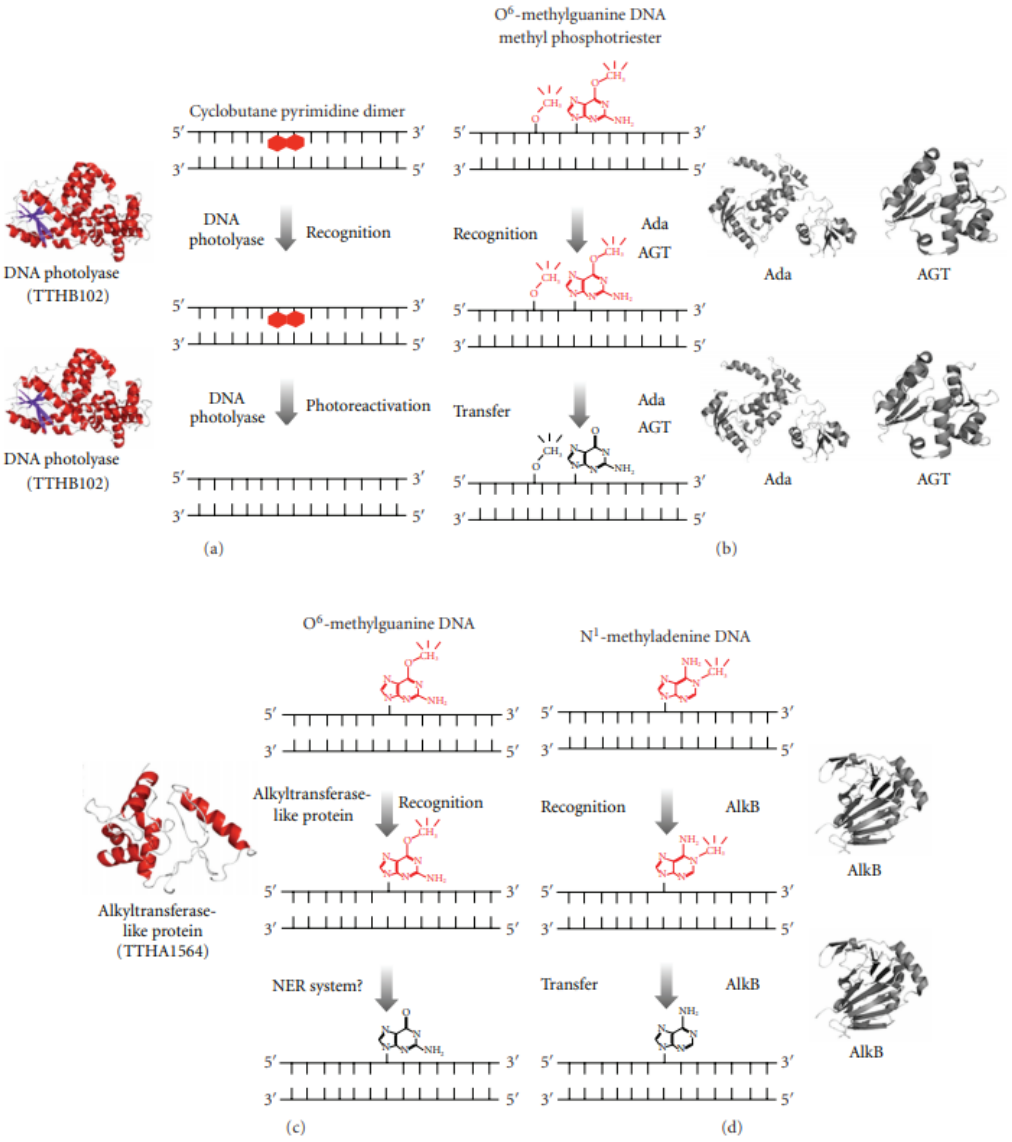


Figure 1-4: Schematic models for direct reversal of DNA damage. (a) Recognition and repair of cyclobutene pyrimidine dimers by DNA photolyase. (b) Reversal of O^6-alkylated guanines by O^6-alkylguanine-DNA alkyltransferase (AGT), or by (c) alkyltransferase-like (ATL) proteins. (d) Reversal of N-alkylated by AlkB family dioxygenase (Morita et al., 2010).

UV light induces the formation of additional bonds between two neighbouring pyrimidine groups of DNA. Two major photoproducts: cyclobutane pyrimidine dimer (CPD)- and (6-4) photoproducts are produced, and without DNA repair, would result in mutagenesis and cell death (Ma et al., 2019). The lesions are repaired by two different subgroups of

photolyases: CPD- and (6-4) photolyases, which both bind as single proteins to the lesions. Photolyases contain a central FAD chromophore which must be fully reduced to $FADH^-$ for DNA repair to be carried out. Oxidized FAD is converted to $FADH^-$ by a process called photoreduction, which is activated by light. Light-activated photolyases bind to the DNA lesions, where electrons are transferred between the lesion and FADH-. This restores the pyrimidines to their monomeric forms, as the pyrimidine dimer and cyclobutene bond are broken (Ma et al., 2019; Morita et al., 2010).

Alkylating agents can bind to DNA and form alkylation adducts, which without DNA repair could induce mutations and cell apoptosis (Morita et al., 2010). Two direct reversal pathways are utilised by bacteria to overcome alkylation damage. The first pathway, shown in Figure 1-4 (b), involves O^6-alkylguanine-DNA alkyltransferase (AGTs) which reverses O^6-alkylated guanines. The second pathway, shown in Figure 1-4 (d), utilizes AlkB family dioxygenase, which reverses *N*-alkylated based adducts and lesions which can interfere with Watson-Crick pairings (Morita et al., 2010; Yi & He, 2013). AGTs transfer alkyl groups from the DNA to a cysteine residue in its active site, resulting in deactivation of the AGT and repair of the DNA. Some species of bacteria contain homologues of AGT called alkyltransferase-like (ATL) proteins, where the cysteine residue is replaced by a different amino acid residue incapable of nucleophilic attack (Margison et al., 2007). ATL's instead bind to O^6-alkyl guanine adducts (Figure 1-4 (c)) and bend the adduct to create a bulky lesion which is repaired via nucleotide excision repair (Tubbs & Tainer, 2010). AlkB family dioxygenases carry out oxidative-demethylation of N^1-methyladenine and N^3-methylcytosine adducts by utilizing Fe(II), O_2, and α-ketoglutarate. Oxidative-demethylation results in the oxidation of the alkyl groups, release of succinate, formaldehyde, and CO_2, and the regeneration of undamaged DNA (Fedeles et al., 2015; Morita et al., 2010; Yi & He, 2013).

1.2.3 Single-stranded Damage Repair

Often there will be damage to only one of the strands making up DNA, which means the undamaged strand can act as a template for repairing single-stranded breaks. Three repair systems are employed by bacteria: base excision repair (BER), nucleotide excision repair (NER), and mismatch repair (MMR).

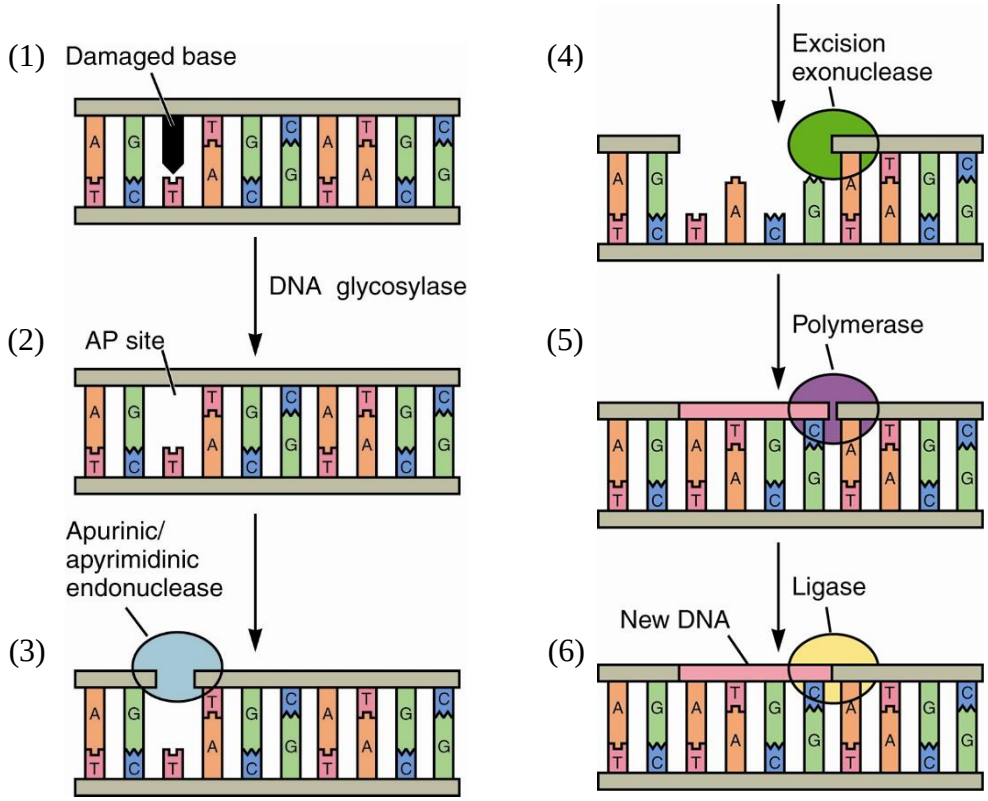


Figure 1-5: Base excision repair pathway in prokaryotes. (1) DNA glycosylase recognizes the recognizes the damaged base and (2) hydrolyses the N-glycosidic bond between the base and deoxyribose sugar, producing an abasic (AP) site. (3 and 4) Apurinic/apyrimidinic endonucleases target the AP site to produce a 3' OH and 5' phosphate end. (5) DNA polymerase I adds the required nucleotide. (6) DNA ligase A seals the nick (Garrett & Grisham, 2017).

BER is a commonly used DNA repair pathway, as single-base lesions that arise from deamination of bases, depurination, depyrimidination, and reactive oxygen species can easily be replaced (shown in Figure 1-5). BER is initiated by a DNA glycosylase which recognizes the damaged base and hydrolyses the N-glycosidic bond between the base and deoxyribose sugar. DNA glycosylases can be either monofunctional or bifunctional, but both result in the generation of 3' OH and 5' phosphate ends of the abasic (AP) site produced by the DNA glycosylase. The AP site generated by a monofunctional DNA glycosylase is targeted by apurinic/apyrimidinic endonucleases (APE), which produce 3' OH and 5' dRP ends. The 5' dRP end is processed by dRPase to give a 5' phosphate end. In contrast, the AP site generated by the AP lyase activity of the bifunctional DNA

glycosylases has a 3'-α,β-unsaturated aldehyde (PA), or a 3' phosphate end, and a 5' phosphate end. The 3' ends are produced as the result of β-elimination or β/δ-elimination respectively. APE is then required to produce the 3' OH end. Once the 3' OH and 5' phosphate ends have been produced, DNA polymerase I then adds the required nucleotide, and DNA ligase A seals the nick (Garrett & Grisham, 2017; Kurthkoti et al., 2020; Morita et al., 2010).

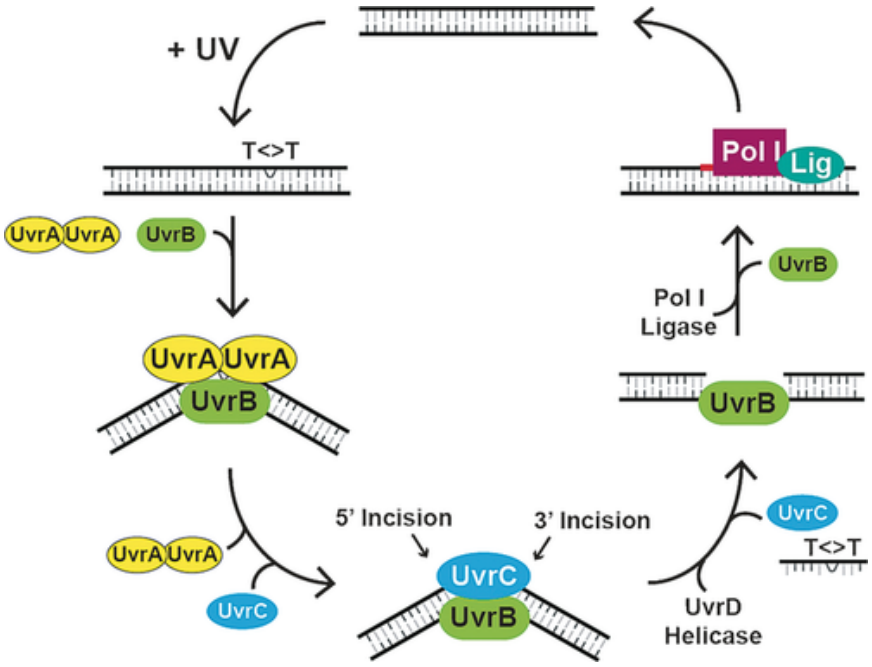


Figure 1-6: Nucleotide excision repair in *E. coli*. The damage is recognized by a (UvrA)2 homodimer, which recruits UvrB to the damage site. An ATP hydrolysis-dependent reaction promotes the formation of a UvrB − DNA complex. This complex recruits UvrC, which incises 5′ and 3′ to the damage. UvrC and the excised dodecamer (12-mer) are released by UvrD helicase. PolII then synthesizes a new strand that is complementary to the excised oligomer, which is then ligated by a DNA ligase (Sancar, 2016).

Prokaryotic NER is an essential repair mechanism, shown in Figure 1-6, that is used to remove UV-induced DNA lesions, such as pyrimidine dimers and bulky adducts. The damage site is recognized via one of two mechanisms: global genomic repair (GGR) or transcription-coupled repair (TCR). During GGR, DNA lesions are recognized by UvrAB, and repair is initiated. In contrast, TCR repair is initiated by a delay in RNA polymerase activity. Both pathways lead to the UvRC incising both sides of the lesion to remove a 12-13nt oligomer, which is released by UvrD helicase. PolII then synthesizes a new strand that is complementary to the excised oligomer, which is then ligated by a DNA ligase (Hu et al., 2017; Morita et al., 2010; Sancar, 2016; Truglio et al., 2006).

Sometimes during DNA replication, the wrong base may be introduced and needs to be replaced, warranting the process of MMR. The enzymes involved during MMR can vary between species, so MMR in *E. coli* has been briefly outlined. During MMR a protein called MutS binds to the mismatched base to form a MutS-mismatch complex. MutL binds to the MutS-mismatch complex to stabilize it, and a restriction endonuclease named MutH is activated. MutH nicks the mismatch-containing unmethylated strand of DNA to create an excision entry point for a DNA helicase, which excises the mismatch-containing region of DNA. DNA polymerase III then synthesizes a new strand which is then ligated by a DNA ligase (Morita et al., 2010).

1.2.4 Double-stranded Break Repair

Double-stranded breaks (DSB) are repaired by two main pathways in prokaryotes: nonhomologous end-joining (NHEJ) and homologous recombination (HR). HR is a universal pathway and is the main DSB pathway used by bacteria, shown in Figure 1-8. In contrast, NHEJ, shown in Figure 1-7, is used by a small number of bacterial species that have dormant phases, such as *Bacillus* and *Streptomyces*, or spore-forming phases, like *Mycobacterium* and *Sinorhizobium*. During these phases, bacteria are "single genome compartments", meaning that HR is not possible (Bertrand et al., 2019).

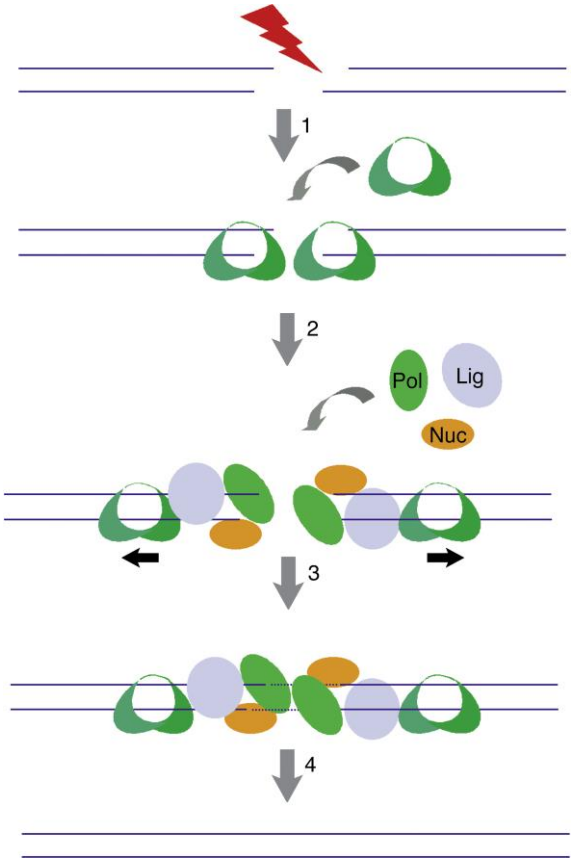


Figure 1-7: NHEJ repair pathway in prokaryotes. (1) The Ku complex binds to the DSB site. (2) LigD is recruited to the damage site. (3) DNA end-processing, gap-filling, and nucleolytic activities are carried out. The 3' overhangs are resected by the nuclease domain of LigD, while PolDom carries out resynthesis of the 3' ends. (4) The break is sealed by LigD and the NHEJ complex dissociates (Bowater & Doherty, 2006).

Unlike in eukaryotes, NHEJ in prokaryotes involves only two proteins: Ku and a multifunctional ATP-dependent ligase, LigD. Ku binds to the 3' overhanging ends of the break and recruits LigD via its polymerase domain (PolDom). PolDom binds to the terminal 5'-phosphate group of the DSB and promotes the synapsis and resection of the 3' ends of the break. If the overhangs of the DSB are complementary, the overhangs can be aligned and religated by the ligase domain of LigD. If the overhangs are not complementary, microhomology-mediated DSB repair is carried out by Ku-LigD, where areas of homology are aligned. The 3' overhangs are resected by the nuclease domain of LigD, while the PolDom carries out resynthesis of the 3' ends and ligation of nicks to seal the break (Bowater & Doherty, 2006; Pitcher et al., 2007).

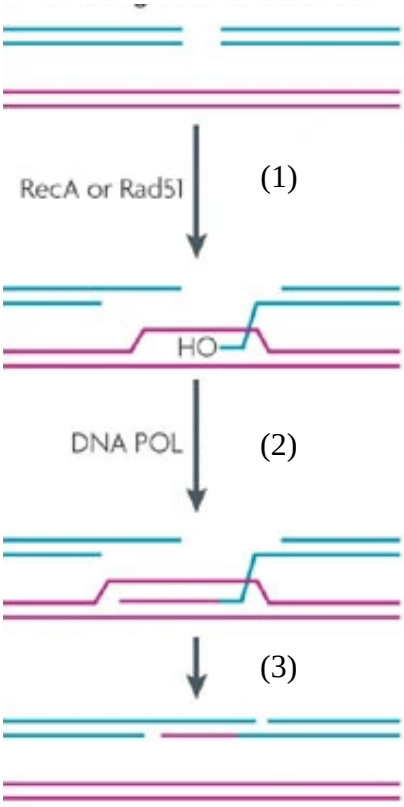


Figure 1-8: HR repair pathway in prokaryotes. Cyan strand of DNA duplex has a DSB that is repaired using the sister chromatid (magenta) as a template. (1) RecA catalyses the invasion of the sister chromatid by the damaged chromatid. (2) DNA polymerase replicates the in-tact strand to repair the DSB. (3) The duplexes separate and LigA seals the nicks of the damaged chromatid (Shuman & Glickman, 2007).

HR uses an intact copy of the damaged chromosome as a template for DNA synthesis to repair the break. One or both ends of the DSB is resected by a 5' to 3' exonuclease to produce a 3' ss-tail. A RecA protein catalyses the invasion of the template sister chromatid by the 3' ss-tail. The invading damaged strand pairs with the in-tact strand using complementary base-pairing. DNA polymerase uses the damaged strand as a primer and replicates the in-tact chromatid to fill and repair the DSB. Any remaining nicks in the repaired DNA duplex are sealed by LigA (Shuman & Glickman, 2007).

1.3 NAD and ATP-dependent Ligases

DNA ligases join broken DNA and catalyse the phosphodiester bond formation between 5' PO_4 and 3' OH ends of adjacent strands that form the DNA duplex (Lohman et al., 2011; Shuman, 2009; Timson et al., 2000; Williamson & Leiros, 2020). DNA ligases are split into two categories: ATP-dependent or NAD-dependent, which is determined by which AMP donor is used during the 3-step phosphodiester bond formation reaction (Timson et al., 2000). During the first step, the α-phosphate of either ATP or NAD+ undergoes nucleophilic attack by a lysine residue from the catalytic motif of the ligase, forming lysine-AMP and releasing pyrophosphate (PP_i) or nicotinamide mononucleotide (NMN) (Timson et al., 2000; Williamson & Leiros, 2019, 2020). AMP is then transferred to the 5' PO_4 end of the DNA to form a pyrophosphate bond. This activates the 5' phosphate for nucleophilic attack by the 3' OH, forming a new phosphodiester bond and releasing AMP (Williamson & Leiros, 2019).

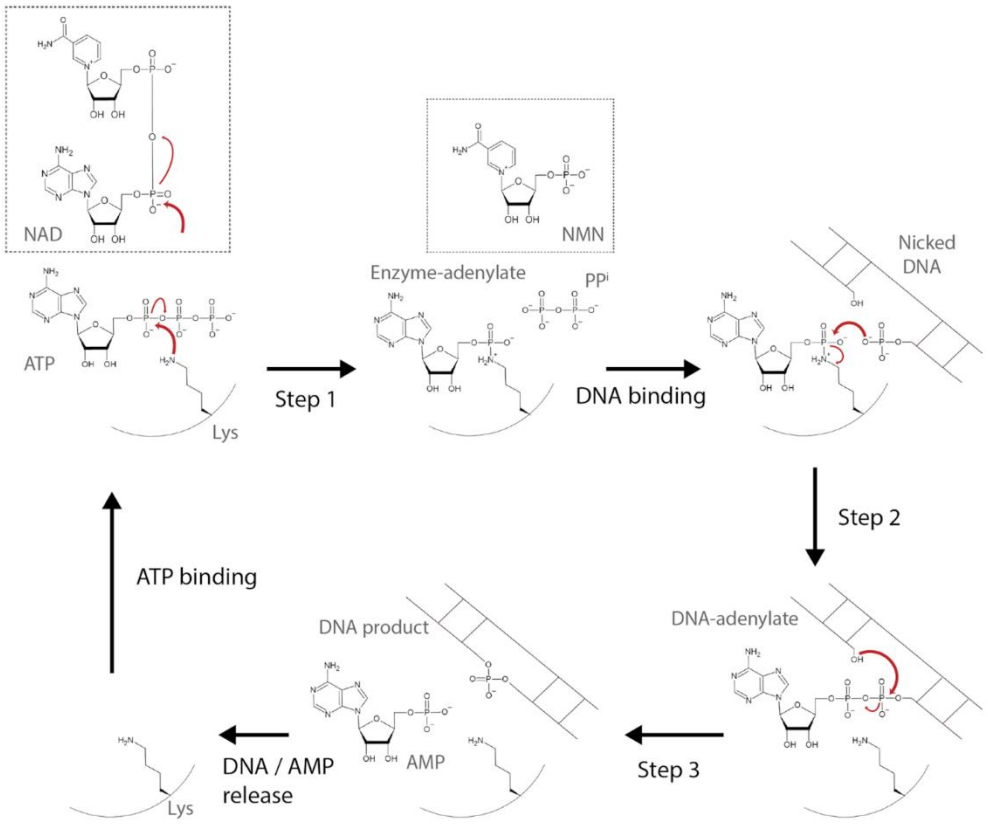


Figure 1-9: The three catalytical steps of the nick-sealing reaction catalysed by DNA ligase. Step 1 shows both the ATP and NAD-dependent pathways, which require ATP or NAD+ respectively. Steps 2 and 3 are common to both pathways and result in the formation of a new phosphodiester bond and the release of AMP (Williamson & Leiros, 2020).

All bacteria contain NAD+-dependent DNA ligases which are essential for DNA replication, with some species also containing ATP-dependent ligases that are used for DNA repair (Williamson & Leiros, 2020). ATP-dependent ligases are also used by eukaryotes and archaea for replication. The difference in ligase utilization and replication machinery is one of the key differences between bacterial and archaeal or eukaryotic cell lineages (Doherty & Suh, 2000; Hjerde et al., 2020; Makarova & Koonin, 2013). It remains unknown as to why bacteria favour NAD+-dependent ligases (Hjerde et al., 2020).

1.3.1 DNA Ligases in Prokaryotes

Prokaryotes contain a variety of DNA ligases that are involved in specific DNA repair pathways, summarized in Figure 1-10.

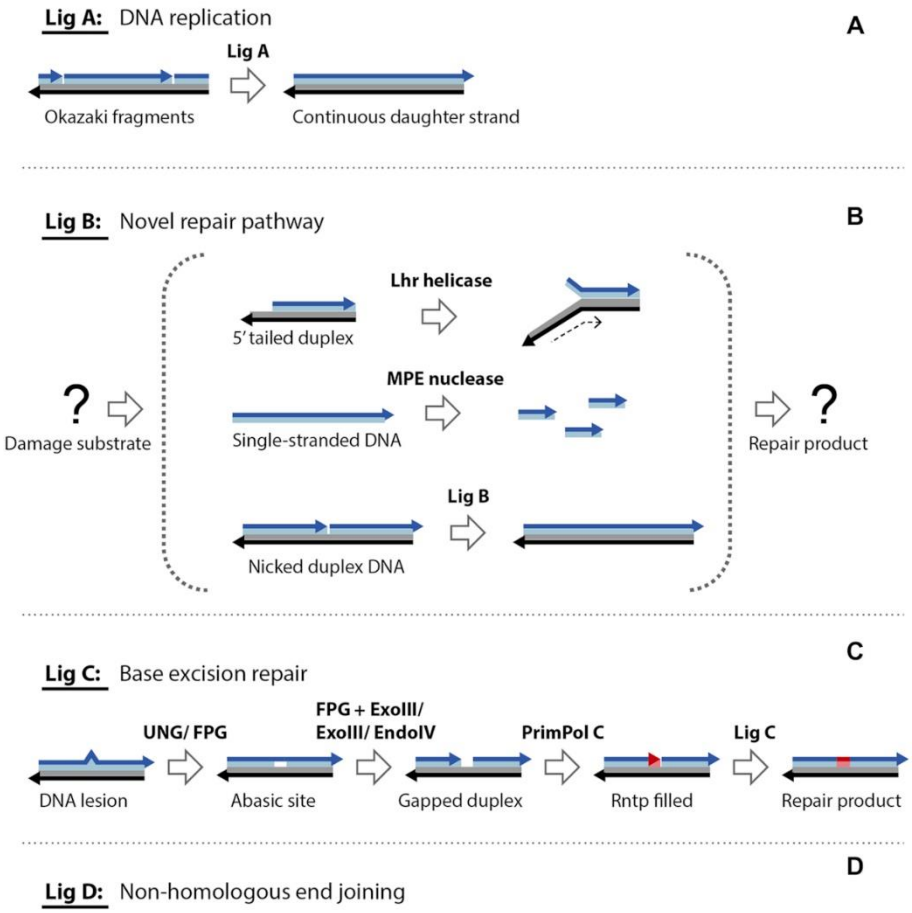


Figure 1-10: DNA modification pathways in bacteria involving DNA ligases. (A) LigA joins Okazaki fragments during DNA replication. (B) Proposed repair pathway including LigB. (C) Base excision repair includes LigC in the final ligation step. (D) Different domains of the multifunctional LigD are involved at different steps during the repair of double-stranded breaks. (E) The function of LigE and the repair pathway it is involved is yet to be determined (Williamson & Leiros, 2020).

All known bacteria contain a highly conserved NAD+-dependent ligase, LigA, which joins Okazaki fragments during DNA replication (Williamson & Leiros, 2020). *Haemophilus influenzae* was one of the first species of bacteria found to contain ATP-dependent ligases in addition to NAD+-dependent ligases (Cheng & Shuman, 1997; Wilkinson et al., 2001). Other human pathogens such as *Neisseria meningitidis*, *Yersinia pestis*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, and *Mycobacterium tuberculosis* contain ATP-dependent DNA ligases (Gong et al., 2004). *M. tuberculosis* encodes three ATP-dependent ligases (LigB, LigC, and LigD), which were successfully characterized and have therefore been used as a leading organism in ATP-dependent ligase studies (Gong et al., 2004). LigB is efficient at sealing nicks, but the repair pathway involving LigB remains to be determined and is expected to be different from existing repair pathways. LigB has been found in gene clusters with a novel Lhr-helicase, binuclear metallophosphoesterase, and putative exonuclease, suggesting that these enzymes will each have a role in a novel repair pathway (Ejaz & Shuman, 2018; Gong et al., 2004; Williamson & Leiros, 2020). LigC seals nicks at the end of base-excision repair, which is carried out during the stationary-phase of the bacterium. LigD, as previously discussed, is a multifunctional ATP-dependent ligase involved in re-joining double-stranded DNA breaks. The function of another ATP-dependent ligase, LigE, is still yet to be determined, but has an affinity for sealing single-stranded breaks (Williamson & Leiros, 2020).

1.3.2 Ligases in *P. marinus* MIT9312

A phylogenomic study of the distribution of ATP-dependent ligases among bacteria (Williamson et al., 2016) found that some strains of *P. marinus* had up to three ATP-dependent ligases in addition to the highly conserved NAD+-dependent replicative DNA ligase (Wilkinson et al., 2001). However, subunits involved with repair pathways that utilise ATP-dependent ligases in other organisms are absent (Williamson et al., 2016). A comparative genomic study by our group (Hjerde et al., 2020) has shown that ATP-dependent ligases are not present in low-light strains, while LigP and LigB are present in all high-light adapted strains. The genes encoding these high-light specific ligases are an example of flexible genes which help these strains to survive in the upper layers of the ocean. At the commencement of this master's project, LigW appeared to only be present in the MIT9312 strain – summarised in Figure 1-11.

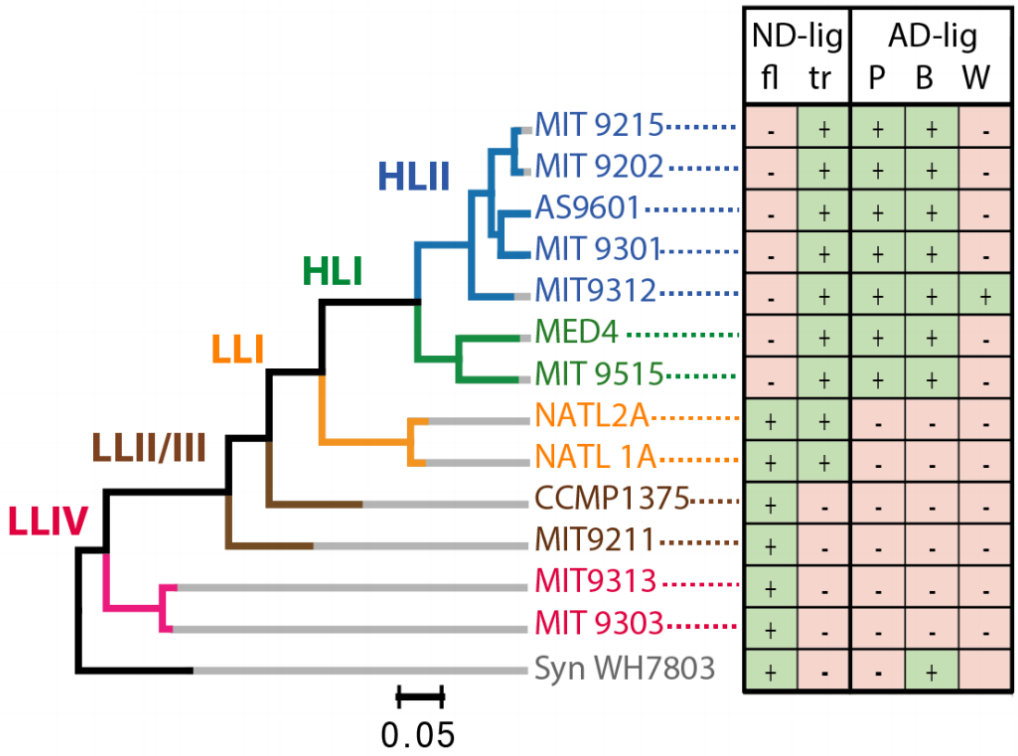


Figure 1-11: "Presence and absence of ATP- and NAD-dependent DNA ligases in the 13 closed genomes of *P. Marinus* strains with *Synechococcus* WH7803 as an outgroup". Strain evolution was determined using DNA gyrase B as a gene marker. Ecotypes are classified as high-light (HL) or low-light (LL) according to their assignment in ProPortal. Truncated (tr) and full-length (fl) NAD-dependent ligases (ND-lig) and ATP-dependent ligases (AD-lig) P, B, and W have been identified in different *Prochlorococcus* strains (indicated by a +). Data is from a pre-print by our research group (Hjerde et al., 2020).

LigB is similar to the LigB found in *M. tuberculosis*, while the other two ligases are unique. These two ligases have been cloned, expressed, and characterized (Williamson & Leiros, 2019). One of the ligases, LigP, was found to have a similar structure to other ATP-dependent ligases, with a similar alpha-helical N-terminal DNA binding domain to the T4 DNA ligase. Lig P was found to also be similar to T4 in being highly active in sealing single-strand breaks, but LigP was not effective in sealing double-strand breaks (Williamson & Leiros, 2019). The second ligase, LigW, was named as such because of its predicted WGR domain at the N-terminus.

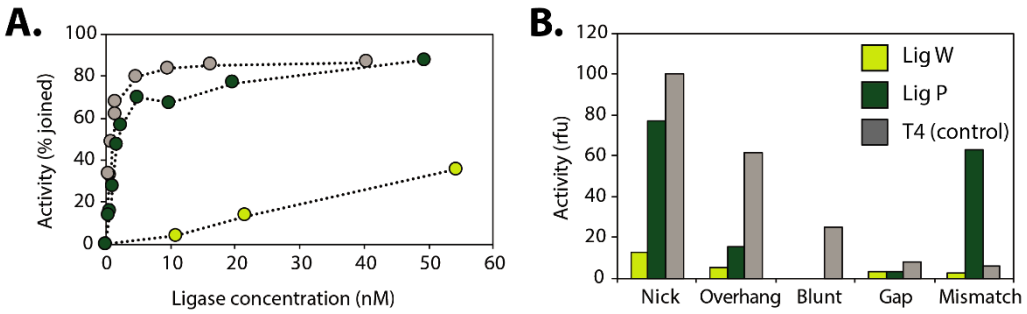


Figure 1-12: Unpublished data comparing rates of ligation in two of the ATP-dependent ligases of *P. marinus* MIT9312. (a) LigW has low joining activity compared to LigP and T4. Activity increases with increasing ligase concentration, but not as significantly as another ATP-dependent ligase, LigP. (b) LigW has comparatively low activity and no evident affinity for different types of DNA damage.

LigW was found to have low DNA ligase activity on nicked DNA in comparison to LigP and T4 DNA ligase, shown in Figure 1-12 (Williamson, Unpublished). The low levels of activity shown by LigW make its retention in the small *Prochlorococcus* genome unusual. It is suspected that LigW could achieve higher rates of ligation by interacting with other proteins. This would be like some DNA ligases such as LigD in other species, which interacts with the Ku protein present on the LigD operon (Zhu & Shuman, 2007). *P. marinus* does not have a Ku homologue (Williamson et al., 2016) but LigW appears to be in an operon with other proteins (shown in Figure 1-13) that are predicted to promote LigW activity or have their own DNA modifying activities (Williamson, Unpublished). This operon that appears to be unique to the MIT9312 strain contains genes that are part of the flexible genomic island. The three proteins of interest have been annotated as Pmar3, Pmar4, and Pmar5.

1.4 DNA-modifying proteins in *P. marinus* MIT9312

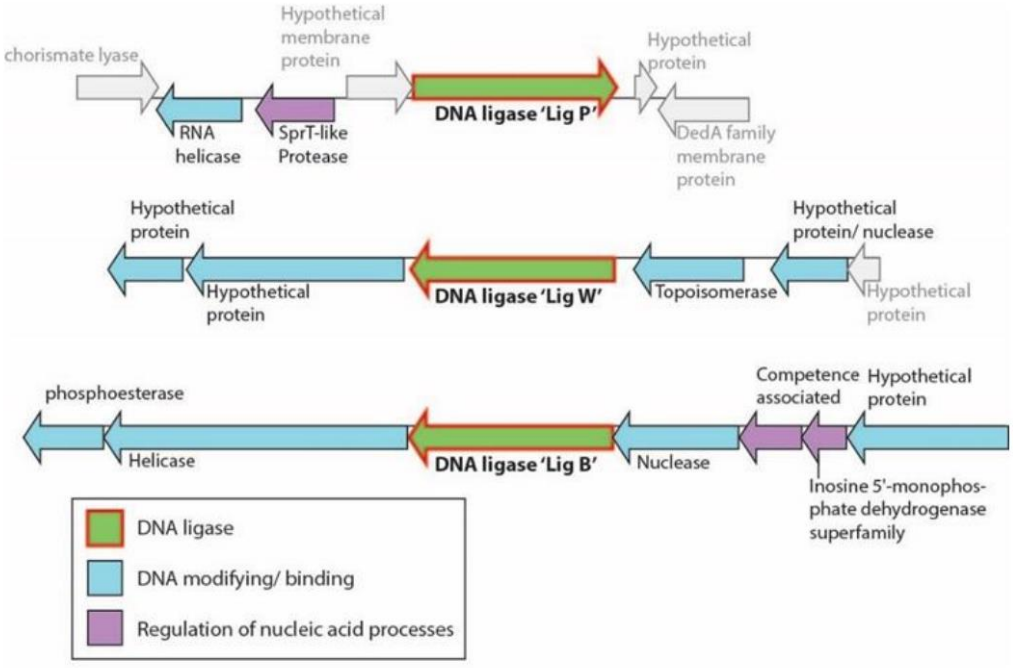


Figure 1-13: Putative DNA repair operons of *P. marinus* MIT9312 including ATP-dependent DNA ligases (green) (Williamson, Unpublished).

At the commencement of this research, Pmar5 was annotated as a potential topoisomerase, based on sequence homology, whereas the function of Pmar3 and Pmar4 remained to be determined. The current details and sequences of these proteins are shown in Appendix A. Each protein has a 6-Histidine tag (highlighted in blue), TEV site (highlighted in green), and cleavage site (written in red). The 6-Histidine tag at the N terminal gives each protein a high binding affinity to nickel. This allows the protein to be separated from chaperone *Escherichia coli* proteins during purifications (Bornhorst & Falke, 2000; Janson, 2012).

1.5 Research Objectives

The aim of my master's research was to recombinantly produce and biochemically characterize three of the proteins adjacent to LigW and assess their activities *in vitro*. The following objectives were set to meet this aim:

1) Recombinantly produce Pmar3, Pmar4, and Pmar5. This would involve optimizing the expression of each protein using an *E. coli*-based system and

optimizing the large-scale purification of each protein using immobilized metal affinity chromatography.

2) Biochemically characterize Pmar3, Pmar4, and Pmar5 to determine the independent function of each protein. This would be achieved through a series of function-specific gel-based assays that assess activities including nuclease and topoisomerase activity, as well as the ability to bind damaged DNA.
3) Determine the relationship of each protein with LigW. This would involve ligation and gel-based assays that assess how each protein stimulates or modifies the activity of LigW.

Characterizing these proteins will help us to understand the significance and purpose of LigW. Once the LigW-containing DNA repair system is characterized, then its role and distribution in the environment can be more easily identified and understood in other organisms where it may not be detected at present. It is hypothesised based on the missing Ku homologue and other repair elements, that LigW and the other ATP-dependent ligases in *P. marinus* MIT9312, are involved in a novel repair pathway, and may carry out unique DNA repair functions (Williamson et al., 2016).

Chapter 2

Protein Expression and Purification

2.1 Introduction

Pmar3, Pmar4, and Pmar5 are unique proteins to *P. marinus* MIT9312 that are proposed to be essential for DNA repair. We wanted to explore the function of each of these genes individually *in vitro* to better our understanding of their purpose and retention in the minimal *Prochlorococcus* genome.

The expression system chosen was an *Escherichia coli*-based system. This was chosen because of its successful application in the recombinant expression of bacterial DNA repair proteins in previous studies (Williamson & Pedersen, 2014), including LigP from *P. marinus* (Williamson & Leiros, 2019). The synthetic constructs were cloned into expression vectors pDEST17 and pHMGWA with N-terminal hexahistidine tags and TEV sites. The N-terminal hexahistidine tag allows the expressed protein to be purified using nickel-immobilized affinity chromatography (Ni-IMAC). TEV protease is able to bind to and cleave the TEV site positioned next to the solubility fusion tags (Histidine [His] and Maltose Binding Protein [MBP]), allowing fusion tag removal during Ni-IMAC (Williamson & Pedersen, 2014).

The plasmids containing the genes encoding the proteins of interest were cloned into a DH5α strain, which has an endA1 mutation that inhibits the endonuclease that degrades plasmid DNA during plasmid purifications, and a recA mutation which helps to stabilize the plasmid by inhibiting homologous recombination (Taylor et al., 1993). The plasmid was then transformed into BL21(DE3)pLysS cells for protein expression. These were used due to containing the DE3 lysogen, which expresses T7 RNA polymerase, which is controlled by the lac UV5 promoter and lacIq (Studier & Moffatt, 1986). Isopropyl-β-D thiogalactopyranoside (IPTG) is an analogue of lactose that is used as an alternative to lactose to induce T7 RNA polymerase expression by binding to the T7 promoter. When IPTG binds, the lac repressor changes shape, dissociates, and allows T7 RNA polymerase transcription to proceed. IPTG cannot be broken down by cells and was therefore used instead of lactose in the expression system, as it would remain at a constant concentration during the expression process. T7 RNA polymerase allows the induction of the otherwise repressed expression from T7 promoter containing plasmids such as the two expression

plasmids used (pDEST17 and pHMGWA). BL21(DE3)pLysS also contains the pLysS plasmid which decreases pre-induction expression by expressing low levels of T7 lysozyme which controls the T7 promoter (Studier & Moffatt, 1986). Because the pLysS and expression plasmids (pDEST17 and pHMGWA) contain chloramphenicol and ampicillin resistance genes respectively, culture media was supplemented with chloramphenicol and ampicillin to maintain plasmid-containing cells during growth.

Soluble expression of Pmar3, Pmar4, and Pmar5 was attempted by using different fusion tags (His and MBP), induction temperatures (15^{o}C and 20^{o}C), and IPTG concentrations (0-1mM) as this strategy had previously overcome solubility issues with difficult proteins (Williamson & Pedersen, 2014). His and MBP fusion tags were used due to their previous success in improving soluble expression of LigW (Williamson, Unpublished) and other proteins transformed into *E.coli* (Niiranen et al., 2007). Soluble expression of recombinant proteins can be difficult to achieve, so work was carried out before the beginning of this project which focused on identifying favourable expression conditions for the proteins of interest.

Table 2.1: Details of the Proteins of Interest and Previous Results.

Protein Name	Expression Plasmid	Fusion Tag	Molecular Weight (kDa)	Present in Pellet?	Soluble Expression?
Pmar3	-	-	16.97	-	-
Pmar4	-	-	46.13	-	-
Pmar5	-	-	23.13	-	-
Pmar3-His	pDEST17	His-attB-His-TEV	21.17	Y	N
Pmar4-His	pDEST17	His-attB-His-TEV	50.33	Y	N
Pmar5-His	pDEST17	His-attB-His-TEV	27.33	Y	N
Pmar3-MBP	pHMGWA	His-MBP-attB-His-TEV	63.67	Y	Y
Pmar4-MBP	pHMGWA	His-MBP-attB-His-TEV	90.83	N	N
Pmar5-MBP	pHMGWA	His-MBP-attB-His-TEV	67.83	Y	Y

During a 10-week undergraduate research project, I focused on the expression and solubility screening of Pmar3, Pmar4, and Pmar5. Six constructs were tested with two different IPTG concentrations at two different temperatures (24 samples in total) at a small-scale. 5 modified constructs were selected for midi-scale expression and solubility screening based on the small-scale results. Nickel pulldowns and midi-scale IMAC's (shown in Figure 2-1) were performed successfully on both Pmar3-MBP and Pmar5-MBP. This prior research provided an early indication that Pmar3 and Pmar5 could successfully be expressed and purified under the following conditions: an MBP fusion tag is used, the culture is grown at 37°C until an OD_{600} of 0.3 is reached and is moved to 15°C after being induced with a final 0.5mM IPTG concentration.

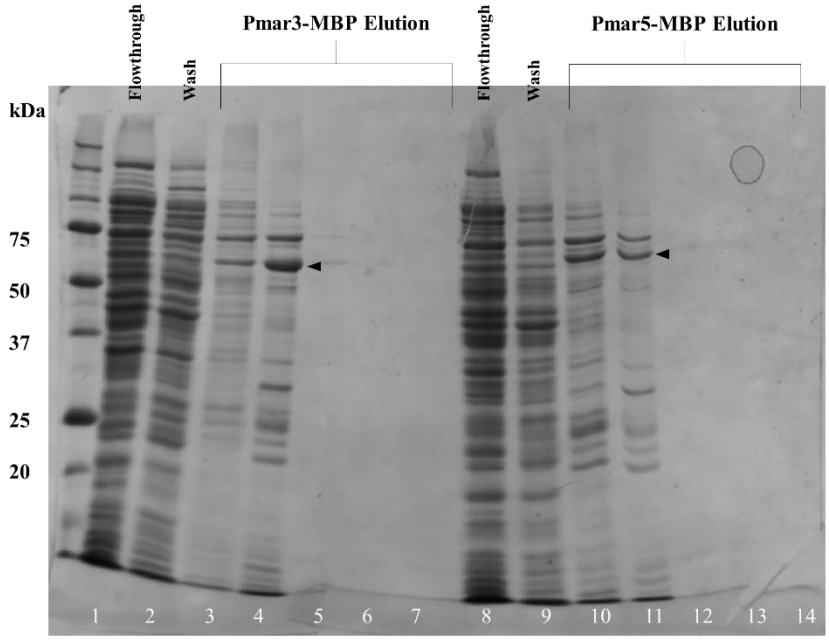


Figure 2-1: 12% SDS-PAGE gel from the Midi-scale IMAC's run of Pmar3-MBP and Pmar5-MBP.

This chapter uses the findings from this pilot project as a starting point for the optimization of large-scale soluble expression and purification of these proteins.

2.2 Materials and Methods

2.2.1 Cloning Pmar3 (WT), Pmar4, Pmar5, and LigW

Genes encoding Pmar3, Pmar4, and Pmar5 were ordered pre-cloned into pDONR221 vectors with codon optimization for *E. coli* from Twist Bioscience. Before the commencement of the current research project, the synthetic constructs were cloned into

expression vectors pDEST17 and pHMGWA with N-terminal hexahistidine-tags and TEV sites. Each plasmid was transformed into *E. coli* DH5α cells for plasmid storage and *E. coli* BL21(DE3) plysS cells for protein expression.

2.2.2 Cloning Pmar3-mut

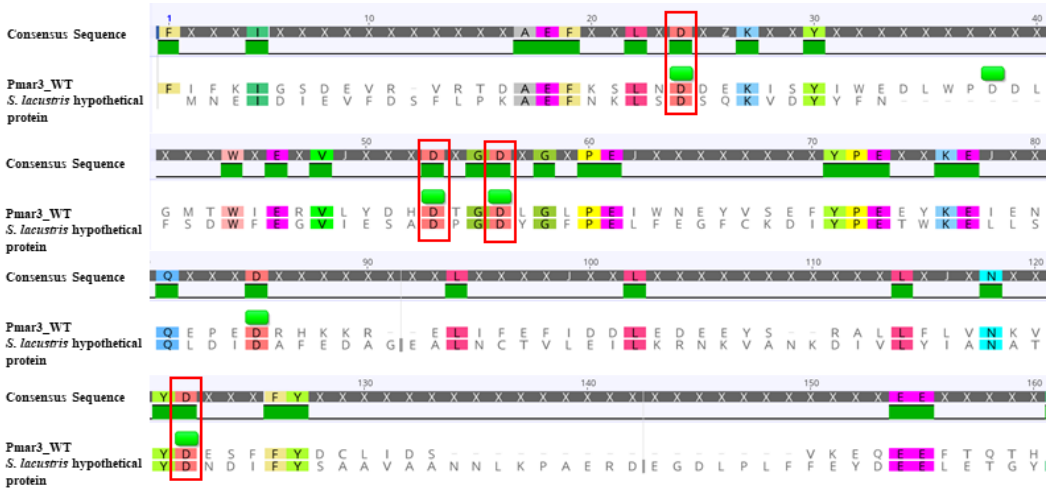


Figure 2-2: Pairwise alignment of wild-type Pmar3 and a hypothetical protein from *Synechococcus lacustris*, which returned the most sequence similarity to Pmar3 (24.1%) from other organisms. Mutations from aspartate to alanine at positions 46, 75, 78, and 140 are highlighted with red boxes.

A mutant of Pmar3 was ordered (hereafter Pmar3-mut) with four point-mutations of aspartate to alanine (positions 46, 75, 78, and 140). The Pmar3-mut gene was ordered pre-cloned into a pDONR221 vector from Twist Bioscience. The Pmar3-mut synthetic construct was sub-cloned into a pHMGWA expression vector with an N-terminal hexahistidine-tag and TEV site using the Gateway LR clonase enzyme according to the manufacturer's instructions.

The Pmar3-mut:pHMGWA plasmid was transformed by heat shock into chemically competent *E. coli* DH5α cells for plasmid storage and was plated onto Luria-Bertani (LB)-agar supplemented with 100µg ml^{-1} ampicillin. A seeder culture of the transformed cells was prepared by inoculating a single colony from a transformation plate of LB agar into 5ml of LB broth supplemented with 100µg ml^{-1} ampicillin. The seeder culture was incubated overnight at 37ºC, shaking at 180rpm. 25% glycerol stocks of the incubated seeder culture were prepared and stored at -80ºC for long-term use.

The plasmid was purified using a QIAprep® Spin Miniprep Kit (Bio-strategy) according to the manufacturer's instructions. The plasmid was transformed into *E. coli* BL21(DE3)pLysS cells for protein expression as per the methods above. A seeder culture of transformed cells was prepared by inoculating a single colony from a transformation plate of LB agar into 5ml of LB broth supplemented with 100µg ml^{-1} ampicillin and 35µg ml^{-1} chloramphenicol. The seeder culture was incubated and 50:50 (w/v) glycerol stocks were prepared as per the methods above.

2.2.3 Colony Polymerase Chain Reaction

Successful transformations were assessed by carrying out colony PCR, using colonies from transformation plates, and an MBP forward primer and T7 reverse primer. PCR reactions were set up by touching a pipette tip to a colony and dipping it into the master-mix (shown in Table 2.2), then carried out using an Eppendorf PCR machine (shown in Table 2.3) then assessed on a 1% TAE agarose gel.

Table 2.2 Colony PCR Conditions

Component	Concentration
5x Mastermix (Hotfire Pol Blend)	1x
10mM Forward Primer	0.3mM
10mM Reverse Primer	0.3mM

Table 2.3 Colony PCR Cycling Protocol

PCR Cycles	Step	Temperature	Time (min:sec)
1	Initial Denaturation	95°C	15:00
	Denature	95°C	0:30
30	Anneal	55°C	0:20
	Extend	72°C	0:30
1	Final Extension	72°C	7:00

The product of the PCR reaction was electrophoresed on a 1% TAE agarose gel to determine its size and therefore whether the transformation was carried out successfully. The 1% TAE agarose gel was made by dissolving 0.3g agarose in a flask of 30ml 1xTAE running buffer (40mM tris-acetate, 1mM EDTA). The flask was microwaved for 30 second periods and swirled gently, repeating until the agarose was fully dissolved. 3µl

SYBRSafe was added and gently swirled to mix, before pouring into a small gel caster. A gel comb was inserted, and the gel was left to set for 30 minutes. The set gel was placed into a small gel tank which was filled with 1xTAE running buffer. The assay samples were prepared for loading by pipette mixing 10µl sample with 2µl 5x DNA loading dye (25% glycerol, 0.2% bromophenol blue). 10µl 1kb Plus DNA Ladder bp was loaded into the first lane of the gel, and the following lanes were loaded with 10µl of sample per lane. Gels were imaged using an iBright imager (Invitrogen).

2.2.4 Midi-scale Expression Testing

2.2.4.1 Buffer Screening – Pmar4

Prior research included small-scale screening, which showed that Pmar4 was insoluble in all the conditions tested. Therefore, a range of lysis buffer conditions was tested to determine whether Pmar4 could be solubilized using different pH's, reducing agents, and salt concentrations. Additionally, both Pmar4-His and Pmar4-MBP were grown at 15^{o}C and 20^{o}C. Overnight cultures of Pmar4-His and Pmar4-MBP were prepared at a midi-scale by adding 5ml of 1xTB broth supplemented with ampicillin (100µgml^{-1}) and chloramphenicol (35 µgml^{-1}) to two 15ml loose-capped tubes. Each tube was inoculated with one of the Pmar4 glycerol stocks which were thawed on ice. The flasks were incubated at 37^{o}C, shaking at 180rpm overnight. Four 150ml flasks were filled with 60ml 1x concentration TB broth supplemented with ampicillin (100µgml^{-1}) and chloramphenicol (35µgml^{-1}). Each flask was inoculated with 1.5ml ONC to give 2 flasks per ONC and was put for incubation at 37^{o}C shaking at 180rpm. All four flasks were induced with 60µl 0.5M IPTG (final concentration of 0.5mM IPTG) and were transferred to 15 or 20^{o}C incubators to incubate overnight, shaking at 180rpm.

Each flask was split into 6x10ml falcon tubes which were centrifuged at 4570g at 4^{o}C for 15 minutes. The supernatants were discarded, and the pellets were resuspended with 2.5ml of relevant buffer (Table 2.4) per 10ml sample volume, and the tubes were transferred to ice.

Table 2.4: Lysis buffer compositions for resuspending Pmar4 expression trial pellets

Name	Composition
Lysis Buffer A	50mM tris pH8.0, 750mM NaCl, 5% glycerol, 10mM $MgCl_2$
Low pH	50mM MOPS pH 6.5, 500mM NaCl, 10% glycerol
High pH	50mM tris pH 9.0, 500mM NaCl, 10% glycerol
Low Salt	50mM tris pH 8.0, 50mM NaCl, 10% glycerol
High Salt	50mM tris pH 8.0, 2M NaCl, 10% glycerol
Reducing	50mM tris pH 8.0, 500mM NaCl, 10% glycerol, 1mM DTT

Samples were sonicated (1/16" tip, amplitude 5, for 3 minutes total, 2 seconds on/2 seconds off) and centrifuged at 11521g at 4°C for 15 minutes. The pellets and supernatants were processed as per methods outlined in section 2.2.4.3.

2.2.4.2 Expression and Solubility Screening – Pmar3-mut

Overnight cultures of Pmar3-mut were prepared at a midi-scale. 18x 30ml cultures were expressed as per the methods outlined in section 2.2.4.1, except falcon tubes were used. Each falcon tube was induced with IPTG (outlined in Table 2.5) and incubated overnight with shaking at 180rpm.

Table 2.5 Pmar3-mut Growth Conditions

Sample	Temperature	OD_{600} for induction	IPTG Final Concentration
1	15°C	0.3	0.5mM
2	15°C	0.3	0.75mM
3	15°C	0.3	1mM
4	15°C	0.5	0.5mM
5	15°C	0.5	0.75mM
6	15°C	0.5	1mM
7	15°C	0.9	0.5mM
8	15°C	0.9	0.75mM
9	15°C	0.9	1mM
10	20°C	0.3	0.5mM
11	20°C	0.3	0.75mM
12	20°C	0.3	1mM
13	20°C	0.5	0.5mM

Sample	Temperature	OD_{600} for induction	IPTG Final Concentration
14	20°C	0.5	0.75mM
15	20°C	0.5	1mM
16	20°C	0.9	0.5mM
17	20°C	0.9	0.75mM
18	20°C	0.9	1mM

Post-incubation, falcon tubes were centrifuged at 4570g at 4°C for 15 minutes. The supernatants were discarded, and the pellets were resuspended with 5ml of lysis buffer, and the tubes were transferred to ice. Samples were sonicated (1/16" tip, amplitude 5, 3 minutes total, 2 seconds on/2 seconds off) and centrifuged at 11500g at 4°C for 15 minutes. The pellets and supernatants were processed as per methods outlined in section 2.2.4.3.

2.2.4.3 Nickel Pulldown

The Nickel pull-down method was used for midi-scale samples to provide an early and quick indication of successful protein expression. 20µl of nickel beads per sample were washed in 1.5ml Eppendorf tubes by adding 1ml lysis buffer A, centrifuging using the Eppendorf MiniSpin® plus centrifuge for 1 minute at 1000rpm and disposing of the supernatant twice. 200µl lysate was pipette-mixed with 800µl lysis buffer and was added to the bead-containing tube for incubation in a thermomixer at 20°C, shaking at 1000rpm for 15 minutes. The tubes were briefly spun down to pellet the beads, and the supernatant fractions were collected. The beads were washed twice by adding lysis buffer and centrifuging for 1 minute at 1000rpm, and the fractions were collected. The beads, wash fractions, pellets, and supernatants were loaded onto SDS-PAGE gels (12% resolving gel, 5% stacking gel for LigW and 12% or 15% resolving gel, 5% stacking gel for Pmar3-MBP + Pmar5-MBP).

2.2.5 Large-scale Expression

This protocol was used for the production of proteins Pmar3 (with MBP tag), Pmar5 (with MBP tag), and LigW (with His-tag) as these were demonstrated to be expressed in soluble form. An overnight 120ml culture (ONC) of the protein of interest was prepared by inoculating 1xTB broth, supplemented with ampicillin (100µg/ml) and chloramphenicol (35µg/ml), and incubating at 37°C 180rpm overnight. 4x 1L flasks were filled with 1L 1x

concentration TB broth, 100µg ampicillin, and 35µg chloramphenicol. Each flask was inoculated with 30ml ONC to give 4 flasks per ONC and was put for incubation at 37^{o}C 180rpm. OD_{600} measurements were taken hourly from one flask per culture. Each set of flasks were moved to 15^{o}C to incubate overnight with shaking at 180rpm once an OD_{600} of 0.3 was reached. The flasks were left to equilibrate for 30 minutes before inducing with 0.5mM IPTG per flask. The total culture was equally poured into pots and centrifuged at 4570g at 4^{o}C for 20 minutes. The supernatant was discarded, and the pellets were collected and weighed in falcon tubes. Tubes were labelled and stored at -20^{o}C for future use.

2.2.6 Lysis of Cell Pellets for Purification

Pre-prepared frozen pellets (see section 2.2.5) were re-suspended on ice by adding 5ml lysis buffer A per 1g of pellet. The pellet was transferred to a beaker and sonicated (1/4", amplitude 3, 5 minutes total, 1 second on/1 second off). Sonicated samples were transferred to centrifuge tubes and were centrifuged at 11521g for 1 hour at 4^{o}C. The supernatant was decanted and collected for purification. The pellet was transferred to a fresh falcon tube to be prepared for loading onto SDS-PAGE gels. The supernatant was filtered through 1.2 µm, 0.45 µm, and 0.2 µm Minisart filters.

2.2.7 Purification of LigW

2.2.7.1 IMAC of LigW

A LigW-His expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6). The lysate had 10mM final concentration of imidazole added to help decrease the background expression of non-specifically bound *E. coli* proteins, before being loaded into a 50ml super-loop. The filtered supernatant was loaded onto a 5ml HisTrap HP column (GE Healthcare Life Sciences) pre-equilibrated with purification buffer A (50 mM Tris pH 8.0, 750 mM NaCl, 5% glycerol, 10 mM imidazole), connected to the ÄKTAprime plus (GE Healthcare Life Sciences) for immobilized metal affinity chromatography using a super-loop, with a flow rate of 2.5ml min^{-1}. The HisTrap HP column was washed with 5 column volumes of purification buffer A to remove any unbound proteins. Bound proteins were eluted by running a 50ml purification buffer B (50 mM Tris pH 8.0, 750 mM NaCl, 5% glycerol, 500mM imidazole) gradient from 0-

100% at 2.5ml min^{-1}. The crude insoluble, crude soluble, flow-through, and eluted fractions were loaded onto SDS-PAGE gels (see section 2.2.11.1).

2.2.7.2 De-salting and TEV Cleavage

Protein-containing fractions were pooled and loaded onto a pre-equilibrated HiPrep™ 26/10 de-salting column (GE Healthcare Life Sciences) connected to the ÄKTAprime plus (GE Healthcare Life Sciences) at 4.0ml min^{-1} with TEV cleavage buffer (50mM tris pH 8.0, 100mM NaCl, 5% glycerol, 1mM DTT, 0.5mM EDTA. Hereafter buffer C). Fractions from the A280 peak on the chromatogram were pooled into falcon tubes, and a sub-sample for loading onto SDS-PAGE was collected. The pooled de-salted fractions had 1mg TEV protease added. TEV protease was thawed on ice before addition to the sample. A pre-cleavage sub-sample for loading onto SDS-PAGE was also collected. The sample was incubated on a tilting mixer at 4°C overnight.

2.2.7.3 Reverse IMAC of LigW

After incubation, the sample was divided into Eppendorf tubes and centrifuged in a Thermo Scientific™ Fresco™ 21 Microcentrifuge at 13,500rpm at 4°C for 10 minutes to remove any precipitated protein. The supernatant was collected and pooled to load onto a 5ml HisTrap HP column (GE Healthcare Life Sciences) pre-equilibrated with buffer C, connected to the ÄKTAprime plus (GE Healthcare Life Sciences) for reverse immobilized metal affinity chromatography using a super-loop, with a flow rate of 2.5ml min^{-1}. The HisTrap HP column was washed with 5 column volumes of buffer C to elute the de-tagged protein. The His-tag, TEV protease, and residual contaminating proteins were then eluted by running a 25ml purification buffer B gradient to 0-100% at 2.5ml min^{-1}. The desalted pooled sample, pre-cleavage sample, overnight cleaved sample insoluble and insoluble samples, reverse His-trap flow-through and reverse His-trap elute were loaded onto SDS-PAGE gels (see section 2.2.11.1).

Successfully purified flow-through fractions were pooled and concentrated to a final volume of 5ml using an Amicon® Ultra-4 10000 MWCO Centrifugal Filter (Merck), pre-equilibrated with buffer C. The fractions were loaded into the concentrator and were centrifuged repeatedly at 3500rpm in 5-minute increments at 4°C using a Heraeus Multifuge 3 S-R Refrigerated Centrifuge with a swinging bucket rotor until a volume of less than 5ml of concentrated protein was reached.

2.2.7.4 Size Exclusion of LigW

The concentrated flow-through fractions were loaded onto a Superdex75 16/60 gel filtration column (GE Healthcare Life Sciences) pre-equilibrated with buffer C, connected to the ÄKTAprime plus (GE Healthcare Life Sciences), for size exclusion. The column was washed with 1 column volume of buffer C at 0.5ml min^{-1}. The pooled reverse IMAC sample, concentrated reverse IMAC sample, and fractions from each size-exclusion peak were loaded onto SDS-PAGE gels (see section 2.2.11.1).

2.2.7.5 Storage of Purified LigW

Purified protein from size exclusion was concentrated to 5.54mg/ml to be used for future assays, using an Amicon® Ultra-4 10000 MWCO Centrifugal Filter (Merck). Approximately 1ml pure protein was stored using two methods. 1) 500µl of protein was diluted with an equal volume of glycerol and was split into 100µl aliquots in PCR tubes. 2) The remaining 500µl was directly transferred into PCR tubes in 100µl aliquots and was flash-frozen using liquid nitrogen. All PCR tubes were labelled and stored at -80°C. The final concentration of purified protein was quantified from A280 measurements taken using a Nanodrop™.

2.2.8 Optimization of Purification of Pmar3

Optimization of the purification of Pmar3 was required to achieve a high yield of concentrated, pure protein. Adjustments were made to the purification method to help increase MBP-tag cleavage, de-tagged Pmar3 yield, and concentration of Pmar3.

2.2.8.1 IMAC Purification with Dialysis

A Pmar3-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6). IMAC of Pmar3-MBP was carried out as per section 2.2.7.1, except a 1ml His-trap was used with a flow rate of 1ml min^{-1}. De-salting of Pmar3-MBP was carried out after the initial IMAC using dialysis methods to try to remove some contaminants. The 5ml lysate from protein-containing fractions was placed in a dialysis bag with a 6-8kDa molecular weight cut-off, which was incubated in 500ml buffer C on a stir-plate, overnight at 4°C. The next day, 0.4mg TEV protease was added to the dialysed

sample and was left to incubate for 6 hours at 4°C. Sub-samples were taken every 2 hours to load onto SDS-PAGE to observe TEV cleavage progression. A reverse IMAC was run after 6 hours of incubation with TEV according to section 2.2.7.3.

2.2.8.2 IMAC Purification with De-salting

A Pmar3-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6). Purification of Pmar3-MBP was carried out as per section 2.2.7.1, including buffer exchange via desalting with the HiPrep 26/10. The only modification to this method was that the post-TEV incubated sample was aliquoted into Eppendorf tubes, which were centrifuged in a Thermo Scientific™ Fresco™ 21 Microcentrifuge at 13,500rpm for 10 minutes at 4°C to pellet precipitants, and only the lysate was loaded on to reverse IMAC.

2.2.8.3 TEV Cleavage Optimization

A Pmar3-MBP expression culture pellet was prepared for IMAC from a 250ml culture (see sections 2.2.5 and 2.2.6). IMAC and de-salting of Pmar3-MBP were carried out as per sections 2.2.7.1 and 2.2.7.2. TEV cleavage samples were set up in 40µl reaction volumes, with 1:5 and 1:10 TEV protease:Pmar3-MBP ratios, 4°C, RT, and 37°C incubation temperatures, and 30 minutes, 1 hour and overnight timepoints (see Appendix B), to identify which conditions provide the best yield of cleaved Pmar3.

2.2.8.4 Optimized Purification of Pmar3

Pmar3 was purified using the methods outlined in section 2.2.8.2. The elution peak from the reverse IMAC was processed through an additional second TEV cleavage step at 4°C overnight and a second reverse IMAC. These steps were repeated to help increase the cleavage of the MBP-tag from Pmar3-MBP and increase the yield of cleaved Pmar3.

2.2.9 Re-folding and IMAC Purification of Pmar4 with Urea

A Pmar4-His expression culture pellet (see section 2.2.5, except 500ml was grown) was prepared for purification. The pellet was re-suspended on ice by adding 20ml lysis buffer (50mM tris pH8.0, 750mM NaCl, 5% glycerol, 10mM imidazole). The pellet was transferred to a beaker and sonicated (1/4”, amplitude 12, 3 minutes total, 1 second on/1

second off). Sonicated samples were transferred to small pots and were centrifuged at 11500g for 1 hour at 4ºC. The supernatant was decanted and collected for SDS-PAGE. The pellet was transferred to a fresh falcon tube, re-suspended in 20ml wash buffer (lysis buffer + 1% Triton-X), and centrifuged (4570g, 20min, 4ºC). The supernatant was decanted and collected for SDS-PAGE. The pellet was dissolved in 5ml denaturation buffer (lysis buffer + 8M urea) and centrifuged (4570g, 1hr, 4ºC). The supernatant was collected in a syringe and was dropwise diluted into a beaker of 45ml (10x supernatant volume) lysis buffer to give a final 0.8M urea concentration, and to re-fold Pmar4-His. The supernatant was filtered through 1.2 µm, 0.45 µm, and 0.2 µm Minisart filters.

IMAC of Pmar4-His was run as per section 2.2.7.1, except the ÄKTA purifier at RT was used to prevent the crystallisation of urea. Protein-containing fractions were pooled, concentrated (see section 2.2.7.3), and size excluded as per section 2.2.7.4, except a BioRad NGC system at RT was used.

2.2.10 Optimization of Purification of Pmar5

2.2.10.1 IMAC Purifications at 4ºC

A Pmar5-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6), which was carried out using the steps outlined in section 2.2.7. The methods were adjusted to include a 30-minute RT TEV protease incubation step before incubation overnight at 4ºC. The reverse IMAC was repeated with the elution fractions of the first reverse IMAC, as Pmar5 did not appear in the flow-through. The repeated reverse IMAC included re-loading the supernatant, a 25ml buffer C wash, 50ml buffer A wash, and a 50ml gradient from 0 to 100% buffer B. Size exclusion of the flow-through was attempted.

A second purification was carried out at 4ºC using the methods stated above. However, incubation with TEV protease was carried out at 4ºC overnight only and resulted in visible precipitation of the protein. A reverse IMAC was attempted using the supernatant, and size exclusion was not completed.

A third purification was carried out at 4ºC using the methods stated above. Incubation with TEV protease was carried out in Eppendorf tubes on a rotating wheel at 4ºC overnight. The reverse IMAC included loading the pooled supernatant, a 25ml buffer C

wash, 50ml buffer A wash, and a 50ml gradient from 0 to 100% buffer B. Size exclusion was not completed.

2.2.10.2 IMAC Purification at Room Temperature

A Pmar5-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6), except 1µl Benzonase per 10ml lysate was added before sonication to decrease DNA contamination. IMAC was carried out as per sections 2.2.7.1 and 2.2.7.2, except the room-temperature BioRad NGC system was used for the entirety of the purification. The pooled de-salted fractions were incubated with TEV protease for 1 hour at RT, then split into Eppendorf tubes for incubation at 4^{o}C overnight. Reverse IMAC was skipped, and the post-TEV samples were pooled and concentrated (see section 2.2.7.3) to be size excluded (see section 2.2.7.4). Protein-containing fractions from size exclusion were pooled and incubated with TEV protease for 1hr at 37^{o}C, then 4^{o}C for 2 days to attempt further de-tagging of Pmar5-MBP. A reverse IMAC was performed manually using a syringe, where the protein was loaded, followed by a 25ml buffer C wash, and a 25ml buffer B wash.

2.2.10.3 IMAC Purification with Glutathione Redox Buffer

A Pmar5-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6), which was carried out using the steps outlined in section 2.2.7. The methods were adjusted to use a different dialysis buffer; glutathione redox buffer (500mM tris pH 8.0, 100mM NaCl, 5% glycerol, 0.5mM EDTA, 0.3:0.03mM glutathione redox) and was completed at RT using the BioRad NGC system. Size exclusion was not completed.

2.2.10.4 Modified IMAC Purification at 4^{o}C

A Pmar5-MBP expression culture pellet (see section 2.2.5) was prepared for purification (see section 2.2.6). Modifications included two post-sonication centrifuge steps. The first centrifugation was at 11521g, 4^{o}C, 45 minutes. The lysate was decanted and collected to be centrifuged for a further 20 minutes. IMAC was carried out as per section 2.2.7.1. Protein-containing fractions from the IMAC were pooled and concentrated to 4ml final volume for size exclusion (see section 2.2.7.4). The protein-containing fractions were pooled, and half of the protein-containing fractions (approximately 7ml) were incubated

with 0.84mg TEV protease overnight at 4°C. Reverse IMAC was carried out the following day (see section 2.2.7.3).

2.2.11 Protein Analysis Methods

2.2.11.1 SDS-PAGE

A Hoeffer five gel multi-gel caster was used to cast SDS-PAGE. 12% or 15% resolving gel solution was made (Table 2.6) and poured into the gel plates, approximately 2cm from the top. Each gel was overlaid with 2ml isopropanol and was left to set at room temperature for 40 minutes. Once set, the isopropanol was decanted off the gels and rinsed with water, before overlaying each gel with approximately 2ml 5% stacking gel (shown in Table 2.6). Gel combs were inserted into each gel, and the gels were left to set at room temperature for 30 minutes.

Table 2.6 Recipe for SDS-PAGE Gels

	12% Resolving Gel	15% Resolving Gel	5% Stacking Gel
ddH_2O	10.05ml	7.05ml	8.5ml
30% Acrylamide	12.00ml	15.00ml	2.125ml
Resolving Buffer (1.5M Tris, pH 8.8)	7.50ml	7.50ml	-
Stacking Buffer (1.0M Tris, pH 6.8)	-	-	1.6ml
10% SDS	0.30ml	0.30ml	0.125ml
10% APS	0.15ml	0.15ml	0.063ml
TEMED	0.015ml	0.015ml	0.0063ml

Samples were prepared for loading onto SDS-PAGE by mixing 5µl 4xSDS loading dye (250mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 10% mercaptoethanol, 0.025% bromophenol blue) with 15µl sample in PCR tubes and boiling at 95°C for 10 minutes. 10µl Precision Plus Protein™ standard was loaded in the first lane of the gel, and the following lanes were loaded with a 10µl sample per lane. The gels were run at 30mA using 1x TG-SDS running buffer (25mM Tris, 250mM glycine, 0.1% SDS).

Once run, the gels were post-stained with Fairbanks A stain (0.05% Coomassie Blue, 25% isopropanol, 10% acetic acid). Fairbanks A stain was poured over the gels, boiled by

microwaving for 30 seconds, and was left to stain by shaking for 5 minutes at room temperature. Background staining was then removed by de-staining with Fairbanks de-staining solution (10% acetic acid), which was also boiled for 30 seconds and left to shake for 5 minutes. Gels were imaged using an iBright imager (Invitrogen).

2.2.11.2 Native-PAGE

Native-PAGE gels were poured as per the same method for SDS-PAGE gels, except with 12% acrylamide (i.e., without the 5% stacking layer) and without the addition of SDS. Samples were prepared for loading by mixing 4µl 5x native loading dye (300mM Tris-HCl pH 6.8, 50% glycerol, 0.05% bromophenol blue) with 16µl sample in PCR tubes. 10µl Precision Plus Protein™ standard was loaded in the first lane of the gel, and the following lanes were loaded with a 10µl sample per lane. The gels were run at 30mA using 1xTris-Glycine running buffer (25mM Tris, 250mM glycine). Gels were stained and visualised as per section 2.2.11.1.

2.2.12 Mass Spectrometry

Liquid chromatography–mass spectrometry (LC-MS/MS) was used to confirm the protein identity of the final purified product of the Pmar3 and Pmar5 purifications. This service was carried out by MS^3 Solutions according to company protocols. In summary, protein bands of interest were excised from SDS-PAGE gels and were processed using an in-gel tryptic digestion method. The tryptic peptides were extracted and detected by LC-MS/MS. The spectral data was searched against a database of *E. coli* protein sequences and against the Pmar3 and Pmar5 sequences.

2.3 Results and Discussion

2.3.1 Pmar4 Expression Trial

Before the commencement of this research, attempts were made to express Pmar4-His and Pmar4-MBP but resulted in inclusion bodies. This led to a midi-scale expression trial using six different buffers (lysis buffer A, low pH, high pH, low salt, high salt, and reducing) in place of lysis buffer A. This was done to see if using different buffers to lyse Pmar4 containing cells could help stabilize the protein better than lysis buffer A, which was used for both Pmar3 and Pmar5. A nickel pull-down was performed for all twenty-

four constructs, and the pellets, supernatants, washes, and beads were loaded on to 12% SDS-PAGE gels.

Figure 2-3 shows that Pmar4-His was consistently observed in the insoluble fractions but did not show any solubilization or nickel bead binding, despite the use of different lysis buffers. There was no expression of Pmar4-MBP observed in any of the constructs at any point. This suggests that Pmar4 may need to be re-transformed into *E. coli* BL21 plysS cells using a different fusion tag in the future to increase the solubility of Pmar4. Alternatively, different growth and lysis methods could be attempted in the future with Pmar4-His, as insoluble expression was observed.

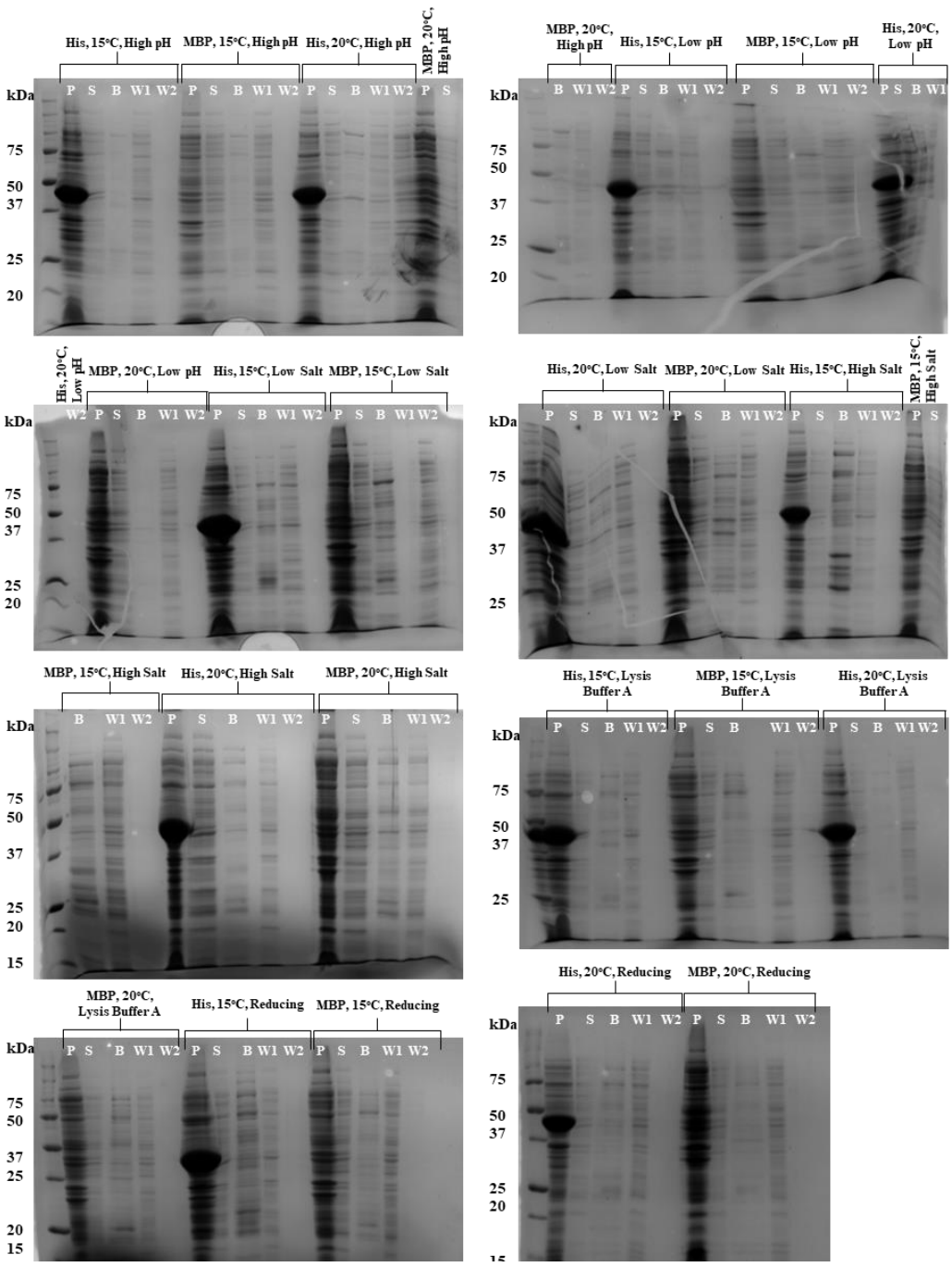


Figure 2-3: 12% SDS-PAGE gels of the Expression Trial and Nickel Pulldowns for Pmar4-His (50kDa) and Pmar4-MBP (91kDa). P = Pellet, S = Supernatant, B = Ni Beads, W1 = Wash 1, W2 = Wash 2.

2.3.2 Pmar3(mut) Transformation and Expression Trial

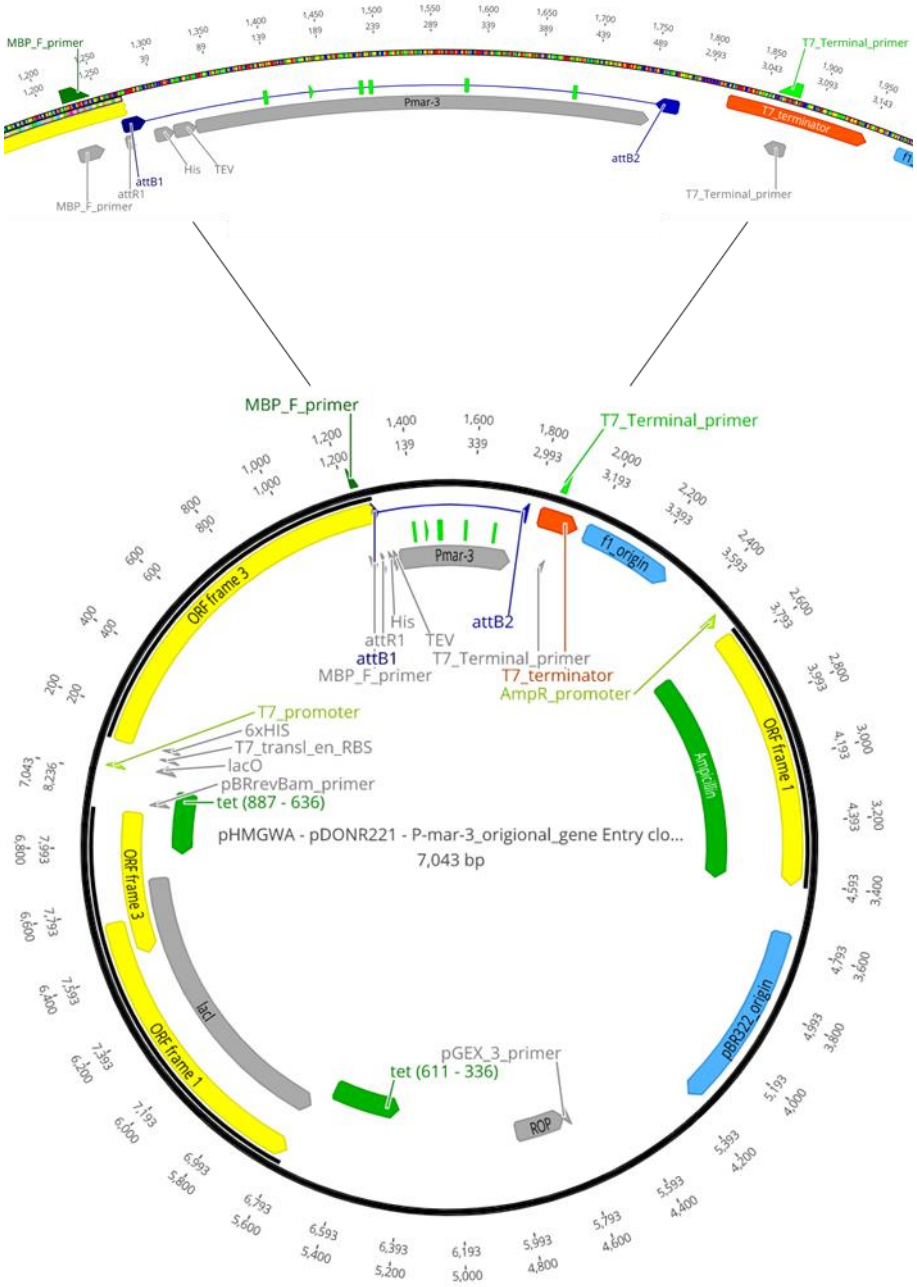


Figure 2-4: pHMGWA-pDONR221-Pmar3 expression clone. Pmar3-mut was designed with an MBP forward primer, T7 terminal primer, and flanking attB sites, allowing the PCR product to be cloned by pDONR221 to generate an entry clone. An LR reaction of Pmar3-mut-pDONR221 was carried out to express Pmar3-mut in a pHMGWA destination vector. The expression clone was then transformed in to *E. coli* DH5α and BL21plysS cells.

The Pmar3 mutant where four aspartic acid residues were mutated into alanine was subcloned into pHMGWA and then successfully transformed into *E. coli* BL21(DE3)pLysS cells. This was confirmed by colony PCR, where an MBP forward

primer and T7 reverse primer were used to amplify the DNA from the transformed DH5α and BL21(DE3)plysS cells. Figure 2-5 shows a 1% TAE agarose gel of the PCR products.

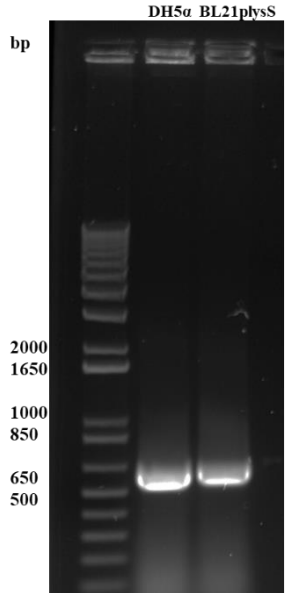


Figure 2-5: 1% TAE Agarose gel of Pmar3-mut PCR products.

The PCR product size for the successful transformation of Pmar3-mut into BL21plysS cells was predicted, using Geneious Prime software, to be 660bp. Figure 2-5 shows that the PCR product produced was the correct size and therefore, Pmar3-mut was successfully transformed.

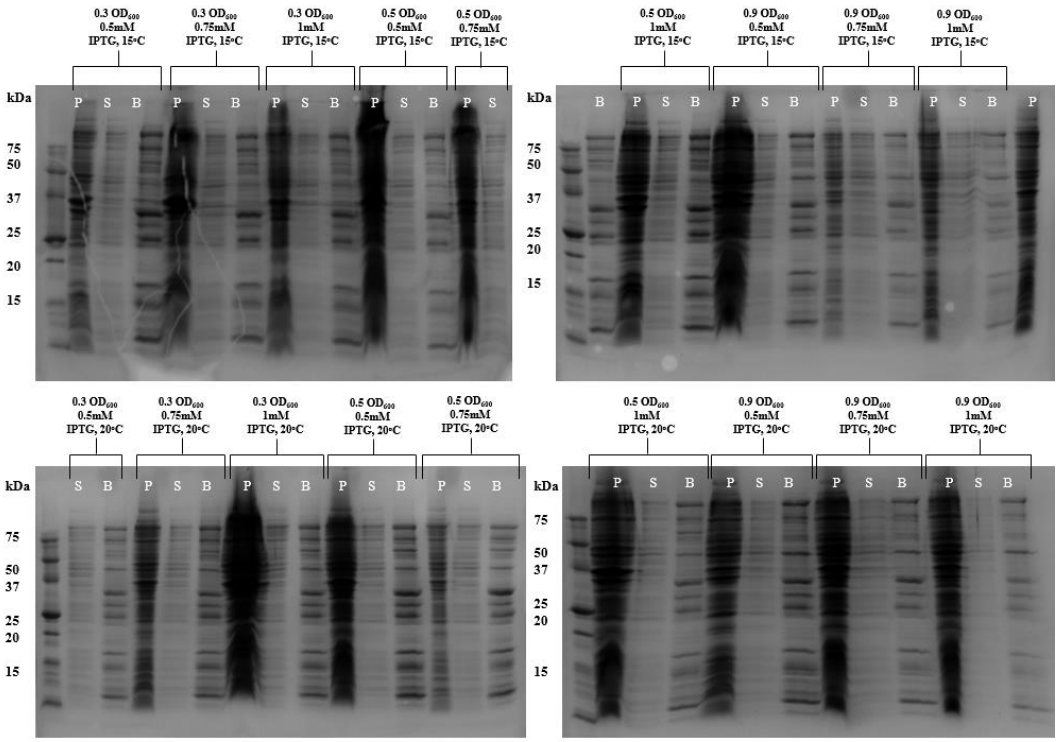


Figure 2-6: 15% SDS-PAGE gels of the Expression Trial and Nickel Pulldowns for Pmar3-mut-MBP (64kDa). P = Pellet, S = Supernatant, B = Ni Beads.

An expression trial of Pmar3-mut including 18 conditions was carried out and 18 nickel pulldowns were performed. The conditions tested included the conditions that had been proven to be successful for Pmar3-MBP, as well as a range of temperatures (15^{o}C and 20^{o}C), and OD_{600} levels (0.3, 0.5, and 0.9) before inducing with different IPTG concentrations (0.5, 0.75, and 1mM). Different OD_{600} levels and IPTG concentrations were used to test how induction during different bacterial growth phases may affect protein expression. Figure 2-6 shows that Pmar3-mut was not successfully expressed in the pellet or supernatant, and therefore did not bind to the nickel beads. The pellets appeared to contain a large amount of cellular material such as non-specific proteins, which make it difficult to resolve the bands, and it is possible that insoluble expression of the mutant could be hidden by these. Further expression trials are required to identify favourable conditions for the expression of Pmar3-mut. In these future trials, DNase will be added to help degrade contaminating DNA, and the pellets will be diluted before loading on to SDS-PAGE to help resolve protein bands.

2.3.3 Purification of LigW

To understand how LigW functions and how it interacts with other proteins, purification of stable, soluble LigW was required. LigW had been purified successfully in previous

work before the commencement of this project, and a new purification was conducted to provide protein for the present project. LigW has previously been purified and kept stable at cold temperatures, so purification was carried out at 4°C to increase stability. LigW was successfully purified by IMAC, reverse IMAC, and size exclusion.

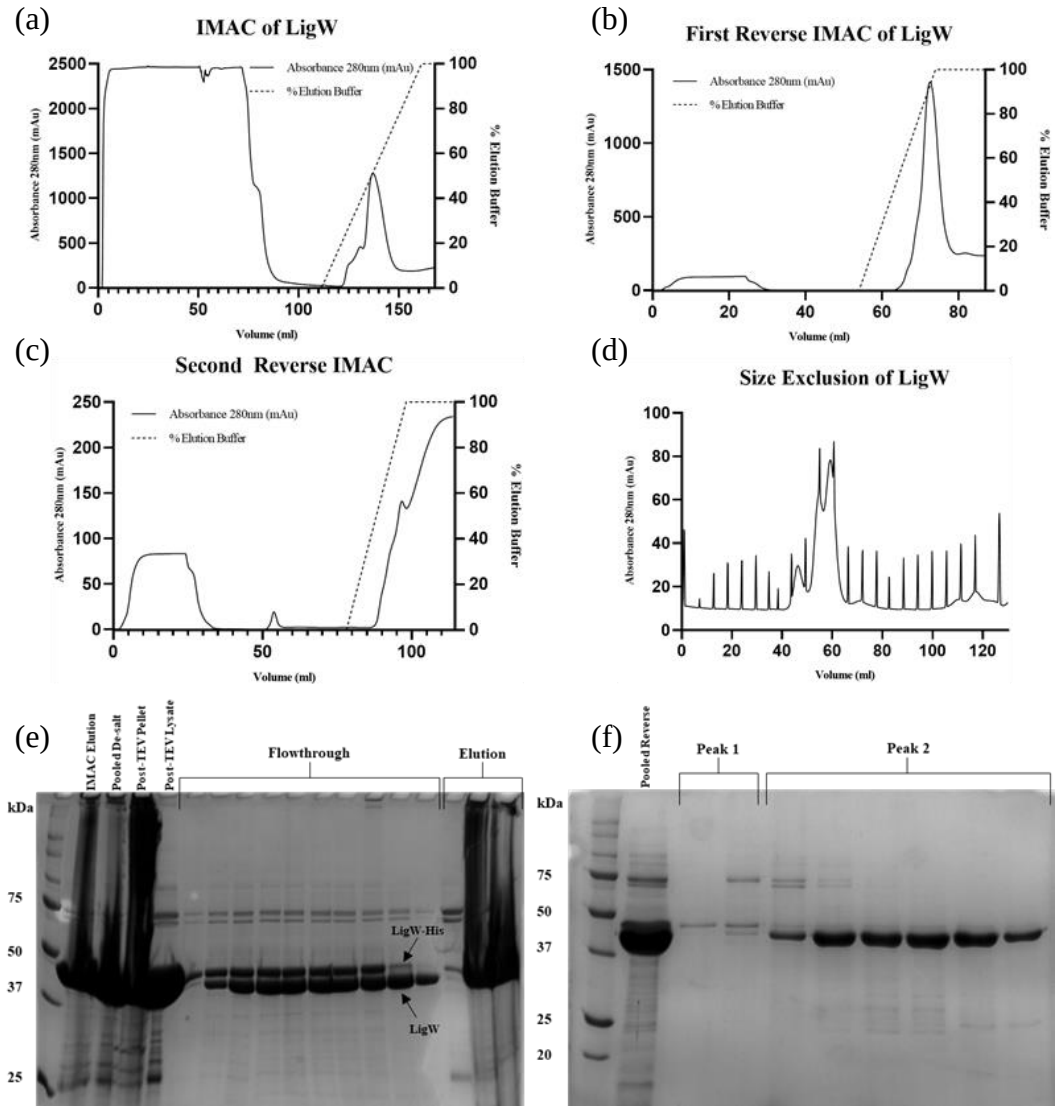


Figure 2-7: IMAC purification and protein analysis of LigW. (a) Chromatogram from IMAC of LigW show the UV absorbance (280nm) and elution of LigW-His at 40-50% buffer B. (b, c) Chromatograms from sequential reverse IMAC's of LigW (b) excluding and (c) including a buffer A wash step to remove contaminating proteins. Chromatograms show UV absorbance (280nm) of LigW in the flowthrough, and the His-tag and other contaminating proteins in the elution peak. (d) Chromatogram showing the UV absorbance (280nm) of LigW (peak 2) and contaminating proteins (peaks 1, 3 and 4). Spikes in the chromatogram are due to electronic noise from the purification instrument. (e) 12% SDS-PAGE gel showing the successful de-tagging of LigW using TEV protease (flowthrough) and separation from remaining LigW-His and contaminating proteins (elution). (f) 12% SDS-PAGE gel showing the size exclusion of LigW (43kDa) from remaining LigW-His (47kDa).

The IMAC chromatogram shown in Figure 2-7 (a) shows that LigW-His eluted at approximately 40-50% buffer B, which corresponds to 200-250mM imidazole. The flowthrough and first 40% of the buffer B gradient contained unbound, non-specific proteins, which continued to be removed with each step of the purification method. The protein-containing fractions (from 40-50% buffer B) were collected, pooled together, and run through a de-salting column using buffer C (see Appendix C), before being incubated with 1mg TEV protease at 4^{o}C overnight. Visible precipitation was observed in the sample after overnight incubation, so the lysate was separated from precipitants by centrifugation and was used for the reverse IMAC. The chromatogram shown in Figure 2-7 (b) corresponds to the first reverse IMAC run of the de-tagged LigW lysate. The SDS-PAGE gel in Figure 2-7 (e) shows that a mixed population of LigW (43kDa) and LigW-His (47kDa) was present in the flowthrough, and the eluted peak contained bound LigW-His and non-specific proteins. The mixed population in the flowthrough suggests that some LigW-His was unable to bind to the Ni column. The flowthrough was collected, pooled together, and run through a second reverse IMAC to attempt to separate the LigW and LigW-His populations. Figure 2-7 (c) shows the chromatogram of the second reverse IMAC. This included a buffer A wash step (from 39-79ml total volume) to help recover any residual de-tagged LigW that may have non-specifically bound to the Ni column under low-salt conditions created by buffer C. The flowthrough from the second reverse IMAC was collected, pooled together, and concentrated to give a total volume of 3ml which was used for size exclusion. Figure 2-7 (d and f) shows that LigW (peak 2) was successfully separated from LigW-His (peak 1) and non-specific proteins (peaks 3 and 4, see Appendix C). Peak 2 (55-63ml run volume) was concentrated to a final concentration of 5.54mg/ml and was stored for future use in characterization assays.

2.3.4 Purification of Pmar3

Table 2.7: Summary of Purification Specifications of Pmar3-MBP

Purification No.	Culture Volume (cell pellet weight)	Lysis Method	TEV Cleavage Method	Reverse IMAC	Second TEV cleavage	Second reverse IMAC	Final Concentration	Uses
1	2L (25.46g)	127.3ml LBA. Sonicated: ¼", Amp 2, 5min, 1s on/1s off. 2.8ml buffer B added to lysate.	1ml TEV added. 4°C overnight.	0-100% 20ml buffer B gradient.	Reverse IMAC flowthrough was split. 500µl TEV was added to one half. 500µl TEV, 50mM $MgCl_2$, and 0.8µl benzonase were added to the second half. 1hr at RT, then 4°C overnight.	Ran separately for lysate with and without benzonase. 0-100% 20ml buffer B gradients.	0.44mg/ml	LC-MS.
2	1L (15.77g)	78.85ml LBA. Sonicated: ¼", Amp 3, 5min, 1s on/1s off.	1ml TEV added. 30min at RT then 4°C. overnight.	0-100% 20ml buffer B gradient.	-	Repeated with a 0-100% 15ml buffer B gradient.	0.39mg/ml	Analytical SEC, topoisomerase, and thermal-shift assays. Nuclease time series assays.
3	2L (28.8g)	144ml LBA. Sonicated: ¼", Amp 3, 5min, 1s on/1s off.	2ml TEV added. 4°C overnight.	25ml buffer A wash. 0-100% 20ml buffer B gradient.[1]	Reverse IMAC elution peak (12.5ml) was incubated with TEV at 4°C overnight.	0-100% 20ml buffer B gradient.[2]	1-P3 = 3.54mg/ml 2-P3 = 1.86mg/ml	Nuclease assays: batch comparison, dilution series. Electrophoretic mobility shift and coupled ligation assays, pH stability thermal-shift assay.

[1] Flowthrough was concentrated to a final volume of 4ml and run through size exclusion. Size exclusion product was concentrated to give 1-P3.
[2] Flowthrough was concentrated to a final volume of 4ml and run through size exclusion. Size exclusion product was concentrated to give 2-P3.

Attempts to purify Pmar3 at a large-scale had been made before the commencement of this master's project, yet none were successful. This meant that optimization of the purification protocol to obtain stable, soluble, and active Pmar3 was required (summarized in Table 2.7). Pmar3 was successfully purified using IMAC, reverse IMAC, and size exclusion at 4^{o}C.

Several attempts were made to obtain sufficient Pmar3 for characterization experiments. The first large-scale purification yielded sufficient pure protein for liquid chromatography-mass spectrometry (LC-MS). The details for this purification are shown in Appendix C.

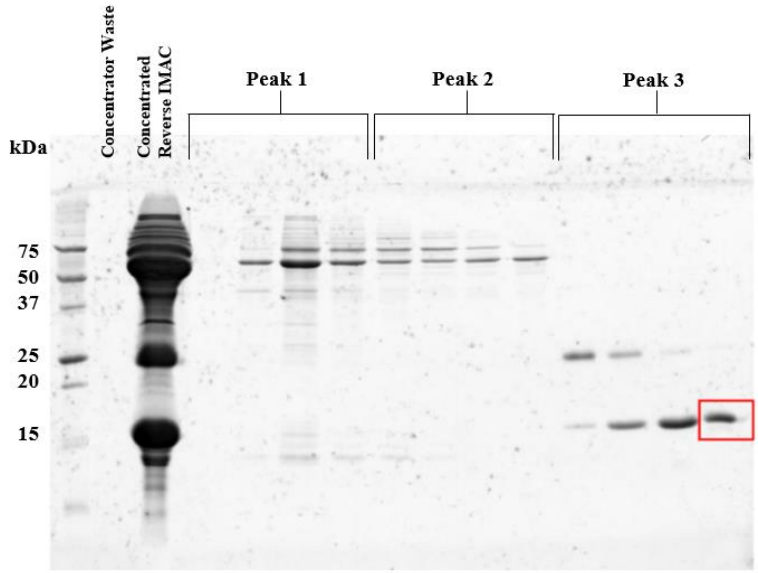


Figure 2-8: 12% SDS-PAGE gel for the size exclusion of Pmar3. The highlighted band at approximately 18kDa was excised for LC-MS. Peaks 1 and 2 show Pmar3-MBP present at 64kDa.

The protein obtained from attempting to purify Pmar3 initially was confirmed to be Pmar3 by carrying out LC-MS. The spectra data obtained for the protein band in lane 14 of the SDS-PAGE gel shown in Figure 2-8 matched the amino acid sequence for Pmar3. Given this confirmation, a second attempt was made at purifying Pmar3 from a Pmar3-MBP culture. Purification of Pmar3 was repeated a further two times to try and increase the yield and purity of the final protein so it could be characterized as accurately and extensively as possible.

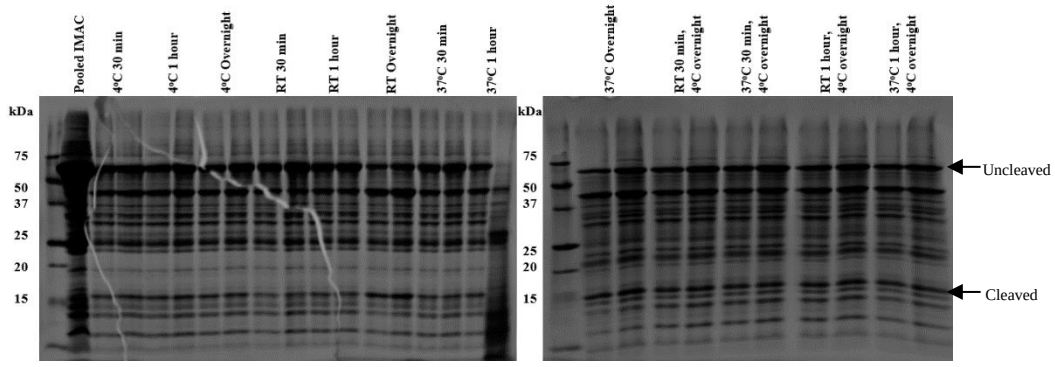


Figure 2-9: 15% SDS-PAGE gel of 13 different TEV cleavage conditions. Ratios of 1:5 then 1:10 TEV Protease:Pmar3-MBP were incubated under each set of conditions. Pmar3-MBP/uncleaved (64kDa) and Pmar3/cleaved (17kDa).

To combat obtaining a low yield of Pmar3 from a large volume of Pmar3-MBP culture, a TEV cleavage assay was carried out to determine which incubation conditions of Pmar3-MBP with TEV protease would give the highest yield of cleaved Pmar3 product. The Pmar3-MBP used in this assay was processed through an IMAC and de-salt step before TEV incubation. Protein analysis by SDS-PAGE in Figure 2-9 shows consistent MBP-tag cleavage across all 26 trialled conditions. Despite processing the Pmar3-MBP lysate through IMAC and de-salt steps before incubation with TEV protease, the IMAC elution peak shown in lane 1 appears to contain many contaminating non-specific proteins. The contaminants may have interfered with the access of TEV protease to the TEV cleavage site on Pmar3-MBP. It was concluded that the next purification of Pmar3 from Pmar3-MBP would be carried out using incubation conditions of 4^{o}C overnight, as this was used successfully with LigW-His. Future steps to optimize the yield of Pmar3 could include a repetition of this experiment.

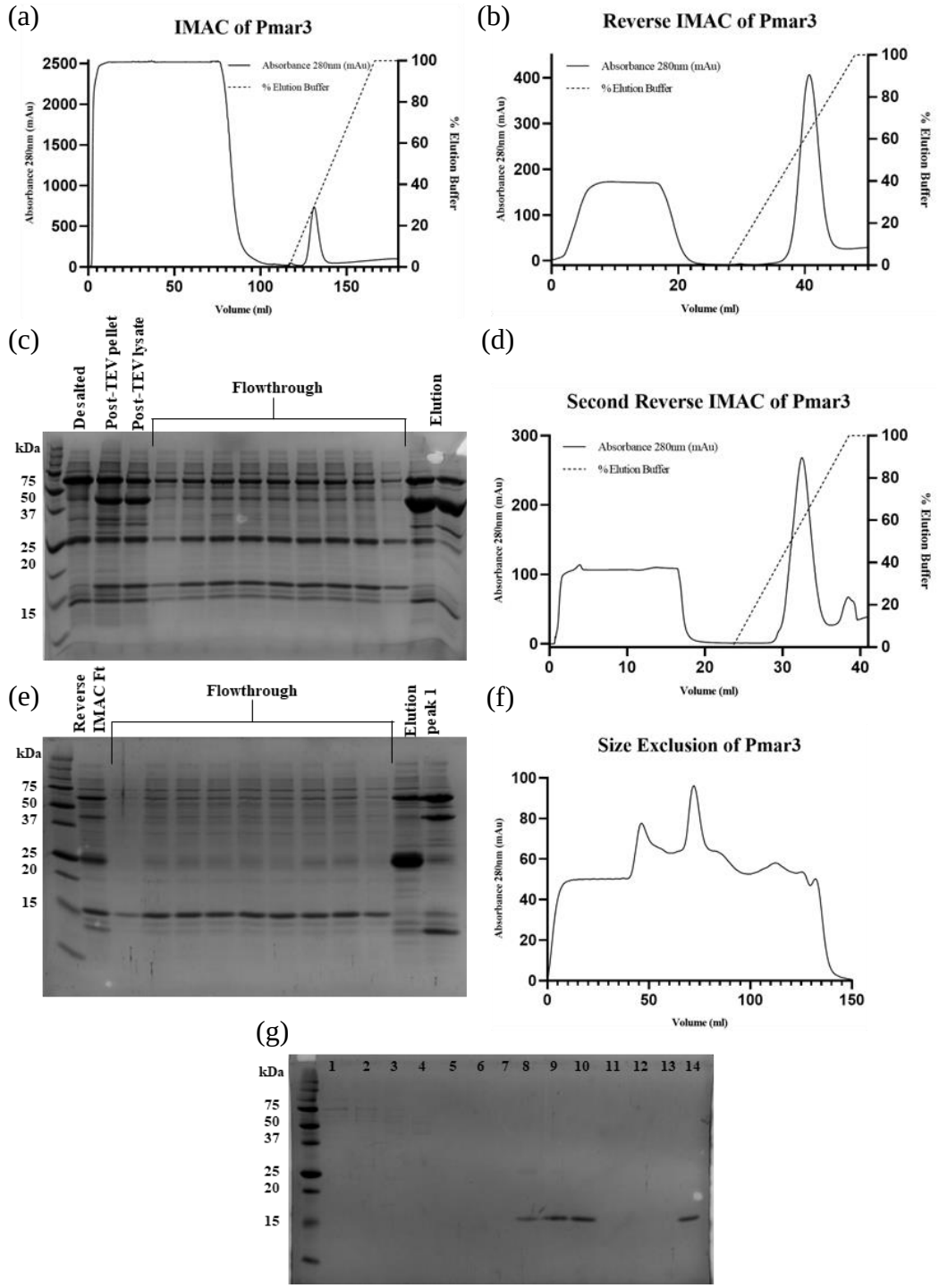


Figure 2-10: Second IMAC purification and protein analysis of Pmar3. (a) Chromatogram from IMAC of Pmar3 showing the UV absorbance (280nm) and elution of Pmar3-MBP at 17-40% buffer B. (b) Chromatogram from the first reverse IMAC of Pmar3, showing the UV absorbance of Pmar3 in the flowthrough, and the MBP-tag, Pmar5-MBP and other contaminating proteins in both the flowthrough and elution peaks. (c) 15% SDS-PAGE gel showing a low yield of isolated Pmar3 and the retention of non-specific proteins in the flowthrough. (d) Chromatogram of the second reverse IMAC showing the UV absorbance of Pmar3 in the flowthrough, and the MBP-tag, Pmar3-MBP and most other contaminating proteins in the elution peak. (e) 15% SDS-PAGE gel showing the separation of Pmar3 (17kDa) from Pmar3-MBP (64kDa). Bands are marked with arrows. (f) Chromatogram showing the UV absorbance of Pmar3 (Pmar3-containing peak is indistinguishable due to instrument malfunction). (g) 15% SDS-PAGE gel showing the isolation of Pmar3 in lanes 8-10 and 14.

A second purification of Pmar3-MBP was attempted. The IMAC chromatogram shown in Figure 2-10 (a) shows that Pmar3-MBP eluted at approximately 17-40% buffer B, which corresponds to 85-200mM imidazole. The flowthrough contained unbound, non-specific *E. coli* proteins. The Pmar3-MBP-containing fractions (from 124-138ml total run volume) were collected, pooled together, and run through a de-salting column using buffer C (see Appendix C), before being incubated with 1mg TEV protease at RT for 30 minutes, then 4^{o}C overnight. Some Pmar3 appeared to precipitate overnight, so the lysate was separated from precipitants by centrifugation and was used for the reverse IMAC. A reverse IMAC of the lysate was carried out unsuccessfully. The SDS-PAGE gel in Figure 2-10 (c) shows a mixed population present in the lysate of tagged and de-tagged Pmar3, suggesting that the TEV cleavage step needed to be optimized to increase the yield of de-tagged Pmar3 in future purifications. A comparison of the flowthrough fractions to the elution fractions shows that Pmar3 did not successfully separate from non-specific proteins, MBP-tags, and any remaining Pmar3-MBP. The flowthrough was collected and pooled together to repeat the reverse IMAC with a new, smaller 1ml Ni column, and a shorter 15ml elution gradient. A new column was used to help increase the affinity of unwanted proteins to the column so Pmar3 would become better separated from contaminants in the flowthrough. The absorbance in the second reverse IMAC, shown in the chromatogram in Figure 2-10 (d), is lower than the absorbance in the first reverse IMAC, shown in the chromatogram in Figure 2-10 (b). This is to be expected as some contaminants, Pmar3-MBP, MBP-tag, and TEV protease were eluted in the first reverse IMAC. Figure 2-10 (e) shows that Pmar3 was mostly retained in the flowthrough. The elution peak contained contaminants and tagged Pmar3 that bound to the Ni column, yet some contaminants remained in the flowthrough. The flowthrough (1.5-16.5ml total run volume) was pooled together and concentrated to give a total volume of 5ml, which was used for size exclusion. The chromatogram for size exclusion, shown in Figure 2-10 (f), shows 2 major peaks with no other defined separated peaks, suggesting that Pmar3 and any contaminants were not able to be easily separated by size. It cannot be concluded which peak contained Pmar3 and if it was isolated from contaminants, as an instrument error occurred that caused fractions to be combined randomly. Protein analysis by SDS-PAGE of each collection tube was carried out to check if Pmar3 was salvageable from any of the combined fractions. Figure 2-10 (g) shows that four fractions (lanes 8-10 and 14) contained Pmar3 (17kDa), while some other fractions contained traces of contaminant bands. The four fractions that appeared to contain Pmar3 were pooled and concentrated to a final concentration of 0.39mg/ml and were stored for use in characterization assays.

The third purification of Pmar3-MBP is shown in Appendix C.

2.3.5 Purification of Pmar4

Soluble expression of Pmar4-His and Pmar4-MBP was unable to be achieved during expression trials and buffer screens, however, insoluble expression of Pmar4-His was achieved. An alternative approach to purifying Pmar4 was taken where I attempted to denature and re-fold Pmar4. Triton-X was used in a wash step to help remove the *E. coli* membrane, cell wall, inclusion bodies, and precipitants while extracting protein (Palmer & Wingfield, 2012). 8M urea is a common denaturant that was used to unfold Pmar4-His. Urea directly interacts with the peptide backbone of the protein by forming hydrogen bonds and creating an extended, unfolded conformation which exposes residues that would otherwise not be exposed to lysis conditions (Candotti et al., 2013).

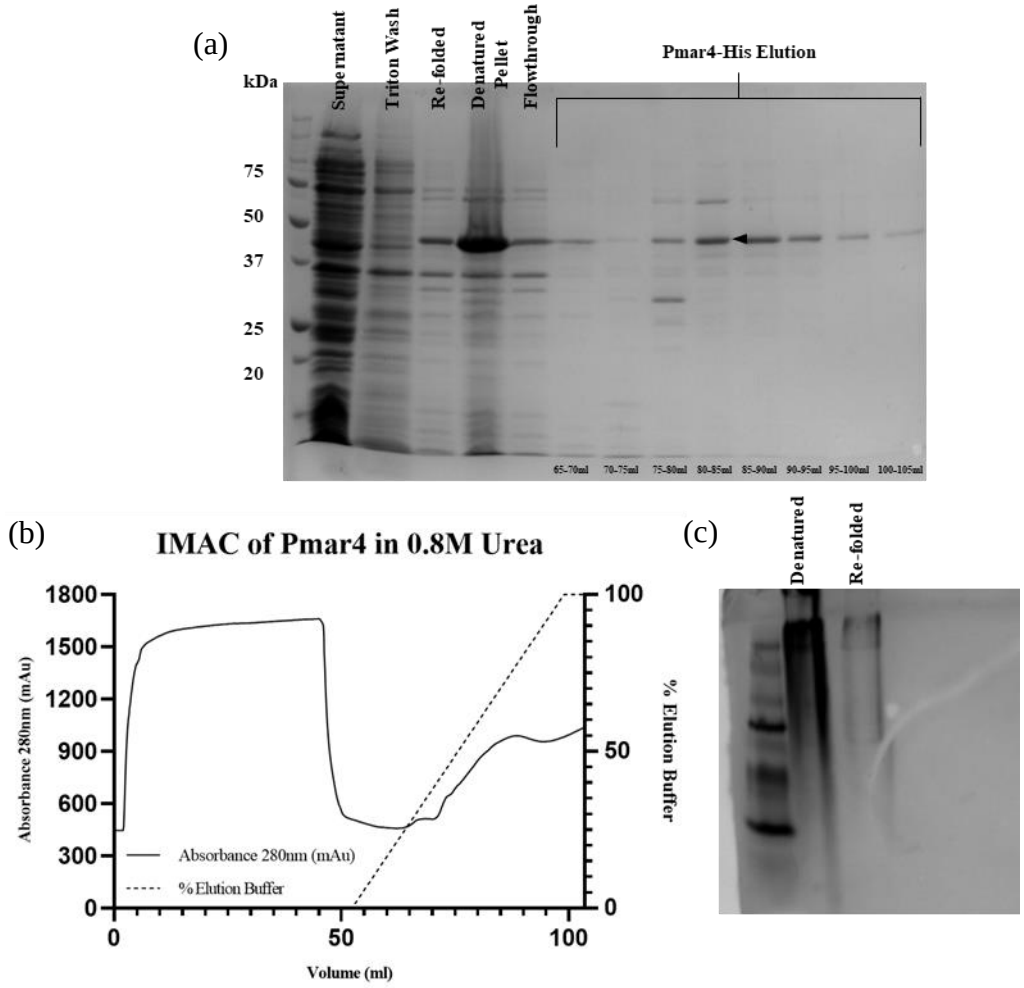


Figure 2-11: Purification and Protein Analysis of Pmar4-His. (a) 12% SDS-PAGE gel of the IMAC of re-folded Pmar4-His (b) Chromatogram of the IMAC of re-folded Pmar4-His. (c) Native PAGE gel of the urea denatured and re-folded lysates.

The IMAC chromatogram shown in Figure 2-11 (b) shows that Pmar4-His eluted at approximately 30-90% buffer B, which corresponds to 15-45mM imidazole. This was confirmed by SDS-PAGE (Figure 2-11[a]), which shows that banding which likely corresponded to Pmar4-His at 52kDa, was consistently present throughout the elution gradient. The protein-containing fractions (from 70-100ml total run volume) were collected, pooled together, and concentrated using a 10kDa cut-off concentrator, to give a total volume of 5ml which was used for size exclusion. The size exclusion of Pmar4-His showed no protein in the elution profile. This suggests that Pmar4-His did not completely re-fold and may have precipitated and stuck to the filter used to load the lysate on to the column. A native gel was run of the denatured and re-folded lysates to check whether the protein re-folded correctly. Figure 2-11(c) shows no clear, distinct banding in the "re-folded" lysate, indicating that Pmar4-His remained misfolded. This purification

was repeated and yielded identical results, concluding that Pmar4-His was unable to be purified using this method.

2.3.6 Purification of Pmar5

Table 2.8: Summary of Purification Specifications of Pmar5-MBP

Purification No.	Culture Volume	Lysis Method	Purification Temperature	TEV Cleavage Method	Reverse IMAC	Size Exclusion?
1	1L (8.37g)	41.85ml LBA. Sonicated: ¼", Amp 2, 3min, 1s on/1s off. 900µl buffer B added to lysate.	Prime, 4°C	1ml TEV added. 30min at RT then 4°C. overnight.	0-100% 20ml buffer B gradient.[1]	Yes
2	1L (9.14g)	45.7ml LBA. Sonicated: ¼", Amp 3, 3min, 1s on/1s off. 1ml buffer B added to lysate.	Prime, 4°C	1ml TEV added. 4°C overnight.	0-100% 50ml buffer B gradient.	-

[1] Second reverse IMAC step added. First IMAC elution peak (5ml) was diluted with buffer C to a final volume of 50ml. 50ml buffer A wash and 0-100% 50ml buffer B gradient ran.

Purification No.	Culture Volume	Lysis Method	Purification Temperature	TEV Cleavage Method	Reverse IMAC	Size Exclusion?
3	1L (12.04g)	60.2ml LBA + 6µl benzonase. Sonicated: ¼", Amp 3, 5min, 1s on/1s off.	NGC, RT	6ml size-exclusion product had 1ml TEV added, 1hr at 37°C, 2 days at 4°C.	Manually run (25ml buffer C wash, 25ml buffer B wash) after TEV cleavage of size exclusion product.	Size Exclusion run first.
4	1L (10.91g)	54.55ml LBA. Sonicated: ¼", Amp 3, 5min, 1s on/1s off.	NGC, RT	1ml TEV added. 1hr at RT, 4°C overnight.[2]	Yes. 0-100% 50ml buffer B gradient.	-
5	1L (12.75g)	60ml LBA. Sonicated: ¼", Amp 3, 5min, 1s on/1s off.	Prime, 4°C	7ml size exclusion product had 1ml TEV added, 4°C overnight.	Yes. 0-100% 25ml buffer B gradient.	Size Exclusion run first.[3]

[2] De-salted lysate with buffer C containing 0.3:0.03mM glutathione redox in place of 1mM DTT.

[3] 1-P5 = 0.53mg/ml, 2-P5 = 0.59mg/ml.

Five attempts were made at purifying Pmar5 at a large-scale, summarized in Table 2.8. The first purification was carried out using the same lysis and IMAC protocol that was used to successfully purify LigW. A 30-minute RT TEV incubation step was included, as this was one of the successful conditions identified when purifying Pmar3. A second reverse IMAC was used for this purification only, as Pmar5 appeared to non-specifically bind to the His-trap column.

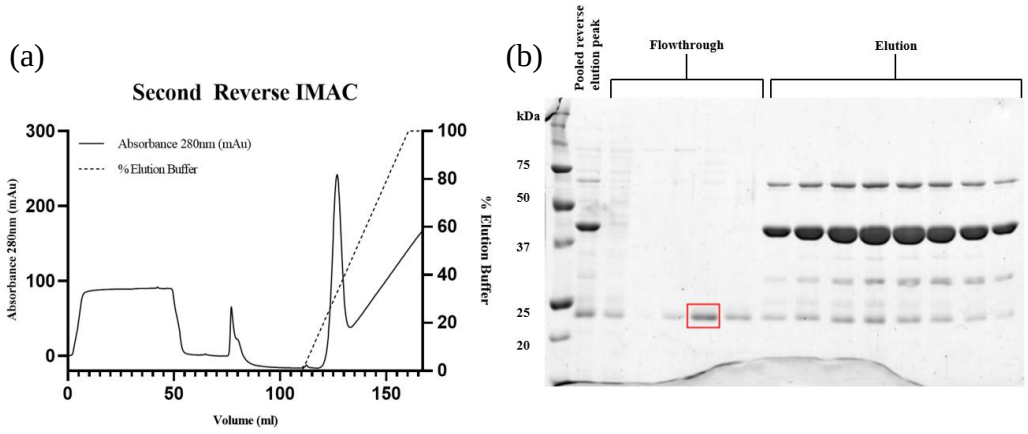


Figure 2-12: First Purification and Protein Analysis of Pmar5-MBP. (a) Chromatogram of the second reverse IMAC of Pmar5-MBP. (b) 12% SDS-PAGE gel of the second reverse IMAC. Highlighted band in lane 5 at approximately 23 kDa was excised for LC-MS.

The cause of this remains unknown, however, a small amount of cleaved Pmar5 was obtained from the second reverse IMAC (Figure 2-12). Despite an unsuccessful size exclusion (shown in Appendix C), the highlighted band in lane 5 of Figure 2-12 (b) was confirmed to be Pmar5 by carrying out liquid chromatography-mass spectrometry. The spectra data obtained for the protein band matched the amino acid sequence for Pmar5 and some minor *E. coli* proteins of similar molecular weight. Given this confirmation, further attempts were made to purify Pmar5 from a Pmar5-MBP culture.

Adjustments were made to the purification protocol to overcome low rates of MBP-tag cleavage, low yield of de-tagged Pmar5, and to address a consistent issue where a mixed population of Pmar5-MBP and Pmar5 was obtained after reverse IMAC and size exclusion. The second purification was not continued beyond the reverse IMAC as the protein had de-stabilized and crashed overnight when incubated with TEV protease at 4^{o}C.

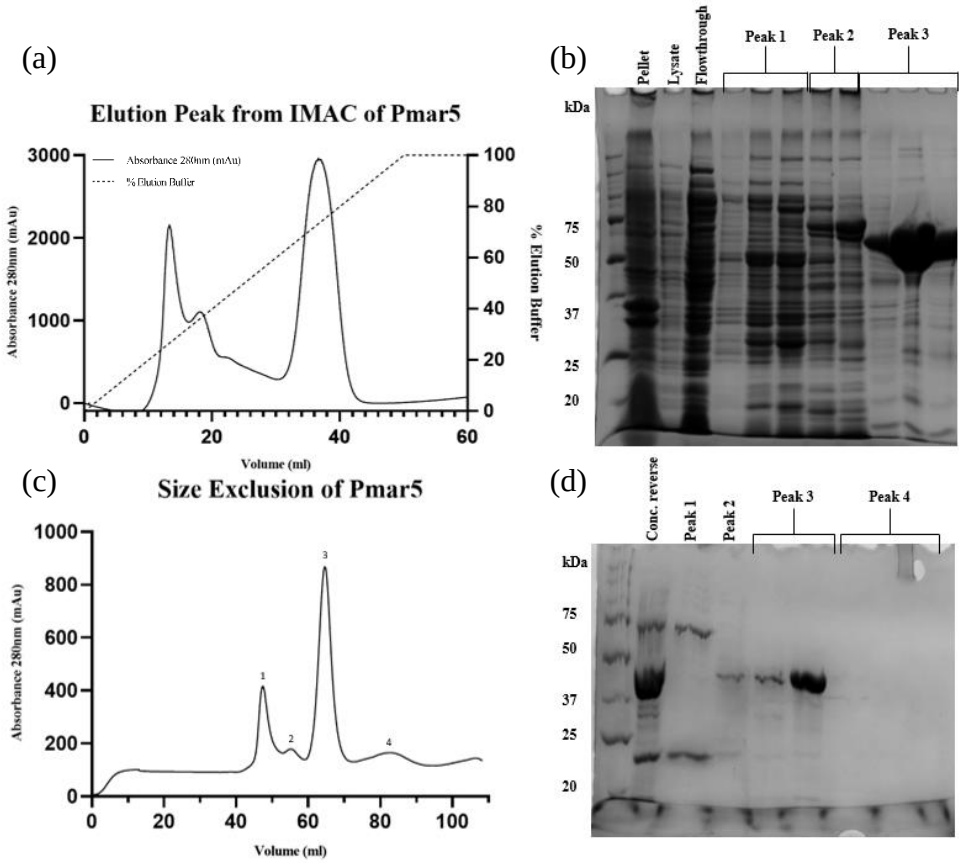


Figure 2-13: Third IMAC Purification and Protein Analysis of Pmar5-MBP. (a) Chromatogram of the IMAC of Pmar5-MBP. (b) 12% SDS-PAGE gel of the IMAC. (c) Chromatogram of the size exclusion of Pmar5-MBP. (d) 12% SDS-PAGE gel of the size exclusion. Peak 1 shows bands corresponding to Pmar5-MBP (68kDa) and Pmar5 (23kDa).

The third purification was carried out at RT to test whether the stability of Pmar5 was affected by temperature. Benzonase was added during the lysis step to help degrade contaminating DNA. Pmar5 remained stable when purified at RT and did not precipitate when incubated with TEV protease at RT for 1 hour, followed by 4^{o}C overnight. The SDS-PAGE gel of the first IMAC shown in Figure 2-13 (b) showed that Pmar5-MBP appeared to separate from the majority of non-specific *E. coli* proteins during the first step, so the cleaved sample was concentrated for size exclusion. The chromatogram shown in Figure 2-13 (c) and the SDS-PAGE gel in Figure 2-13. (d) shows that Pmar5 was successfully separated from contaminant proteins. However, Pmar5 was unable to be separated from Pmar5-MBP by size exclusion. I began to suspect that the de-tagged Pmar5 may have continued to interact with Pmar5-MBP in a way that could not be disrupted. I hypothesised that Pmar5 could be a multimeric protein, and the MBP-tag was not fully cleaved from all sub-units.

The fourth purification was carried out using an alternative buffer C which contained a glutathione redox (0.3mM reduced glutathione to oxidized glutathione) in place of the reducing agent DTT. This buffer was trialled to assist with MBP-tag cleavage, as TEV is a cysteine protease that requires a reducing agent to achieve high activity. Pmar5 contains 7 cysteine residues and is predicted to have up to three disulphide bonds form between the thiol groups of cysteine residues to stabilize the protein structure. Pmar5-MBP had aggregated during some purifications, which was potentially caused by breaking disulphide bonds that were essential for protein stability. The glutathione redox buffer was proposed as a suitable alternative that would still allow TEV protease to function.

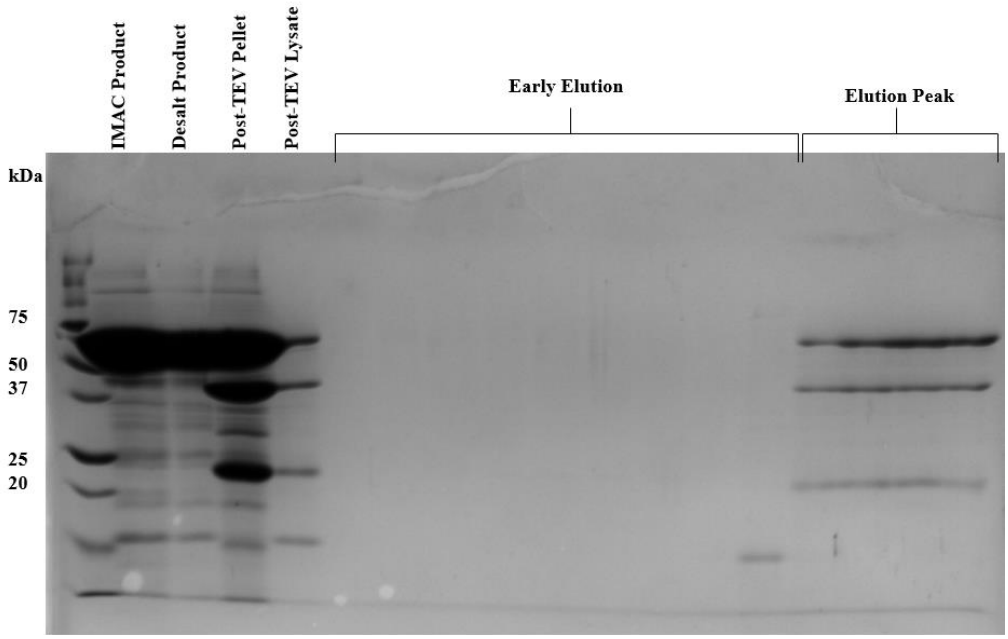


Figure 2-14: 12% SDS-PAGE gel of the reverse IMAC for the fourth purification of Pmar5-MBP.

Although Pmar5-MBP did not precipitate overnight, the SDS-PAGE gel of the reverse IMAC (Figure 2-14) shows that Pmar5-MBP was not successfully cleaved. This suggests that TEV protease activity was not sufficiently stimulated by the glutathione redox, so buffer C containing DTT continued to be used in future purifications.

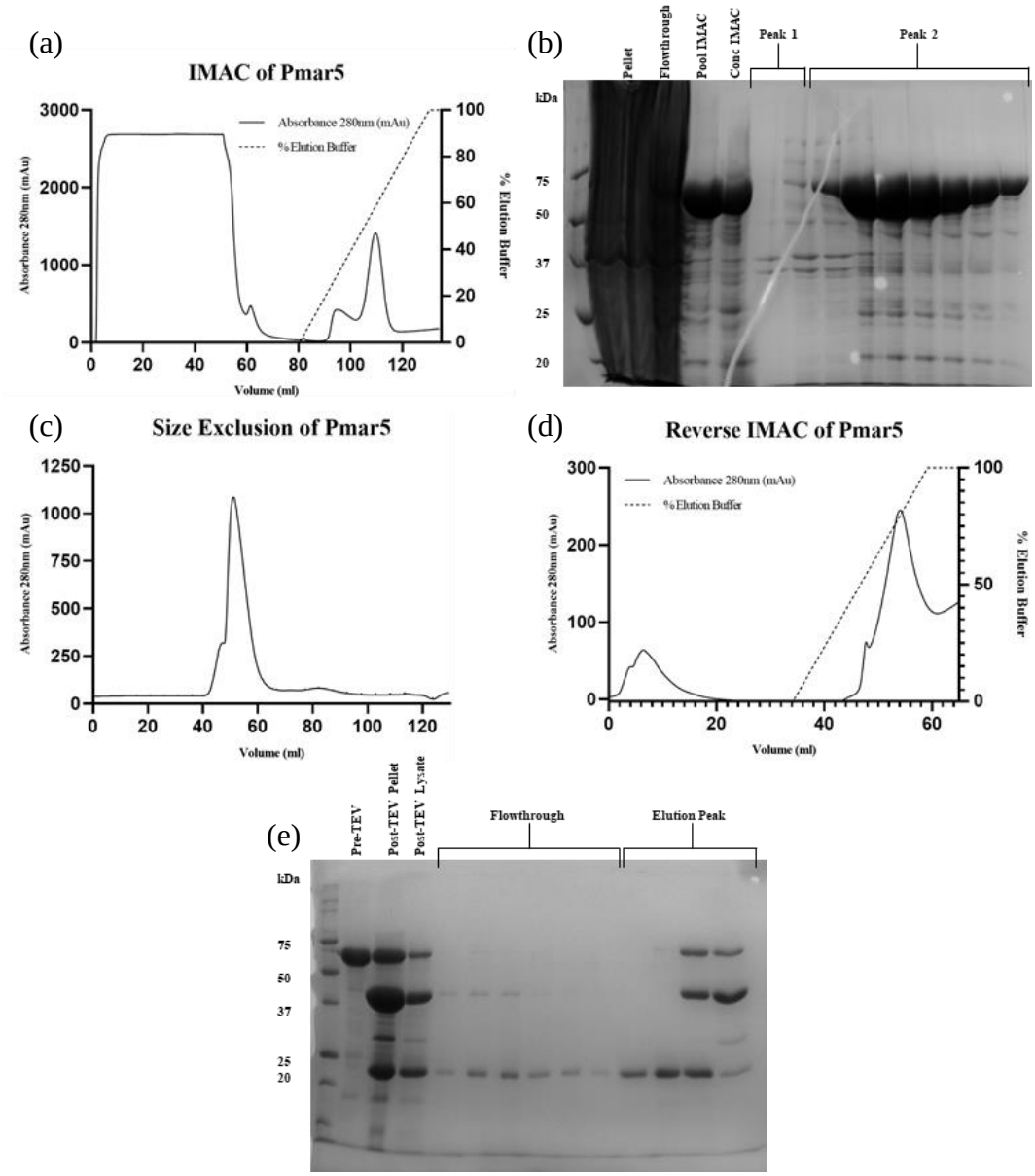


Figure 2-15: Fifth IMAC purification and protein analysis of Pmar5. (a) Chromatogram from IMAC of Pmar5 showing the UV absorbance (280nm) and elution of Pmar5-MBP at 48-68% buffer B. (b) 12% SDS-PAGE gel showing a high yield of Pmar5-MBP and removal of non-specific proteins in the IMAC elution peak. (c) Chromatogram from the size exclusion of Pmar5-MBP, showing the UV absorbance (280nm) of Pmar5-MBP between 49-59 run volume. (d) Chromatogram of the reverse IMAC showing the UV absorbance of Pmar5 in the flowthrough, and the MBP-tag, Pmar5-MBP and other contaminating proteins in the elution peak. (e) 12% SDS-PAGE gel showing the separation of Pmar5 (23kDa) from Pmar5-MBP (68kDa).

The fifth purification of Pmar5-MBP resulted in the isolation of Pmar5 from Pmar5-MBP. In summary, the IMAC chromatogram shown in Figure 2-15 (a) and the SDS-PAGE gel shown in Figure 2-15 (b) shows that Pmar5-MBP eluted at approximately 48-68% buffer B, which corresponds to 240-340mM imidazole. The flowthrough and first 48% of the buffer B gradient contained unbound, non-specific proteins, which continued to be removed with each step of the purification method. The protein-containing fractions

(from 104-114ml total run volume) were collected, pooled together, and concentrated to a final volume of 4ml for size exclusion. Figure 2-15 (b and c) shows that Pmar5-MBP (peak 2) was successfully separated from non-specific proteins (peak 1). Peak 2 (49-59ml total run volume) was pooled together, and 7ml was incubated with 1mg TEV protease at 4^{o}C overnight, while the remaining 3ml was stored in 100µl aliquots at a final concentration of 6.06mg/ml for future use in assays. The incubated lysate was run through a reverse IMAC. The chromatogram shown in Figure 2-15 (d) and the SDS-PAGE gel in Figure 2-15 (e) shows that Pmar5 (23kDa) was successfully isolated in the flowthrough from Pmar5-MBP (68kDa) in the elution peak. The shoulder of the elution peak appeared to contain most of the cleaved Pmar5, so these fractions and the flowthrough were pooled together and concentrated in two batches to final concentrations of 0.53mg/ml (1-P5) and 0.59mg/ml (2-P5) which was stored for future use in characterization assays.

2.4 Conclusions

LigW, Pmar3, and Pmar5 can be successfully expressed and purified using the methods that were optimized and outlined in this chapter. Pmar5 remains challenging to purify, as the MBP-tagged and de-tagged protein appear to interact and remain as a mixed population during purifications, suggesting that Pmar5 may be a multimeric protein. A low yield of de-tagged Pmar5 is attainable, but further optimization is required to obtain a higher concentration of Pmar5. Soluble expression and purification of Pmar4 remains to be achieved. Insoluble expression of Pmar4 was consistently achieved in a pDEST17 expression vector, however, no expression was observed in a pHMGWA expression vector.

Sufficient yields of LigW, Pmar3, and Pmar5 were obtained for characterization, however, further optimization is required for Pmar4. The next chapter summarizes the steps that were taken to characterize Pmar3 and Pmar5 as individual proteins, and their ability to enhance LigW activity was explored.

2.5 Future Research

The expression of Pmar4 will be continued to be tested and optimized to ideally achieve soluble expression. Pmar4 will be transformed into a pDEST15 expression vector which has a glutathione S-transferase tag that assists with the expression and purification of

proteins (Francis & Page, 2010). Although an MBP-tag should assist with soluble expression (as seen with Pmar3 and Pmar5) no expression of Pmar4-MBP was observed. The His-tag appeared to allow expression of Pmar4, so a His-tag will be placed on the C-terminus in an attempt to improve soluble expression. Pmar3 and Pmar5 will have spacers introduced between the protein and TEV binding site to help increase TEV cleavage efficiency and increase the yield of de-tagged protein, which will be used for crystallization trials.

We suspect that Pmar3-mut may have been present in the insoluble pellets of the expression trial. Further expression trials of Pmar3-mut will be carried out using DNase to degrade contaminating DNA and improve gel samples, to identify favourable expression conditions, and to clarify whether Pmar3-mut showed insoluble expression. If expression is not achieved, Pmar3-mut will be transformed into pDEST15 and pDEST17 expression vectors and further expression trials will be carried out. Alternatively, a new mutant may be created with one aspartate to alanine point mutation, as the stability of Pmar3 may have been disrupted by the four point-mutations.

Chapter 3

Protein Characterization

3.1 Introduction

Although hypothesised as DNA-modifying proteins, the exact function of Pmar3, Pmar4, and Pmar5 needed to be determined. Biochemical characterization of proteins involves a series of activity assays that are necessary to determine how each protein functions under different conditions and in response to different substrates. In addition to determining the molecular weight and thermal stability of Pmar3 and Pmar5, each protein's ability to bind damaged DNA and each protein's interactions with LigW were explored through gel-based assays.

Electrophoretic mobility shift assays were used to assess each protein's ability to recognize and bind to different length DNA substrates. Successful binding to a substrate would create larger protein-DNA complexes that can be observed as a gel-shift on electrophoresed native TBE gels. Nuclease activity was assessed using a fluorescence nuclease assay, where nucleases would bind to and degrade DNA substrates of different lengths. Degradation of DNA was also visualized as a gel-shift on electrophoresed native TBE gels. Native TBE gels show the native conformation of the DNA, allowing differentiation between structurally conserved double-stranded and single-stranded products. Nuclease assay samples were treated with proteinase K to digest the protein by hydrolysing the peptide bonds, allowing visualization of only the DNA. The substrates were labelled with 5-carboxyfluorescein (5-FAM) labels to allow visualization of products on the gels and to quantify the fluorescence of different-sized products (Williamson et al., 2014).

At the commencement of this research, Pmar5 had been annotated as a predicted topoisomerase. Topoisomerases are enzymes capable of overwinding or unwinding DNA to relieve supercoiling during DNA replication and repair (Champoux, 2001). The potential topoisomerase activity of both Pmar3 and Pmar5 was explored by incubating each protein with a supercoiled, circular, double-stranded pUC19 plasmid. Unlike the gel-shift assays, the product of topoisomerase assays does not change composition, and instead, a change in the topology of pUC19 is to be expected. There are multiple classes of topoisomerase that introduce single-strand or double-strand cuts in the DNA, which

can be visualized as different sized bands on a 1% TAE agarose gel with lower linking numbers (i.e. less supercoiling) in the plasmid running with greater apparent size (Champoux, 2001; Terekhova et al., 2012). The assay reactions were proteinase K treated similarly to the nuclease assay.

The characterization of Pmar3 and Pmar5 entails determining each protein's role in the LigW operon. Both proteins are adjacent to LigW and are hypothesised to enhance LigW's poor ligation activity (Williamson, Unpublished). Each protein was incubated with LigW and a 5-FAM labelled nicked DNA substrate to see if and to what extent the single-stranded breaks were being sealed. Reactions were visualized on 7M urea gels, which denature proteins and DNA, allowing visualization of only the 5-FAM labelled single strand. Single nucleotide differences such as sealed nicks can be seen on a urea gel as distinct different-sized bands (Summer et al., 2009).

This chapter focuses on the biochemical characterization of Pmar3 and Pmar5.

3.2 Materials and Methods

3.2.1 Analytical Size Exclusion

The ENrich 650 analytical gel filtration column (Bio-Rad Laboratories, USA) was calibrated to determine the buffer-specific molecular weights and column void volume, according to the manufacturer's instructions. In summary, the column was equilibrated with 2 column volumes of buffer C before loading 250µl Bio-Rad size-exclusion standards (catalogue #151-1901) or blue dextran respectively, followed by 1 column volume of buffer C. The molecular weight of proteins was determined by loading 200-500µl concentrated protein onto the pre-equilibrated column. The sample was loaded onto the column by manual injection using a syringe. The column was run with 1 column volume of buffer C at 1ml min^{-1}. Eluted protein fractions were collected and analysed on SDS-PAGE. The elution volumes (V_e) of the standards, the column void volume (V_o), and the total column volume (V_c) were used to calculate the gel phase distribution coefficient (K_{av}), shown in Equation 3-1.

Equation 3-1: Formula for calculating the gel phase distribution coefficient.

$$K_{av} = \frac{(V_e - V_o)}{(V_c - V_0)}$$

Calibration standard curves of K_{av} relative to molecular weight were made. Protein elution volumes were plotted against the calibration standard curves to determine the molecular weight of each protein, shown in Equation 3-2.

Equation 3-2: Formula for calculating the molecular weight.

$$Molecular\ Weight = e^{\frac{K_{av} - y\ intercept}{Slope}}$$

3.2.2 Thermal-Shift Assay

The melting temperature and pH stability of proteins were determined by performing a thermal-shift assay. 25µl reactions involving varying concentrations of protein, 3µl 5X SYPRO, and 14µl Britten Robinson Universal Buffer (40mM H_3BO_3, 40mM H_3PO_4, 40mM CH_3COOH, pH 5.0-9.0) were set up in PCR tubes at RT. PCR tubes were loaded into the Corbett Rotor-Gene™ 6000, and rotor-gene Q series software was used to run and analyse a melt from 25°C-99°C.

3.2.3 Electrophoretic Gel Mobility Shift Assay

Electrophoretic gel mobility shift assays were used to assess each protein's ability to bind to damaged DNA. 20µM of Pmar3, Pmar5, or LigW was incubated with 80nM DNA substrate (see Table 3.2 and Table 3.3) at 15°C for 30 minutes. Reaction compositions are summarized in Appendix D. 5µl loading dye (100mM EDTA, 0.25% bromophenol blue, 25% v/v glycerol) was added to each reaction, and 10µl was loaded and run on 10x native TBE gels (cast using a Hoeffer five gel multi-gel caster, see Table 3.1) at 5mA per gel using 1x TBE buffer (90mM tris-base, 90mM boric acid, 2mM EDTA). Gels were imaged using an iBright imager (Invitrogen).

Table 3.1 Recipe for 10x Native TBE Gels

	1 Gel (10ml)
ddH_2O	6.5ml
40% acrylamide/bis-acrylamide (29:1)	2.5ml
10x TBE Buffer[4]	1ml
10% APS	100µl
TEMED	3µl

Table 3.2: Sequences of the oligonucleotides used to make electrophoretic mobility shift assay substrates.

Description	**Oligonucleotide sequence (5' to 3')**	**Modification**
Nick FAM	5'-AGGCCATGGCTGATATCGCA-3'	5' FAM
Nick Phos	5'-TAGGCATTCGAGCTCCGTCG-3'	5' Phosphate
Nick Complement	5'-CGACGGAGCTCGAATGCCTA TGCGATATCAGCCATGGCCT-3'	-
Linear	5'-AGGCCATGGCTGATATCGC ATAGGCATTCGAGCTCCGTCG-3'	5' FAM
Blunt Complement	5'-CGACGGAGCTCGAATGCCTA-3'	-

Table 3.3: Oligonucleotide components annealed to make electrophoretic mobility shift assay substrates.

Substrate	**FAM-labelled Oligo**	**Complement**	**Phosphate-labelled Oligo**
Double-strand Matched	Linear	Nick Complement	-
Single-strand	Linear	-	-
3'-Tail	Linear	Blunt Complement	-
5'-Tail	Nick FAM	Nick Complement	-
Nick	Nick FAM	Nick Complement	Nick Phos

3.2.4 Topoisomerase Assay

Topoisomerase activity of Pmar3, Pmar5, and LigW was assessed by incubating each protein with a pUC19 plasmid. pUC19 purified from *E. coli* exists in a supercoiled conformation and can be relaxed by topoisomerases to give a relaxed circular product, or

[4] Tris-borate-EDTA buffer. 108g Tris-base, 55g boric acid, 7.5g EDTA, ddH_2O to 1L.

cuts can be made by nucleases. pUC19 was prepared using a QIAprep® Spin Miniprep Kit (Bio-strategy) according to the manufacturer's instructions. 5µl of proteins of varying concentrations were incubated at RT with 2µl pUC19 (81.2-98.6 ng/µl) in PCR tubes for time periods of 30 minutes to 18 hours. Metal-dependency and substrate-dependency for topoisomerase activity was measured by adding 5mM $MgCl_2$, 5mM $MnCl_2$, 10mM $ZnCl_2$, 1mM ATP, or 1mM GTP to reactions. 2µl proteinase K was added to each reaction and incubated at 37°C for 1 hour. The reaction was stopped by adding 2.5µl 10x topoisomerase stop buffer (100mM EDTA, 0.25% bromophenol blue, 25% v/v glycerol). 1% agarose gels were used to degrade the protein, which revealed any changes in pUC19 supercoiling.

1% agarose gels were made and run according to section 2.2.3, but without the addition of SYBR-safe during casting. Gels were post-stained for 20 minutes in 100ml 1xTAE and 10µl thiazole orange was added. A post-stain was performed instead of adding thiazole orange to the agarose gel to preserve any supercoiled topology. Thiazole orange inter-chelates with bases and causes supercoiled pUC19 to appear the same as pUC19 with a relaxed topology. Agarose gels were imaged using an iBright Imager (Invitrogen).

3.2.5 Fluorescence Nuclease Assay

Fluorescence-based nuclease assays were used to assess the nuclease activity of Pmar3. 2.5µl of Pmar3 of varying concentrations was incubated with one of six 80nM DNA substrates (see Table 3.4 and Table 3.5) at RT for time periods of 30 minutes to 17 hours. Reaction compositions are summarized in Appendix D. 2µl proteinase K was added to each reaction and incubated at 37°C for 1 hour. The reaction was stopped by adding 2.5µl nuclease stop buffer (100mM EDTA, 0.25% bromophenol blue, 25% v/v glycerol). 10µl was loaded and run on 10x native TBE gels at 5mA per gel using 1x TBE buffer. Gels were imaged using an iBright Imager (Invitrogen).

Table 3.4: Sequences of the oligonucleotides used to make fluorescence nuclease assay substrates.

Description	Oligonucleotide sequence (5' to 3')	Modification
Nick FAM	5'-AGGCCATGGCTGATATCGCA-3'	5' FAM
Nick Complement	5'-CGACGGAGCTCGAATGCCTA TGCGATATCAGCCATGGCCT-3'	-
Linear	5'-AGGCCATGGCTGATATCGC ATAGGCATTCGAGCTCCGTCG-3'	5' FAM
Blunt Complement	5'-CGACGGAGCTCGAATGCCTA-3'	-

Table 3.5: Oligonucleotide components annealed to make fluorescence nuclease assay substrates.

Substrate	FAM-labelled Oligo	Complement
Double-strand Matched	Linear	Nick Complement
Single-strand	Linear	-
3'-Tail	Linear	Blunt Complement
5'-Tail	Nick FAM	Nick Complement
20+20 Control	Nick FAM	Blunt Complement
Single-strand (20) Control	Nick FAM	-

3.2.6 Coupled Ligation Assay

Coupled assays of Pmar3 and Pmar5 with LigW were conducted to observe the effect that each protein has on the ligation activity of LigW. Varying concentrations of protein were incubated for 1 hour at 25ºC with a constant concentration of 5µM LigW and nicked DNA master mix (80nM DNA substrate, 1mM ATP, 50mM NaCl, 50mM tris pH 8.0, 1mM DTT). Reactions were quenched with 5µl quench buffer (95% formamide, 1% EDTA, 1% bromophenol blue) and further incubated for 5 minutes at 95ºC. 7M urea PAGE gels (cast using a Hoeffer five gel multi-gel caster, see Table 4.4) were pre-run for 30 minutes at 5mA per gel at 47ºC with 1xTBE buffer. 10µl samples were loaded onto pre-run 7M urea PAGE gels and were ran at 5mA for approximately 1.5hrs. Gels were visualised using an iBright imager (Invitrogen).

Table 3.6 Recipe for 7M Urea Gels

	1 Gel
20% Acrylamide/7M Urea	10ml
10% APS	100µl
TEMED	3µl

3.3 Results and Discussion

3.3.1 Analytical Size Exclusion

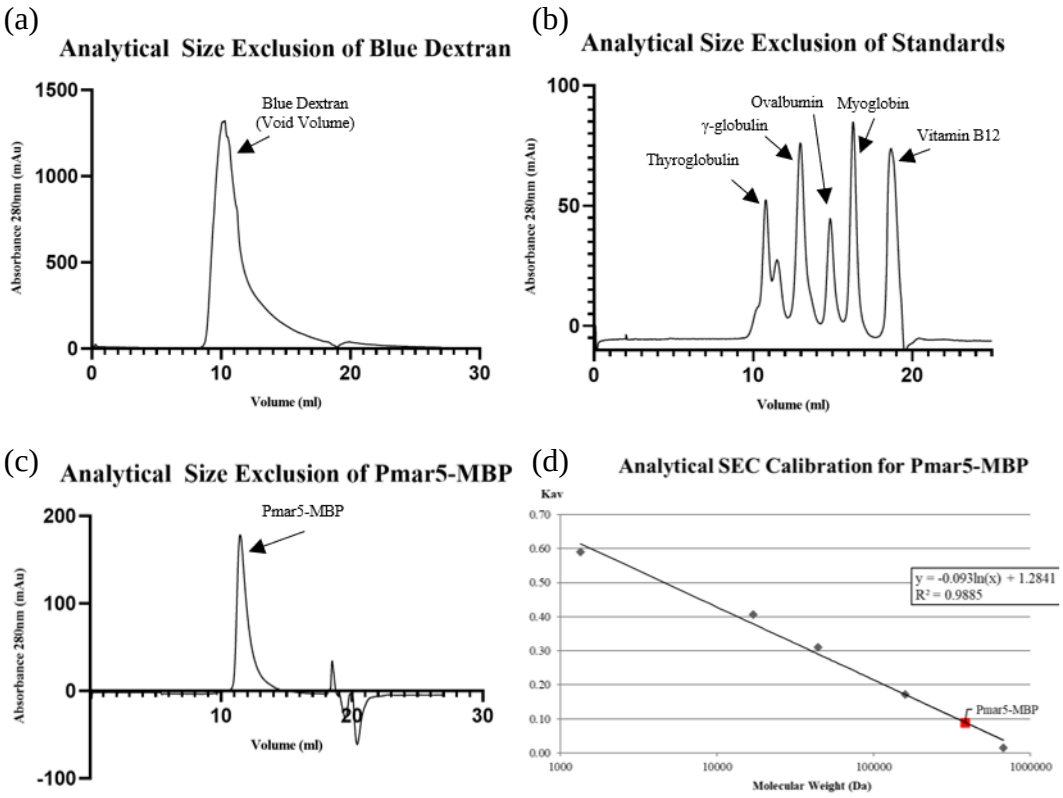


Figure 3-1: Determination of the molecular weight of Pmar5-MBP. (a) Chromatogram for blue dextran to determine the void volume of the ENrich 650 analytical gel filtration column (Bio-Rad Laboratories, USA). (b) Chromatogram for Bio-Rad size-exclusion standards (catalogue #151-1901) to determine the gel phase distribution coefficient (K_{av}). (c) Chromatogram for the size exclusion of Pmar5-MBP. (d) Calculated calibration curve of molecular weight against K_{av}. The K_{av} of Pmar5-MBP (red) was plotted and used to determine the molecular weight of Pmar5-MBP.

The molecular weight of Pmar5-MBP in glutathione redox buffer (500mM tris pH 8.0, 100mM NaCl, 5% glycerol, 0.5mM EDTA, 0.3:0.03mM glutathione redox) was determined using analytical size exclusion (shown in Figure 3-1). A calibration curve (Figure 3-1 [d]) was plotted based on the elution volumes and corresponding K_{av} values, summarized in Table 3.7.

Table 3.7: Bio-Rad size-exclusion standards (catalogue #151-1901) and the corresponding elution volumes (V_e) and K_{av} values when using glutathione redox buffer.

Protein	Molecular Weight (Da)	Elution Volume (ml)	K_{av}
Vitamin B12	1350	18.35	0.59
Myoglobin	17000	15.83	0.41
Ovalbumin	44000	14.5	0.31
γ-globulin	158000	12.59	0.17
Thyroglobulin	670000	10.45	0.02

Figure 3-1 (c) shows that Pmar5-MBP eluted at 11.46ml, with a K_{av} of 0.09. The molecular weight of Pmar5-MBP (calculated using Equation 3-2) is equal to 379.6kDa. According to the sequence of Pmar5-MBP, the molecular weight is predicted to be approximately 67.8kDa. Based on the analytical SEC results, Pmar5-MBP is predicted to be a pentameric protein, as the molecular weight was calculated to be approximately 5 times larger than the sequence-predicted molecular weight. The multimeric nature of Pmar5-MBP could account for the difficulties encountered when attempting to de-tag Pmar5 and when separating the mixed population of tagged and de-tagged Pmar5 by size exclusion. If Pmar5 is a multimeric protein, the MBP-tag may be difficult to cleave successfully from all the domains. The calculated molecular weight is more than 5 times larger than predicted, which also suggests that the protein may be asymmetric with a larger Stokes radius than a sphere of the same molecular weight, causing it to elute earlier on the column (Burgess, 2018). The molecular weight and multimeric nature of Pmar5 could be confirmed in the future by running de-tagged Pmar5 through analytical size exclusion, as well as attempting crystallography to solve the protein's structure.

The molecular weight of Pmar3 in buffer C (50mM tris pH 8.0, 100mM NaCl, 5% glycerol, 1mM DTT, 0.5mM EDTA) was attempted to be determined using analytical size exclusion (shown in Figure 3-2 [c]). A calibration curve (Figure 3-2 [a]) was plotted based on the elution volumes and corresponding K_{av} values, summarized in Table 3.8.

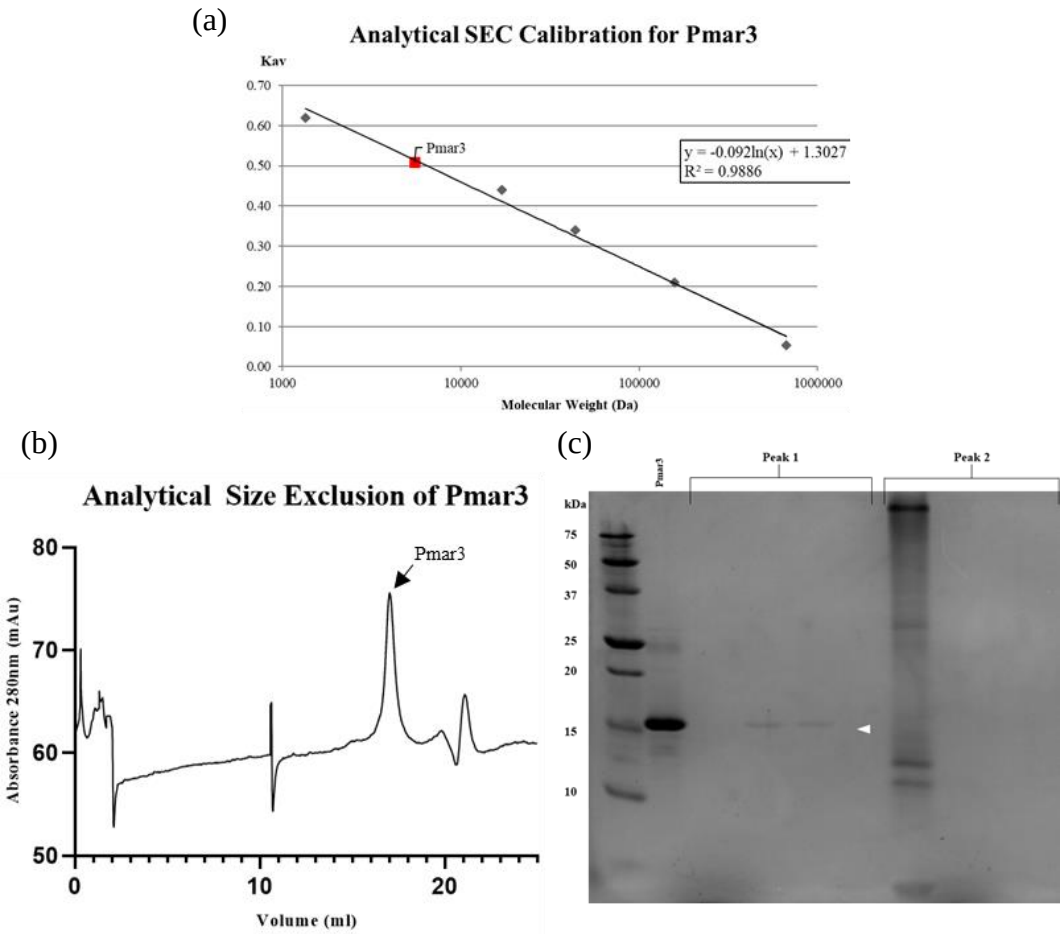


Figure 3-2: Determination of the molecular weight of Pmar3. (a) Calculated calibration curve of molecular weight against K_{av}. The K_{av} of Pmar3 (red) was plotted and used to determine the molecular weight of Pmar3. (b) Chromatogram for the size exclusion of Pmar3. (c) 12% SDS-PAGE gel for collected fractions from size exclusion of Pmar3. Pmar3 was present in peak 1 (indicated by the white arrow).

Table 3.8: Bio-Rad size-exclusion standards (catalogue #151-1901) and the corresponding elution volumes (V_e) and K_{av} values when using buffer C.

Protein	Molecular Weight (Da)	Elution Volume (ml)	K_{av}
Vitamin B12	1350	18.58	0.62
Myoglobin	17000	16.04	0.44
Ovalbumin	44000	14.61	0.34
γ-globulin	158000	12.75	0.21
Thyroglobulin	670000	10.54	0.05

Figure 3-2 (b) shows that Pmar3 eluted at 17.03ml, with a K_{av} of 0.51. The molecular weight of Pmar3 (calculated using Equation 3-2) is equal to 5.5kDa. According to the sequence of Pmar3, the molecular weight is predicted to be approximately 17.0kDa and was expected to elute at approximately 16ml. Based on the analytical SEC results, Pmar3 is predicted but cannot be accurately concluded, to be a monomer. The calculated

molecular weight from analytical SEC is less than the sequence-predicted molecular weight. Pmar3 may be asymmetrical, but in contrast to Pmar5-MBP, it may have a smaller Stokes radius than a sphere of the same molecular weight, causing it to elute later on the column. Alternatively, Pmar3 could have interacted with the resin or column tube, interfering with its movement down the column and also causing it to elute later (Burgess, 2018). The molecular weight of Pmar3 could be confirmed in the future by running a native PAGE gel, as this would allow us to visualize its size and native structure. Attempts could also be made to crystallize Pmar3 to solve the protein's structure.

3.3.2 Thermal-Shift Assays

Thermal-shift assays were run of all three proteins of interest over a pH scale of 5.0-9.5, and a temperature range of 25-99°C. The purpose of these assays was to identify the melting temperatures (T_m) of each protein and the pH that promoted protein stability. Fluorescence of Sypro was used to detect the protein states: folded, unfolded, or aggregated. An increase in fluorescence is observed upon denaturation when SYPRO Orange binds to the protein's exposed hydrophobic groups. Fluorescence decreases as the protein aggregates with increasing temperature (Bai et al., 2019).

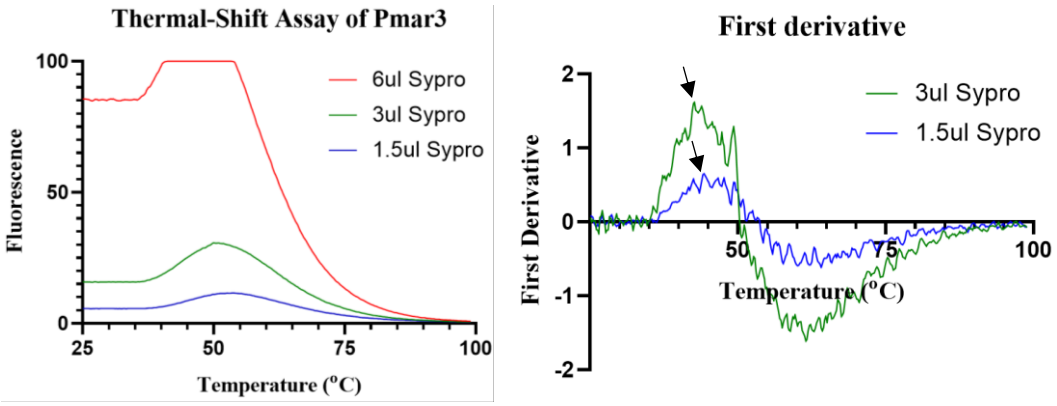


Figure 3-3: Thermal-shift assay of 17.40µM Pmar3. Reaction compositions: 17.40µM Pmar3 (19µl), Sypro + buffer C (6µl). (a) Chromatogram showing a thermal-shift assay of Pmar3 from 25-99°C. (b) First derivatives of assay data to determine the T_m of Pmar3. Arrows point to the maximum point of each curve. 3µl Sypro sample T_m = 42.7°C, 1.5µl Sypro sample T_m = 44.3°C.

A thermal-shift assay (Figure 3-3) was run of three samples of a final concentration of 17.4µM Pmar3 (Purification 2). Varying concentrations of Sypro tested for optimal fluorescence detection as the protein unfolded. The 6µl Sypro sample was excluded from analysis, as the maximum detectable fluorescence on the instrument was exceeded. Figure

3-3 (b) shows the first derivative calculations for the remaining two samples. The maximum points of the first derivative curves correspond to the T_m of the samples. The 3µl Sypro sample had a T_m of 42.7°C, while the 1.5µl Sypro sample had a T_m of 44.3°C. An additional thermal-shift assay was completed of Pmar3 across the pH scale (5.0-9.5) to confirm the T_m and preferred pH that promotes protein stability.

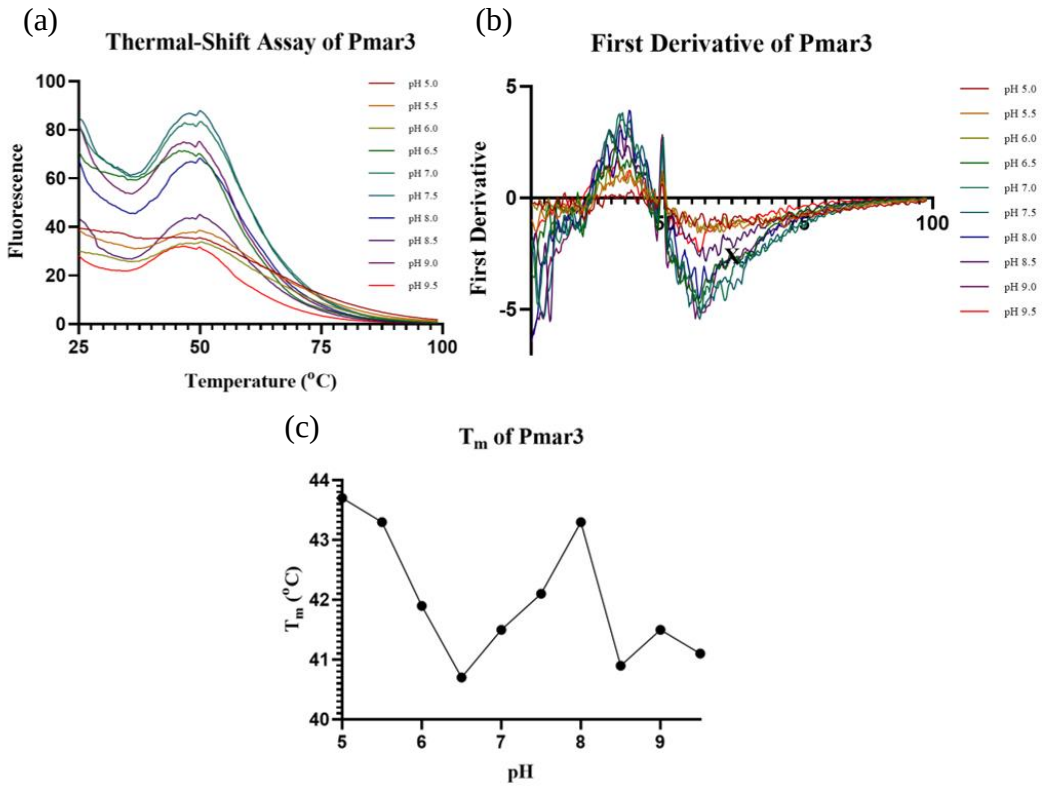


Figure 3-4: Thermal-shift assay of 35.13µM Pmar3 from 25-99°C in Britten Robinson universal buffer (40mM H_3BO_3, 40mM H_3PO_4, 40mM CH_3COOH, pH 5.0-9.5). Reaction compositions: 35.13µM Pmar3 (8µl), Sypro (3µl) + pH buffer (14µl). (a) Chromatogram showing a thermal-shift assay of Pmar3 from 25-99°C. (b) First derivatives of assay data to determine the T_m of Pmar3. (c) Plot of T_m of Pmar3 for pH 5.0-9.5.

Figure 3-4 shows that Pmar3 has a T_m range of approximately 41-44°C and no conclusive pH that provides protein stability. The findings from this assay were used to help select the temperature used to deactivate Pmar3 in fluorescence nuclease assays.

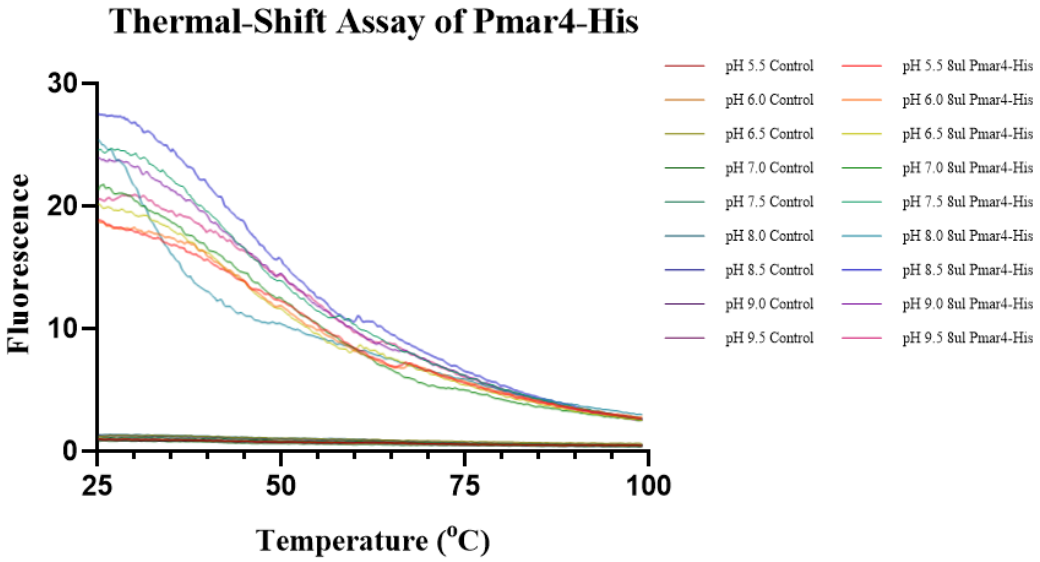


Figure 3-5: Chromatogram showing a thermal-shift assay of Pmar4-His from 25-99°C in Britten Robinson universal buffer (40mM H_3BO_3, 40mM H_3PO_4, 40mM CH_3COOH, pH 5.0-9.0).

A thermal-shift assay was run of Pmar4-His in addition to the native PAGE gels after attempting an IMAC purification, to confirm whether Pmar4-His had re-folded correctly. Figure 3-5 shows that the reactions containing Pmar4-His fluoresced before any heat treatment. This confirms that Pmar4-His did not re-fold correctly, as fluorescence is observed upon denaturation when SYPRO Orange binds to the protein's exposed hydrophobic groups (Bai et al., 2019). As discussed in the previous chapter, further optimization of soluble expression or protein re-folding is required to characterize Pmar4.

The thermal-shift assays were conducted with 13.64 and 7.27µM Pmar5 to determine the T_m and T_m range between a pH of 5.0-9.5 respectively.

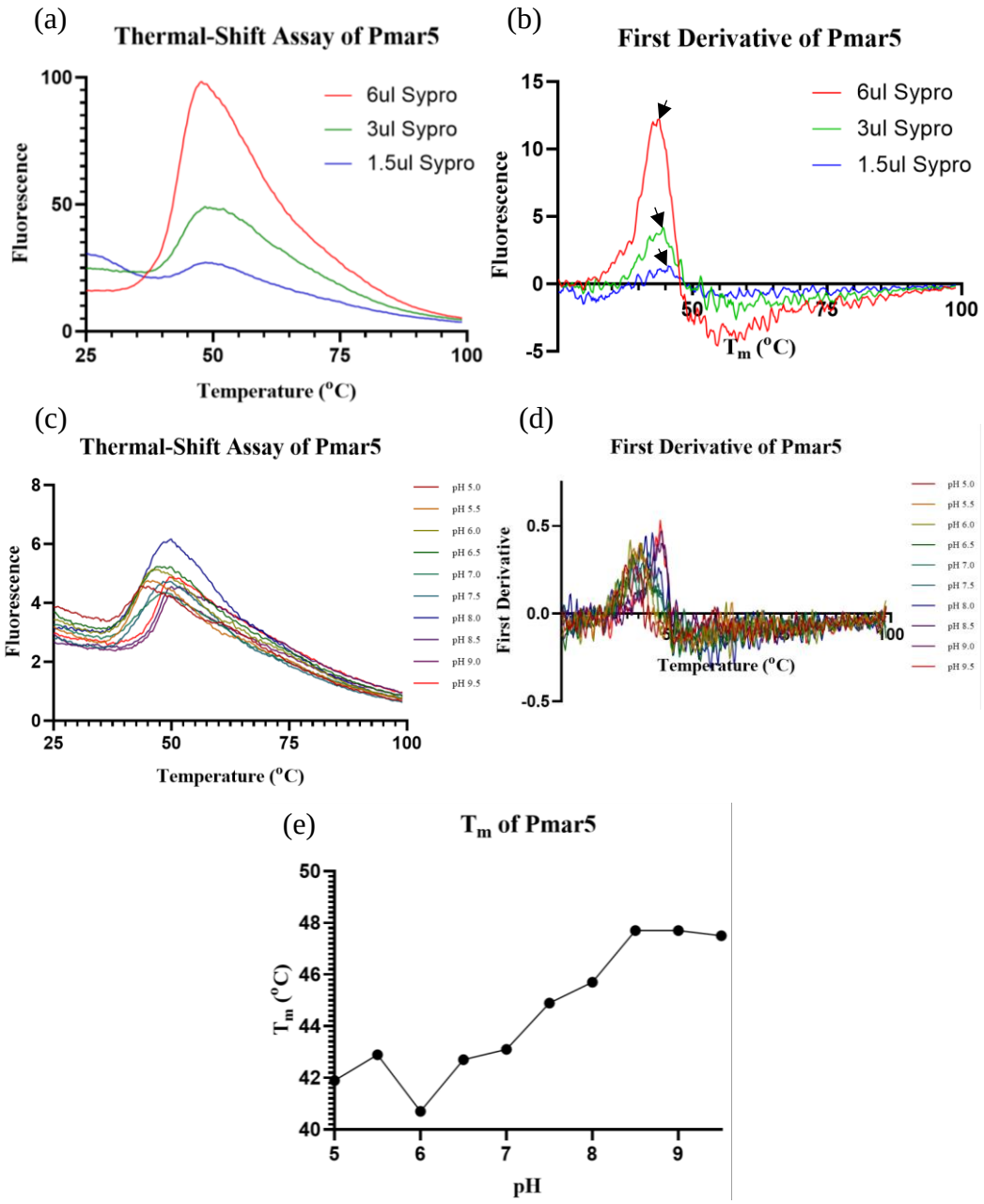


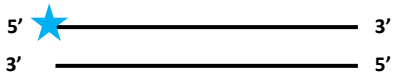
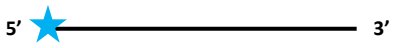
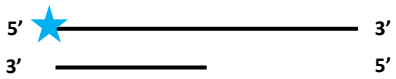
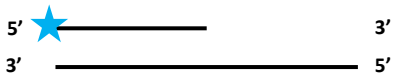
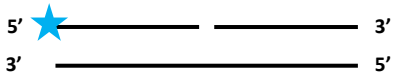
Figure 3-6: Thermal-shift assays of Pmar5. (a, b) Reaction compositions: 13.64µM Pmar5 (15µl), Sypro + buffer C (10µl). (a) Chromatogram showing a thermal-shift assay of Pmar5 from 25-99ºC. (b) First derivatives of assay data to determine the T_m of Pmar5. Arrows point to the maximum point of each curve. 6µl Sypro sample T_m = 47.6ºC, 3µl Sypro sample T_m = 48.4ºC, 1.5µl Sypro sample T_m = 48.4ºC. (c-e) Reaction compositions: 7.27µM Pmar3 (8µl), Sypro (3µl) + pH buffer (14µl). 25-99ºC in Britten Robinson universal buffer (40mM H_3BO_3, 40mM H_3PO_4, 40mM CH_3COOH, pH 5.0-9.5). (c) Chromatogram showing a thermal-shift assay of Pmar5 from 25-99ºC. (b) First derivatives of assay data to determine the T_m of Pmar5. (c) Plot of T_m of Pmar5 for pH 5.0-9.5.

A low protein concentration was used comparatively to Pmar3, which introduced some background noise in the data, which is particularly obvious in the second derivative curves. However, it can be concluded from Figure 3-6 that the T_m of Pmar5 increases by

almost 8 °C with increasing pH. Buffer C has a pH of 7, which explains why a T_m of approximately 48°C was observed in Figure 3-6 (a and b).

3.3.3 Electrophoretic Mobility Shift Assays

Table 3.9: Diagrams of the substrates used in electrophoretic mobility shift assays.

Substrate	Substrate Diagram
Double-strand Matched	
Single-strand	
3’-Tail	
5’-Tail	
Nick	

Protein-nucleic acid interactions were assessed through electrophoretic mobility shift assays (EMSA), where each protein’s ability to bind to damaged DNA was tested using double-stranded matched, single-stranded, 3’ tail, 5’ tail, and nicked substrates (see Table 3.9). These assays were selected as Pmar3 and Pmar5 were hypothesised to be DNA-modifying proteins, and we wanted to investigate if either of these proteins had an affinity for different types of damage. A variety of substrates were used to help predict the functions of Pmar3 and Pmar5.

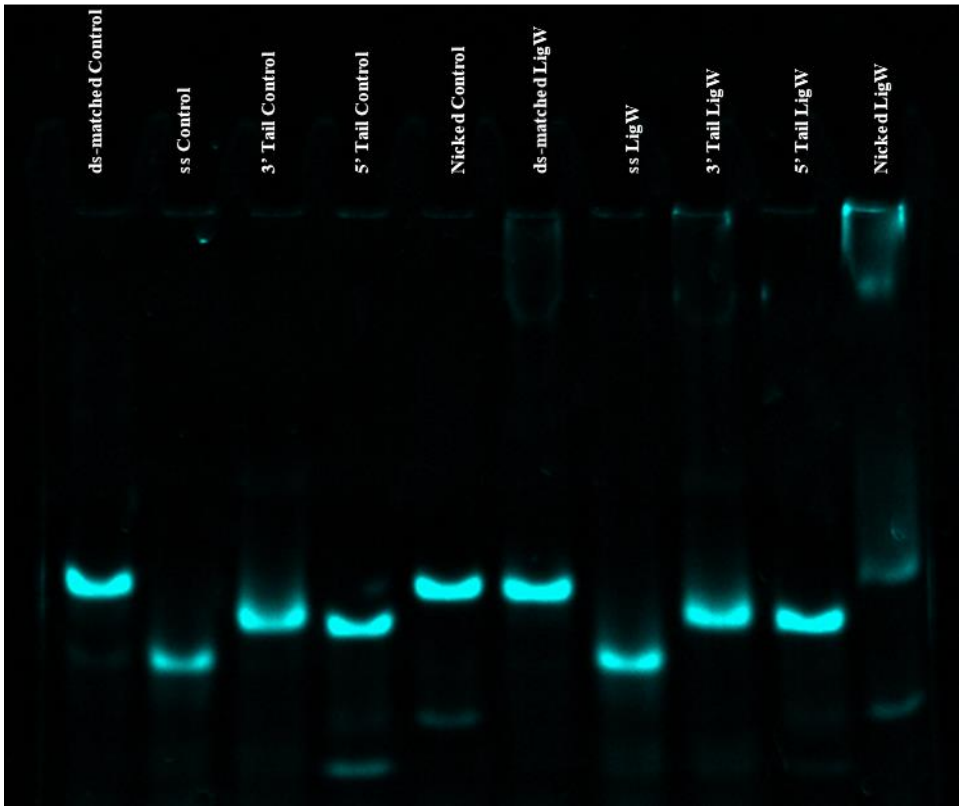


Figure 3-7: Native 10% TBE gel of the Electrophoretic Mobility Shift Assay of 20µM LigW.

LigW was used as a positive control to test whether the conditions were suitable to show DNA binding. Ligases are expected to recognize and bind to nicked DNA, so a gel shift of the nicked substrate should be observed. A gel shift in lane 10 of Figure 3-7 shows that LigW successfully bound to the nicked substrate, indicating that LigW recognizes nicked damage and that the assay conditions were appropriate.

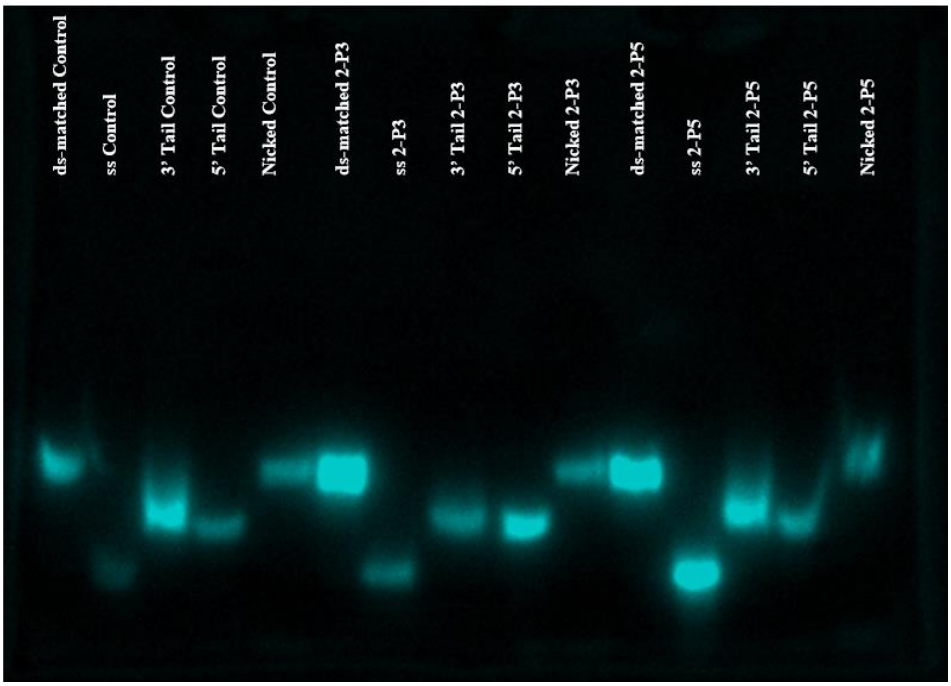


Figure 3-8: Native 10% TBE gel of the Electrophoretic Mobility Shift Assay of 20µM Pmar3 (lanes 6-10) and 20µM Pmar5 (lanes 11-15).

Unlike LigW, Figure 3-8 shows that Pmar3 and Pmar5 did not bind to any of the substrates used. It is likely that both Pmar3 and Pmar5 do not recognize or prefer the substrates tested. Although the proteins did not bind, it is suspected that the proteins can recognize other sequences of DNA or other types of damage. It is also possible that the proposed DNA-binding activity of Pmar3 and Pmar5 may be dependent on metals or other additives that were omitted from this assay. Pmar3 and Pmar5 may have weak binding activity, which can be explored in the future by repeating this assay with a higher concentration of protein compared to DNA substrate (Hellman & Fried, 2007). The potential functions of Pmar3 and Pmar5 were further explored through a series of nuclease and topoisomerase assays, which confirmed that protein inactivity was not the cause for the lack of substrate binding.

3.3.4 Assessing Topoisomerase Activity of Pmar3

Although hypothesised to be a DNA-modifying protein, the function of Pmar3 remained to be determined through a series of activity assays. Determination of the function of Pmar3 began by assessing its ability to act as a topoisomerase, where it would relieve the supercoiling of a pUC19 plasmid.

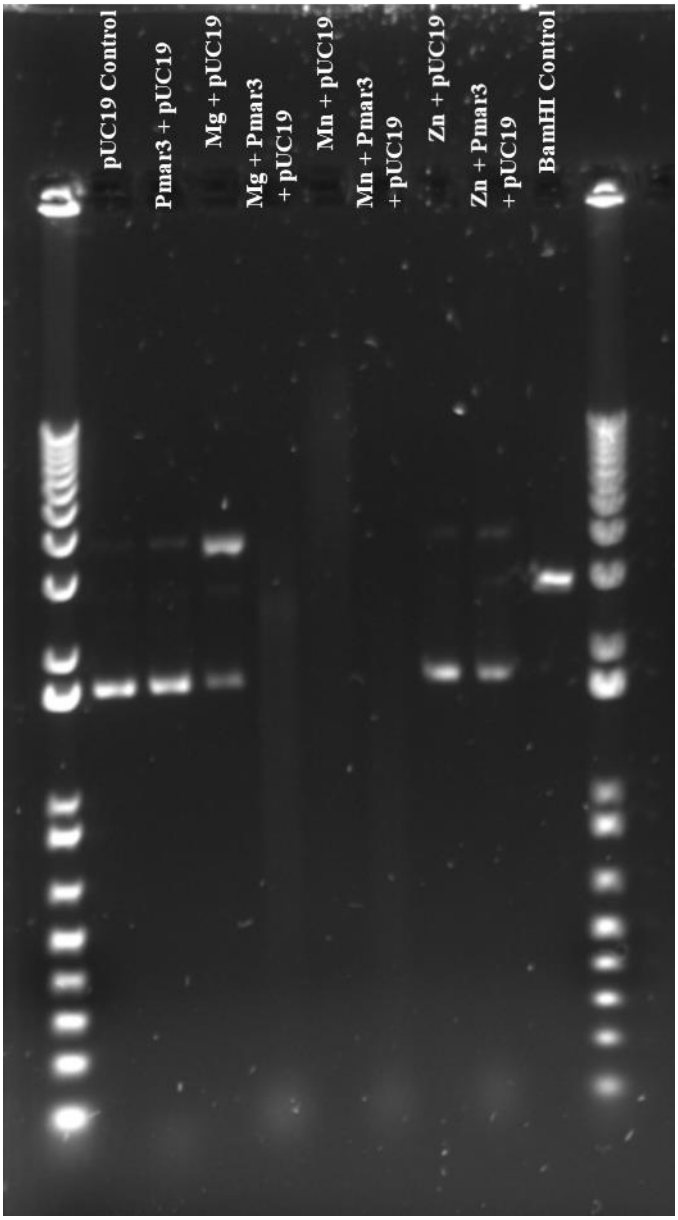


Figure 3-9: 1% TAE agarose gel of a topoisomerase assay where 2.29µM Pmar3 was incubated with 162.4ng pUC19 for 18 hours. Metal-dependent activity was investigated by incubating Pmar3 and pUC19 with 5mM $MgCl_2$, 5mM $MnCl_2$ or 10mM $ZnCl_2$. BamHI was included as a restriction enzyme control to show the difference in apparent size when pUC19 is cut.

Firstly, metal dependency was investigated by incubating Pmar3 and pUC19 with Mg, Mn, and Zn. Mg and Zn have been required for DNA relaxation in different classes of topoisomerase (Champoux, 2001), and were therefore predicted to be essential for Pmar3 activity. Figure 3-9 shows that Pmar3 did not appear to interact with pUC19. There was no increase in size consistent with topoisomerase activity; however complete degradation was observed with Pmar3 + Mg, which was far greater than the control which had a mixture of supercoiled and nicked plasmid. There was no change in the apparent size of the bands observed in the Zn-containing reactions, indicating that Pmar3 does not rely on

Zn to be active. The Mn control showed complete degradation of pUC19 without Pmar3 present, so no conclusions could be drawn on the activity of Pmar3 and Mn with plasmid DNA. Mg alone appeared to cut pUC19 without Pmar3 present, shown by a difference in apparent band size in lane 2. This was later attributed to the homemade topoisomerase buffer containing low-grade Bovine Serum Albumin (BSA), which was omitted from the buffer used in the assays assessing the topoisomerase activity of Pmar5. The degradation of pUC19 suggests that Pmar3 could be displaying nuclease activity.

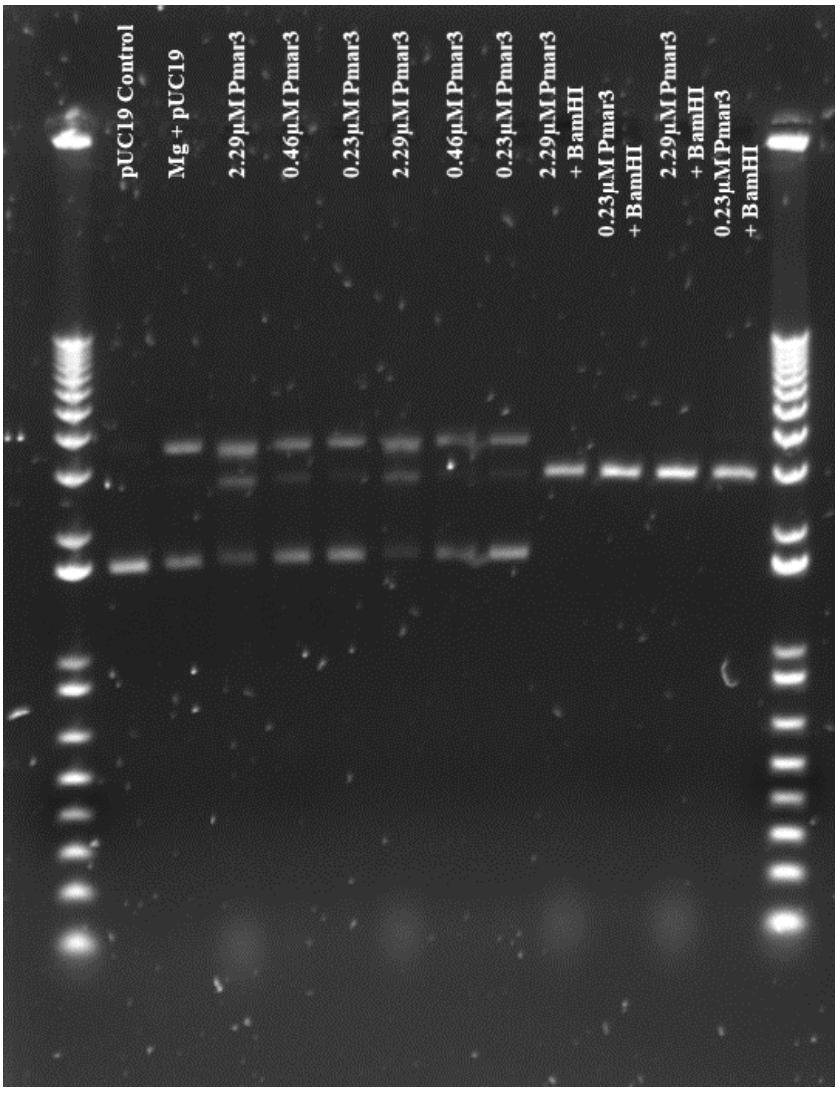


Figure 3-10: 1% TAE agarose gel of a topoisomerase assay where a dilution series of 2.29µM (1x), 0.46µM (5x dilution), and 0.23µM (10x dilution) Pmar3 was incubated with 162.4ng pUC19 and 5mM $MgCl_2$ for 2 hours. Pmar3 was incubated with the restriction enzyme BamHI to check for potential exonuclease activity. All reactions contained pUC19.

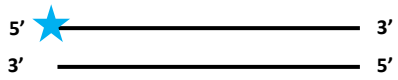
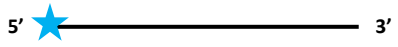
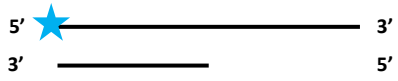
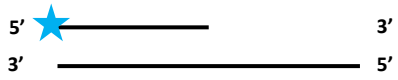
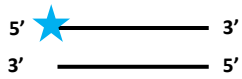
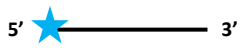
A second assay, shown in Figure 3-10, was run for a shorter time of 2 hours to see if the supercoiling of pUC19 was being relaxed by Pmar3. 5mM $MgCl_2$ was included as a co-factor as it appeared to promote Pmar3 activity in the initial assay. Three dilutions of

Pmar3: 2.29µM (1x), 0.46µM (5x dilution), and 0.23µM (10x dilution), were incubated with pUC19 to observe how the topology of pUC19 changed with different concentrations of Pmar3. The ratio of cut pUC19 to supercoiled pUC19 appeared to decrease with decreasing Pmar3 concentration, suggesting that higher concentrations of Pmar3 are required to relieve supercoiling. Pmar3 was incubated with BamHI, a restriction enzyme, to check if Pmar3 was displaying exonuclease activity. There was no shift in banding compared to the BamHI control reaction, allowing us to conclude that Pmar3 is not an exonuclease. The Pmar3 containing reactions have banding at the bottom of the agarose gel, which shows that pUC19 has continued to be degraded by Pmar3.

It was concluded that Pmar3 is unlikely to be a topoisomerase, as it appears to be the cutting pUC19 plasmid instead of relaxing the supercoiling. The degradation of pUC19 observed in these assays led to the hypothesis that Pmar3 may be a nuclease. The potential nuclease activity of Pmar3 was investigated through a series of fluorescence nuclease assays.

3.3.5 Assessing Nuclease Activity of Pmar3

Table 3.10: Diagrams of the substrates used in fluorescence nuclease assays.

Substrate	Substrate Diagram
Double-strand Matched	
Single-strand	
3’-Tail	
5’-Tail	
20+20 Control	
Single-strand (20) Control	

The nuclease activity of Pmar3 was assessed through a series of fluorescence nuclease assays, where the affinity to and degradation of double-stranded matched, single-stranded, 3’ tail, and 5’ tail substrates (see Table 3.10) by Pmar3 was explored. Pmar3 appeared to degrade the supercoiled pUC19 plasmid and was hypothesised to also be capable of degrading other nucleotide sequences.

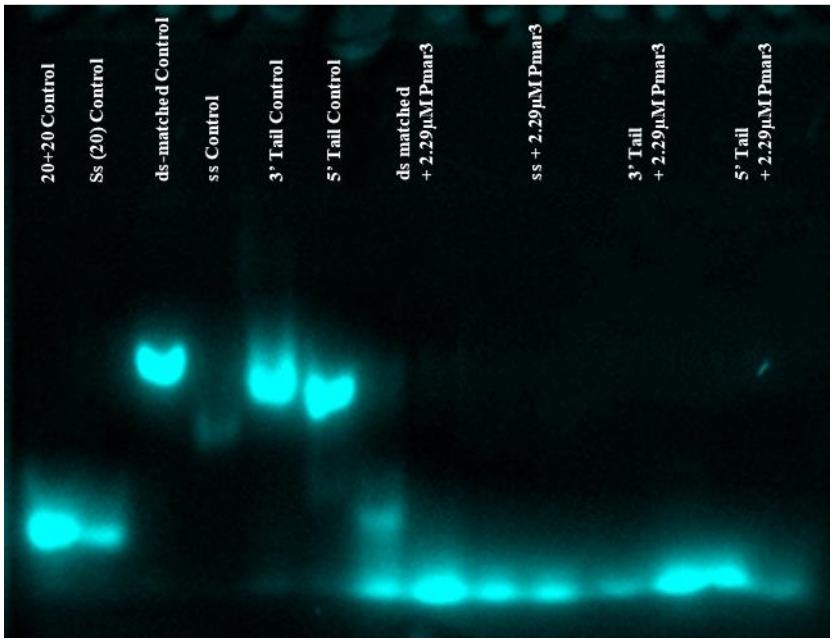
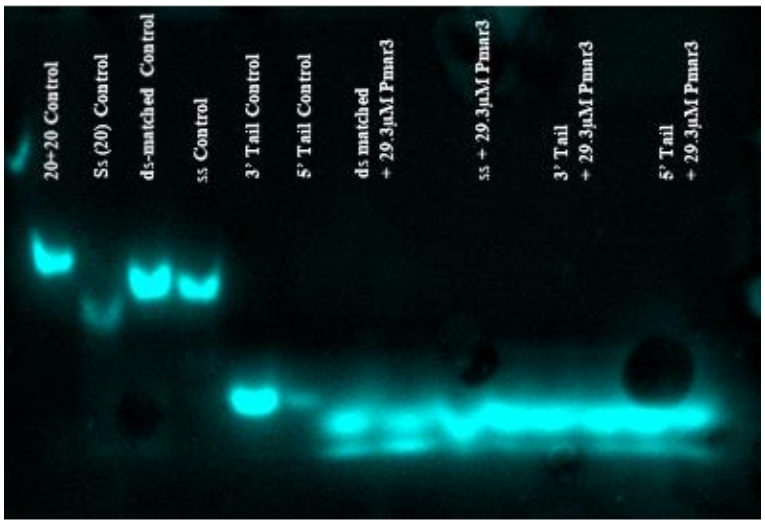


Figure 3-11: 10x Native TBE gel of a fluorescence nuclease assay. 2.29µm Pmar3 was incubated overnight with double-strand matched, single-strand, 3' tail and 5' tail 5'-FAM labelled substrates in duplicates.

An initial assessment of potential nuclease activity was made by incubating Pmar3 with the substrates overnight (19 hours) in duplicates. The first six lanes of Figure 3-11 show the expected size of the substrates without any nuclease activity, and include 20nt double-stranded and single-stranded controls in lanes 5 and 6 respectively. Pmar3 successfully and consistently degraded all the tested substrates, further suggesting that Pmar3 could function as a nuclease. The batch of Pmar3 used in this assay (purification 2) was of low concentration, so this assay was repeated with a 6-hour incubation period, using two higher concentration purified batches of Pmar3 (purification 3) to check for consistent nuclease activity.

(a)



(b)

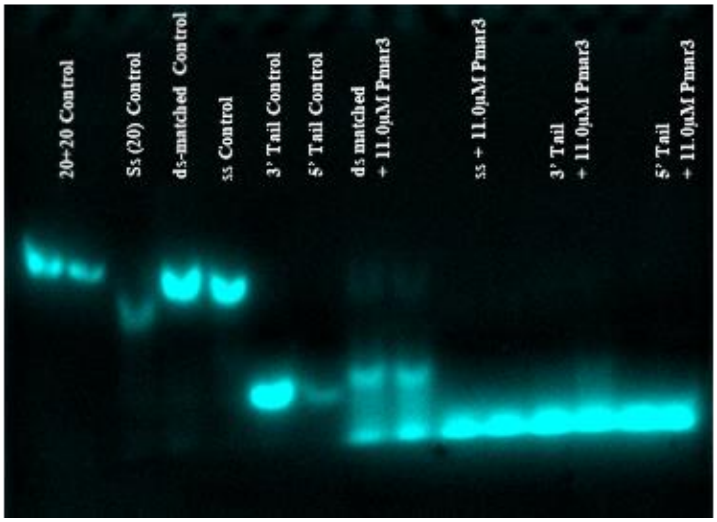


Figure 3-12: 10x Native TBE gels of 6-hour fluorescence nuclease assays comparing the two batches of Pmar3 from purification 3. (a) 1-P3, (b) 2-P3.

The substrates were all consistently degraded by different batches of Pmar3, (shown in Figure 3-12) which further suggests that Pmar3 may be a nuclease. The ds-matched substrate was not completely degraded by the lower concentration (11.0μM) batch but was completely degraded by the higher concentration (29.3μM) batch of Pmar3.

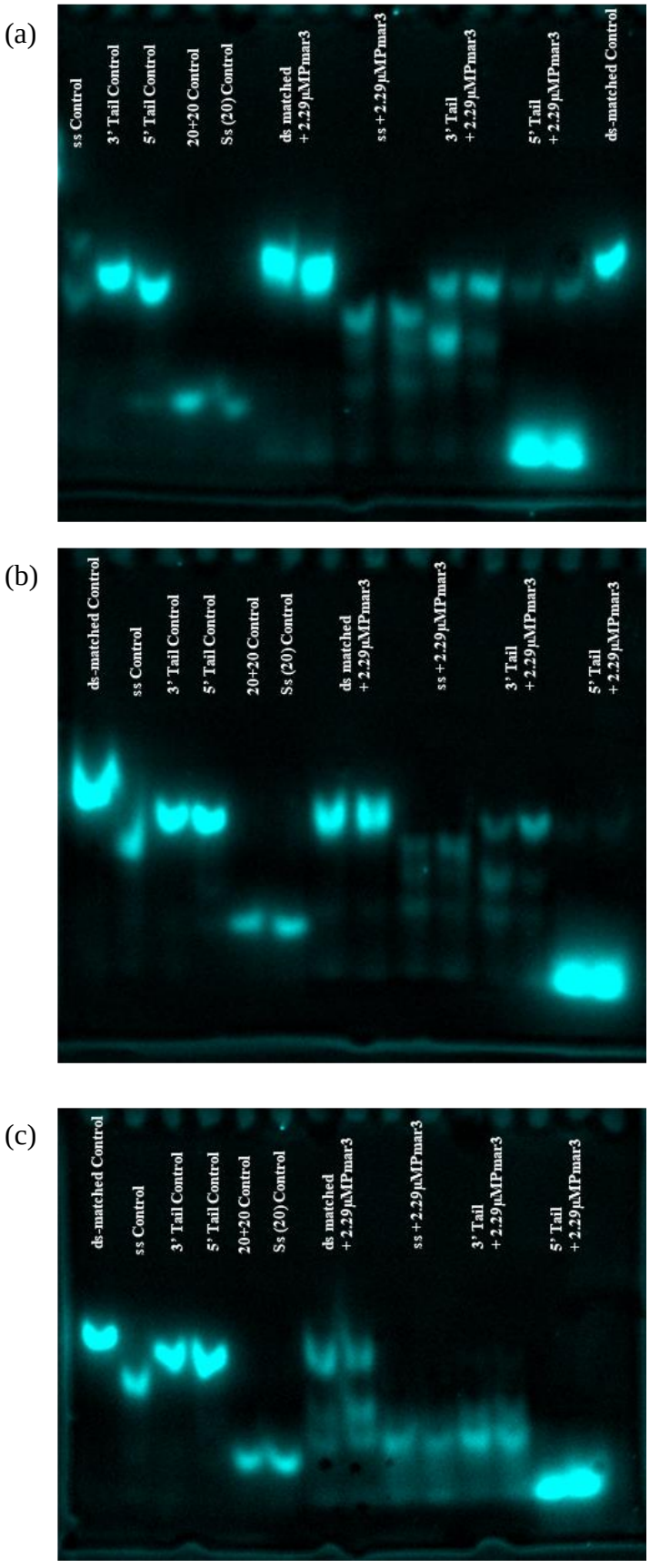


Figure 3-13: 10x Native TBE gels of fluorescence nuclease assays of Pmar3. 2.29µm Pmar3 was incubated with double-strand matched, single-strand, 3' tail and 5' tail 5'-FAM labelled substrates in duplicates for (a) 1 hour, (b) 2 hours, and (c) 4 hours.

The nuclease substrates appeared to be completely degraded when incubated overnight with Pmar3 (purification 2), and for 6 hours with higher concentration Pmar3 (purification 3). The affinity of Pmar3 to different substrates was investigated by observing the changes in the degradation of substrates over different time periods. Figure 3-13 shows the native TBE gels for fluorescence nuclease assay reactions that were incubated for 1, 2, and 4 hours. An increase in degradation over time was observed for all four substrates. Pmar3 appeared to have an affinity for the 5' tail substrate, as this was almost completely degraded after 1 hour of incubation. In contrast, Pmar3 appears to act slower on the ds-matched substrate. Different sized product bands can be observed in the ss and 3'-tail substrate samples in Figure 3-13 (b) which corresponds to both 20 and 40nt products. The affinity towards the ss-matched substrate seen in Figure 3-12 was confirmed by performing a 6-hour assay of a dilution series of Pmar3 incubated with the ds-matched and ss-matched substrates (Figure 3-14).

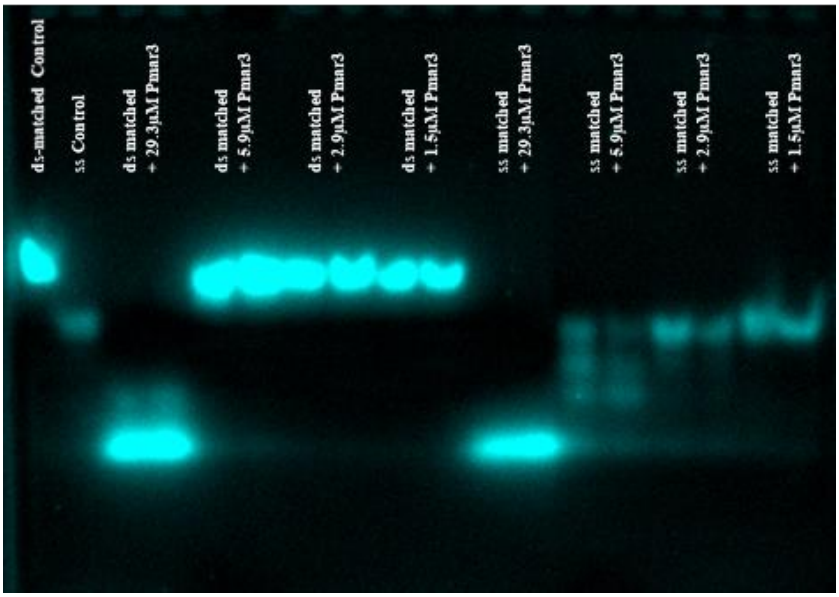


Figure 3-14: 10x Native TBE gel of a 6-hour fluorescence nuclease assay of Pmar3. A dilution series was performed with 1x, 1/5, 1/10 and 1/20 dilutions of 1-P3 (29.3, 5.9, 2.9, 1.5µM).

It was concluded that Pmar3 shows consistent nuclease-like activity. However, nucleases can be difficult to remove from the solution during purifications. To further prove that Pmar3 may be a nuclease, Pmar3 (purification 2) was deactivated and then incubated with the nuclease substrates for 8 hours parallel to active Pmar3. The thermal-shift assay results for Pmar3 indicated that heating Pmar3 above its melting temperature of 41-44^{o}C would deactivate the protein. To ensure deactivation, Pmar3 was heated at 55^{o}C. Parallel assays were completed and compared to observe if any residual nucleases in solution may be responsible for the observed nuclease activity in previous assays.

(a) (b)

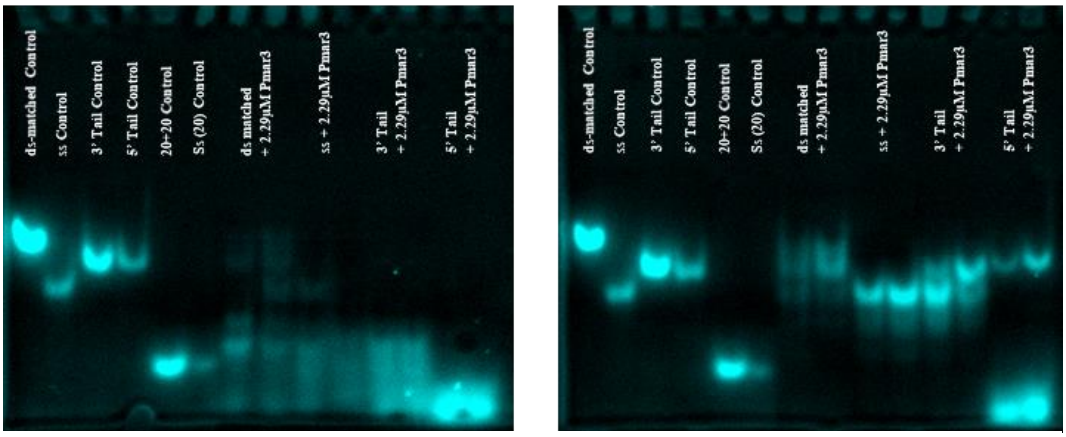


Figure 3-15: 10x Native TBE gels of fluorescence nuclease assays of Pmar3. (a) 2.29μm Pmar3 was incubated with double-strand matched, single-strand, 3' tail and 5' tail 5'-FAM labelled substrates in duplicates for 8 hours. (b) Pmar3 was de-activated by heating at 55^{o}C for 5 minutes. 2.29μm de-activated Pmar3 was incubated with double-strand matched, single-strand, 3' tail and 5' tail 5'-FAM labelled substrates in duplicates for 8 hours.

Figure 3-15 shows a decrease in nuclease activity in the deactivated Pmar3 reactions compared to the active Pmar3 containing reactions. Although the substrates appear to still be slightly degraded in Figure 3-15 (b), the decreased nuclease-like activity upon deactivation shows that Pmar3 is likely to be a nuclease. The trace activity observed could be because not all the Pmar3 was unfolded at 55^{o}C, or Pmar3 potentially re-folded after heating. In the future, a higher melting temperature could be used to increase the successful deactivation of Pmar3.

Observed nuclease activity could be due to residual contaminants that could not be removed during IMAC purification. To exclude this possibility, a mutant of Pmar3 was designed with four point-mutations of aspartate to alanine (positions 46, 75, 78, and 140), which removes the acidic residues that are often conserved in the catalytic domain, and

are essential for nuclease activity (Nishino & Morikawa, 2002). As discussed in the previous chapter, further work is required to optimize the expression of Pmar3-mut. Pmar3-mut will be used in future fluorescence nuclease assays to investigate whether the observed nuclease activity can likely be attributed to Pmar3.

3.3.6 Assessing Topoisomerase Activity of Pmar5

At the commencement of this research, Pmar5 had been annotated as a topoisomerase. This was due to its sequence similarity to a viral topoisomerase. The function of Pmar5 was yet to be confirmed, so we began by assessing its ability to act as a topoisomerase, where it would relieve the supercoiling of a pUC19 plasmid.

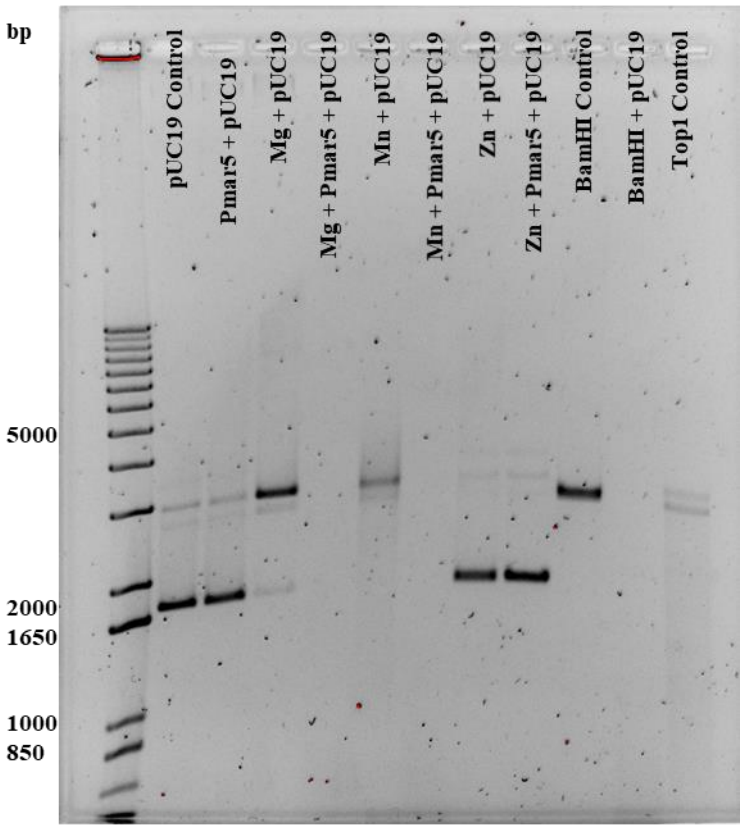


Figure 3-16: 1% TAE agarose gel of a topoisomerase assay where 6.38µM Pmar5 was incubated with 162.4ng pUC19 for 18 hours. Metal-dependent activity was investigated by incubating Pmar5 and pUC19 with 5mM $MgCl_2$, 5mM $MnCl_2$ or 10mM $ZnCl_2$. BamHI was included as a restriction enzyme control to show the gel-shift when pUC19 is cut.

Similarly, to Pmar3, metal-dependency was investigated first by incubating Pmar5 and pUC19 with Mg, Mn, and Zn. Figure 3-16 shows no banding in the fractions containing Pmar5 and pUC19 with Mg, Mn, or BamHI. There was no shift in apparent size observed

in the Zn-containing reactions, indicating that Pmar5 does not rely on Zn to be active. Mg and Mn alone appeared to cut pUC19 without Pmar5 present, shown by an apparent size shift in lanes 3 and 5. Top1 was included as a topoisomerase control but appeared to not relieve the supercoiling of pUC19 as it was expected to. Topoisomerase activity can be visualized as a set of bands decreasing in size appearance, which corresponds to the different topologies of pUC19. The pUC19 control shows the expected running size of the supercoiled topology, and the BamHI control shows the expected running size of the nicked topology. Additional bands between the two controls are expected to be seen in topoisomerase-containing samples. These two bands, decreasing in apparent size, are for the linear and relaxed topologies. A new batch of pUC19 plasmid was purified and used in the remaining assays, given that Top1 did not show clear topoisomerase activity.

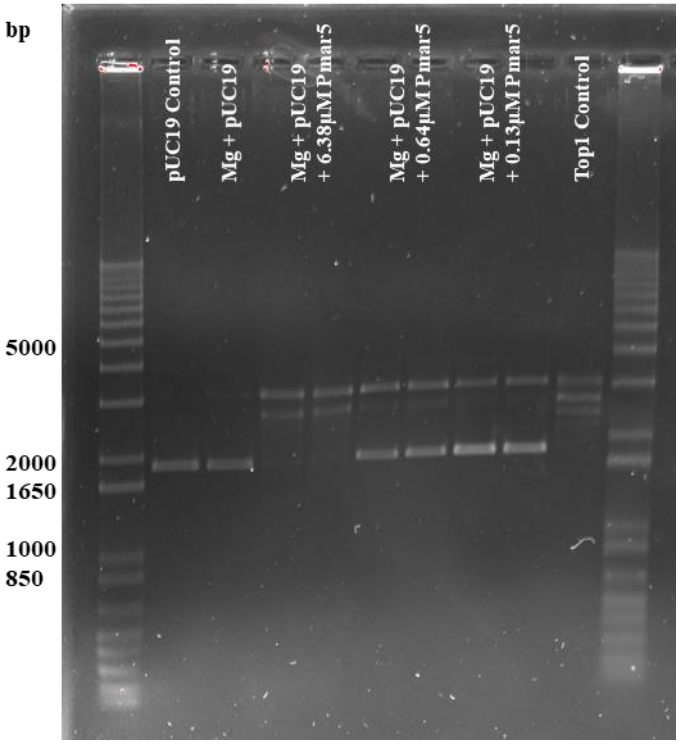


Figure 3-17: 1% TAE agarose gel of a topoisomerase assay where a dilution series of 6.38µM (1x), 0.64µM (10x dilution), and 0.13µM (50x dilution) Pmar5 was incubated with 201.0ng pUC19 and 5mM $MgCl_2$ for 2 hours.

It was suspected that Pmar5 may have Mg-dependent activity, so a second assay shown in Figure 3-17 was run for a shorter time period of 2 hours to see if the supercoiling of pUC19 was being relaxed by Pmar5. Three dilutions of Pmar5: 6.38µM (1x), 0.64µM (10x dilution), and 0.13µM (50x dilution), were incubated with 5mM $MgCl_2$ and 201.0ng pUC19. The ratio of nicked and linear pUC19 to supercoiled pUC19 appeared to decrease

with decreasing Pmar5 concentration, indicating that higher concentrations of Pmar5 are necessary to relieve supercoiling.

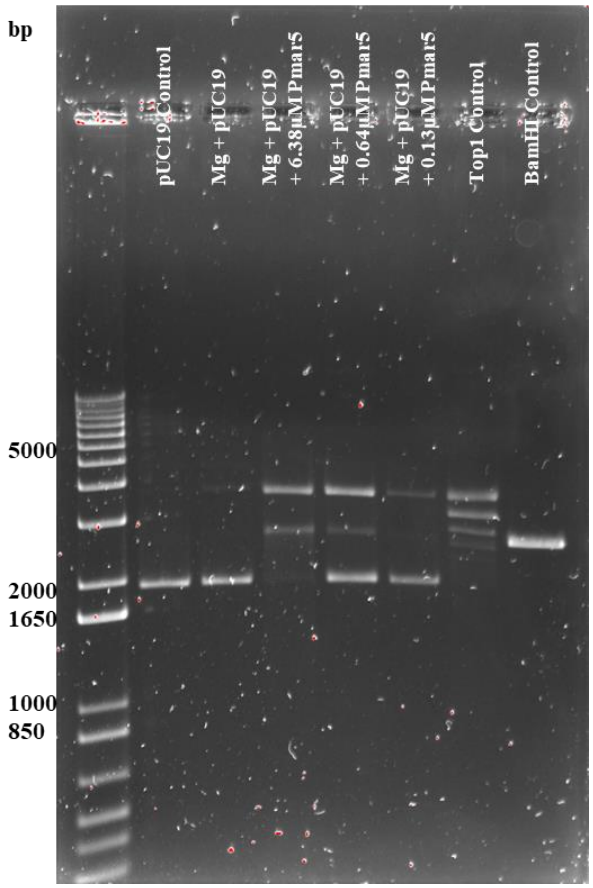


Figure 3-18: 1% TAE agarose gel of a topoisomerase assay where a dilution series of 6.38µM (1x), 0.64µM (10x dilution), and 0.13µM (50x dilution) Pmar5 was incubated with 201.0ng pUC19 and 5mM $MgCl_2$ for 2 hours. A restriction enzyme control (BamHI) was included to confirm the topologies observed in the Pmar5-containing reactions.

The 2-hour assay was repeated with a restriction enzyme control (BamHI) to confirm the topology of the bands in the Pmar5-containing reactions. Regardless of Pmar5 concentration, Pmar5 does not appear to have the same function as Top1, based on the banding observed in Figure 3-18. There are two distinct bands in the 1xPmar5 sample: the top band shows that Pmar5 may be nicking pUC19, and the bottom band shows that Pmar5 appears to be linearizing pUC19 like BamHI. It was hypothesised that Pmar5 may be capable of nicking pUC19 but cannot re-join the DNA to create a relaxed plasmid topology. Some topoisomerases (e.g. Type IIA) are ATP-dependent, so substrate-dependency was investigated by incubating Pmar5 with ATP and GTP.

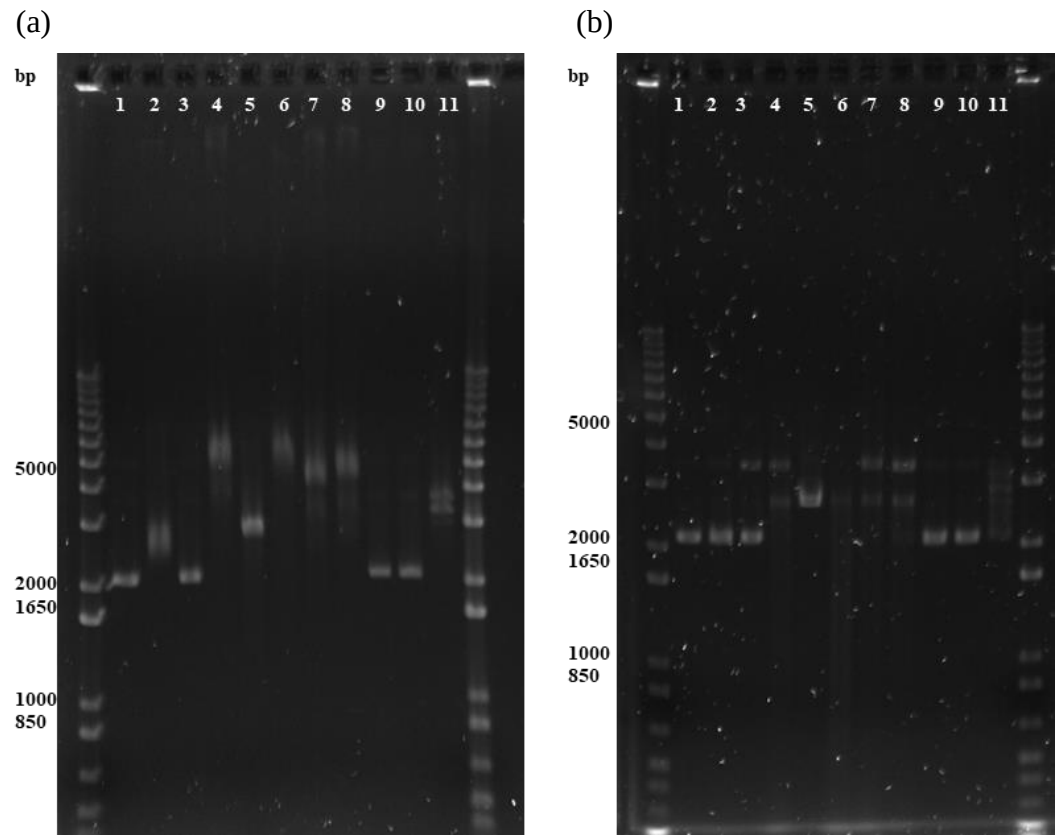


Figure 3-19: 1% TAE agarose gel of a topoisomerase assay where 6.38µM Pmar5 was incubated with 197.2ng pUC19 for 2 hours. Lane 1 = pUC19 control, 2 = Pmar5+pUC19, 3 = Mg control, 4 = Pmar5+Mg+pUC19, 5 = BamHI control, 6 = Pmar5+BamHI, 7 = Pmar5+ATP+Mg+pUC19, 8 = Pmar5+GTP+Mg+pUC19, 9 = ATP+Mg+pUC19, 10 = GTP+Mg+pUC19, 11 = Top1 control. (a) Assay run without proteinase K incubation step. (b) Assay run with proteinase K incubation step.

Figure 3-19 (a and b) compares assays that were run excluding and including a proteinase K step respectively. Topoisomerase assays include a proteinase K step to degrade the protein and leave only the native structure of the plasmid visible on the gel. By excluding proteinase K, we can see any protein-DNA complexes that have formed. Figure 3-19 (a) confirms that Pmar5 is binding to pUC19. Figure 3-19 (b) shows that Pmar5 activity is not enhanced by ATP or GTP, as lanes 7 and 8 (ATP and GTP) have similar banding ratios to lane 4 (no substrates). Pmar5 consistently recognizes the coiled topology of pUC19 and cuts the plasmid. However, Pmar5 does not seem to act further on pUC19 with a linear topology to fully relieve the supercoiling like Top1. A more extensive time series of plasmid-based reactions will be carried out in the future to identify if Pmar5 has the potential to carry out complete topoisomerase activity after a certain period of time.

We hypothesised that Pmar5 may be part of a more extensive repair pathway and may be capable of relieving supercoiling as part of a multi-protein system, and thus form a complex with the DNA substrate. To test this hypothesis, Pmar5 was incubated with

LigW and Pmar3 to see if there were any changes in apparent DNA size observed. Mg was omitted so binding and metal-independent interaction could be observed.

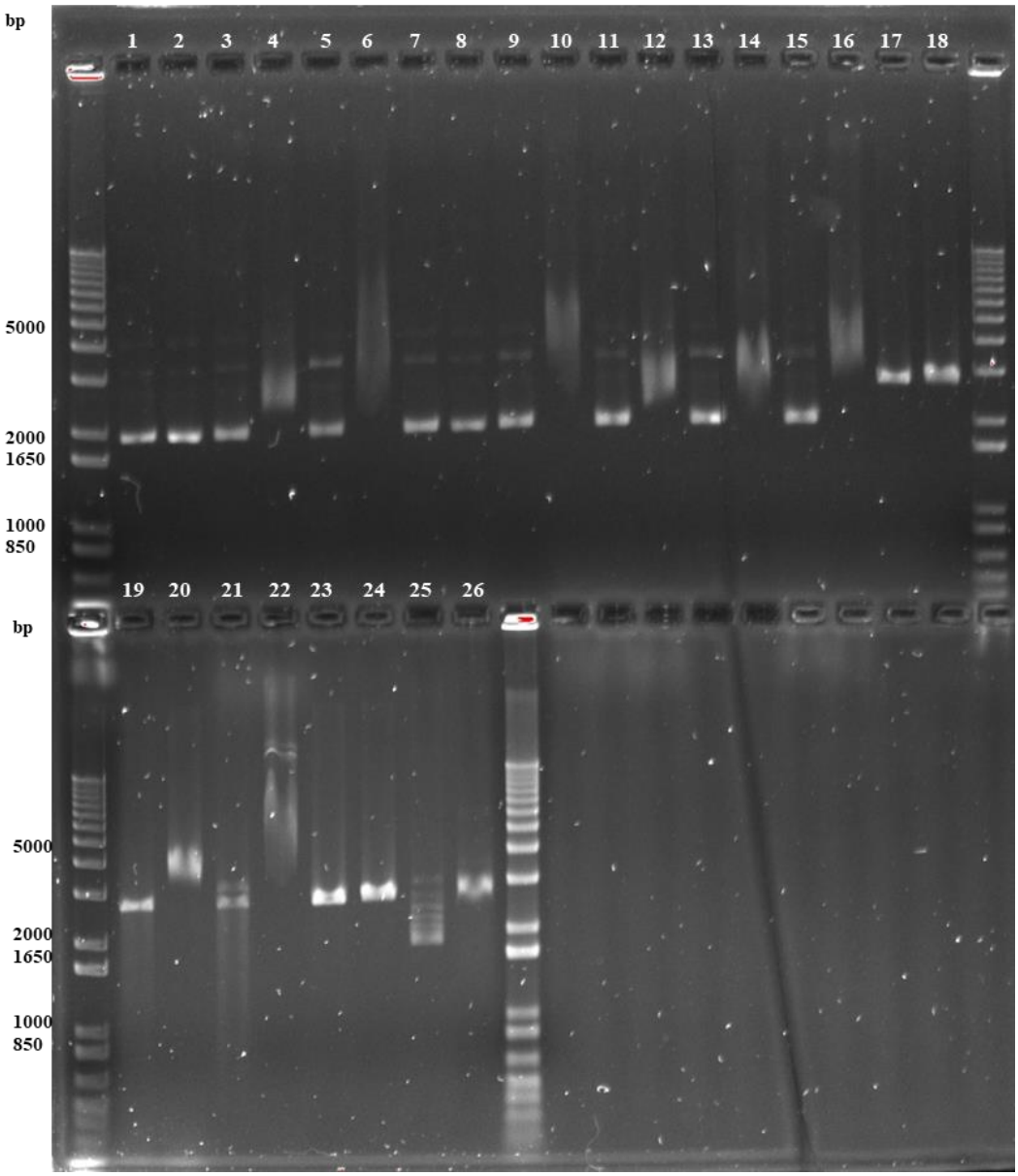


Figure 3-20: 1% TAE agarose gel of a topoisomerase assay where 5µM Pmar3, Pmar5 or LigW was incubated with 180.3ng pUC19 for 2 hours. Samples were loaded in order of with proteinase K incubation, then without proteinase K incubation. 1 & 2 = pUC19 control, 3 & 4 = Pmar5 + pUC19, 5 & 6 = LigW + pUC19, 7 & 8 = Pmar3 + pUC19, 9 & 10 = Pmar5, LigW + pUC19, 11 and 12 = Pmar5, Pmar3 + pUC19, 13 & 14 = LigW, Pmar3 + pUC19, 15 & 16 = Pmar5, LigW, Pmar3 + pUC19, 17 & 18 = BamHI + pUC19, 19 & 20 = Pmar5 + BamHI + pUC19, 21 & 22 = LigW + BamHI + pUC19, 23 & 24 = Pmar3 + BamHI + pUC19, 25 & 26 = Top1 + pUC19.

In Figure 3-20, Top1 (lane 25) successfully relaxed the supercoiled plasmid. The pUC19 control (lanes 1 and 2) shows that the pUC19 plasmid had no degradation that would be caused by nuclease contamination, and there was no gel-shift activity caused by proteinase K (lane 1) or the buffer conditions. Lanes 4 shows that Pmar5 bound to pUC19,

which was expected given the previous results. No cutting or topoisomerase activity was observed in the absence of $MgCl_2$, confirming that Pmar5 activity is Mg-dependent. In contrast to Pmar5, LigW bound to pUC19 and caused a gel-shift only (lane 6), whereas Pmar3 did not bind or cut the plasmid, which shows that both LigW and Pmar3 are also Mg-dependent. Pmar3 did not bind to the plasmid in the presence of LigW or Pmar5. Lane 10 shows a similar shift of Pmar5 and LigW compared to LigW alone, meaning that LigW may be blocking Pmar5 from binding to pUC19. However, in lane 12, Pmar3 does not appear to block Pmar5 from binding to pUC19. A shift is observed in lane 14 that is likely due to LigW binding. However, this shift appears to be smaller than with LigW alone (lane 6), suggesting Pmar3 may be possibly blocking LigW from fully binding to pUC19. Lanes 15 and 16 contain Pmar3, Pmar5, and LigW. There appears to be no change in activity compared to Pmar5 alone, and the shift in lane 16 is similar to the shift observed with only LigW. BamHI cut pUC19 but did not remain bound, as there is no gel-shift in lane 18. Lanes 19 and 20 show that BamHI cut pUC19 to produce a linear plasmid, and Pmar5 successfully bound to the linear plasmid. This shows that Pmar5 can bind to both supercoiled and linear plasmids. LigW and BamHI appeared to work together by cutting and re-ligating the plasmid, shown by the double-banding in lane 21. In contrast, Pmar3 did not bind to the linear plasmid.

Despite assessing Pmar3 for topoisomerase activity previously, it appears that Pmar3 does not bind to pUC19 or enhance the binding and activity of Pmar5. Both LigW and Pmar5 bound to pUC19 separately yet did not appear to bind together. LigW did not enhance the topoisomerase activity of Pmar5. These findings suggest that an undetermined co-factor or protein may be required for pUC19 to relieve supercoiling to completion, if this is in fact its biological function. In the future, this assay will be repeated with Mg present, as Pmar3, Pmar5, and LigW have metal-dependent activity. Once expressed and purified, Pmar4 will also be tested in combination with Pmar5 to investigate the relationship between these proteins.

3.3.7 Coupled Ligation Assays

LigW has been found to have significantly lower rates of ligation activity compared to other ATP-dependent ligases in *P. marinus* MIT9312 (Williamson, Unpublished). This led to the hypothesis that the proteins on the LigW operon (Pmar3, Pmar4, and Pmar5)

may play a role in enhancing the ligation activity of LigW. The ligation activity of Pmar3 and Pmar5 was assessed by carrying out coupled ligation assays with LigW.

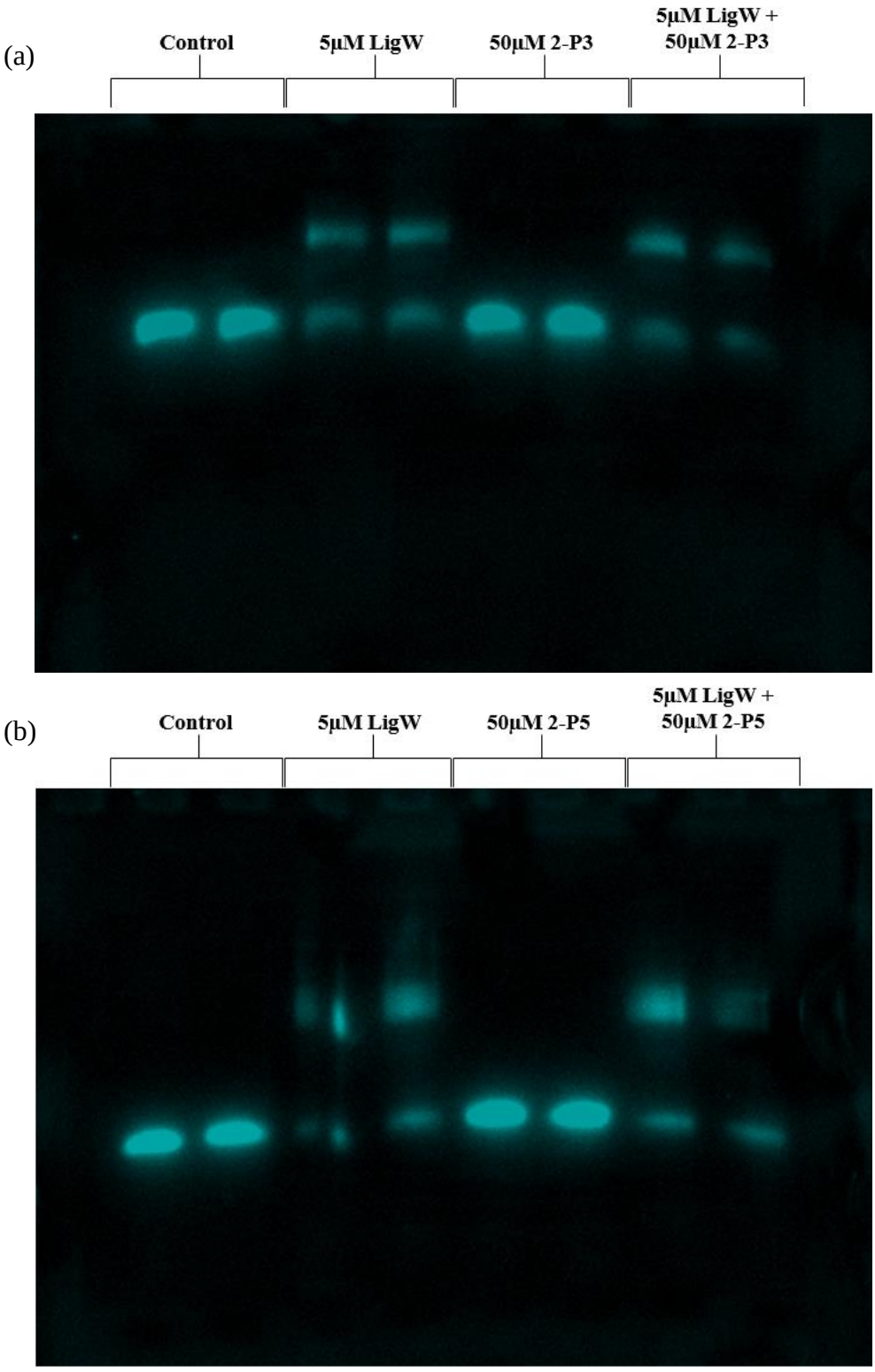


Figure 3-21: 7M Urea gels of coupled ligation assays. LigW was incubated with a nicked substrate and (a) Pmar3, and (b) Pmar5, to assess how each adjacent protein interacts with LigW and enhances ligation activity.

Figure 3-21 shows that LigW successfully sealed the nick in the nicked substrate, shown by a gel shift upwards in the LigW-containing reactions. Pmar3 and Pmar5 appear to not display any ligation activity individually, which was expected given the proposed nuclease and topoisomerase activities observed respectively. The ratio of nicked to repaired substrate appeared to be identical and unchanged by both Pmar3 and Pmar5, confirming that ligation activity is not enhanced by these proteins. Other co-factors may be required to enhance LigW activity and to activate Pmar3 or Pmar5. Co-factor dependency will be explored in the future by repeating these assays with metals and ATP. The tested conditions potentially were not favourable for LigW, so this assay will be repeated over a range of temperatures and for different time periods. Pmar4 will be tested with LigW in the future in addition to testing combinations of Pmar3, Pmar4, and Pmar5, to identify any relationships between the proteins that may enhance ligation.

3.4 Conclusions

Based on the characterization completed during this research, Pmar3 is suspected to be a nuclease and Pmar5 shows similar activity to a topoisomerase. Pmar3 appears to be a monomeric nuclease with an affinity for single-stranded DNA. Pmar3 had Mg-dependent DNA cutting activity and therefore may require Mg or other undetermined co-factors to help increase LigW activity. Pmar5 appears to be a pentameric topoisomerase-like protein. Pmar5 can recognize and cut pUC19 with a supercoiled topology, but it doesn't relieve supercoiling entirely to give a linear product. Like Pmar3, Pmar5 has Mg-dependent activity, and therefore may also require Mg or other undetermined co-factors to help increase LigW activity.

3.5 Future Research

The suspected nuclease activity of Pmar3 will be confirmed by repeating fluorescence nuclease assays with Pmar3-mut to confirm whether the nuclease activity is caused by Pmar3 or contaminants. Biochemical characterization of Pmar4 will be carried out in the future once Pmar4 is successfully purified. Analytical size-exclusion will be carried out to determine the molecular weight of Pmar4. Coupled ligation assays involving Pmar4 and LigW in addition to combinations of all three proteins with LigW will be completed to conclude if any of the proteins are responsible for increasing LigW's ligation activity.

The DNA repair pathway involving Pmar3, Pmar4, Pmar5, and/or LigW remains unknown. Further topoisomerase assays will be completed which include Pmar4 coupled with Pmar5. It is hypothesised that Pmar5 and Pmar4 may be components of a two-protein mechanism for recognizing supercoiled and linear topology respectively. Alternatively, Pmar5 may achieve complete topoisomerase function with an unidentified co-factor. Given the inconclusive relationship of Pmar5 and LigW, and that Pmar5 can bind to pUC19 but not the EMSA nucleotide substrates, the DNA-binding activities of Pmar5 will be investigated further. Gel filtration of Pmar5, LigW, and pUC19 will be run to check whether both Pmar5 and LigW bind to pUC19.

Chapter 4

Conclusions and Future Research

Prochlorococcus marinus, is one of the most abundant photosynthetic organisms in the ocean. *P. marinus* is thought to contribute towards approximately 8.5% of global ocean primary productivity (Laurenceau et al., 2020). *P. marinus* is of interest because of the cell's simple and unique photosynthetic system, efficient carbon concentrating mechanism, and minimal streamlined genome. *P. marinus* strain MIT9312 thrives in the upper euphotic zone where there are minimal nutrients and high levels of UV exposure, which would typically put an organism at risk. Little is known about the DNA repair genes, proteins, and pathways utilised by *Prochlorococcus* that enable it to survive in these conditions. Researching how DNA repair enzymes among others within *Prochlorococcus* function could help us to understand how it survives under increasingly more harmful conditions, such as those caused by climate change.

LigW was found to have low DNA ligase activity on nicked DNA in comparison to other ATP-dependent ligases in *P. marinus* strain MIT9312. The low levels of activity shown by LigW make its retention in the small *Prochlorococcus* genome unusual. LigW, an ATP-dependent ligase, appears to be in an operon with other proteins which were predicted to promote LigW activity or have their own DNA modifying activities (Williamson, Unpublished). The three proteins of interest were annotated as Pmar3, Pmar4, and Pmar5, and were the focus of this research. The LigW operon is unique to the MIT9312 strain and is part of the flexible genome that makes up the *Prochlorococcus* pan-genome. Since the commencement of this project, 13 new HLII isolates of *Prochlorococcus* were discovered in the western Pacific Ocean and the South China Sea by a group from the China University of Geoscience (NCBI accession number: PRJNA611498). LigW, Pmar3, and Pmar4 were identified in the sequence of one of these strains, named the XMU1402 strain. Pmar5 does not appear to be present in any of the new strains. This presents new opportunities for comparative research between the two XMU1402 and MIT9312 strains and highlights how active and fast-moving this field of research is.

Pmar3 was successfully expressed and purified for characterization. The DNA-specific binding activity of Pmar3 remains to be understood, as Pmar3 did not bind to supercoiled

or linear pUC19, or the double-stranded, single-stranded, 3'-tail, and 5'-tail DNA substrates. However, Pmar3 showed consistent nuclease activity by degrading double-stranded, single-stranded, 3'-tail, and 5'-tail DNA substrates across multiple purification batches and concentrations. A mutant of Pmar3 was designed with four point-mutations of aspartate to alanine, to remove the acidic residues that are often conserved in the catalytic domain, and are essential for nuclease activity (Nishino & Morikawa, 2002). Although successfully transformed, Pmar3-mut was not able to be expressed, meaning comparative nuclease assays could not be completed to confirm the nuclease activity of Pmar3. The DNA-binding activity and specificity of Pmar3 will be explored further. The nuclease activity of Pmar3 will be confirmed by repeating fluorescence nuclease assays with Pmar3-mut to confirm whether the nuclease activity is caused by Pmar3 or contaminants. Soluble expression and purification of Pmar3-mut will be optimized to identify favourable conditions.

Challenges were encountered when attempting to express and purify Pmar4. Soluble expression of Pmar4 was not achieved, but the insoluble expression of Pmar4-His was. Attempts were made to denature and re-fold Pmar4-His using a Triton-X and urea-based IMAC protocol (Palmer & Wingfield, 2012). Pmar4 was unable to be re-folded under the tested conditions and therefore was not characterized. Optimization of soluble expression and purification of Pmar4 will be continued. Pmar4 will be transformed into different expression vectors, to promote soluble expression. Biochemical characterization of Pmar4 will be carried out in the future once Pmar4 is successfully purified.

Pmar5 was successfully expressed and purified for characterization. Extensive optimization of the Pmar5 purification protocol was essential, as de-tagged Pmar5 was unable to be separated from MBP-tagged Pmar5 by size exclusion. This was later attributed to Pmar5 being a pentameric protein, which was calculated and confirmed by analytical size exclusion. At the commencement of this project, Pmar5 was annotated as a topoisomerase based on sequence homology. Characterization of Pmar5 confirmed that Pmar5 has partial Mg-dependent topoisomerase activity. Pmar5 consistently bound to and nicked supercoiled pUC19 but was unable to completely relax the supercoils like a fully functional topoisomerase due to its inability to re-join the DNA. Complete topoisomerase activity was not achieved with ATP or GTP co-factors or with Pmar3 and LigW. Further topoisomerase assays will be completed which include Pmar4 coupled with Pmar5 and other co-factors to help identify what Pmar5 requires to completely relieve the

supercoiling of DNA. The DNA-binding activity and specificity of Pmar5 will also be investigated further.

Pmar3, Pmar4, and Pmar5 were hypothesised to have DNA-modifying functions, including potentially enhancing LigW's poor ligation activity. Coupled ligation assays of Pmar3 and Pmar5 with LigW showed no increase in ligation activity under the tested conditions. Undetermined co-factors may be required to promote ligation, but at present, the relationship of LigW with each protein remains to be determined. Coupled ligation assays involving Pmar4 and LigW in addition to combinations of all three proteins with LigW will be completed to conclude if any of the proteins are responsible for increasing LigW's ligation activity. Additionally, crystallization would provide further insight into the possible function of these proteins and would allow us to detect more distant relationships that these proteins may have. Time constraints did not allow for the structural characterization of any of the *Prochlorococcus* proteins. Crystallization requires high concentrations of pure protein, and therefore the optimization of each protein's purification protocol will be the first step taken towards structural characterization. Crystallization trials will begin with robot screens to identify any favourable conditions that can be used for obtaining crystals for all three proteins.

Attempts to understand each protein's function beyond DNA repair will be made by attempting to construct functional *in vivo* knockouts of each protein in different strains of *P. marinus*. This will be done to assess the impact of how these changes may affect the organism's ability to survive under light stress. A protocol for gene knockouts in *P. marinus* is currently being developed by The Chisholm Lab with success still to be achieved (Laurenceau et al., 2020). An alternative approach that may be explored is carrying out *in vivo* knock-ins of each protein into *Synechococcus*. This could allow us to observe how the proteins impact the organism's ability to respond to environmental stress. Cyanobacteria have proven difficult to manipulate compared to other species of bacteria such as *Escherichia coli* or *Bacillus subtilis*. This is due to their circadian rhythms, restriction and modification systems, and the diversity in ecology, genetics, physiology, and morphology that can be seen within a single genus, such as *Prochlorococcus* (Taton et al., 2014). Target gene regulation and expression in cyanobacteria have been studied in *Synechococcus* sp. strain PCC 7002. A promising anhydrotetracycline-dependent inducible sRNA system has been developed, which could be useful for manipulating cyanobacteria to express a variety of synthetic plasmids (Zess et al., 2016).

Synechococcus elongatus strain PCC7942 is another genetically manipulable strain, which uses homologous recombination to integrate heterologous sequences into the chromosome at knock-in target sites (Taton et al., 2014).

The novel DNA repair pathway involving Pmar3, Pmar4, Pmar5, and/or LigW remains yet to be determined. Biochemical and structural characterization of each protein is still required, but Pmar3 and Pmar5 are likely to be a nuclease and topoisomerase respectively. The research undertaken during this project and the future steps that will be taken will make significant contributions towards fully characterizing the unique LigW and its adjacent proteins. LigW at present has been successfully characterized as an ATP-dependent ligase unique to *P. marinus* MIT9312, yet the extent of its role in DNA repair remains to be understood. This research is expected to lead to the discovery of other unidentified proteins in *Prochlorococcus* that aid its survival, as well as in other bacterial species. Understanding the function of these proteins is significant for being able to identify the DNA repair pathways used by *Prochlorococcus*. Understanding how DNA repair enzymes among others within *Prochlorococcus* function could help us to utilize and engineer bacteria to survive under increasingly more harmful conditions, such as those caused by climate change. Once the LigW DNA repair system is characterized, then this can be more easily identified and understood in other organisms where it may not be detected at present.

References

Bai, N., Roder, H., Dickson, A., & Karanicolas, J. (2019). Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Scientific Reports, 9*(1), 2650. https://doi.org/10.1038/s41598-018-37072-x

Bertrand, C., Thibessard, A., Bruand, C., Lecointe, F., & Leblond, P. (2019). Bacterial NHEJ: a never ending story. *Molecular microbiology, 111*(5), 1139-1151.

Biller, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). Prochlorococcus: the structure and function of collective diversity. *Nature Reviews Microbiology, 13*(1), 13-27.

Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods Enzymol, 326*, 245-254. https://doi.org/10.1016/s0076-6879(00)26058-8

Bowater, R., & Doherty, A. J. (2006). Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS genetics, 2*(2), e8.

Bryant, D. A. (2003). The beauty in small things revealed. *Proceedings of the National Academy of Sciences, 100*(17), 9647-9649.

Burgess, R. R. (2018). A brief practical review of size exclusion chromatography: rules of thumb, limitations, and troubleshooting. *Protein expression and purification, 150*, 81-85.

Candotti, M., Esteban-Martín, S., Salvatella, X., & Orozco, M. (2013). Toward an atomistic description of the urea-denatured state of proteins. *Proceedings of the National Academy of Sciences, 110*(15), 5933-5938.

Cassier-Chauvat, C., & Chauvat, F. (2015). Responses to oxidative and heavy metal stresses in cyanobacteria: recent advances. *International journal of molecular sciences, 16*(1), 871-886.

Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry, 70*(1), 369-413.

Cheng, C., & Shuman, S. (1997). Characterization of an ATP-Dependent DNA Ligase Encoded by Haemophilus Influenzae. *Nucleic Acids Research, 25*(7), 1369-1374. https://doi.org/10.1093/nar/25.7.1369

Coleman, M. L., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., Delong, E. F., & Chisholm, S. W. (2006). Genomic islands and the ecology and evolution of Prochlorococcus. *Science, 311*(5768), 1768-1770. https://doi.org/10.1126/science.1122050

Doherty, A. J., & Suh, S. W. (2000). Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res, 28*(21), 4051-4058. https://doi.org/10.1093/nar/28.21.4051

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., Le Gall, F., Makarova, K. S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I. B., Scanlan, D. J., de Marsac, N. T., Weissenbach, J., Wincker, P., Wolf, Y. I., & Hess, W. R. (2003). Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences, 100*(17), 10020-10025. https://doi.org/10.1073/pnas.1733211100

Ejaz, A., & Shuman, S. (2018). Characterization of Lhr-Core DNA helicase and manganese-dependent DNA nuclease components of a bacterial gene cluster encoding nucleic acid repair enzymes. *Journal of Biological Chemistry, 293*(45), 17491-17504.

Fedeles, B. I., Singh, V., Delaney, J. C., Li, D., & Essigmann, J. M. (2015). The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *Journal of Biological Chemistry, 290*(34), 20734-20742.

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., & Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc Natl Acad Sci U S A, 110*(24), 9824-9829. https://doi.org/10.1073/pnas.1307701110

Francis, D. M., & Page, R. (2010). Strategies to optimize protein expression in E. coli. *Current protocols in protein science, 61*(1), 5.24. 21-25.24. 29.

Garrett, R., & Grisham, C. M. (2017). *Biochemistry*. Cengage Learning.

Gong, C., Martins, A., Bongiorno, P., Glickman, M., & Shuman, S. (2004). Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *Journal of Biological Chemistry, 279*(20), 20594-20606.

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature protocols, 2*(8), 1849-1861.

Hess, W. R. (2011). Cyanobacterial genomics for ecology and biotechnology. *Current opinion in microbiology, 14*(5), 608-614.

Hjerde, E., Maguren, A., Rzoska-Smith, E., Kirby, B., & Williamson, A. (2020). DNA ligases of Prochlorococcus marinus: an evolutionary exception to the rules of replication. *BioRxiv*.

Hu, J., Selby, C. P., Adar, S., Adebali, O., & Sancar, A. (2017). Molecular mechanisms and genomic maps of DNA excision repair in Escherichia coli and humans. *Journal of Biological Chemistry, 292*(38), 15588-15597.

Janson, J.-C. (2012). *Protein purification: principles, high resolution methods, and applications* (Vol. 151). John Wiley & Sons.

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., & Chisholm, S. W. (2006). Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. *Science, 311*(5768), 1737-1740.

Kent, A. G., Dupont, C. L., Yooseph, S., & Martiny, A. C. (2016). Global biogeography of Prochlorococcus genome diversity in the surface ocean. *The ISME journal*, *10*(8), 1856-1865.

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., & Johnson, J. (2007). Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS genetics*, *3*(12), e231.

Kreuzer, K. N. (2013). DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harb Perspect Biol*, *5*(11), a012674. https://doi.org/10.1101/cshperspect.a012674

Kurthkoti, K., Kumar, P., Sang, P. B., & Varshney, U. (2020). Base excision repair pathways of bacteria: new promise for an old problem. *Future medicinal chemistry*, *12*(04), 339-355.

Laurenceau, R., Bliem, C., Osburne, M. S., Becker, J. W., Biller, S. J., Cubillos-Ruiz, A., & Chisholm, S. W. (2020). Toward a genetic system in the marine cyanobacterium Prochlorococcus. *Access microbiology*, *2*(4).

Little, J. W., & Mount, D. W. (1982). The SOS regulatory system of Escherichia coli. *Cell*, *29*(1), 11-22.

Lohman, G. J. S., Tabor, S., & Nichols, N. M. (2011). DNA Ligases. *Current Protocols in Molecular Biology*, *94*(1), 3.14.11-13.14.17. https://doi.org/https://doi.org/10.1002/0471142727.mb0314s94

Ma, H., Holub, D., Gillet, N., Kaeser, G., Thoulass, K., Elstner, M., Krauß, N., & Lamparter, T. (2019). Two aspartate residues close to the lesion binding site of Agrobacterium (6 - 4) photolyase are required for Mg2+ stimulation of DNA repair. *The FEBS journal*, *286*(9), 1765-1779.

Makarova, K. S., & Koonin, E. V. (2013). Archaeology of eukaryotic DNA replication. *Cold Spring Harbor perspectives in biology*, *5*(11), a012963.

Margison, G. P., Butt, A., Pearson, S. J., Wharton, S., Watson, A. J., Marriott, A., Caetano, C. M., Hollins, J. J., Rukazenkova, N., & Begum, G. (2007). Alkyltransferase-like proteins. *DNA repair*, *6*(8), 1222-1228.

Maslowska, K. H., Makiela - Dzbenska, K., & Fijalkowska, I. J. (2019). The SOS system: a complex and tightly regulated response to DNA damage. *Environmental and molecular mutagenesis*, *60*(4), 368-384.

Michel, B. (2005). After 30 years of study, the bacterial SOS response still surprises us. *PLoS biology*, *3*(7), e255.

Moore, L. R., Rocap, G., & Chisholm, S. W. (1998). Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. *Nature*, *393*(6684), 464-467.

Morita, R., Nakane, S., Shimada, A., Inoue, M., Iino, H., Wakamatsu, T., Fukui, K., Nakagawa, N., Masui, R., & Kuramitsu, S. (2010). Molecular mechanisms of the

whole DNA repair system: a comparison of bacterial and eukaryotic systems. *Journal of nucleic acids, 2010*.

Niiranen, L., Espelid, S., Karlsen, C. R., Mustonen, M., Paulsen, S. M., Heikinheimo, P., & Willassen, N. P. (2007). Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli. *Protein expression and purification*, *52*(1), 210-218.

Nishino, T., & Morikawa, K. (2002). Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene*, *21*(58), 9022-9032. https://doi.org/10.1038/sj.onc.1206135

Palmer, I., & Wingfield, P. T. (2012). Preparation and extraction of insoluble (inclusion‐body) proteins from Escherichia coli. *Current protocols in protein science*, *70*(1), 6.3. 1-6.3. 20.

Partensky, F., & Garczarek, L. (2010). Prochlorococcus: advantages and limits of minimalism. *Annual Review of Marine Science*, *2*, 305-331.

Partensky, F., Hess, W. R., & Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology and molecular biology reviews*, *63*(1), 106-127.

Pathak, J., Singh, P. R., Häder, D. P., & Sinha, R. P. (2019). UV-induced DNA damage and repair: A cyanobacterial perspective. *Plant Gene*, *19*, 100194.

Paul, S., Dutta, A., Bag, S. K., Das, S., & Dutta, C. (2010). Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria Prochlorococcus. *BMC genomics*, *11*(1), 1-15.

Pitcher, R. S., Brissett, N. C., & Doherty, A. J. (2007). Nonhomologous end-joining in bacteria: a microbial perspective. *Annu. Rev. Microbiol.*, *61*, 259-282.

Prabha, R., Singh, D. P., Gupta, S. K., & Rai, A. (2014). Whole genome phylogeny of Prochlorococcus marinus group of cyanobacteria: genome alignment and overlapping gene approach. *Interdisciplinary Sciences: Computational Life Sciences*, *6*(2), 149-157.

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., & Chisholm, S. W. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*, *424*(6952), 1042-1047. https://doi.org/10.1038/nature01947

Roth-Rosenberg, D., Aharonovich, D., Luzzatto-Knaan, T., Vogts, A., Zoccarato, L., Eigemann, F., Nago, N., Grossart, H.-P., Voss, M., & Sher, D. (2020). Prochlorococcus cells rely on microbial interactions rather than on chlorotic resting stages to survive long-term nutrient starvation. *mBio*, *11*(4), e01846-01820.

Sancar, A. (2016). Mechanisms of DNA repair by photolyase and excision nuclease (Nobel Lecture). *Angewandte Chemie International Edition*, *55*(30), 8502-8527.

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., Post, A. F., Hagemann, M., Paulsen, I., & Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. *Microbiology and molecular biology reviews, 73*(2), 249-299.

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology, 11*(8), 2078-2093.

Shuman, S. (2009). DNA ligases: progress and prospects. *Journal of Biological Chemistry, 284*(26), 17365-17369.

Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews Microbiology, 5*(11), 852-861. https://doi.org/10.1038/nrmicro1768

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology, 189*(1), 113-130.

Summer, H., Grämer, R., & Dröge, P. (2009). Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *JoVE (Journal of Visualized Experiments)*(32), e1485.

Sun, Z., & Blanchard, J. L. (2014). Strong genome-wide selection early in the evolution of Prochlorococcus resulted in a reduced genome through the loss of a large number of small effect genes. *PLoS One, 9*(3), e88837.

Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., & Golden, S. S. (2014). Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Research, 42*(17), e136-e136.

Taylor, R. G., Walker, D. C., & McInnes, R. (1993). E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research, 21*(7), 1677.

Terekhova, K., Gunn, K. H., Marko, J. F., & Mondragon, A. (2012). Bacterial topoisomerase I and topoisomerase III relax supercoiled DNA via distinct pathways. *Nucleic Acids Research, 40*(20), 10432-10440.

Timson, D. J., Singleton, M. R., & Wigley, D. B. (2000). DNA ligases in the repair and replication of DNA. *Mutation Research/DNA Repair, 460*(3-4), 301-318.

Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical reviews, 106*(2), 233-252.

Tschoeke, D., Salazar, V. W., Vidal, L., Campeão, M., Swings, J., Thompson, F., & Thompson, C. (2020). Unlocking the genomic taxonomy of the Prochlorococcus collective. *Microbial Ecology, 80*, 546-558.

Tubbs, J. L., & Tainer, J. A. (2010). Alkyltransferase-like proteins: molecular switches between DNA repair pathways. *Cell Mol Life Sci, 67*(22), 3749-3762. https://doi.org/10.1007/s00018-010-0405-8

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

Williamson, A., Hjerde, E., & Kahlke, T. (2016). Analysis of the distribution and evolution of the ATP‐dependent DNA ligases of bacteria delineates a distinct phylogenetic group 'L ig E'. *Molecular microbiology, 99*(2), 274-290.

Williamson, A., & Leiros, H.-K. S. (2019). Structural intermediates of a DNA–ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation. *Nucleic Acids Research, 47*(14), 7147-7162.

Williamson, A., & Leiros, H.-K. S. (2020). Structural insight into DNA joining: from conserved mechanisms to diverse scaffolds. *Nucleic Acids Research, 48*(15), 8225-8242.

Williamson, A., & Pedersen, H. (2014). Recombinant expression and purification of an ATP-dependent DNA ligase from Aliivibrio salmonicida. *Protein expression and purification, 97*, 29-36.

Williamson, A., Rothweiler, U., & Leiros, H.-K. (2014). Enzyme–adenylate structure of a bacterial ATP-dependent DNA ligase with a minimized DNA-binding surface. *Acta Crystallographica Section D: Biological Crystallography, 70*(11), 3043-3056.

Yan, W., Feng, X., Zhang, W., Zhang, R., & Jiao, N. (2020). Research advances on ecotype and sub-ecotype differentiation of Prochlorococcus and its environmental adaptability. *Science China Earth Sciences, 63*(11), 1691-1700.

Yi, C., & He, C. (2013). DNA repair by reversal of DNA damage. *Cold Spring Harbor perspectives in biology, 5*(1), a012575.

Zess, E. K., Begemann, M. B., & Pfleger, B. F. (2016). Construction of new synthetic biology tools for the control of gene expression in the cyanobacterium Synechococcus sp. strain PCC 7002. *Biotechnology and bioengineering, 113*(2), 424-432.

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

Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J., & Chisholm, S. W. (2006). Prochlorococcus ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology, 72*(1), 723-732.

Appendices

Appendix A: Protein Information

Protein sequences including N-hexahistidine tag (highlighted in blue) and TEV-site (highlighted in green). TEV cleavage occurs between Q and G (red). Mutations from aspartate to alanine in Pmar3-mut are highlighted in yellow.

Pmar3 Protein Sequence

HHHHHHENLYFQGFIFKIGSDEVRVRTDAEFKSLNDDEKISYIWEDLWPDDLGMTWI
ERVLYDHDTGDLGLPEIWNEYVSEFYPEEYKEIENQEPEDRHKKRELIFEFIDDLEDEE
YSRALLFLVNKVYDESFFYDCLIDSVKEQEEFTQTH

Pmar3-mut Protein Sequence

HHHHHHENLYFQGFIFKIGSDEVRVRTDAEFKSLNADEKISYIWEDLWPDDLGMTWI
ERVLYDHATGALGLPEIWNEYVSEFYPEEYKEIENQEPEDRHKKRELIFEFIDDLEDEE
YSRALLFLVNKVYAESFFYDCLIDSVKEQEEFTQTH

Pmar4 Protein Sequence

HHHHHHENLYFQGKPVSIYSFYIHLKAYESEICTYFHEKRISSATYSAISIFLNDSDPD
AYGVDETFLDEEFSSLFDWVSTYEGVLDYLETRYQEAVADPLEGDNETWLFYIGDT
YSCPYKSYEEFKSELDIEMILKIFKKCWERSCKNEYPNGRQLPKDFPEYKIDYSKNKK
LRGDELLKKIKSDILPKYENIKKYGYISLNKNGEEIANLDEFLKEHRNALETAKNNGQ
SIKDHLFPSAQKIGKLSKGKYIIGDLSKCLPIADKNLVDELKALERFKSFDEVDEDLGL
PIYIEDLNDNENKKLDIYKSKKDNNSQDFYLCGLKKKSDGSSFLYILPMIDESSFQDD
KGGRYFQSSNCLICIKVDDEYINREANINSYLFEEDFECIYYDCYDTLKLGDFHVYAD
Y

Pmar5 Protein Sequence

HHHHHHENLYFQGKEKDLAQLNIKVSKELHSKLKFLAQNSDKTMEKYINSILKQSA
EVDFIDEGVKSSVIEDINKRLTNLEKVFLGNNINSLKGTAFTDQEAKNCTLFMKGLFN
KIFQERNYTKINQCWKDFIPLVEKFSQWNHFYTLRLKEVMLIDDPDPWTGDELNELC
LGNKCPCPIYSALREWSGCNDFPTQQIICDKGSDLVDSIEIK

Appendix B: TEV Cleavage Trials

Table B.1: Pmar3-MBP TEV Cleavage Trial Conditions

Conditions	Pmar3-MBP (6.88mg ml^{-1})	TEV Protease (1.05mg ml^{-1})	Buffer C
4°C, 30min (1:5)	10µl	14µl	16µl
4°C, 30min (1:10)	10µl	7µl	23µl
4°C, 1hr (1:5)	10µl	14µl	16µl
4°C, 1hr (1:10)	10µl	7µl	23µl
4°C overnight (1:5)	10µl	14µl	16µl
4°C overnight (1:10)	10µl	7µl	23µl
RT, 30min (1:5)	10µl	14µl	16µl
RT, 30min (1:10)	10µl	7µl	23µl
RT, 1hr (1:5)	10µl	14µl	16µl
RT, 1hr (1:10)	10µl	7µl	23µl
RT overnight (1:5)	10µl	14µl	16µl
RT overnight (1:10)	10µl	7µl	23µl
37°C, 30min (1:5)	10µl	14µl	16µl
37°C, 30min (1:10)	10µl	7µl	23µl
37°C, 1hr (1:5)	10µl	14µl	16µl
37°C, 1hr (1:10)	10µl	7µl	23µl
37°C overnight (1:5)	10µl	14µl	16µl
37°C overnight (1:10)	10µl	7µl	23µl
RT 30min, 4°C overnight (1:5)	10µl	14µl	16µl
RT 30min, 4°C overnight (1:10)	10µl	7µl	23µl
37°C 30min, 4°C overnight (1:5)	10µl	14µl	16µl
37°C 30min, 4°C overnight (1:10)	10µl	7µl	23µl
RT 1hr, 4°C overnight (1:5)	10µl	14µl	16µl
RT 1hr, 4°C overnight (1:10)	10µl	7µl	23µl
37°C 1hr, 4°C overnight (1:5)	10µl	14µl	16µl
37°C 1hr, 4°C overnight (1:10)	10µl	7µl	23µl

Appendix C: Additional Purification Results

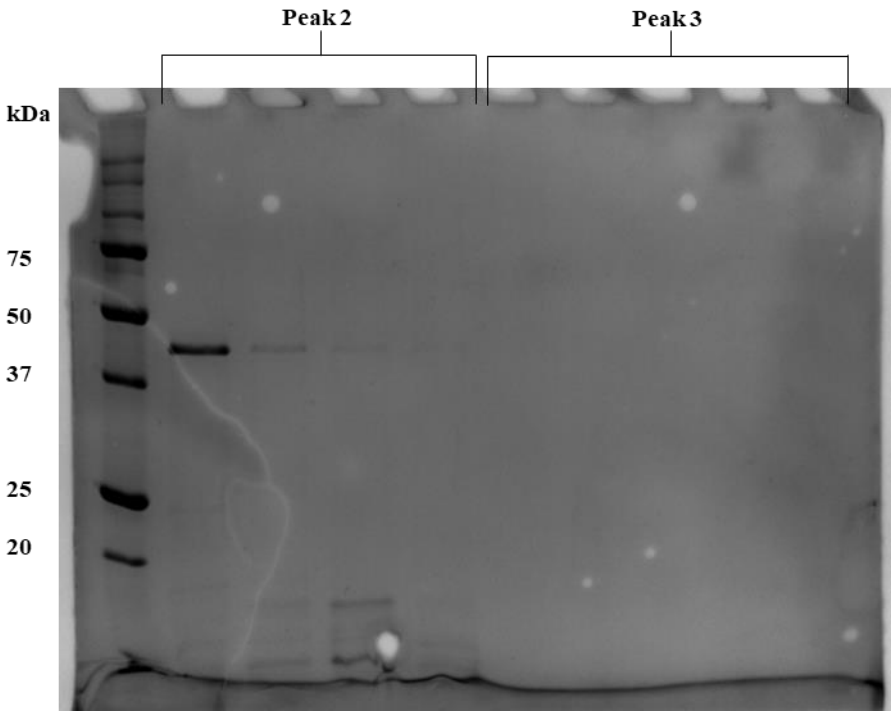


Figure C-1: Second 12% SDS-PAGE from the size exclusion of LigW. Fractions comprising peak 2 which contained LigW (43kDa) carried over into lanes 1-4. Lanes 5-9 contain contaminants that were excluded in peak 3.

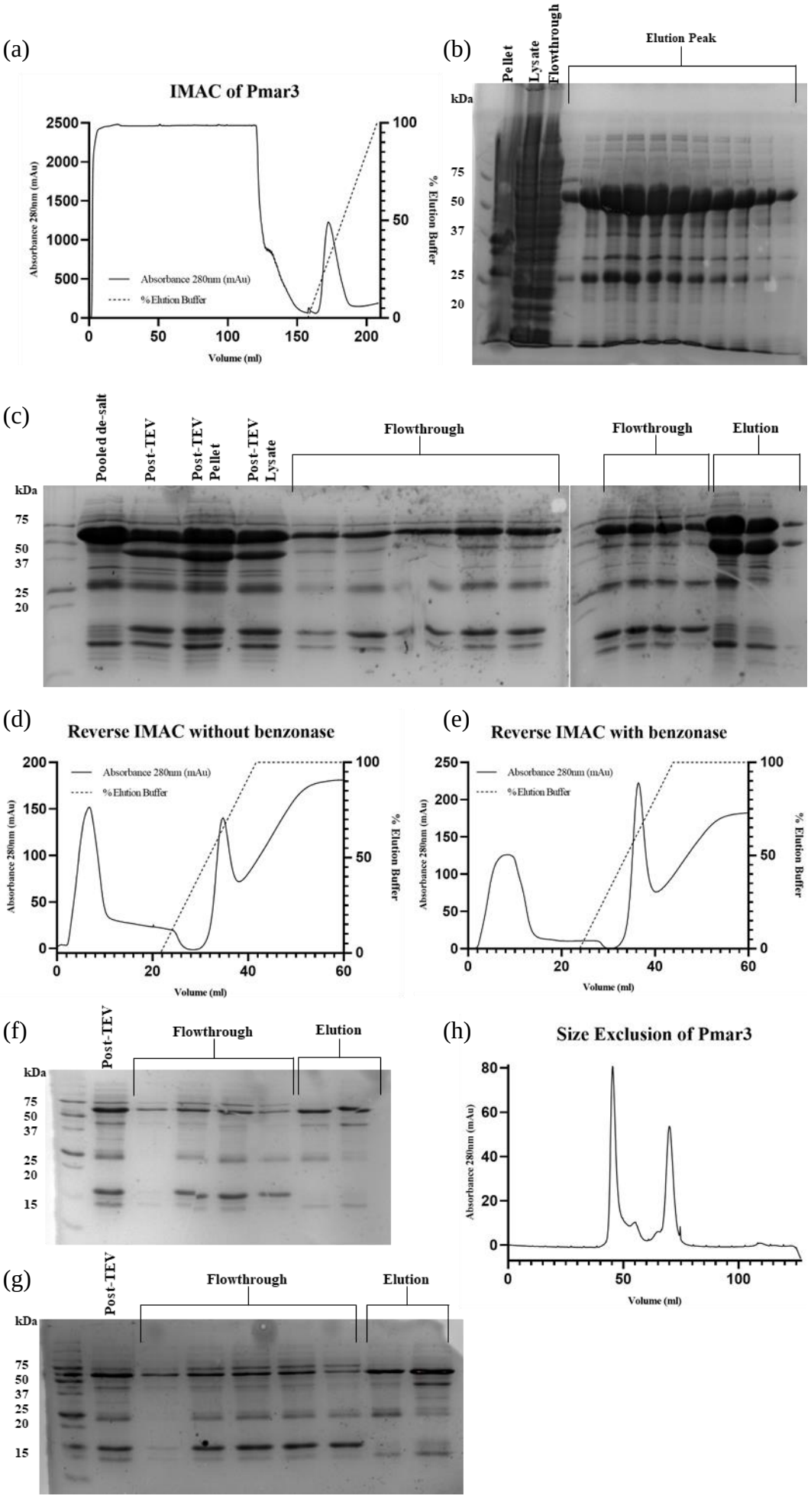

(a)
IMAC of Pmar3
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(b)
Pellet
Lysate
Flowthrough
Elution Peak
kDa
75
50
37
25
20
(c)
Pooled de-salt
Post-TEV
Post-TEV Pellet
Post-TEV Lysate
Flowthrough
Flowthrough
Elution
kDa
75
50
37
25
20
(d)
Reverse IMAC without benzonase
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(e)
Reverse IMAC with benzonase
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(f)
Post-TEV
Flowthrough
Elution
kDa
75
50
37
25
20
15
(g)
Post-TEV
Flowthrough
Elution
kDa
75
50
37
25
20
15
(h)
Size Exclusion of Pmar3
Absorbance 280nm (mAu)
Volume (ml)

Figure C-2: First IMAC purification and protein analysis of Pmar3-MBP. (a) Chromatogram from IMAC of Pmar3-MBP shows the UV absorbance (280nm) and elution of Pmar3-MBP at 15-30% buffer B. (b) 15% SDS-PAGE gel shows the presence of Pmar3-MBP (64kDa). All fractions were collected, pooled together, and run through a desalting column using buffer C, before being incubated with 1mg TEV protease at 4oC overnight. The protein appeared to precipitate overnight, so the lysate was separated from precipitants by centrifugation and was used for the reverse IMAC. (c) 15% SDS-PAGE gels show a low yield of de-tagged Pmar3 (17kDa) was obtained, with inefficient MBP-tag cleavage. Flowthrough was pooled and split into two fractions. 1µl benzonase per 10ml lysate and a final concentration of 5mM $MgCl_2$ was added to half of the pooled reverse IMAC flowthrough, 500µl TEV was added to both halves and was incubated at RT for 30 minutes, then overnight at 4ºC. (d and e) Reverse IMAC chromatograms for both conditions. (f and g) 15% SDS-PAGE gels for fractions from both reverse IMAC's. Bands are visible in the flowthrough of both lysate at approximately 17kDa and were suspected to be Pmar3. Protein-containing fractions from both lysates were pooled together and concentrated to give a total volume of 5ml which was used for size exclusion. (h) Chromatogram for size exclusion of Pmar3. Peak 1-2 = Pmar3-MBP and non-specific proteins, Peak 3 = Pmar3.

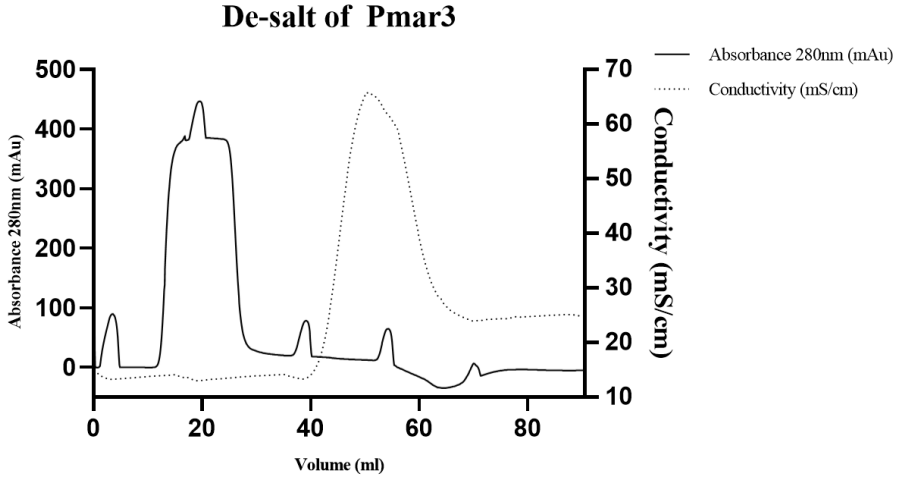


Figure C-3: De-salt chromatogram from the second purification of Pmar3-MBP.

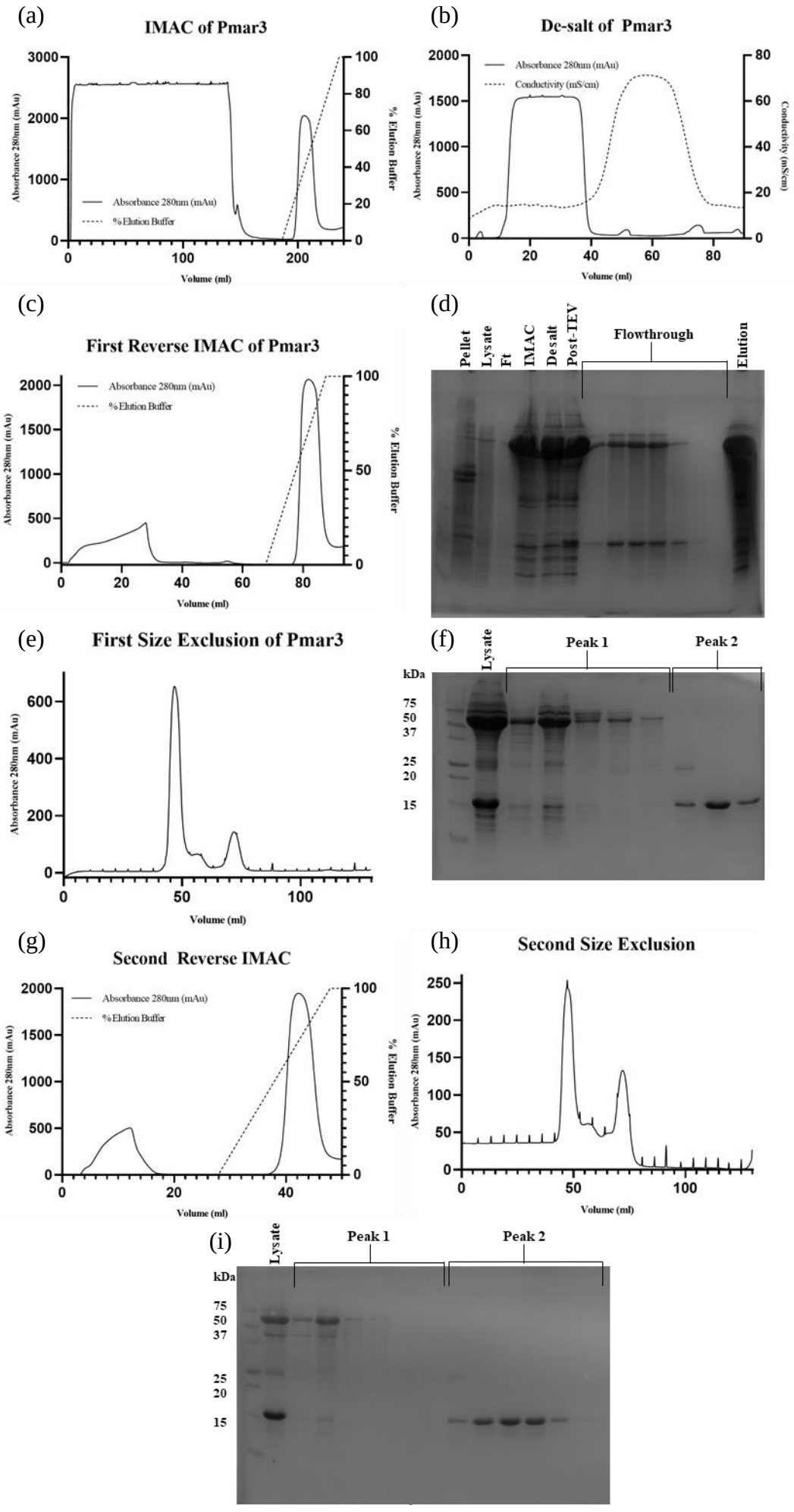

(a)
IMAC of Pmar3
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(b)
De-salt of Pmar3
Absorbance 280nm (mAu)
Conductivity (mS/cm)
Volume (ml)
(c)
First Reverse IMAC of Pmar3
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(d)
Pellet
Lysate
Ft
IMAC
Desalt
Post-TEV
Flowthrough
Elution
(e)
First Size Exclusion of Pmar3
Absorbance 280nm (mAu)
Volume (ml)
(f)
Lysate
Peak 1
Peak 2
kDa
75
50
37
25
20
15
(g)
Second Reverse IMAC
Absorbance 280nm (mAu)
% Elution Buffer
Volume (ml)
(h)
Second Size Exclusion
Absorbance 280nm (mAu)
Volume (ml)
(i)
Lysate
Peak 1
Peak 2
kDa
75
50
37
25
20
15

Figure C-4: Third IMAC purification and protein analysis of Pmar3-MBP. (a) Chromatogram from IMAC of Pmar3-MBP shows the UV absorbance (280nm) and elution of Pmar3-MBP at 15-55% buffer B. (b) Chromatogram from de-salt of Pmar3-MBP. (c) Chromatogram from the first reverse IMAC of Pmar3. This included a buffer A wash before the 20ml 0-100% buffer B gradient. (d) 15% SDS-PAGE gel of the flowthrough and elution peak from the reverse IMAC. The protein ladder was not visible in this gel. Elution peak contained a high concentration of Pmar3-MBP, so the fractions were collected, pooled together, and run through a desalting column using buffer C, before being incubated with 1mg TEV protease at 4ºC overnight. (e) Chromatogram showing the size exclusion of the concentrated flowthrough from the reverse IMAC. (f) 15% SDS-PAGE gel showing the separation of Pmar3 (17kDa) from Pmar3-MBP (64kDa) and other contaminants. Lanes 8-9 = fractions pooled to give batch "1-P3". (g) Chromatogram from the reverse IMAC of the batch that had a second TEV cleavage step. (h) Chromatogram showing the size exclusion of the concentrated flowthrough from the second reverse IMAC. (i) 15% SDS-PAGE gel showing the separation of Pmar3 (17kDa) from Pmar3-MBP (64kDa) and other contaminants. Lanes 9-13 = fractions pooled to give batch "2-P3".

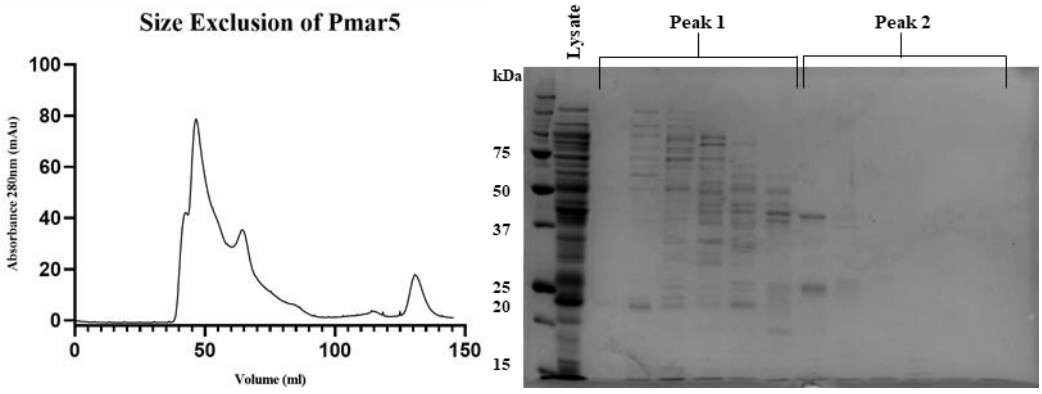


Figure C-5: First IMAC purification and protein analysis of Pmar5. (a) Chromatogram from size exclusion of Pmar5. (b) 12% SDS-PAGE gel of size exclusion fractions, confirming the purification was unsuccessful.

Appendix D: Assay Compositions

Table D.1: DNA Substrate Compositions for Fluorescence Nuclease Assays

	(1) Double-strand matched	**(2) Single-strand**	**(3) 3'-tail**	**(4) 5'-tail**	**(5) 20+20 Control**	**(6) Single-strand control**
FAM-Labelled Oligo	4µl Linear (0.5µM)	4µl Linear (0.5µM)	4µl Linear (0.5µM)	4µl Nick FAM (0.5µM)	4µl Nick FAM (0.5µM)	4µl Nick FAM (0.5µM)
Complement	4µl Nick Complement (0.7µM)	-	4µl Blunt Complement (0.7µM)	4µl Nick Complement (0.7µM)	4µl Blunt Complement (0.7µM)	-
10x Buffer (500mM Tris ph8.0, 500mM NaCl, 10mM DTT)	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl
$MgCl_2$ (1M)	0.25µl	0.25µl	0.25µl	0.25µl	0.25µl	0.25µl
MQ	9.25µl	13.25µl	9.25µl	9.25µl	9.25µl	13.25µl
Total	20µl	20µl	20µl	20µl	20µl	20µl

Table D.2: DNA Substrate Compositions for Electrophoretic Mobility Shift Assays

	(1) Double-strand matched	**(2) Single-strand**	**(3) 3'-tail**	**(4) 5'-tail**	**(5) Nicked**
FAM-Labelled Oligo	4µl Linear (0.5µM)	4µl Linear (0.5µM)	4µl Linear (0.5µM)	4µl Nick FAM (0.5µM)	4µl Nick FAM (0.5µM)
Complement	4µl Nick Complement (0.7µM)	-	4µl Blunt Complement (0.7µM)	4µl Nick Complement (0.7µM)	-
5' Phos Oligo	-	-	-	-	4µl Nick Phos (2.5µM)
Long Oligo	-	-	-	-	4µl Linear (0.5µM)
EDTA (1M)	0.25µl	0.25µl	0.25µl	0.25µl	0.25µl
ATP (100mM)	0.25µl	0.25µl	0.25µl	0.25µl	0.25µl
10x Buffer (500mM Tris ph8.0, 500mM NaCl, 10mM DTT)	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl
MQ	9.25µl	13.25µl	9.25µl	9.25µl	5µl
Total	20µl	20µl	20µl	20µl	20µl