The Structural and Mechanistic Characterisation of Thermophilic RNA Ligases

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

Aliyah Kamal

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Abstract

MicroRNAs (miRNAs) are short regulatory RNAs that play critical roles in post-transcriptional gene regulation. Dysregulation of miRNA expression can be implicated in numerous diseases, including cancer, making miRNAs valuable diagnostic and prognostic biomarkers. Next-generation sequencing (NGS) enables quantitative profiling of miRNA populations, but a key limitation lies in the adapter ligation step required for library preparation. Current commercial RNA ligases, typically active between 14-37 °C, display strong biases based on RNA sequence and secondary structure resulting in a skewed representation of miRNA populations. At low temperatures, miRNA secondary structures can obscure 5'-phosphate, and 3'-hydroxyl ends, preventing efficient ligation. Thermophilic RNA ligases with high thermostability and distinct substrate preferences may therefore reduce ligation bias and improve NGS accuracy. The archaeal *Paleococcus pacificus* RNA ligase (PpaRnl) represents a valuable model for exploring the structural and mechanistic features of thermophilic RNA ligases.

This study aimed to further investigate the substrate preference of a PpaRnl mutant; K238G, a point mutation of lysine at position 238, a key residue in the coordination of the AMP and in turn the binding of the Mg^{2+} ions in PpaRnl. Alongside the structural and mechanistic characterisation of several PpaRnl mutants to explore the role of a newly identified Mg^{2+} binding site (Mg^{2+} B) close to the active site. Key residues for coordinating the Mg^{2+}B ion were mutated in the PpaRnl enzyme (D94A, D248A, E155A, and the double mutant D94A-D248A) as characterisation of these mutants would allow for a deeper understanding of this mechanism.

Mutants were expressed in *Escherichia coli* and purified using a combination of immobilised metal affinity chromatography (IMAC) and gel-filtration chromatography. Expression and purification outcomes varied significantly between constructs, while some mutants were successfully obtained in sufficient yields, others demonstrated poor or inconsistent expression, often not even seeing improvement after optimisation of growth conditions and large-scale culture volumes. In particular, the E155A mutant displayed extremely slow growth, while the D94A-D248A double mutant produced negligible results across expression trials. Further characterisation of the mutants, K238G, D94A, and D248A, had shown that these enzymes retain robust activity between 60-65 °C, with measurable

activity extending up to 80 °C. Distinct sequence and cofactor preferences were observed. When combined with adapters, these mutants efficiently produced RNA ligation products with minimal side reactions and, in several cases, displayed improved activity compared to the wild-type enzyme.

The structures of PpaRnl D94A and D248A were solved via X-ray crystallography and although there was similarity in the structures compared to that of the wild-type enzyme and K238G, it was found that a number of the residues coordinating the ligands present within the active-site had changed their orientation. This resulted in the Mg^{2+}B ion not being present in D94A, and no AMP cofactor, Mg^{2+}A or Mg^{2+}B present within the D248A mutant.

Overall, this work highlighted the variability in expression and crystallisation outcomes among the second Mg^{2+} binding site mutants of PpaRnl with notable differences observed in expression, activity, and structure, highlighting that the Mg^{2+}B ion and its coordination have an impact of the efficiency of the enzyme. This unique mechanism is still not fully understood but its importance and impact has been highlighted within this research, reiterating the importance of construct specific optimisation in structural studies and the potential utility of engineered thermophilic RNA ligases, as tools to reduce ligation bias in NGS based miRNA sequencing. Allowing for the possible discovery of biomarkers within viral diseases, and cancer.

Acknowledgements

I would like to start off by thanking Dr Joanna Hicks for the amazing opportunity to complete my master's under your supervision. From the first lecture I attended with you in my Bachelors, your passion, and drive to see your students succeed was admirable. Your kindness, patience, and constant, unwavering belief in my ability to complete this degree even when I could not see it, did not go unnoticed. I want to thank you for all the time you gave, and all the knowledge I have received. Working with you have been an absolute privilege.

I would also like to thank the soon-to-be Dr. Meghan Rousseau. While completing your own PhD, you somehow found the time and energy to guide me through the many challenges of beginning my master's. Patiently answering my endless questions and helping me navigate all that was new. Your patience, kindness, and occasional tough love were instrumental in helping me build both the confidence in myself and the ability to conduct an independent research project.

I would also like to thank everyone in the C2 lab for reminding me that we were all in this together. I will always look back with fondness on the memories we made, the laughter we shared, and, of course, the caramel slices that kept us going. To those in the C.2.03 student offices in particular; there was truly never a dull moment. My time with you all will always be remembered with happiness (and just a little stress). I have no doubt that each of you will go on to achieve incredible things.

Lastly, the biggest thank you of all to my Mum, Jane, the woman whose unwavering support has been a constant source of strength. Through every success, no matter how small, and during times of doubt, you have stood by me through it all. You are the reason I am the woman I am today, and there will never be enough words to describe how grateful I am to have you in my life. Getting to call you my mum will forever be my greatest accomplishment.

Table of Contents

List of Figures

List of Tables

Glossary

Abbreviation	Definition
3'-OH	3'-hydroxyl
5'-PO_4	5'-monophosphate
A	Alanine (amino acid)
ADP	Adenosine diphosphate
Ala	Alanine (amino acid)
AMP	Adenosine monophosphate
APS	Ammonium persulphate
Arg	Arginine (amino acid)
Asn	Asparagine (amino acid)
Asp	Aspartate (amino acid)
ATP	Adenosine triphosphate
C	Cysteine (amino acid)
cDNA	Complementary DNA
C-terminal	Carboxy-terminal
Cys	Cysteine (amino acid)
D	Aspartate (amino acid)
DNA	Deoxyribonucleic acid
E	Glutamate (amino acid)
F	Phenylalanine (amino acid)
FD	First derivative
G	Glycine (amino acid)
GDP	Guanosine diphosphate
Gln	Glutamine (amino acid)
Glu	Glutamate (amino acid)
Gly	Glycine (amino acid)
H	Histidine (amino acid)
His	Histidine (amino acid)
I	Isoleucine (amino acid)
Ile	Isoleucine (amino acid)
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropylthio-β-D-galactosidase

K	Lysine (amino acid)
kDa	Kilodalton
L	Leucine (amino acid)
LB	Luria-Bertani broth
Leu	Leucine (amino acid)
Lys	Lysine (amino acid)
M	Methionine (amino acid)
Met	Methionine (amino acid)
min	Minute
miRNA	MicroRNA
mRNA	Messenger RNA
MthRnl	Methanobacterium thermoautotrophicum RNA ligase
MW	Molecular weight
N	Asparagine (amino acid)
NAD^+	Nicotinamide adenine dinucleotide
NGS	Next-generation sequencing
nm	Nanometre
NTase	Nucleotidyltransferase
N-terminal	Amino terminus
P	Proline (amino acid)
PabRnl	Pyrococcus abyssi RNA ligase
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethylene glycol
PfuRnl	Pyrococcus furiosus RNA ligase
Phe	Phenylalanine (amino acid)
PpRnl	Palaeococcus pacificus RNA ligase
Pro	Proline (amino acid)
Q	Glutamine (amino acid)
R	Arginine (amino acid)
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
S	Serine (amino acid)

SDS	Sodium dodecyl sulphate
sec	Seconds
Ser	Serine (amino acid)
SO_4	Sulphate
sRNA	Small RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
T	Threonine (amino acid)
T4 Rnl1	T4 bacteriophage RNA ligase 1
T4 Rnl2	T4 bacteriophage RNA ligase 2
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TEMED	Tetramethyl ethylenediamine
Thr	Threonine (amino acid)
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
Trp	Tryptophan (amino acid)
V	Valine (amino acid)
v/v	Volume per volume
Val	Valine (amino acid)
W	Tryptophan (amino acid)
w/v	Weight per volume
XDS	X-ray detector software
Y	Tyrosine (amino acid)

Chapter One: Introduction and Literature Review

The biological world operates according to the central dogma, in which genetic information flows from DNA to RNA and ultimately to protein, through transcription and translation (Koonin, 2015). While this framework has been fundamental in the understanding of molecular biology, it is now well established that not all RNA serves as a simple intermediate in protein synthesis. Non-coding RNA such as long non-coding RNAs (lncRNAs), small interfering RNAs (siRNAs), and micro RNAs (miRNAs) all play essential roles in the regulation of gene expression, revealing a much more complex layer of post-translational and epigenetic control (Ratti et al., 2020).

1.1 MicroRNAs and their Importance

Non-coding RNAs known as microRNAs (miRNAs) play a key role in regulating post-transcriptional gene expression, modifying messenger RNA (mRNA) and its stability, and repressing and activating of genes through partial complementarity to the target sequence. With such a significant role in gene regulation variations in miRNA can in some cases result in gene dysregulation and differential gene expression. (S.M Hammond et al., 2015; McGeary et al., 2019). Like many other biological processes, the miniscule margin of error means that when things go wrong it tends to result in a larger scale issue. For example, the dysregulation and off-target behaviour of miRNAs have been linked to cancer and viral infections (Cui et al., 2025; Girardi et al., 2018; Zhang et al., 2025) as well as range of other conditions. Identification of miRNAs linked with various diseases can now be used as biomarkers, giving a non-invasive method to identify differential miRNA expression and provide the basis for the synthesis of drugs and other potential therapeutics though the discovery of new miRNA targets.

Identification of miRNAs is key in the ability to detect and profile miRNAs in different disease states. This can be done with a range of methodologies, including hybridisation-based microarrays, quantitative reverse transcription PCR (RT-qPCR), and next-generation sequencing (NGS) technologies (Git et al., 2010). Microarrays were once the predominant tool for small RNA profiling, though they presented several technical limitations. These methods rely on radiolabelled

oligonucleotide probes that bind to target RNA molecules under fixed-temperature conditions. Consequently, careful consideration of probe melting temperatures (T_m) is essential, often requiring probe length adjustment and T_m normalization. However, the short length of miRNAs constricts probe design, frequently necessitating the use of the full miRNA sequence as the probe. While alternative probe design strategies have emerged to address this issue (Li & Ruan, 2009), hybridisation-based methods still face limitations. One major drawback is their limited specificity, particularly when distinguishing between closely related miRNA family members or identifying novel variants. This limits the accuracy of expression measurements and impairs the method's ability to provide absolute quantification. Furthermore, microarrays require large quantities of RNA input, which can be problematic when working with low-yield sources like liquid biopsies (Benesova et al., 2021). RT-qPCR offers a more sensitive alternative, widely regarded as the gold standard for quantifying gene expression. It requires minimal RNA input, enables fast and targeted analysis, and can be optimized for absolute quantification. However, RT-qPCR is generally limited to known miRNAs and is less suitable for high-throughput screening due to its cost and labour intensity, making it better suited for validation studies or diagnostic assays. The development of next generation sequencing (NGS) technologies has become an impactful tool for profiling miRNAs.

1.2 Small RNA Sequencing of microRNAs

Small RNA sequencing allows for comprehensive identification and quantification of novel and known miRNAs, with both high sensitivity and resolution. Compared to traditional methods, like microarrays, that rely on the prior knowledge of miRNA sequences, NGS can detect miRNA isoforms, rare variants and species-specific miRNAs with no need for specific probes (Motameny et al., 2010). NGS does have its challenges; biases can be introduced during the preparation of the library, there is a need for accurate reference databases, and the computational demands for data interpretation. Despite its shortcomings NGS has had a significant impact on the advancement of miRNA research, providing a deeper understanding of their expression profiles and regulatory roles in a diverse range of biological contexts (Motameny et al., 2010). The standard workflow involves several steps: adaptor ligation to the small RNA molecules, enrichment of miRNAs, reverse transcription into complementary DNA (cDNA), PCR amplification, and high-throughput

sequencing, This is followed by extensive bioinformatic processing, including adaptor trimming, sequence alignment to allow for the referencing of genomes or databases, and prediction of secondary structures to support novel miRNA discovery (Benesova et al., 2021) (Figure 1.1).

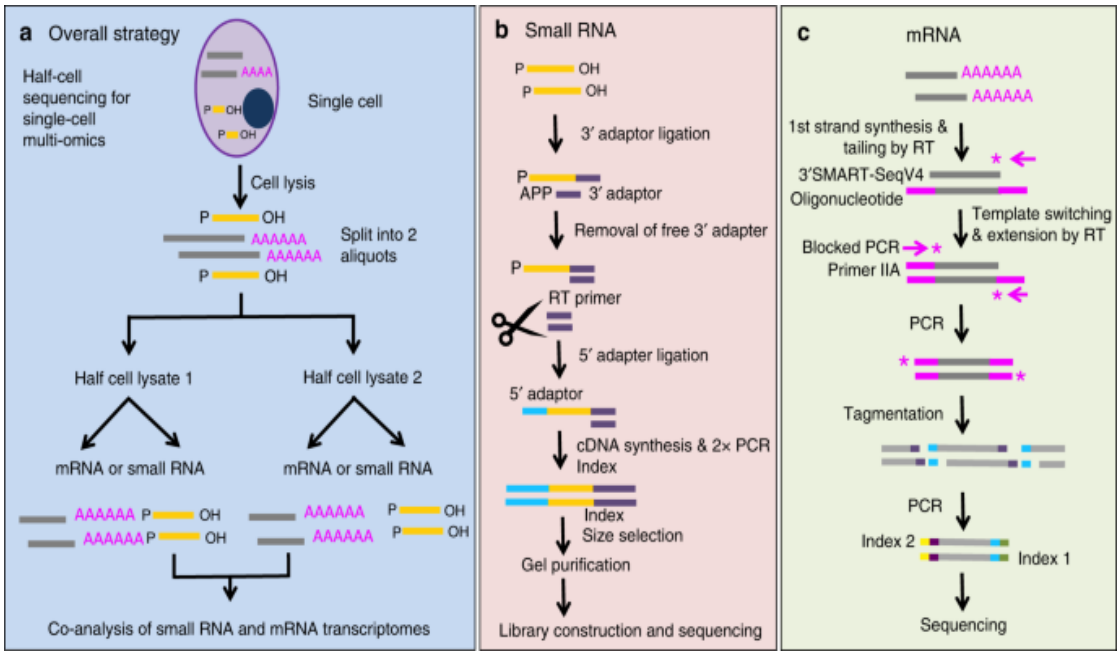

Figure 1.1 *Experimental workflow and library construction of miRNAs and mRNAs (Wang et al., 2019)*

1.3 RNA Ligases in Sequencing

Within small RNA sequencing, RNA ligases are essential during the adaptor ligation steps. These ligation events are critical for enabling subsequent reverse transcriptions, amplification, and sequencing. The efficiency and specificity of RNA ligases in these steps significantly impact the accuracy and completeness of miRNA libraries generated for NGS. RNA ligases act to join RNA molecules and have a wide range of uses in various biotechnology applications. They catalyse formation of the phosphodiester bond between the 5'-PO4 and 3'-OH groups on opposing ends of RNA strands.

1.3.1 RNA-Ligase Dependent Bias

Like with most chemical reactions there is not only the desired products formed, but also those formed during the simultaneous side reactions that are also taking place. These side reactions usually involving AMP, can lead to the formation of

primer dimers, or modified RNA molecules, that can cause interference with the ligation reaction. These side reactions result in the removal of the 5′-adenylated adapter and the unintended adenylation of the RNA substrate's 5′-phosphate group (Hafner et al., 2011). Furthermore, circularisation and/or concatenation are common due to the quick ligase-catalysed adenylate transfer from the phosphorylated 3'-adaptor intermediate to the miRNA, which lowers adapter ligation efficiency and implicates micro-RNA (miRNA) complementary DNA (cDNA) production. Despite its necessity, adapter ligation introduces substantial biases into miRNA libraries due to enzyme substrate specificity and structural constraints.

During small RNA library preparation, several biochemical challenges can affect ligation efficiency and introduce bias into miRNA quantification. One issue involves AMP transfer side reactions, where RNA ligases, unintentionally remove the AMP group from pre-adenylated adapters and transfer it to RNA substrates, priming them for self-ligation (Zhuang et al., 2012). This can lead to the formation of circularised or concatenated RNA molecules, especially when the RNA ends are brought into proximity due to secondary structure-induced folding (Hafner et al., 2011). These secondary structures, such as hairpin loops and base-paired termini, can physically obstruct ligase access to the 5′ and 3′ ends of miRNAs, significantly reducing ligation efficiency (Jayaprakash et al., 2011). Additionally, enzyme specificity contributes to sequence and structure-dependent biases. Ligases such as T4 RNA ligase 1 (Rnl1) and T4 RNA ligase 2 (Rnl2) show preferences based on the nucleotide composition, secondary structure, and terminal chemistry of their RNA substrates (Hafner et al., 2011; Sorefan et al., 2012). These factors collectively skew miRNA abundance profiles and highlight the importance of optimised enzymatic conditions in small RNA sequencing protocols.

1.4 General Mechanism and Families of RNA Ligases

A defining feature of RNA ligases is their conserved nucleotidyltransferase (NTase) core domain, which contains multiple catalytic motifs such as Motifs I, III, and V which are central to their enzymatic mechanism (Shuman & Lima, 2004). These motifs orchestrate the three-step ligation reaction (1) enzyme adenylation, (2) AMP transfer, and (3) formation of the phosphodiester bond. Enzyme adenylation begins with the nucleophilic attack on the α-phosphate located on the nucleotide cofactor next to an active site lysine. This forms the covalent ligase-(lysyl-Nζ)–AMP

intermediate. AMP-transfer then begins by moving from the adenylated ligase to the exposed 5'-PO_4 end of an RNA strand, resulting in the formation of a 5'-adenylated RNA intermediate (5'-aRNA). The phosphodiester bond is then formed through the ligases catalysation of a nucleophilic attack of the RNA 3'-OH group to the α-phosphate of the 5'-aRNA sealing the two ends together as seen in the Figure 1.2.

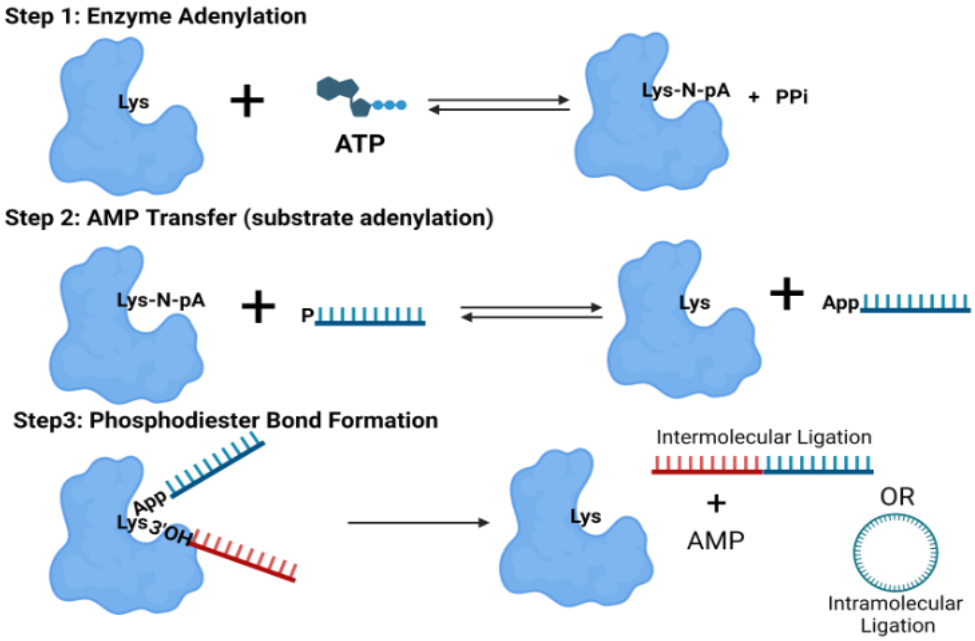

Figure 1.2 *RNA ligase enzymatic mechanism. Step 1 represents adenylation of the active site lysine of the enzyme with cofactor ATP, forming the enzyme adenylate. Step 2 shows transfer of the AMP from the active site lysine to the 5'PO_4 of the RNA molecule. In step 3 the 3'OH of a second RNA molecule or the end of the same molecule acts as a nucleophile attacking the 5'PO_4 of the AMP-RNA intermediate, displacing the AMP and forming the phosphodiester bond joining the RNA molecule(s) together.*

The functional integrity of this mechanism is highly dependent on specific active site residues, such as lysine's (commonly in Motif I), arginine's, and glutamine's, which are crucial for nucleotide binding, stabilisation of the transition state, and precise substrate positioning (Lohman et al., 2014). Even minor alterations to these residues can impair catalysis or prevent ligation altogether, highlighting their indispensable role in enzyme function.

ATP-dependent RNA ligases can be classified into five families that are structurally characterised based on sequence, structure, and phylogenetic relationship: Rnl1 (T4 RNA ligase 1), consist of a single catalytic domain and are commonly used in a range of biological applications due to their 3'OH to 5'-PO_4 ligation activity (Ho et al., 2004). In the Rnl2 family (T4 RNA ligase 2) a conserved N-terminal

nucleotidyltransferase domain and a C-terminal domain contribute to enhanced RNA binding and ligation (Nandakumar et al., 2006). These too, are commonly used in a range of biological applications. Compared to Rnl1 and Rnl2, Rnl3 ligases are found in archaea and thermophilic bacteria. They form atypical homodimers, facilitating the formation of covalently closed circular RNAs (Becker et al., 2019). Their modular architecture, and thermostability, make them well-suited for ligating structured RNAs under harsh conditions, highlighting their evolutionary specialisation in archaeal RNA biology (Becker et al., 2019). The Rnl4 family are part of a complex involved in RNA repair, often working alongside polynucleotide kinase/phosphatase and Hen1-like methyltransferase proteins. They consist of an N-terminal nucleotidyltransferase domain and a C-terminal domain, mediating ATP-dependent sealing of RNA breaks. Mutational analyses highlight conserved residues for catalysis, substrate binding, and complex assembly (Raymond & Shuman, 2007). The final family is Rnl5; the RNA ligases are characterised by the specific organisation of their domain, this includes an N-terminal domain that is critical in adenylation and nick sealing, but when it comes to the phosphodiester bond formation and the pre-adenylated nick, it is dispensable They are similar to T4 Rnl2, but lack a C-terminal appendage to their NTase domain (Uciuleac et al., 2015). Rnl5 ligase have been found to prefer manganese (Mn^{2+}) as a co-factor but will use magnesium (Mg^{2+}) to facilitate the initial adenylation step (Uciuleac et al., 2015). These families represent a structural diversity and functional specialisation of RNA ligases across all domains of life.

1.5 Commercially Available RNA Ligases: T4 RNA Ligase 1 & 2

Beyond the NTase core, the N-terminal and C-terminal domains of RNA ligases contribute significantly to substrate recognition, binding affinity, and catalytic efficiency. These flanking regions modulate the enzyme's interaction with structured or bulky RNA substrates and help tailor its activity for specific biological or experimental contexts. In the case of T4 RNA ligase 1 (Rnl1) a commonly used ligase in small RNA library preparation, removal or truncation of the C-terminal domain has been shown to reduce its binding affinity for structured RNAs, such as tRNAs, as well as to impair its adenylation activity, a critical first step in the ligation reaction (Nandakumar et al., 2006). Furthermore, point mutations in the conserved NTase motifs within the core domain can abolish AMP transfer, effectively disabling the enzyme's catalytic function. These structural dependencies are not

only relevant for understanding native enzyme biology but also have practical implications in high-throughput sequencing, where ligase inefficiencies or biases can lead to distorted RNA abundance profiles. Selecting ligases with favourable substrate compatibility, and engineering modifications to improve activity, or reduce bias, are therefore active areas of research in sequencing technology development.

As briefly mentioned above two common RNA ligases used in RNA sequencing are the mesophilic T4 Rnl1 and Rnl2 both with an operating temperature of 16-37 °C. T4 Rnl1 is a well characterised ATP-dependent enzyme that catalyses the formation of phosphodiester bonds between RNA molecules, specifically ligating the 5′-phosphate of one RNA strand to the 3′-hydroxyl of another. It is primarily used in biochemical applications such as RNA circularisation, adapter ligation in next-generation sequencing, and RNA labelling. Rnl1 features a highly conserved NTase core domain, which consists of the catalytic motifs that are integral to its function (Ho et al., 2004). These motifs form the conserved scaffold present across multiple ligase and polymerase families. The NTase domain is bordered by unique N- and C-terminal extensions that contribute to substrate recognition and enzymatic specificity.

The catalytic mechanism of Rnl1 follows the three-step reaction pathway common to all RNA ligases: (1) self-adenylation of the ligase using ATP to form a covalent enzyme–AMP intermediate, (2) transfer of the AMP to the 5′-phosphate of the RNA donor strand, and (3) nucleophilic attack by the 3′-hydroxyl group of the acceptor strand to form a phosphodiester bond with release of AMP. This mechanism is orchestrated by several key active site residues, such as Lys99, which acts as a nucleophile, to form the enzyme–AMP intermediate, Glu159, which coordinates metal ions for catalysis, and Tyr246, which is thought to stabilise the transition state (Ho et al., 2004).

Importantly, the C-terminal domain of Rnl1, composed of ten α-helices, is critical for substrate binding and enzyme stability. This domain contributes to the recognition of structured RNA, such as tRNAs, by facilitating interactions beyond the ligation junction. Truncation studies have shown that removing the C-terminal region severely impairs the enzyme's ability to bind and ligate full-length tRNAs (Nandakumar et al., 2006). This suggests that while the NTase domain is sufficient for catalysis, the C-terminal domain modulates substrate specificity and enhances ligation efficiency, particularly for structured RNA molecules. This makes Rnl1

especially useful in techniques requiring ligation of single-stranded RNAs with complex secondary structures.

Comparatively T4 RNA ligase 2 (Rnl2) differs substantially from Rnl1 in both substrate preference and mechanistic characteristics. Rnl2 is specialised for ligating nicks within double-stranded RNA (dsRNA) or RNA–DNA hybrid helices, a function that makes it critical in biological processes requiring RNA repair and in vitro methods such as ribosome profiling and strand-specific RNA sequencing (Nandakumar et al., 2004). Like Rnl1 and most other RNA ligases, Rnl2 operates via the same three-step ligation. However, it exhibits a strong requirement for a base-paired nicked duplex, and its efficiency drops sharply in the presence of single-stranded overhangs or gaps, indicating a stringent substrate structure preference (Nandakumar et al., 2004).

T4 RNA ligases one (Rnl1) and two (Rnl2) are widely used in small-RNA library preparation for their known ability to attach pre-adenylated DNA adaptors to RNA 3'-ends. Their efficiency shows variation among different RNA sequences and structures (Zhuang et al., 2012). Using an unbiased ligation-selection approach with randomisation in the RNAs, and high-throughput sequencing, demonstrated that these ligases did not exhibit a primary sequence preference, instead a strong bias against structured RNA ends/adaptor regions (Zhuang et al., 2012). Substrates with less than three unpaired nucleotides at the 3'-end and those that are predicted to form a stable, co-folded, structure with adaptors, are ligated inefficiently (Zhuang et al., 2012). It was found that engineering adaptors with randomised overhangs can disrupt unfavourable secondary structures, and significantly reduce ligation bias, and improve both yield and representation in sequencing libraries (Zhuang et al., 2012). Highlighting the role of the secondary structure over sequence as the main barrier to efficient adaptor ligation by both Rnl1 and Rnl2, guiding improved protocols for small-RNA cloning and quantification.

Structurally, Rnl2 also possesses an NTase domain and an extended C-terminal domain, both of which are essential for function. The NTase domain harbours critical residues such as Lys35, His37, and Glu204, which are directly involved in AMP binding and catalysis (Nandakumar & Shuman, 2004). The C-terminal domain is enriched with residues such as Arg266, Asp292, and Glu296, which form interactions with the minor groove of the RNA duplex, positioning the nicked ends for catalysis. Mutations in any of these residues disrupt the enzyme's ability to

complete the ligation process, underscoring the cooperative function of the two domains (Nandakumar et al., 2006).

Rnl2 also shows a lower tolerance to mismatched or distorted substrates, which, while limiting in some experimental contexts, makes it particularly valuable for applications requiring high-fidelity ligation of correctly paired RNA duplexes (Nandakumar et al., 2006). This rigid substrate recognition ensures that ligation only proceeds on precise RNA structures, contributing to the accuracy of RNA mapping in sequencing workflows.

1.6 Thermophilic RNA Ligases

Thermophilic RNA ligases derived from Archaea, such as PpaRnl from *Palaeococcus pacificus*, have garnered attention for their potential to enhance RNA ligation processes, particularly in the context of next-generation sequencing (NGS) library preparation. These enzymes exhibit optimal activity at elevated temperatures ranging from 60°C to 80°C, conditions that promote the denaturation of stable RNA secondary structures (Rousseau et al., 2024). This thermal property facilitates improved accessibility to the RNA termini, thereby increasing the efficiency of adapter ligation steps critical in NGS.

1.6.1 *Methanobacterium thermoautotrophicum* RNA Ligase

Methanobacterium thermoautotrophicum (MthRnl) contains an ATP-dependent thermophilic RNA ligase that has been found to catalyse intramolecular circularisation of single-stranded RNA through the three-step nucleotidyl-transfer mechanism; ligase to AMP formation, RNA adenylation, and the formation of the phosphodiester bond (Zhuang et al., 2012). Crystal structures for this RNA ligase in both the ATP bound and the lysine-AMP intermediate states, highlighted key active-site residues and a newly proposed RNA -binding interface; this had the involvement of the surface residues Tyr159, Phe281, Glu285 and co-crystallised sulphate ions (Gu et al., 2016). On further mutational analysis of the 36 residues, it was revealed that the motif loop (primarily Thr117 and Arg118) were essential in driving the final step of the phosphodiester bond formation as mutations here shifted the equilibrium back to that of the RNA adenylation intermediate (Gu et al., 2016). Comparatively substitutions that were made at Tyr159, Phe281, and Glu285 resulted in a reduction of the RNA binding affinity, really highlighting their role in substrate engagement (Gu et al., 2016). The overall analysis of structural, kinetic, and mutagenesis data highlighted MthRnl's dimeric architecture, the surface-

exposed residues, active-site configuration, work in a calculated harmony to orchestrate substrate recognition, binding, and multi-step catalysis that is required for the desired RNA circularisation.

1.6.2 *Pyrococcus furiosus* RNA Ligase

In the hyperhalophilic archaeon *Pyrococcus furiosus* there has been found to be a thermostable RNA ligase (PfuRnl) encoded by the gene PF0353, that displays distinctive catalytic capabilities that can be utilised and expanded in RNA manipulation (Yang et al., 2022). When compared to other ATP-dependent RNA ligases, PfuRnl directly adenylates the 5'-hydroxyl (5'-OH) terminal without the need for prior phosphorylation, on top of also completing the traditional adenylation of 5'-phosphorylated RNA and DNA oligonucleotides (Yang et al., 2022). This additional feature is especially effective when PfuRnl is fused to T4 polynucleotide kinase, which forms a bifunctional chimera with the capability of processing 5'-OH substrates in a single step (Yang et al., 2022). From the functional assays that had been performed it showed that PfuRnl was extremely active at temperatures 65–85 °C which was expected considering the extreme heat of its natural environment. It was also shown that PfuRnl had a particular preference for single stranded substrates and supported both intra- and intermolecular ligation. Domain analysis and mutagenesis suggests that although there is retention of a conserved nucleotidyl-transferase fold, there are key variations that support its broader substrate scope and thermal resistance. The ability to generate the 5'-adenylated RNA molecules under extreme conditions makes this RNA ligase a strong competitor for RNA-seq library preparation and ligation-based RNA modifications in which the high temperature steps would be beneficial (Yang et al., 2022).

1.6.3 *Pyrococcus abysii* RNA Ligase

In the hyperthermophilic archaeon *Pyrococcus abysii* there is a 381-residue ATP-dependent RNA ligase (PabRnl) that functions as a homodimer and can effectively catalyse the intramolecular ligation of single stranded RNA or DNA, to covalently form a closed circular molecule (Brooks et al., 2008). PabRnl catalyses the three-step process of ligation of an AMP intermediate, adenylation, and the final formation of the phosphodiester bond. PabRnl exhibits noticeably strong thermostability, with its optimal activity around 65 °C; it's found that above this temperature, the phosphodiester bond can be found to predominate, compared to lower temperature (~35°C) de-adenylation becomes favoured, indicative of a

temperature regulated control over the reaction (Brooks et al., 2008). The C-terminal 127-residue domain, mediated homodimerisation and functions in trans to modulate activity. Truncating this region abolished dimer formation and ligation capability (Brooks et al., 2008). PabRnl shows remarkable substrate promiscuity with the ligation of both RNA and DNA, suggesting the possible representation of an ancestral, dual substrate enzyme that pre-dates the evolution of the more specialised ligases (Brooks et al., 2008). The robust thermal activity, multi-turnover efficiency, structural modularity, and substrate versatility of PabRnl, make it a compelling candidate for ancestral ligases and a possibility in biotechnological applications where there is a demand for thermostable nucleic acid circularisation.

1.6.4 *Palaecoccus pacificus* RNA Ligase

Within *Palaecoccus pacificus* a thermophilic archaeon, there is a homodimeric RNA ligase (PpaRnl) that was extensively characterised both structurally and biochemically to reveal a canonical nucleotidyltransferase fold supplemented by accessory domains that are typical with the RNA ligase family (Rousseau et al., 2024). Through high-resolution crystallography, and mutational analysis, a second Mg^{2+} binding site was discovered. The Mg^{2+} found in the active site is a canonical catalytic metal that helps in the coordination of the RNA during the critical second and third steps of catalysis (AMP transfer and the phosphodiester bond formations). Due to the positive pull of the Mg^{2+} it attracts the negative charge of the RNA. The aspartate residue Asp94 plays a central role in collating this auxiliary metal ion, facilitating the stabilisation of the phosphate tail of ATP and nucleotidyl transfer intermediates (Rousseau et al., 2024). The findings support the possibility of a two-metal-ion catalytic mechanism, where the second Mg^{2+} (Mg^{2+}B) allows enhancement of substrate positioning, and transition state stabilisation, which has the effect of lowering the activation energy for ligation (Rousseau et al., 2024).

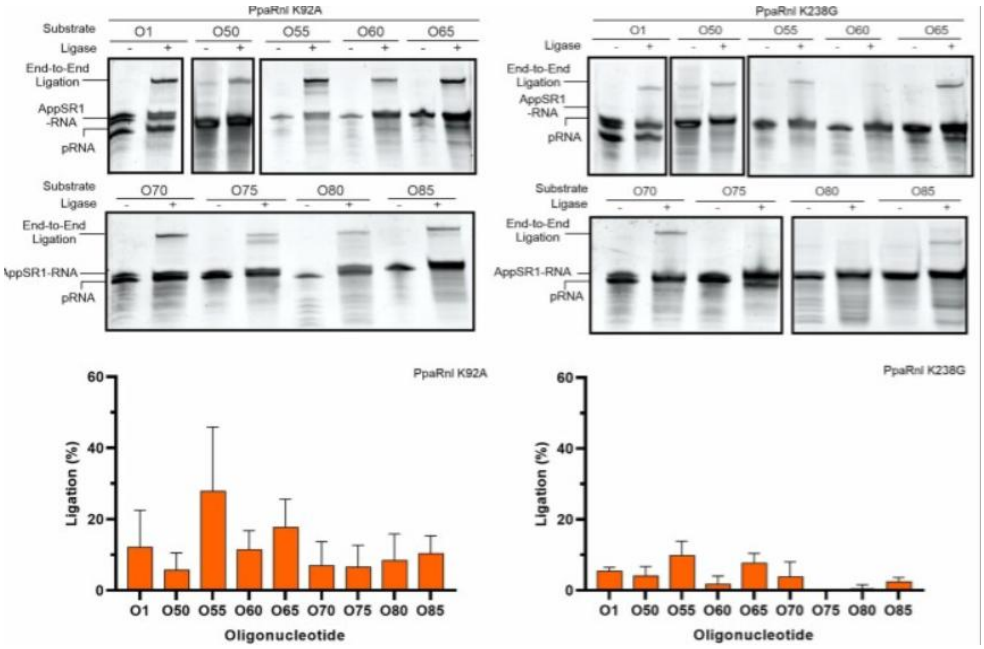

Figure 1.3 *K92A and K238G mutants exhibit 3'-blocked SR1 RNA ligation across an oligonucleotide panel that differs in sequence, secondary structures, and Tm. As stated, a no enzyme control is included in every assay response. The input substrate and the various reaction products are labelled as follows: AppSR1-RNA or AppSR1-DNA (adenylated SR1 RNA or DNA), pRNA/pDNA (input RNA or DNA), etc. Triplicate assays are carried out, and a representative gel picture is displayed. Graphs show the triple average with standard error-representative errors (Rousseau et al., 2024).*

The functional assays performed (Figure 1.3) showed that although the wild-type PpaRnl has the ability to adenylate a diverse variety of substrates, the ligation activity is modest at best; through the engineering of variants like motif I (Lys92 to Ala) and motif V (Lys238 to Gly) this outcome can change, and offer enzymes with an improvement in ATP-dependent/independent ligation, without compromising the engagement of the second Mg^{2+} (Rousseau et al., 2024). This relatively small modification, which led to a noticeable improvement in enzymatic activity, provided valuable insight into the underlying mechanism of how this RNA ligase operates. Specifically, it contributed to a more comprehensive understanding of the enzyme's functional dynamics and revealed new information about its preferences when interacting with RNA substrates. Highlighting whether its activity is more influenced by the nucleotide sequence or the structural conformation of the RNA.

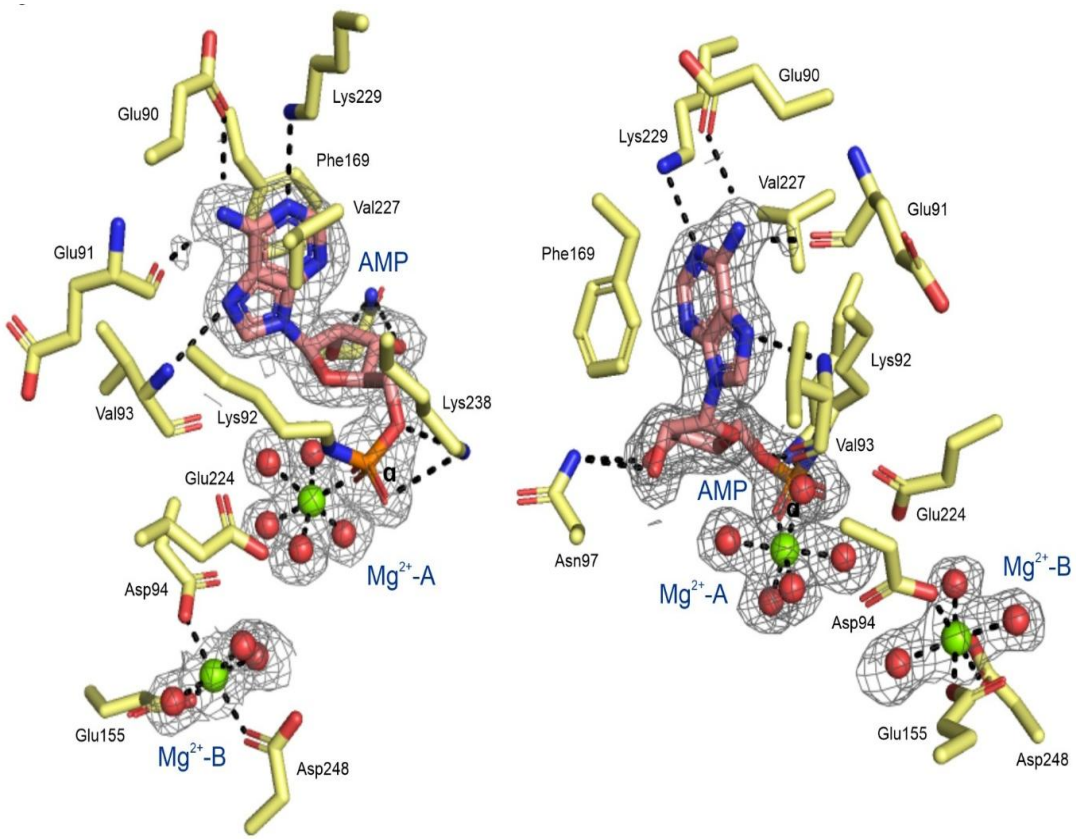

Figure 1.4 *Stereoview of the second Mg^{2}+ binding site (Mg^{2+}-B) with 2Fo – Fc electron density map for AMP (grey) and the PpaRnl wild-type AMP-Mg^{2+}(H_2O)5 complex. Stick models of amino acids and AMP are displayed; the active site's amino acids are coloured yellow, while the AMP is pale pink with orange phosphorus atoms. The interacting fluids and Mg^{2+} ions are shown by red and green spheres, respectively. Dashed lines represent atomic contacts (Rousseau et al., 2024).*

In the figure above (Figure 1.4) the two Mg^{2+} binding sites can be visualised, Mg^{2+}A in the active site and Mg^{2+}B in the secondary location. The proximity to the active site of this second Mg^{2+} encourages the possibility of its role in the positioning of RNA molecules and the ability of this ligase to carry out the three step process of adenylation, AMP transfer and the formation of the phosphodiester bond.

The identification of this second Mg^{2+} and the involved coordination by Asp94 sheds a new light on the understanding of the archaeal RNA ligase catalysis and opens the door to more possibilities in the realm of engineering optimised thermostable ligases, for the application in biotechnology. Structurally, PpaRnl functions as a homodimer and possesses a unique domain architecture distinct from that of mesophilic ligases like T4 RNA ligases. This configuration, along with its diverse cofactor preferences, may contribute to a reduction in ligation bias, a common issue in RNA sequencing applications. Furthermore, engineering efforts have led to the development of mutant forms of PpaRnl that are ATP-independent

and lack self-adenylation activity. These modifications allow for more controlled and efficient adapter ligation by minimising unwanted side reactions and by-products that can distort miRNA quantification. Such advancements position thermophilic archaeal RNA ligases as promising tools for improving the accuracy and reliability of RNA-based analyses.

1.7 Project Aims

With the recent advancements in high-throughput small RNA sequencing technologies, a detailed understanding of the structural and functional properties of RNA ligases is increasingly critical. These enzymes play a central role in the understanding of miRNAs. Insights into their mechanisms have the potential to improve library preparation methods and reduce sequencing biases. Despite their importance, the current repertoire of DNA and RNA ligases remains limited, and comprehensive structural and mechanistic information is sparse.

Even with recent efforts to engineer ATP-independent ligases aimed at enhancing small RNA sequencing, inherent enzyme dependent biases persist. These biases arise from the ligases sequence and structure specific substrate preferences, which can distort the representation of miRNA populations, and complicate downstream analyses. This research aims to gain a deeper mechanistic and structural understanding of these enzymes. This is essential, both to inform the rational design of improved ligases and to optimise small RNA sequencing protocols. Accordingly, this research has three main objectives:

Objective 1:

Examine the influence of RNA sequence and secondary structure on substrate preference of the PpaRnl enzyme.

Objective 2:

Characterise the ligation activity of PpaRnl and the functional contribution of the second magnesium binding site in catalysis.

Objective 3:

Determine the three-dimensional structures of PpaRnl magnesium binding mutants by X-ray crystallography.

Chapter Two: Materials and Methods

2.1 Thermophilic RNA Ligase Expression in *Escherichia coli*

Using Twist Bioscience, the PpaRnl constructs were synthesised and codon modified for expression in *E. coli.* Each of the PpaRnl mutants were ordered cloned into pET28b between the restriction sites NdeI and XhoI, with an N-terminal hexahistidine tag by Dr Joanna Hicks. The sequenced plasmids were transformed into *E. coli* BL21 (DE3) for expression.

2.1.1 Storage of Transformed RNA Ligases into *E. coli* BL21 (DE3)

A single colony from the LB agar plates with successful transformation was inoculated into 10 mL of LB Broth to form the seeder cultures (for the future preparation of 1L expression cultures). 10 µL of 50 µg/ml-1 kanamycin was added and the cultures were incubated at 37 °C at 180 rpm overnight. Glycerol stocks of each mutant were prepared with a 1:1 ratio of 50% (v/v) glycerol to seeder culture and stored at -80 ℃

2.1.2 Large Scale Expression Cultures of RNA Ligases

Prepared seeder cultures were inoculated into 1 L of Terrific Broth. The prepared cultures were incubated at 37 ℃ at an rpm of 180 rpm until at an optical density (OD) of A600nm = 0.4-0.6 was reached. Once within the desired OD, cultures were induced with 1 mL of 1M isopropylthio-β D-galactosidase (IPTG). For mutants D248A and D94A-D248A incubation after induction was overnight at 22 ℃ and 37 ℃ for D94A. After overnight incubation the cells were pelleted via centrifugation at 4,600 rpm for 30 minutes at 4 °C. The supernatant was then be discarded. The pellet was then re-suspended in 30 mL of lysis buffer (50 mM Tris pH 7.4, 200 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol), transferred to a 50 mL Falcon tube, and centrifuged at 9,000 rpm for 30 minutes at 4 °C, again discarding the supernatant. Pellets were the stored at -80 ℃.

2.2 Purification of RNA Ligases

Purification of RNA ligases was carried out via immobilised metal affinity chromatography (IMAC) and gel filtration chromatography. The purity and

solubility of the protein was assessed via electrophoresis SDS-PAGE gels. Measurement of the protein within its folded state was done using gel filtration chromatography and SDS-PAGE gel electrophoresis.

2.2.1 Immobilised Metal Affinity Chromatography

Frozen cell pellets were thawed and resuspended in 30 mL lysis buffer with manual mixing with 1 mL transfer pipette and vortexing. To the re-suspended pellet 1 µL of 25 U/µL benzonase nuclease was added (for a final concentration of 0.83 U/ml), and one cOmplete^{TM}, Mini, EDTA-free Protease Inhibitor Cocktail tablet (Roche) was added to inhibit cellular proteases. After an incubation at 70 ℃ for 30 minutes the pellet was sonicated on ice to promote cell lysing. The QSONICA Q700 sonicator with a ¼ inch microtip probe (12.5-25 mL volume) was set to alternating 1 second bursts at an amplitude of 4 and no sonication for a total of 2 minutes. On completion of sonication the lysate was centrifuged at 9,000 rpm for 30 minutes at 4 °C (if after this cycle the supernatant still appears cloudy, an additional centrifugation step was performed at the same conditions). The supernatant was passed through a 0.2 µm filter into a new falcon tube. IMAC purification was completed using an NGC FPLC system (BioRad). A 5 mL HisTrap^{TM} column (GE Life Sciences) was attached to the NGC and pre-equilibrated into lysis buffer before manual loading of the filtered supernatant. The NGC was programmed to first wash the column with 15 mL lysis buffer, followed by a high salt wash buffer (50 mM Tris pH 7.4, 4 M NaCl, 20 mM Imidazole, 10% (v/v) glycerol) to a total volume of 25 mL to remove any un-bound protein and any occurrence of nucleic acid that coprecipitated with the RNA ligases. Another wash of 15 mL lysis buffer was run through the column to remove remnants of the high salt wash buffer. Elution of the hexahistidine-tagged RNA ligases was done with a gradient of elution buffer (50 mM Tris pH 7.4, 200 mM NaCl, 1 M Imidazole, 10% glycerol) gradually increasing to 100% over 50 ml, collecting 20 x 2 mL fractions from across the gradient. Fractions were collected and analysed via SDS-PAGE.

2.2.2 Gel Filtration Chromatography

Any fractions containing the RNA ligase protein from the IMAC purification were collected, combined, and concentrated using a Amicon Ultra-15 10 kDA Centrifugal Filter, centrifuged at 3000 rpm until the volume was at approximately 5 mL. The injection loop of the FPLC was manually flushed with ~10 mL of gel filtration buffer (50 mM Tris pH 7.4, 200 mM NaCl, 10% (v/v) glycerol), followed

by loading of the concentrated protein to proceed with injection onto a HiLoad 16/60 Superdex Gel Filtration Column (GE Life Sciences) which had been pre-equilibrated with gel filtration buffer with a flow rate of 1 ml/min. The fractions containing the eluted protein were collected and stored at 4 °C and analysed by SDS-PAGE.

2.3 SDS-PAGE gel Electrophoresis

Protein samples were prepared for SDS-PAGE using pre-cast Mini-PROTEAN® TGX™ gels (Bio-Rad). Each sample had approximately 10 µL of the insoluble, soluble, or collected fraction which was mixed in individual tubes with 20 µL of 4× SDS loading dye. For gel loading, 10 µL of Precision Plus Protein™ ladder (Bio-Rad Laboratories) was added to the first lane, followed by 10 µL of each prepared sample loaded into separate wells. The gel tank was filled with 1× Tris-Glycine SDS-PAGE running buffer, and electrophoresis was carried out at 200 V for 30 minutes, or longer if the tracking dye had not yet reached the bottom of the gel**.** To visualise the protein bands, the gel was removed from the cast, placed into a box and fully covered with Coomassie Fairbanks stain. The stained gels were microwaved for 30 seconds and left to sit in the stain for five minutes, shaking at 120 rpm. The excess stain was poured off, gel was rinsed with distilled water, then covered with destaining solution (10% acetic acid). After 30 seconds in the microwave a paper towel was placed on top of the gel, and the gel was left in the destaining solution overnight shaking at 120 rpm. The gels were imaged using an iBRIGHT**™** gel imager (Invitrogen, Thermo Fisher Scientific)

2.4 Determination of Protein Concentration

For the determination of protein concentration, the Denovix ^{TM} 2000 (Life Science Technologies) was used and measures absorbances at 280 nm. The Denovix ^{TM} calculated the concentration of the protein using the Beer-Lambert law ($A = \varepsilon cl$ Beer-Lambert equation. A= absorbance, Ɛ = molar absorption co-efficient (L mol-1 /cm-1), l = pathlength (cm) and c = concentration (molL-1)). The produced Denovix ^{TM} readings were corrected by dividing it by the RNA ligase's molar absorption coefficient (PpRnl = 0.9 L mol-1/cm -1) which was determined via ProtParam.

2.5 Thermal Shift Assays

Thermal stability assays were carried out with a 5x SYPRO orange fluorescence dye (made from a 50x stock) in combination with buffers within a pH of 5-9 (40 mM boric acid, 40 mM phosphoric acid, 40 mM acetic acid). The final concentration of the enzyme in all conditions was 10 µM. Each of the thermal shift assays were prepared in triplicate with a final volume of 25 µL as seen in the table below. Protein samples were then heated at 0.5 °C/min within the range of 25 to 100 °C. through the duration the fluorescence intensity was measured at 470 nm with a gain of 2 in a Rotogene qPCR machine.

Table 1 *Thermal shift assay reactions with a constant enzyme concentration throughout conditions, varying pH buffers, and all components were made using MQ H_2O*

Component	Volume (µL)
Reaction Buffer (40 mM H_3BO_3, 40 mM H_3PO_4, 40 mM CH_3COOH, pH 5-9)	18
Enzyme (2 mg.ml-1)	4 (0 in control samples)
SYPRO Orange Dye (5x stock)	3
Gel Filtration Buffer (50 mM Tris pH 7.4, 200 mM NaCl, 10% glycerol)	0 (4 in control samples)

All fluorescence data was smoothed and normalised. Using GraphPad Prism the first derivative (FD) was calculated. The sigmoidal curves undertook non-linear fitting in GraphPad Prism as per the Boltzmann equation Y= fluorescence emission, X= temperature. Bottom = baseline fluorescence at low temperature; Top = maximal fluorescence; Slope = steepness of the curve; Tm = melting temperature of the protein; to determine the melting point ™ that occurs at the midpoint of the unfolding of the protein.

$$Y = Bottom + (Top - Bottom)/(1 + \exp(T_m - \frac{X}{Slope}))$$

2.6 Ligation Activity Assays

The RNA ligase enzymes that were used in the assays were taken from the previously purified and concentrated stocks. Stocks were made to 20 μM via dilution with gel filtration buffer (same composition as that used in purification). The master stocks of the cofactors and substrates were made to a concentration of 200 mM using Ultrapure DNase/RNase-free distilled water (Thermofisher). Before preparation all labware and working area was treated with RNase Away™ to reduce any RNase contamination.

2.6.1 Activity Assays for Thermophilic Ligases

Each reaction mixture was made to a total volume of 20 μL. 10 μM wild-type or mutant RNA ligase, 5 μM 5'-phosphate-labelled or pre-adenylated RNA or DNA oligonucleotides, and NEBuffer 1.1 (New England BioLabs), were made up in combination with or without 1mM ATP. All reactions were then incubated at 65 °C for 90 minutes (unless otherwise stated) in a T100 Thermal Cycler (BioRad) with the lid temperature set higher than the incubation temperature to reduce evaporation. Reactions were initiated by the addition of the enzyme and subsequently stopped by a 1:1 dilution with 2× formamide loading dye and heated for five minutes at 95 °C in a T100 Thermal Cycler (BioRad). The samples were then either stored overnight at 4 °C or frozen at −20 °C.

Table 2 *Set up of ligation reactions. Any variation in substrate concentrations was altered by the volume of H_2O. Any other variations in oligonucleotides were altered where indicated.*

Component	Volume (μL)
NEBuffer™ 1.1	2
1 mM ATP	1.4
100 μM Oligo	1
Ultrapure DNase/RNase-free H_2O	5.6
Enzyme (20 μM)	10

2.6.2 Urea-PAGE Gel Preparation, Staining and Imaging

The adenylation and ligation products for each enzyme assay were analysed via 20% Urea-PAGE gels, prepared as in the table below. To ensure the quality of the gels was high, they were made and used the same day. 10x TBE Buffer was used in gel preparation but diluted to 1x with Ultrapure H_2O for running. To combine all components to a homogeneous solution they needed to be heated to 50 °C for the Urea to dissolve. Once ~50% of the urea had been dissolved 10% (w/v) APS was added in, once fully combined, it was removed from heat and TEMED added. Solution was poured into pre-prepared multi-gel caster (Hoefer). Fill to the top and ensure any air bubbles have dispersed, then place in ten-well comb into each gel slot. Gels were covered with damp paper towel and left to set at room temperature for 30 minutes.

Table 3 *Components and volumes for the preparation of five 20% Urea-PAGE gels.*

Components	Volume
10x TBE Buffer	2.25 mL
Urea	18.9 g
40% Acrylamide/Bis Solution (29:1)	22.5 ml
H_2O	4.5 ml
10% APS	0.18 ml
TEMED	0.036 ml

After gels were set, they were pre-run at 100 V for 30 minutes in 1x TBE buffer. This was followed by thorough rinsing of each well with 1x TBE buffer to remove any urea that may have settled. ~10 µL of sample was loaded into each well. Urea-PAGE gels were run at 150 V for 2 hours or until the dye reached the bottom of the gel. A solution of 50 mL1x TBE buffer and 5 µL of SYBR™ Gold Nucleic Acid Gel Stain (10,000x concentration, for a final concentration of 1x) was made for staining each gel. The gel plus staining solution was wrapped in tinfoil to prevent

light exposure and incubated for 15 minutes, shaking at 160 rpm. Gels were imaged using an iBRIGHT™ gel imager (Invitrogen, Thermo Fisher Scientific).

2.7 Structure Determination Through Crystallography

2.7.1 High Throughput Crystal Screens

High throughput crystal screens were used to identify any promising crystallisation conditions, which would then undergo further optimisation to produce crystals that would be viable to be used in X-ray diffraction. These were the conditions set up in a low-profile 96-2 well INTELLI-Plate: IndexTM-HR2-144, PEGRxHTTM-HR2-086, CrystalHTTM-HR2- 130, NatrixHTTM-HR2-116/117 and the SaltRxHTTM-HR2-136 (Hampton Research). The Mosquito crystallisation robot (TTP LabTech Ltd) was used to place a drop of the pre-purified protein and reservoir solution at a 1:1 ratio into each well. For each condition there was a reservoir solution (aka crystallisation solution) consisting of 200 µL and 200 nL sitting drops. Screens were checked and scored weekly for any crystallisation.

2.7.2 Hanging-drop Fine Screens

Any conditions that showed crystallisation from the high throughput screens were further optimised through scaling up the conditions to hanging-drop fine screens. The well reservoir (crystallisation solution) was set up in a 24 well VDX tray with a 1 mL volume from concentrated stock solutions. The purified protein was concentrated to ~7 mg/mL (this concentration may vary depending on the PpaRnl mutant). For the hanging drop a 1:1 mix of well reservoir solution and concentrated protein to a total volume of 2 µL was placed on the siliconised cover slip and placed drop side down over the well now lined with silicone grease. pH and precipitant concentration for the crystallisation parameters were optimised for each condition. The checking for crystal growth within the crystallisation drops was done under a dissection microscope daily for the first week and weekly from that point forward. The conditions were based on the NatrixHTTM robot screen (HR2-117, Hampton Research): 0.08 M strontium chloride hexahydrate, 0.02 M magnesium chloride hexahydrate, 0.04 M sodium cacodylate trihydrate pH 7.0, 20% v/v (+/-)-2-Methyl-2,4-pentanediol and 0.012 M Spermine tetrahydrochloride. All crystals were grown at 18 °C

2.7.3 X-ray Diffraction-Crystal Preparation

Using a cryo-loop and cryo-protectant solution which consists of the crystallisation condition of the crystal and 15% glycerol (v/v) each crystal was looped and transferred. Crystals were then flash-cooled and stored in liquid nitrogen.

2.7.4 Data Collection

All X-ray diffraction data was collected at the Australian Synchrotron (Melbourne, Victoria) on the available MX beamline equipped with an EIGER x 16M detector (Dectris). Data collection occurred over 360 ° rotation of the crystal over 360 seconds at a beam attenuation of 60-80%

2.8 Data Processing

2.8.1 Scaling, Indexing, and Integration

The diffraction images were indexed, scaled, integrated, using the XDS (Kabsch, 2010). Crystals were indexed in the space group 20 ($C222_1$) the data sets were manually changed to space group 18 ($P222_12_1$) due to the original group resulting in unnecessarily high R_{work} and R_{Free} values in the produced models; the dimensions of the unit cell were also adjusted accordingly. When merging the reflections, this was done via AIMLESS within the CCP4 suite. AIMLESS also allowed for the assessment of data quality. FreeR flag dataset (5% of dataset reflections not used in refinement and to compute R_{free}) was generated in AIMLESS.

2.8.2 Twinning and Non-crystallographic Symmetry Detection

Collected data was analysed in AIMLESS and *phenix.xtriage* from PHENIX suite for any evidence of twinning and translational non-crystallographic symmetry.

2.9 Structural Analysis

2.9.1 Matthew's Coefficient

To determine the number of monomers in the asymmetric unit the solvent content was calculated using Matthew's coefficient (Matthews, 1968) within the CCP4 suite.

2.9.2 Molecular Replacement

The structure of *P. pacificus* RNA ligase (8UCE) (Rousseau et al., 2024) was taken from the protein data bank (PDB). The single monomer was extracted from the

structure through PyMOL (The PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC) and all ligand and waters were removed. Using *phenix.phaser* molecular replacement was able to be carried out within the PHENIX suite. The 8UCE structure was used for molecular replacement in all PpRnl mutant datasets.

2.9.3 Building and Refining the Model

The process of building and refining was originally done in *phenix.autobulid* in the PHENIX suite. The structure that resulted was manually refined and built using COOT, the 2Fo-Fc and Fo-Fc electron density maps were set to 1σ and 3σ. There were repeated rounds of real-space refinement and auto-building using *phenix.refine* and *phenix.autobuild*.

2.9.4 Ligand Fitting and Refinement

The structure was readied for ligand fitting by the building, refinement, and addition of waters using *phenix.refine*. Any ligands were identified through a process of elimination of a list of ligands that had been identified in similar RNA ligases, used buffers, and crystallisation conditions. *CheckMyBlob*, *CheckMyMetal*, and *phenix.ligandfit*, were used to confirm placed ligands were correct. The manual real-space refinement and overall visual inspections were completed in COOT.

2.9.5 Analysis of Structure

In the PHENiX suite, *MolProbity* was used to assess the Ramachandran outliers. All images and maps of the model were constructed in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC) and COOT. Bond distances were also measured in PyMOL and COOT, classification of hydrogen bonds that were moderate were any within the range of 2.5-3.2Å, weak bonds were within the range of 3.3-4.0Å. A comparison of the wild type structures to the mutants was done in PyMOL using the alignment tool. This calculated the root mean square distances (RMSD) of each structure comparatively.

2.9.6 Structural Analysis

To generate images of the model PyMOL was used. The Hydrogen bonds were classified as being either weak (3.3-4.0Å) or moderate (2.5-3.2Å) distances; these were measured in PyMOL. The comparison of structures was done by calculating the root mean square distances (RMSD) with the alignment tool in PyMOL.

Chapter Three: Results and Discussion

3.1 Substrate Preferences for PpaRnl K238G

Previous research identified that mutation of the active site lysine (K) 238 to glycine (G) resulted in improved activity with increased intramolecular (circularisation) or intermolecular (concatemerisation) ligation. While with the wild-type (WT) PpaRnl enzyme, only adenylation of the substrate was observed. This previous research demonstrated increased ligation activity of PpaRnl K238G with oligonucleotide 65 (Rousseau et al., 2024.) The substrate preference in relation to sequence and RNA secondary structure was explored. For the determination of substrate specificity of the *P. pacificus* RNA ligase K238G mutant a range of ligation assays were carried out, investigating the differing preference depending on structure versus sequence.

3.1.1 Expression and Purification of PpaRnl K238G

The PpaRnl K238G enzyme was purified using previously published methodology which required some optimisation (Rousseau et al., 2024). The plasmid pET28a-PpaRnl K238G was transformed into chemically competent *E. coli* BL21 (DE3). This was initially done with the first incubation on ice occurring for 15 minutes, followed by a 45 second heat shock at 42 ℃, addition of 1 mL of LB broth and incubation for 60 minutes at 37 ℃ before plating. However, there were very few colonies present on the transformation plate. An increase in the first incubation period from 15 minutes to 30 minutes, saw a resolution to this issue.

The next issue was encountered during the purification process. Following existing protocols using LB broth for expression cultures resulted in a low yield of protein after IMAC and size exclusion chromatography. A switch was made from LB broth to Terrific broth in the large-scale expression stage. Cell pellets were found to be much larger in size and resulted in an increased yield of protein after purification (Figure 3.1, Figure 3.2).

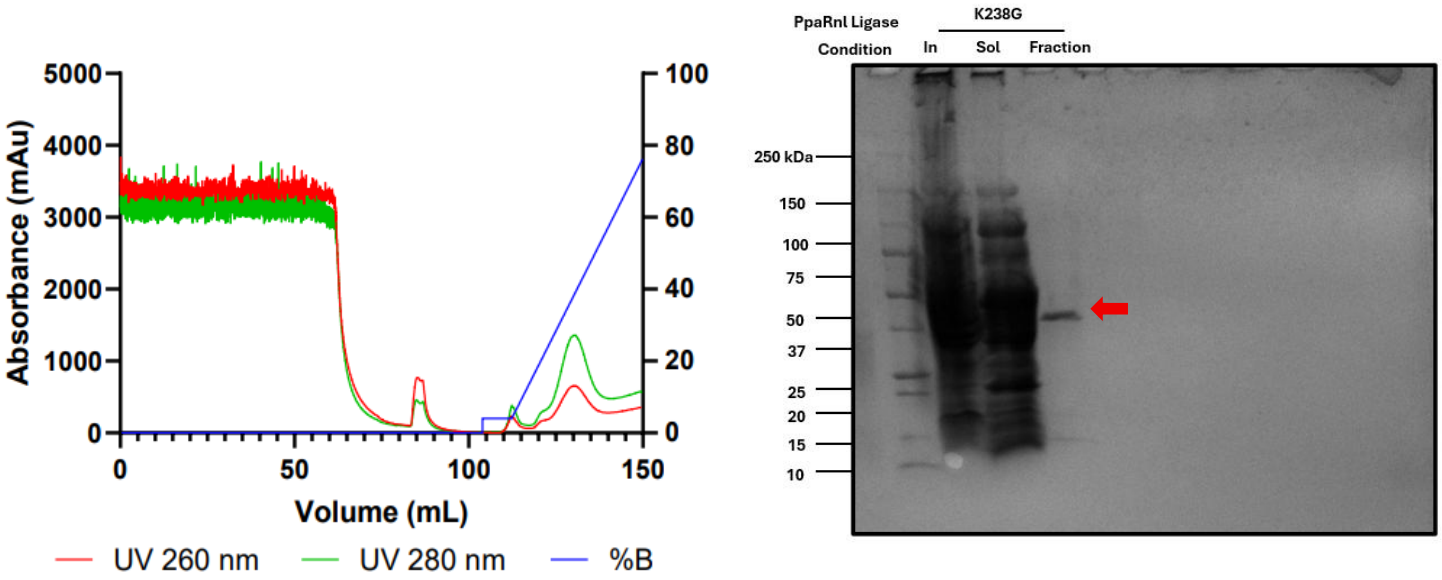

Figure 3.1 *Chromatogram of immobilised metal affinity chromatography (IMAC) PpaK238G purification. The fractions within the peak at ~125 mLwere collected to proceed with gel filtration chromatography. An SDS-PAGE gel was run with the insoluble, soluble, and collected fraction, the first well was loaded with the Bio-Rad Precision Plus Protein™ ladder, with molecular weight markers indicated in kilodaltons (kDa) along the side. The 280 nm peak corresponds to the protein present. With 260 nm representing the RNA.*

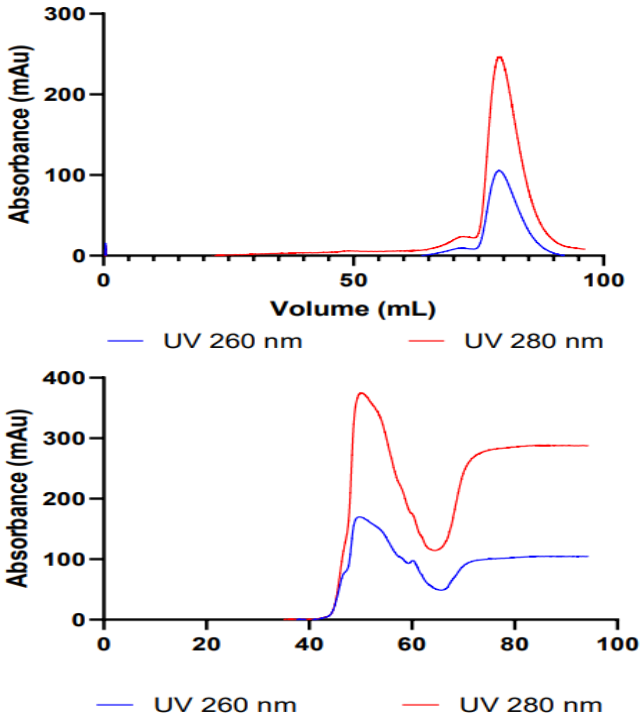

Figure 3.2 *Gel filtration chromatography of PpaRnl K238G. A comparison of the chromatograms from size exclusion performed with a HiLoad 16/60 Superdex Gel Filtration Column (GE Life Sciences) on an NGC FPLC system (BioRad). Yield of PpaRnl K238G indicated by peak height at 280nm indicating increased yield using Terrific broth (bottom) vs LB broth (top panel).*

3.2 Activity Assays

For the determination of substrate preference in the case of structure versus sequence, a range of activity assays were performed that varied in substrates and temperature.

3.2.1 Design of RNA Oligonucleotides to Elucidate Differences in Structure versus Sequence

In previous research there was a focus on the improvement of the wild-type PpaRnl as only adenylation of the RNA substrate was observed in these assays. The behaviour of the RNA ligase was determined with varying substrate sequence compared to substrate secondary structure. Previous research performed activity assays with PpaRnl K238G with the following oligonucleotides: 1, 50, 55, 60, 65, 70, 75, 80, 85. These were designed by Dr Joanna Hicks, and range in sequence, structure and T_m (Figure 3.3). PpaRnl K238G displayed substantially better activity with oligo 65 compared to the other RNA oligos in the panel.

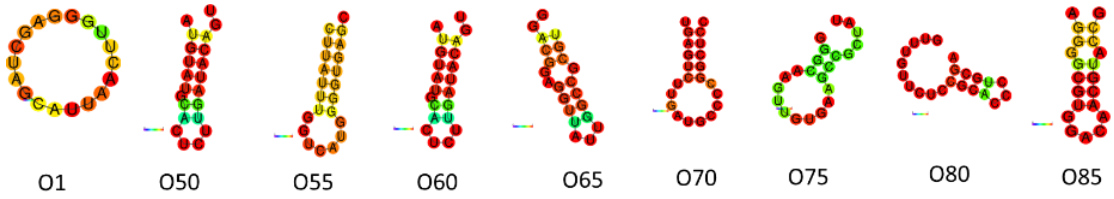

Figure 3.3 *The sequence and structure of the different oligonucleotides designed by Dr Joanna Hicks. From left to right showing; oligo 1, oligo 50, oligo 55, oligo 60, oligo 65, oligo, 70, oligo 75, oligo 80, oligo 85. Number represents the melting temperature of the oligonucleotide i.e. O60 indicates a T_m of 60 ℃.*

Within this research and the desire to understand substrate bias, five new oligonucleotides were designed. Oligo 65 V2 and V3 have a highly similar sequence and structure to oligo 65 but vary at the end of the folded molecule. Oligo 65 has a 5' overhang, whereas Oligo 65 V2 has a 3' overhang, and Oligo 65 V3 has no overhang (blunt end). Both variants of Oligo 65 had the same T_m as the original oligonucleotide sequence. Oligonucleotides N1, N3, and N4 were designed to have differing structures but highly similar sequences (Figure 3.4) with the same T_m. Each oligonucleotide has a different secondary structure but only varies in sequence by one base.

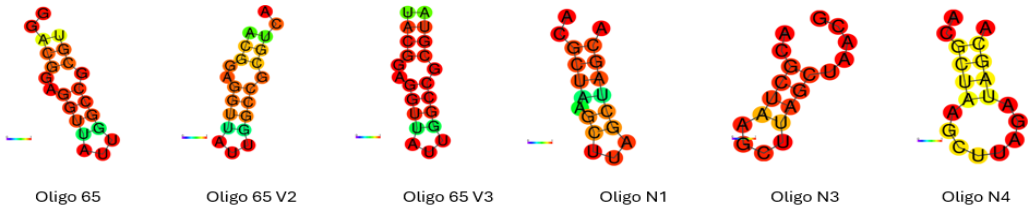

Figure 3.4 *The sequence and structure of the newly designed oligonucleotides with highly similar structure and sequence to oligo 65. Oligos N1-N4 represent RNA substrates with highly similar sequences (vary by one nucleotide) but differing secondary structures.*

3.2.2 Activity Assay Results

Assay reactions were set up as in previously conducted activity assays and analysed via 20% Urea-PAGE gels (Rousseau et al., 2024). In the previous protocol, the loaded gels were run at 200 V for one hour. This was no longer applicable as the gels began to run hot and the samples ran off the gels within that duration. An edit to the run of 150 V for two hours saw this issue resolved. The reactions on each assay were set up at three temperatures: 55 ℃, 65 ℃, 75 ℃. This was to investigate ligation activity when the oligonucleotide was folded (55 ℃), partially folded (65 ℃) and mostly unfolded (75 ℃) (Table 6).

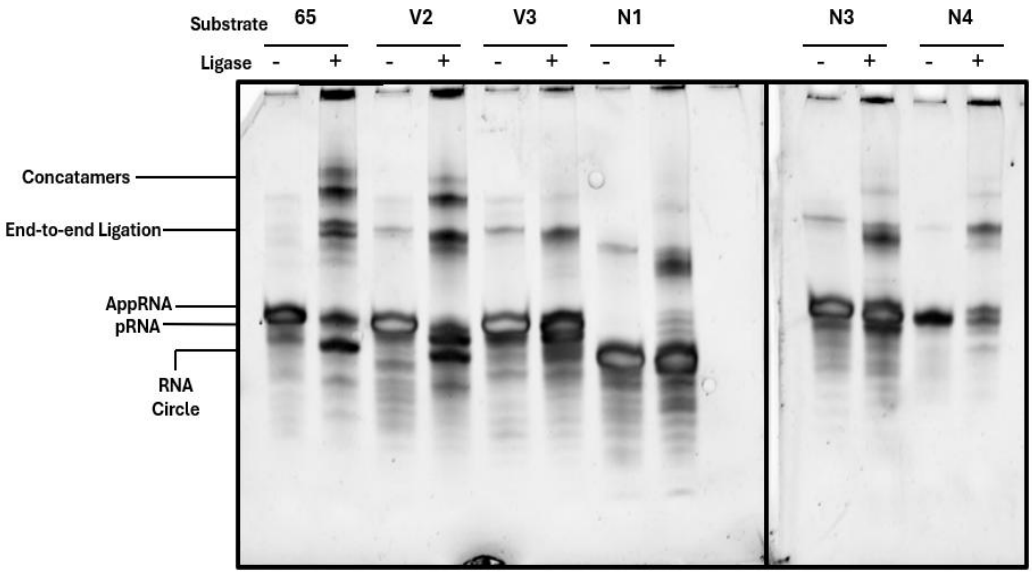

Figure 3.5 *PpaRnl K238G activity assay with varying oligonucleotide substrates at 65 ℃. Bands are a representation of the ligation step corresponding on the side. Each different substrate that has been used is noted at the top; each substrate has a control i.e. no addition of PpaRnl K238G denoted by the "- ".*

With the assay performed at 65 ℃ (Figure 3.5) it was found that there was the expected bias toward oligo 65 (as previously observed). Notably, a comparable level of activity was also observed in the oligo 65 V2 reaction, indicating that this variant retained similar functionality under the assay conditions. There was little to no RNA circularisation present in the remaining oligos (Oligo 65 V3, N1, N3, N4) although end-to-end ligation and adenylation was present. The least amount of activity seen with oligo N4 possibly due to its slight change in sequence.

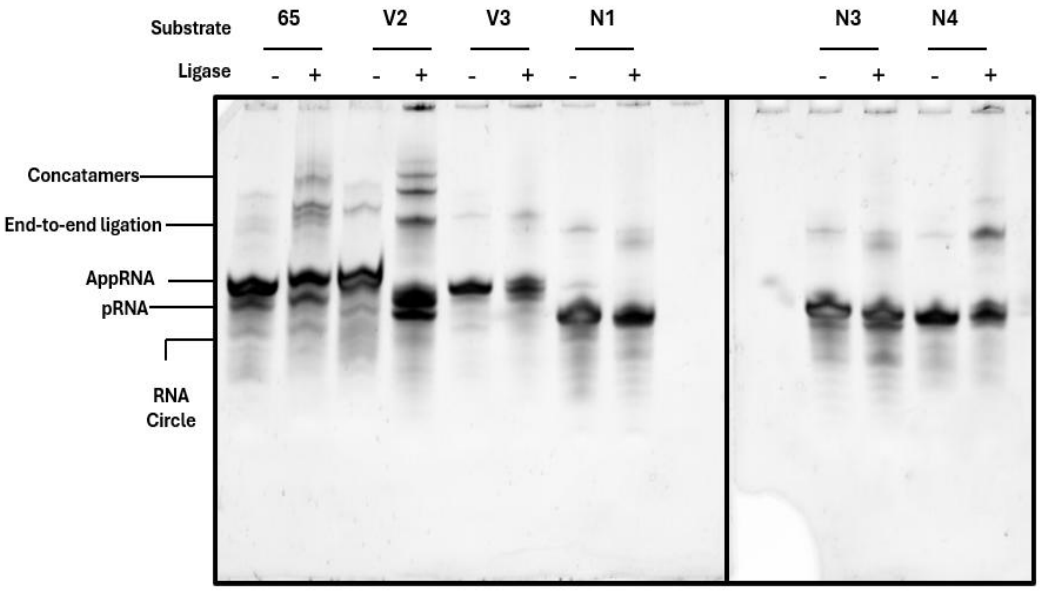

Figure 3.6 *PpaRnl K238G activity assay with varying oligonucleotide substrates at 55 ℃ Bands are a representation of the ligation step corresponding on the side. Each different substrate that has been used is noted at the top; each substrate has a control i.e. no addition of PpaRnl K238G denoted by the "- ".*

It was anticipated that in the assay conducted at a reaction temperature of 55 ℃ (Figure 3.6), enzymatic activity would be reduced. At this lower temperature, the secondary structures of the RNA were expected to remain largely intact, thereby impeding the enzyme's ability to efficiently perform ligation, adenylation, and circularisation. Consistent with previous observations at 65 ℃ it was seen that there was an increase in activity of oligo 65 and oligo 65 V2, with end-to-end ligation and adenylation being carried out. While all other substrates completed adenylation, oligo N4 also appeared to show a significant amount of end-to-end ligation comparatively.

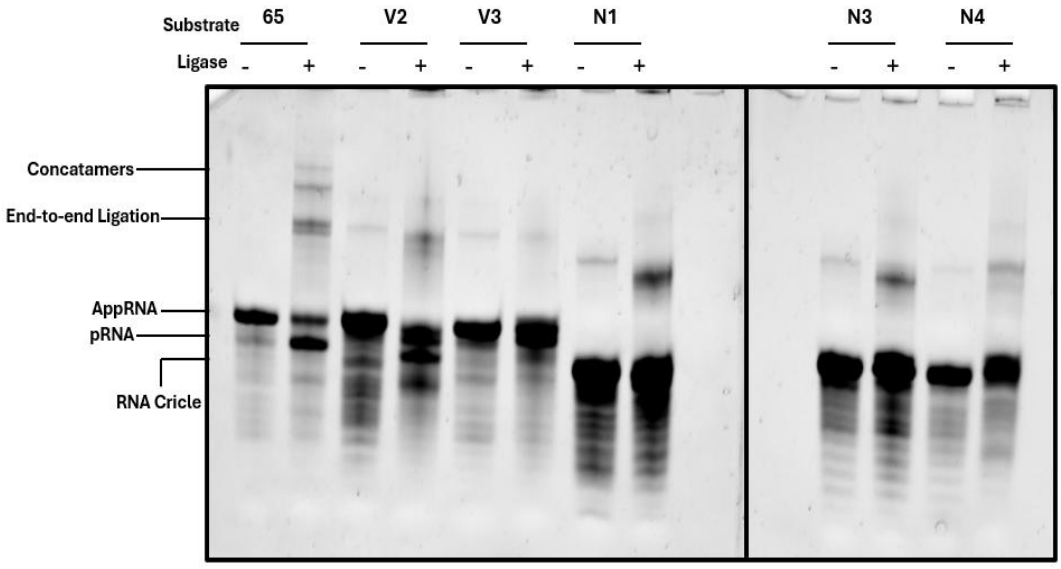

Figure 3.7 *PpaRnl K238G activity assay with varying oligonucleotide substrates at 75 ℃ Bands are a representation of the ligation step corresponding on the side. Each different substrate that has been used is noted at the top; each substrate has a control i.e. no addition of PpaRnl K238G denoted by the "- ".*

In the final assay conducted with a reaction temperature of 75 ℃ (Figure 3.7), the higher temperature was expected to facilitate the complete unfolding of the substrate, thereby eliminating the secondary structure and leaving only the primary sequence available. Under these conditions, the enzyme once again demonstrated a consistent preference for oligo 65, closely followed by oligo 65 V2. In contrast to the trends observed at lower reaction temperatures, an increase in activity was seen for oligos N1 and N3.

3.3 Mutation of Mg^{2+} B Binding Site Residues in PpaRnl

As seen in (Figure 1.4) there is a secondary Mg^{2+} (Mg^{2+} B) in PpaRnl that is located close to the active site but does not interact with the ATP cofactor. Previous research from the Hicks lab hypothesised a role for Mg^{2+} B in the coordination of the RNA and/or the ligation process. To investigate this, mutations were made to the WT with changes to the key amino acids responsible for the configuration of this second Mg^{2+}.

3.3.1 Mutants Designed

A total of four mutants were designed to determine the effect (if any) of the second Mg^{2+} (Mg^{2+} B) on ligation activity. Three of these were single point mutations with acidic, negatively charged residues replaced with the neutral amino acid alanine:

aspartic acid at position 94 was mutated to alanine (D94A), aspartic acid at position 248 to alanine (D248A), and glutamic acid at position 155 to alanine (E155A). The fourth construct was a double mutant, D94A-D248A designed to eliminate coordination of the Mg^{2+} B ion entirely. These mutations were selected to target key residues involved in Mg^{2+} B binding, and the substitution of negatively charged side chains (Asp and Glu) with a neutral residue (Ala) is expected to directly influence the metal ion's binding affinity.

All the above were ordered from Twist Bioscience for expression with an N-terminal His tag.

3.3.2 Expression and Purification of PpaRnl Mg^{2+} Binding Site Mutants

Using methods established for the purification of PpaRnl as in Rousseau et al, issues with protein expression and low yield were encountered. With the amount of protein obtained at a high enough concentration to be used in further experiments was less than 100 µL. To optimise protein expression for increased yield small-scale expression and nickel-pulldowns were carried out. The overnight induction temperature was varied; 22 ℃, 30 ℃, and 37 ℃. All the mutants expressed well at 22 ℃, with very little expression, based on the low yield, and faint bands, seen in D248A, and D94A-D248A at 30-37℃; (Figure 3.8). This however was not the case for D94A, although it showed the same lack of expression at 30 ℃ as the other mutants. The expression of D94A at 37 ℃ was significantly more than that at 22 ℃. This led to a change within the expression protocol, altering the induction temperature to reflect the results observed in the small-scale expression tests.

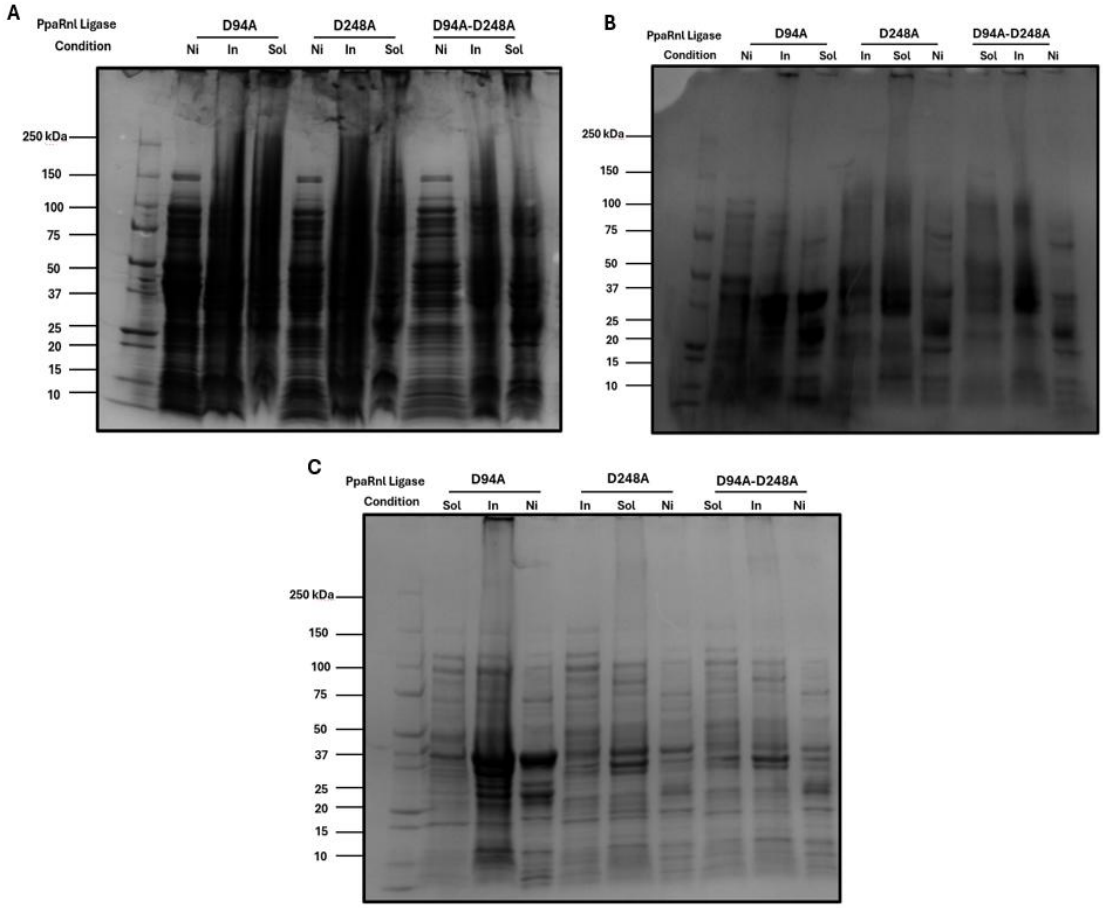

Figure 3.8 *SDS-PAGE gels of the soluble (sol) and insoluble (In) states of the small-scale expressions and nickel pulldowns (Ni) conducted on the PpaRnl mutants; D94A, D248A, D94A-D248A.) the samples were induced at 22 ℃ and incubated overnight. B) the samples were induced at 30 ℃ and incubated overnight. C) the samples were induced at 37 ℃ and incubated overnight. The first well is the Bio-Rad Protein Ladder- Precision Plus Protein™ with corresponding scale indicated on the side in kilodaltons.*

3.3.2.1 PpaRnl E155A

Despite multiple attempts at expression, PpaRnl E155A mutant cultures consistently exhibited poor growth compared to the other Mg^{2+} B site variants. Following a previous transformation into *E. coli* BL21(DE3), six colonies were obtained and used to inoculate two LB seeder cultures. The LB seeder cultures were grown overnight at 37 ℃, and visible growth was observed the following day. These cultures were subsequently used to inoculate 1 L TB expression cultures. Despite standard growth conditions at 37 ℃ with shaking, cell density remained abnormally low. After six hours, the optical density at 600 nm (OD_{600}) measured only 0.040, and after a further hour, had decreased slightly to 0.034. With an additional two hours of incubation (eight hours total), the OD_{600} of the E155A culture was still only 0.037, while control cultures of other mutants had reached approximately 0.500 within the same timeframe. Due to the extremely slow growth

rate, the cultures were discarded. This same pattern was observed in each attempt at growing E155A. With the limited time frame, further attempts at growth of this mutant were no longer pursued.

These observations indicate that the expression protocol effective for other mutants was insufficient for E155A. Possible alternatives to enhance growth include supplementing the growth medium with glucose to provide an additional carbon source and repress any leaky protein expression or using an auto-inducing broth to delay induction until the culture has reached a higher density.

3.3.2.2 PpaRnl D94A-D248A

Although the culture for expression of the PpaRnl D94A-D248A double mutant typically exhibited reasonable growth at 22 ℃, protein yield was extremely low after IMAC purification (less than 0.5 mg/mL in 100 µL from a 2 L culture). Based on expression yields from scaling up culture volumes for other mutants it was estimated that approximately 4 L of culture would be necessary to produce enough purified protein for initial screening. This still yielded a low amount of protein of <1 mg/mL in ~100 µL. Therefore, it was decided to massively scale the culture volume to a total volume of 10 L to obtain enough material for crystallisation or at least have enough to perform an activity assay to assess the effect of the double mutation on ligase activity. Unfortunately, this did not yield enough protein (Figure 3.9) at a concentration above 1mg/mL to allow for characterisation.

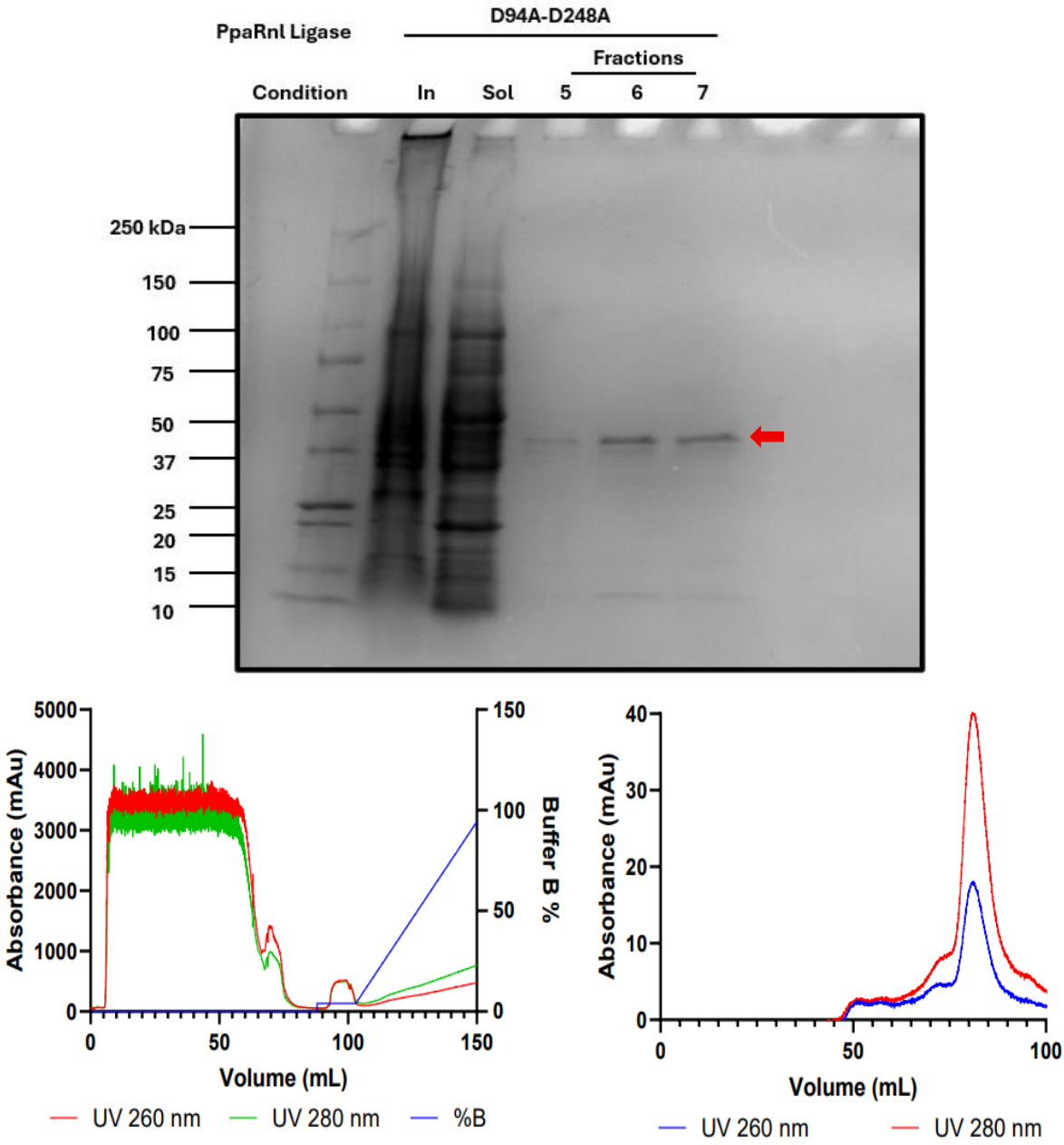

Figure 3.9 *Expression of PpaRnl D94A-D248A. SDS-PAGE gel of the insoluble and soluble samples of the large-scale expression cultures of D94A-D248A and the collected fractions of the completed gel filtration chromatography (right) with the fraction being found within the peaks of the IMAC (left). The first well in the gel is the Bio-Rad Protein Ladder- Precision Plus Protein™ with corresponding scale indicated on the side in kilodaltons.*

This result was unexpected, particularly given that prior expression tests conducted at 22 °C had shown detectable expression under similar induction conditions. Moreover, Terrific Broth (TB) was used for seeder culture preparation, and culture growth appeared comparable, although the pellets were visibly smaller than those of other mutants. To investigate further, a small-scale expression trial was performed in LB medium following the same protocol as previously used, with induction at 22 °C. This trial successfully produced detectable levels of protein,

albeit still at a low yield. Further optimisation, including refinement of induction timing and IPTG concentration, may be necessary to achieve reliable expression at preparative scale in the future. Due to low yield characterisation of the PpaRnl D94A-D248A double mutant was not continued in this project.

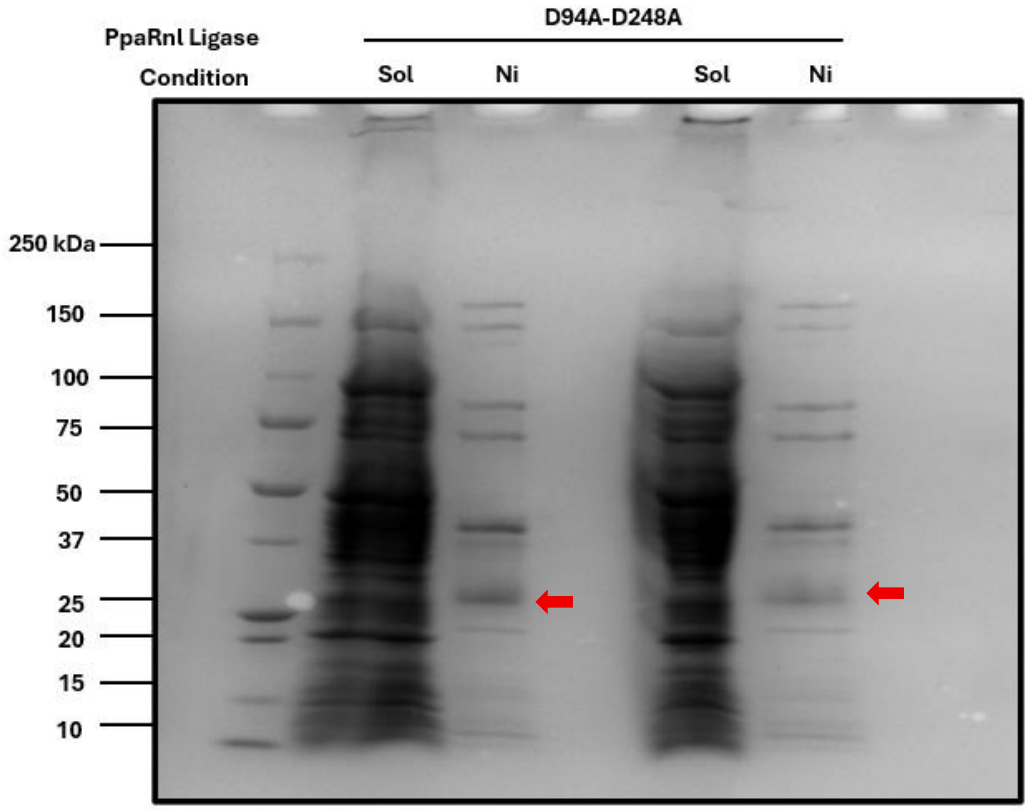

Figure 3.10 *Small-scale expression trials of the PpaRnl D94A-D248A mutant. Ni pull downs of small-scale expression cultures were performed in duplicate alongside soluble fractions collected from expression culture. In the SDS-PAGE analysis, the first well was loaded with the Bio-Rad Precision Plus Protein™ ladder, with molecular weight markers indicated in kilodaltons (kDa) along the side.*

3.3.2.3 PpaRnl D94A

Like other Mg^{2+} B site mutants, the PpaRnl D94A variant exhibited persistent expression challenges throughout the research process. Following transformation into *E. coli* BL21(DE3), only a limited number of colonies were observed on LB agar plates, indicating low transformation efficiency. Subsequent inoculation with seeder cultures with scale-up to large-scale expression resulted in consistently low protein yields, as confirmed by the chromatographic profiles compared to those in K238G (Figure 3.2). However, it was found that on the SDS-PAGE gels the protein was present (Figure 3.11).

Given that all four Mg^{2+} B mutants displayed similar issues, a decision was made to prioritise one mutant at a time for optimisation. Among the variants, D94A showed relatively higher growth and expression potential. Therefore, repeated transformations into BL21(DE3) were performed to obtain plates with sufficient colony density for reliable expression trials. This re-transformation process was repeated three times, each iteration incorporating protocol optimisations, such as adjusted incubation times and freshly prepared competent BL21 cells, before satisfactory results were achieved.

Despite these improvements, large-scale expression and purification continued to yield low quantities of recombinant protein. Although the target band corresponding to PpaRnl D94A was visible on SDS-PAGE gels, the amount of protein eluted from the column was consistently low, and insufficient for most downstream applications. However, with additional rounds of purification, the highest concentration achieved was approximately 7 mg/mL. At this concentration, and with sufficient total volume collected, it became feasible to proceed with subsequent activity assays, temperature melts, and crystallography of the PpaRnl D94A mutant.

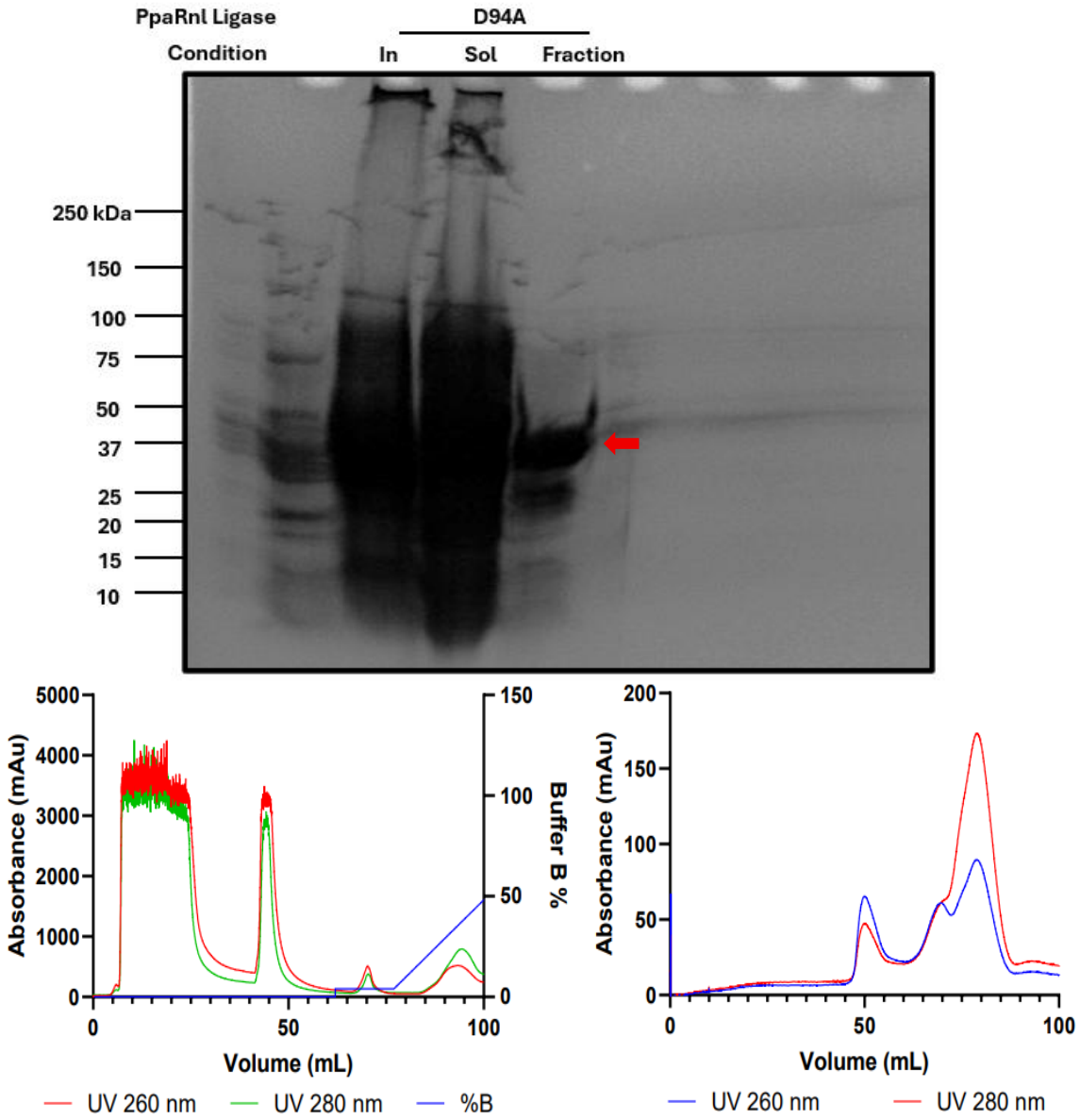

Figure 3.11 *Expression and purification of PpaRnl D94A. SDS-PAGE gel of D94A, with samples from the insoluble (In), soluble (Sol) and the fraction collected after the completed immobilised metal affinity chromatography (right) from the fractions collected after IMAC (bottom left). In the SDS-PAGE analysis, the first well was loaded with the Bio-Rad Precision Plus Protein™ ladder, with molecular weight markers indicated in kilodaltons (kDa) along the side.*

3.3.2.4 PpaRnl D248A

The initial expression and purification of PpaRnl D248A mutant yielded low protein concentrations, with decreases in protein yield noted after immobilised metal affinity chromatography and during the concentration step. Despite these losses, SDS-PAGE analysis of the insoluble, soluble samples, and collected elution fractions revealed a single band at the expected molecular weight, indicating that the protein was present and did indicate to be pure. Although the protein appeared to be expressing, the concentration of the collected protein remained below 1 mg/mL in ~100 ul. Further concentration would have resulted in a volume too small

to carry out further experiments. In response to this, a change in induction temperature was trialled, although as previously seen in (Figure 3.8) the expression at 22 ℃ seemed to be the optimal condition a switch was made to 37 ℃ to solidify this finding. When purified again, the yield was even lower than in the previous attempt resulting in approximately 500 µL of protein at ~1 mg/mL. with further concentration, this was potentially adequate for a limited number of assays but insufficient for crystallisation screening. After IMAC purification, the collected fractions 4 and 6 were analysed on an SDS-PAGE gel to assess the protein collected. It was found that although it was within the peak (Figure 3.12) there was no protein present in fraction 4.

Due to the persistently low protein concentrations obtained from 2 L cultures, with future crystallisation efforts, further purification was planned with the use of 4 L expression cultures. It was then decided to prepare sufficient material, expression was scaled up using two 5 L flasks, with a total volume of 8 L of culture. Based on prior results in which 4 L of culture yielded ~2mg/mL of protein it was anticipated that this increased volume might provide protein at or near the target concentration of ~7 mg/mL which would be ideal for the goal of crystallisation.

To address this issue further purification was scheduled using 8 L of culture. The aim was to obtain sufficient protein at the appropriate concentration for crystallisation and activity assays, although only ~3 mg/mL was collected at a large enough volume.

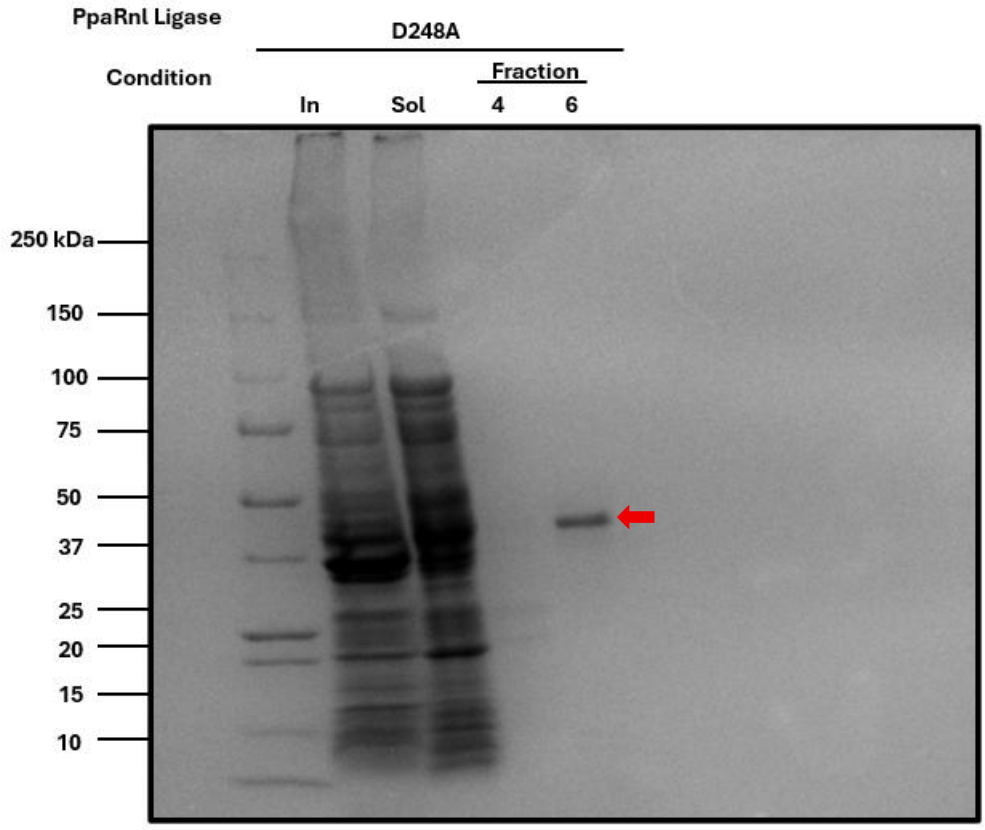

Figure 3.12 *SDS-PAGE gel of D248A, with samples from the insoluble (In), soluble (Sol) and the fractions collected from within the peak after the completed immobilised metal affinity chromatography. In the SDS-PAGE analysis, the first well was loaded with the Bio-Rad Precision Plus Protein™ ladder, with molecular weight markers indicated in kilodaltons (kDa) along the side.*

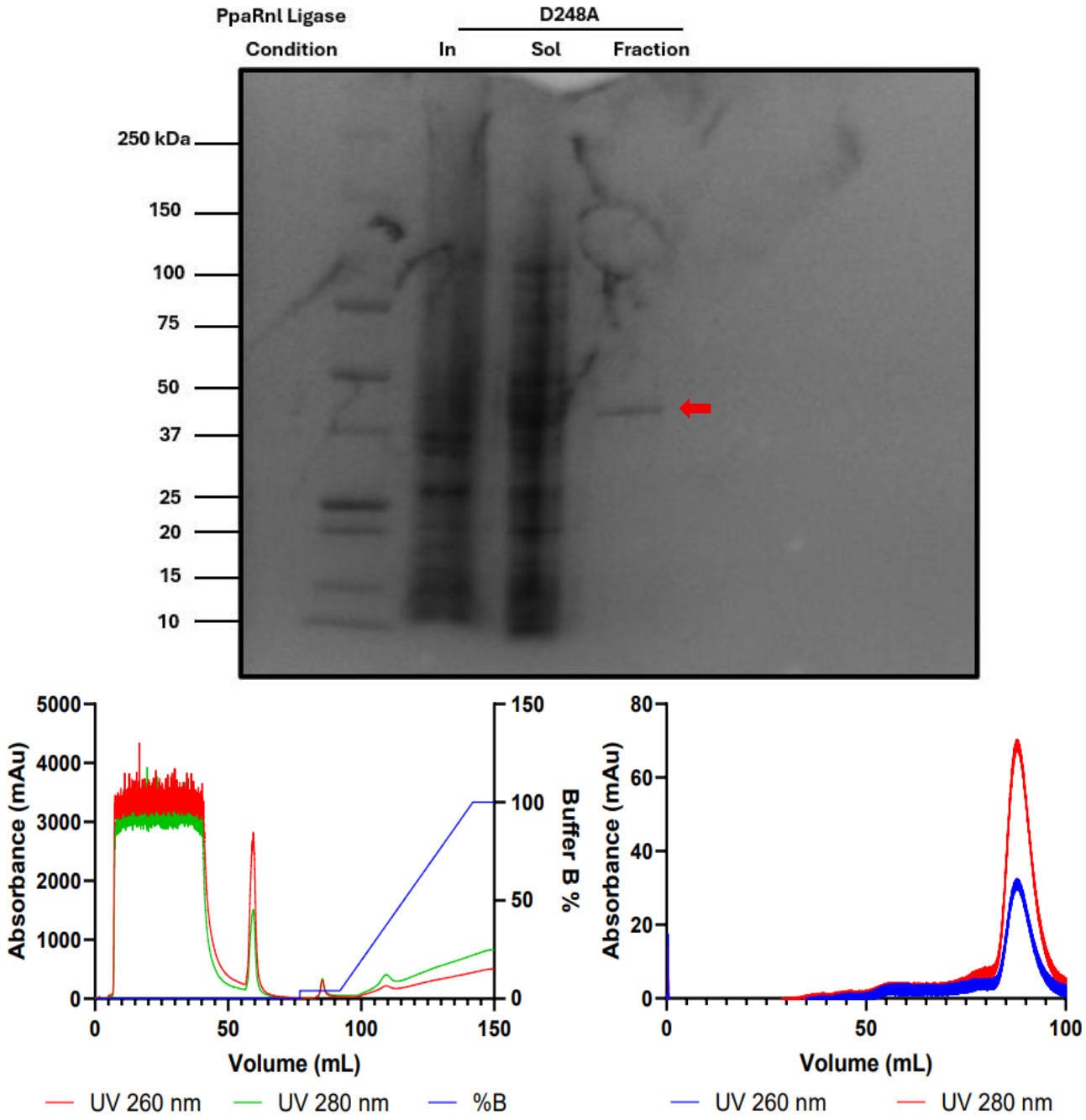

Figure 3.13 *Expression and purification of PpaRnl D248A. SDS-PAGE gel of D248A, with samples from the insoluble (In), soluble (Sol) and the fraction collected after the completed immobilised metal affinity chromatography (bottom right) after completed IMAC (bottom left). In the SDS-PAGE analysis, the first well was loaded with the Bio-Rad Precision Plus Protein™ ladder, with molecular weight markers indicated in kilodaltons (kDa) along the side.*

From the four mutants that this research began with, only two were included in characterisation experiments. The inability to get the E155A mutant cultures to grow to a high density where purification was able to be carried out, hindered its characterisation. A similar issue was encountered with PpaRnl D94A-D248A. Although the expression culture grew, the amount of protein collected after

purification did not provide a large enough volume of a viable concentration to procced with characterisation. Therefore, due to the limited amount of time within this project further research was conducted with the remaining two mutants, D94A and D248A.

3.3.3 Activity Assays with PpaRnl Mg^{2+} Binding Site Mutants D94A and D248A

A series of activity assays were conducted to assess the functional impact of the mutations that were introduced around Mg^{2+} B in PpaRnl. It is hypothesised that this site plays a critical role in coordination of the RNA into the active site (Rousseau et al., 2024.) By evaluating the ligation activity of each mutant, with a comparison to the WT, it allows for a stronger ability to determine whether specific residues contribute to metal ion coordination, substrate binding, or efficiency. These assays provided functional evidence of how each mutation affected the activity and supported structural or mechanistic hypotheses regarding the role of Mg^{2+} B in the RNA ligase function.

Similar to the assays previously performed with the K238G mutant, assays were conducted under the same conditions with the ligation assay protocol. The identical environments were used with three temperatures: 55 ℃, 65 ℃, 75 ℃, to allow for the direct comparison of activity with previous results. The assays utilised the same set of newly designed RNA oligonucleotides: oligo 65, oligo 65 V2, oligo 65 V3, oligo N1, oligo N3, oligo N4. The selection of these oligonucleotides was to investigate the ligase's sequence and structural preferences.

3.3.3.1 Characterisation of PpaRnl D94A RNA Ligation Activity

To begin to understand the effects of the D94A mutant on the activity of the enzyme a series of preliminary control assays were set up. These assays were designed to directly compare the activity of D94A to that of the previously characterised K238G mutant. Within these assays oligo 65 was used as the substrate due to its well-established reactivity with K238G, making it a reliable control for comparative analysis.

The reactions were set up as follows:

- Gel filtration buffer + Oligo 65 (no ATP) (negative)
- K238G + Oligo 65 (no ATP)

- D94A + Oligo 65 (no ATP)
- K238G + Oligo 65 + ATP (triplicates)
- D94A + Oligo 65 + ATP (triplicates)

A 20 μM stock of D94A was prepared from concentrated, purified protein, and used in the assays. The inclusion of reactions without ATP served as negative controls, given that ATP is and essential cofactor in the ligation reaction. However, the enzyme can be purified in an adenylated form due to endogenous ATP in the *E. coli* expression system, sometimes resulting in minor activity in the absence of ATP. The controls with no ATP yielded ligation products indicating the enzyme or at least a proportion of the enzyme is purified in an adenylated form (Figure 3.14). In several reactions degradation of oligo 65 was observed. This observation prompted the preparation of fresh oligonucleotide for future assays. When comparing assay reactions in the presence of ATP, those containing K238G exhibited all three steps of the ligation process: adenylation of the 5′ phosphate, strand ligation (end-to-end joining), and RNA circularisation, noted by the distinct bands on the gel (Figure 3.14). In contrast, the reactions containing D94A and ATP only exhibited substrate adenylation. These results indicate that that the mutation of aspartic acid 94 to an alanine did not improve RNA ligase activity under the tested conditions. Unlike K238G, D94A failed to catalyse end-to-end ligation or RNA circularisation, however adenylation of the substrate was still observed. This adenylation is like that of the wild-type enzyme (Rousseau et al., 2024.) and are indicative that this had no improvement or reduction of activity.

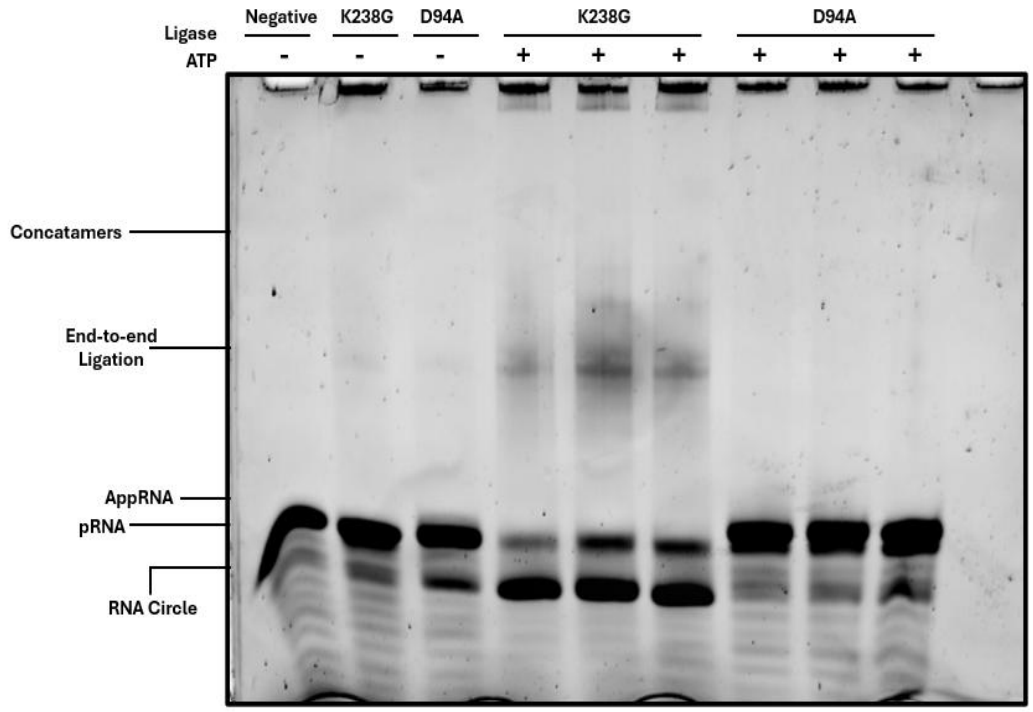

Figure 3.14 *Ligation activity of PpaRnl K238G and PpaRnl D94A with oligonucleotide 65 at 65 °C. Assay analysed via UREA-PAGE gel to compare the activity of D94A with K238G. Reactions were set up with a negative control (no addition of enzyme) and without the ATP cofactor (as indicated on the gel). Ligation products indicated to the left of the gel. Assays presented in triplicate.*

Further assays were carried out with the newly designed oligonucleotides (O65 V2, O65 V3, N1, N3, N4) alongside oligonucleotide 65 to evaluate the activity of PpaRnl D94A under various temperatures with different RNA substrates. These experiments were designed to mirror those previously performed with K238G. Three independent reactions were set up for each substrate and carried out at 55 °C, 65 °C, and 75 °C. At 55 °C (Figure 3.15), the results closely resembled those obtained in earlier assays with K238G (Figure 3.6). While the substrates were successfully adenylated, no additional enzymatic activity was observed. Specifically, there was no evidence of end-to-end ligation or circularisation for any of the substrates.

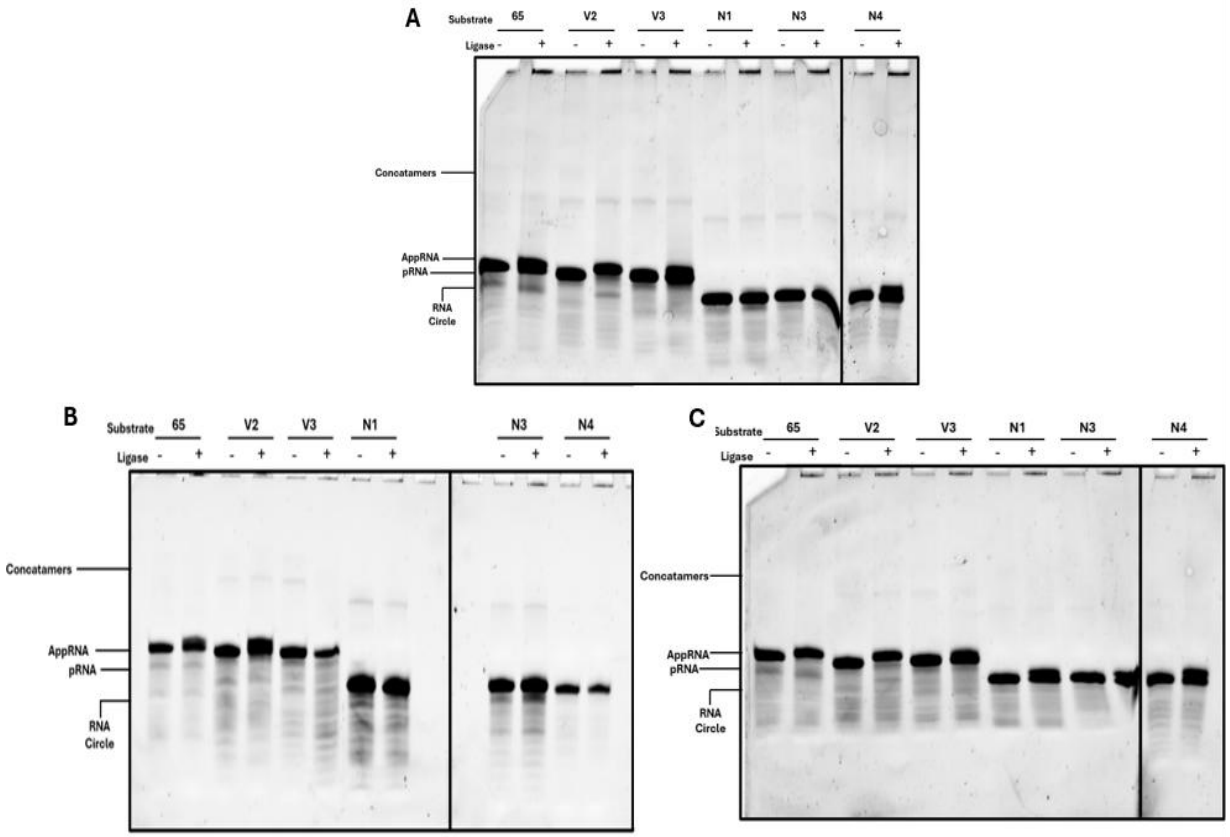

Figure 3.15 *Ligation activity of PpaRnl D94A with varying oligonucleotides at 55°C (A), 65 °C (B), and 75 °C (C). Assays analysed via UREA-PAGE gels. Each reaction with each oligonucleotide (labelled on top of gels), was set up with a negative, containing gel filtration buffer instead of enzyme prior to the reactions with the presence of enzyme (denoted by the "+").*

The last assay had the reaction incubated at 75 °C, this temperature being much higher than the T_m of the substrates, theoretically resulting in the complete unfolding of the secondary structures, thereby maximising accessibility of the RNA for ligation (Figure 3.15). Despite these conditions, the results mirrored those obtained at 55 °C and 65 °C, with no observable improvement in ligation efficiency for the D94A mutant. Although, assays performed at the same temperature using K238G showed a notable reduction in activity when compared to 65 °C (Figure 3.5); K238G still retained significantly greater ligation efficiency (Figure 3.7) compared to D94A under there identical conditions. These findings further emphasise the critical role of the Mg^{2+}B binding and its coordination in facilitating efficient catalysis. The significant loss of activity in D94A even under conditions that have supported ligation in K238G, suggests that proper Mg^{2+} coordination at this site is essential for maintaining enzymatic function.

3.3.3.2 Characterisation of PpaRnl D248A RNA Ligation Activity

As with D94A, preliminary control assays were set up with oligo 65. The reactions were made up with a 20 µM stock of D248A and K238G in the following conditions:

- Oligo 65 + gel filtration buffer + no ATP (Negative)
- Oligo 65 + K238G + no ATP
- Oligo 65 + D248A + no ATP
- Oligo 65 + K238G + ATP (duplicate)
- Oligo 65 + D248A + ATP (triplicate)

All reactions were incubated at 65 °C. In contrast to results seen with D94A (Figure 3.14), the D248A mutant demonstrated evidence of end-to-end ligation and RNA circularisation (Figure 3.16). Distinct bands corresponding to these ligation products were observed. Negative control reactions, which lacked either enzyme of ATP, displayed no additional bands on the gel, suggesting that the observed products were not due to RNA degradation or non-specific enzymatic activity. The absence of "background" bands was an indication that the bands being observed were due to the ligase activity of the enzyme. These results provided a promising indication that the substitution of aspartic acid at position 248 with alanine enhanced the ligation activity of PpaRnl. The observed improvement in ligation efficiency suggested that this residue plays a significant role in regulation catalytic performance and potentially coordination or the RNA into the active site.

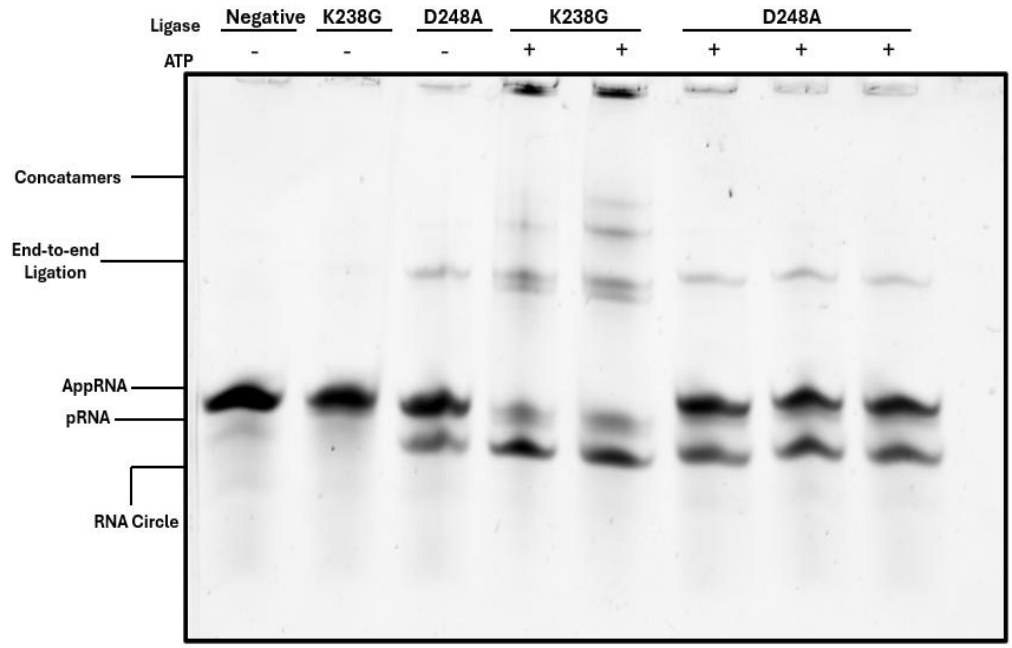

Figure 3.16 *Ligation activity of PpaRnl K238G and PpaRnl D248A with oligonucleotide 65 at 65 °C. Assays were analysed via UREA-PAGE. Reactions were set up with and without the ATP cofactor with multiple reactions set up for each mutant.*

Given the activity of PpaRnl D248A on oligonucleotide 65 we explored its activity using the newly designed oligonucleotides (O65 V2, O65 V3, N1, N3, N4) alongside oligonucleotide 65. Assays were carried out at the same three temperatures: 55 ℃, 65 ℃, and 75 ℃. As described previously, these temperatures were selected for they effect on the secondary structures of the oligonucleotides. 55 ℃ was associated with predominantly intact secondary structures, 65 ℃ corresponded to a partially denatured state, and 75 ℃ was sufficient to induce complete unfolding. This temperature range allowed for the assessment of how the PpaRnl Mg^{2+} B site mutants interacted with oligonucleotides in different conformational states, thereby providing insights into the role of RNA structure in ligation efficiency and substrate recognition versus sequence preference.

At 55 ℃, activity seen with D248A resembled the results previously observed for the K238G variant under the same assay conditions (Figure 3.6) with a lack of activity at this temperature. Notably, a distinct substrate presence was observed with reactions containing oligo 65 and oligo 65 V2 exhibiting significantly higher levels of circularisation compared to reactions containing the other oligonucleotide variants. These results reaffirm that this specific mutation has improved activity.

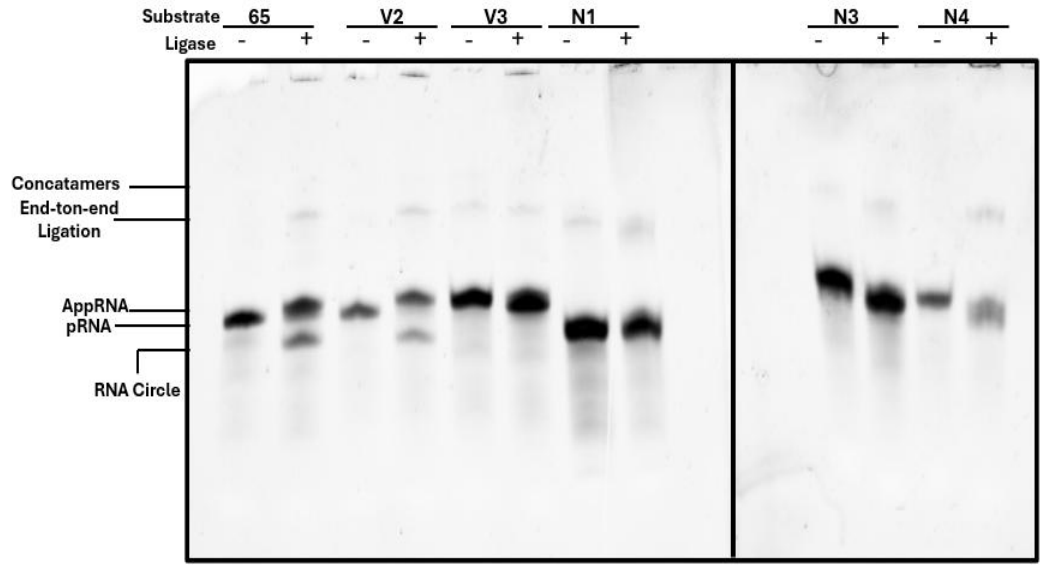

Figure 3.17 *Ligation activity of PpaRnl D248A with varying oligonucleotides at 55 ºC Assays analysed via UREA-PAGE. Each reaction with each oligonucleotide, was set up with a negative, containing gel filtration buffer instead of enzyme prior to the reactions with the presence of enzyme (denoted by the "+").*

As previously, a ligation assay for D248A was conducted at 65 ℃ (Figure 3.18), the temperature that had previously been observed to support the highest activity in assays, as seen with K238G. Consistent with results obtained for K238G, D248A demonstrated enhanced ligation efficiency under these conditions (Figure 3.5). As observed previously, there remained a noticeable substrate preference. Oligo 65 and Oligo 65 V2 continued to show the highest levels of ligation, with additional, comparatively lower activity detected for oligos N1, N3, N4. In contrast, ligation activity toward oligonucleotide 65 V3, which possesses blunt ends, remained substantially reduced, suggesting that the structure significantly influences enzymatic efficiency. These findings indicate that the D248A mutant retains substrate selectivity patterns, similar to those of K238G, and further emphasised the role of RNA secondary structure, terminal configuration, and sequence, in

modulating ligase activity.

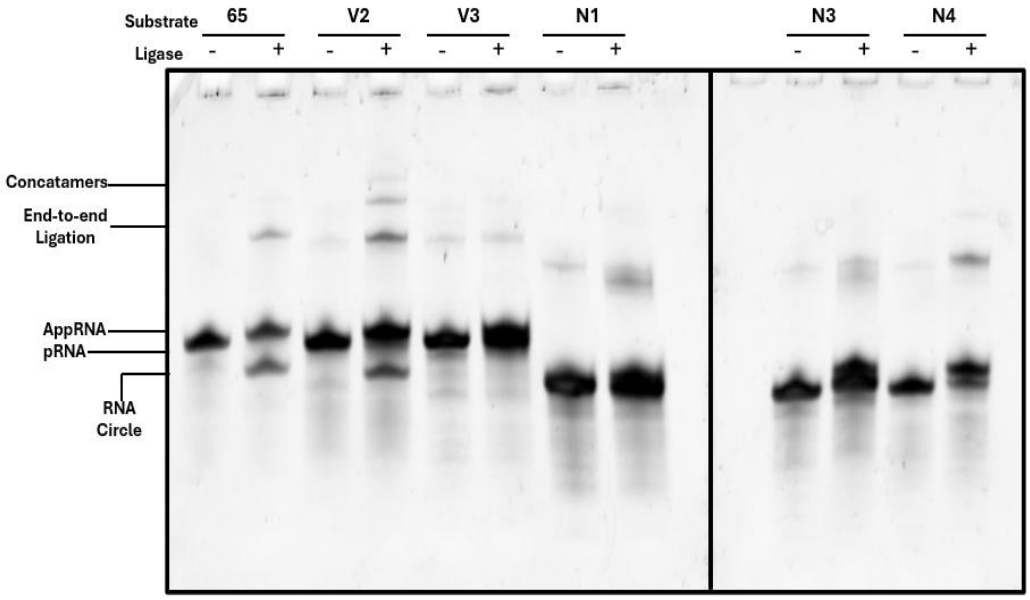

Figure 3.18 *Ligation activity of PpaRnl D248A with varying oligonucleotides at 65 ℃ Assays analysed via UREA-PAGE. Each reaction with each oligonucleotide, was set up with a negative, containing gel filtration buffer instead of enzyme prior to the reactions with the presence of enzyme (denoted by the "+").*

The ligation assay for D248A was conducted at 75 ℃ (Figure 3.19) to allow for comparison across all mutants. Compared to the K238G mutant, which retains detectable ligation activity under these conditions (Figure 3.7), D248A exhibited a significantly reduced amount of activity. The results more closely resembled those observed for the D94A mutant, with the predominant reaction being adenylation of the substates across most of the conditions tested. Only minimal levels of end-to-end ligation was observed at 75 ℃, restricted to reaction containing oligo 65, oligo 65 V2, and N1 substrates. In these cases, the corresponding ligation bands were distinctly faint and far less intense than those observed in assays conducted at 65 ℃. This indicated that the D248A mutation seems to have compromised ligation efficiency at higher temperatures, despite partial adenylation activity being retained. These findings suggest that the mutation of Asp 248 to alanine may also have its roles in facilitating complete ligation activity under thermophilic conditions, although still improves overall efficiency of the enzyme.

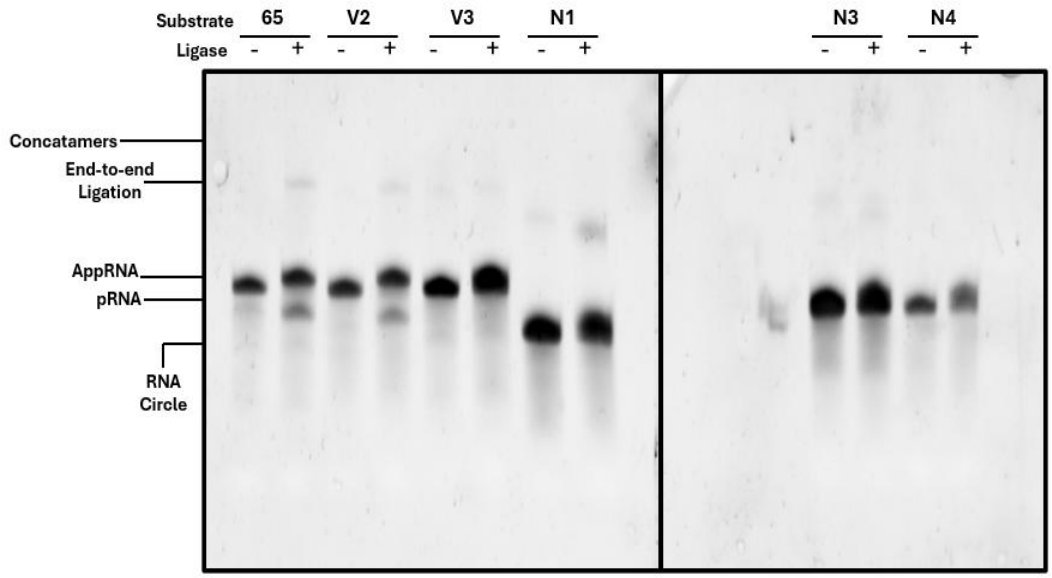

Figure 3.19 *Ligation activity of PpaRnl D248A with varying oligonucleotides at 75 ℃ Assays analysed via UREA-PAGE. Each reaction with each oligonucleotide, was set up with a negative, containing gel filtration buffer instead of enzyme prior to the reactions with the presence of enzyme (denoted by the "+").*

3.3.3.2.1 Activity of D248A with Oligonucleotides of Varying T_m and Secondary Structures

In previous research, it was demonstrated that the K238G mutation in PpaRnl resulted in enhanced enzymatic activity (Rousseau et al., 2024). As a part of that investigation, a series of ligation assays were conducted using a panel of designer oligonucleotides of increasing T_m and sequence variation (Table 6). Oligo 50, oligo 55, oligo 60, oligo 65, oligo 70, oligo 75, oligo 80, oligo 85. Among these substrates, K238G consistently showed a preference for oligo 65 (Figure 3.20), which led to its selection as the standard substrate for the assays performed with these mutants as outlined above. As the D248A mutation also demonstrates improved activity, activity assays were carried out with the same set of designer oligonucleotides to determine whether the preference for oligo 65 observed in K238G, was maintained despite the different mutation. This comparative approach was designed to assess whether the improvement in activity observed in D248A was accompanied by changes in substrate preference, and a possible means to improve activity further.

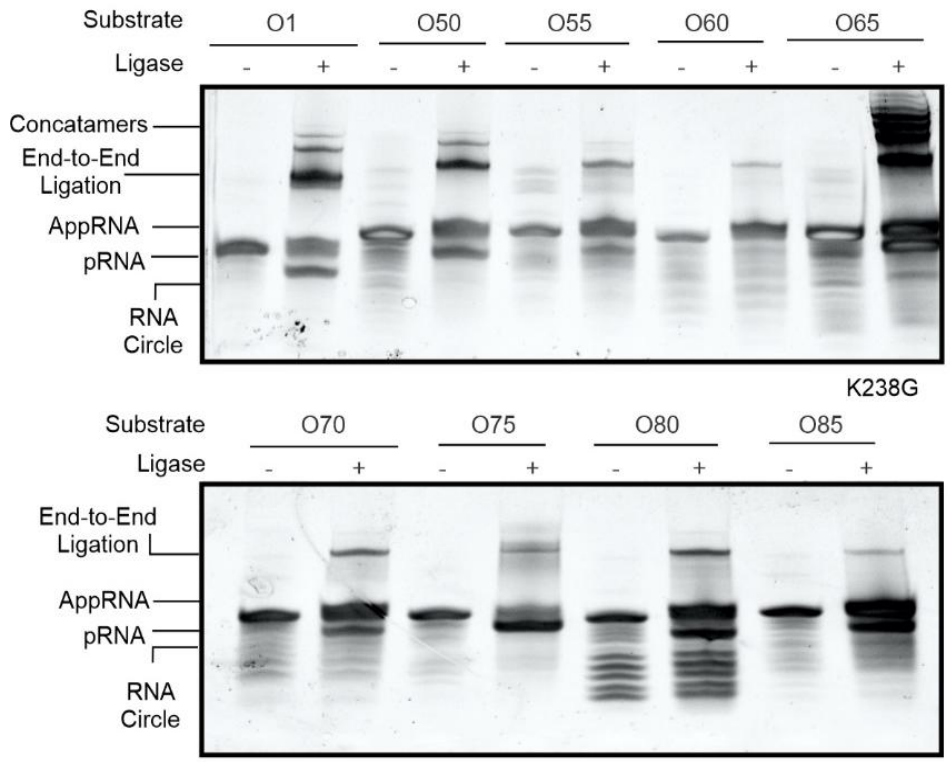

Figure 3.20 *Substrate ligation and adenylation activity of K238G across a test panel of nine-oligonucleotides varying in sequence, secondary structure and T_m in the presence of 70 μM ATP. Adenylation, circularisation and end-joining ligation activity denoted by the corresponding band (Rousseau et al., 2024)*

As the activity of D248A with oligonucleotide 65 had already been established, this specific reaction was not repeated in the present assay. Prior data had demonstrated that oligo 65 yielded the highest activity with this mutant across assays. Activity of D248A with the same oligos showed that in reactions containing oligonucleotides 50, 55, 60, both end-to-end ligation, and RNA circularisation were observed; the efficiency of these reactions was significantly low compared to those previously observed with oligo 65. The ligation with the oligonucleotides with the lower T_m were subtle, suggesting that while D248A retained some capacity to process these substrates, its activity was significantly diminished.

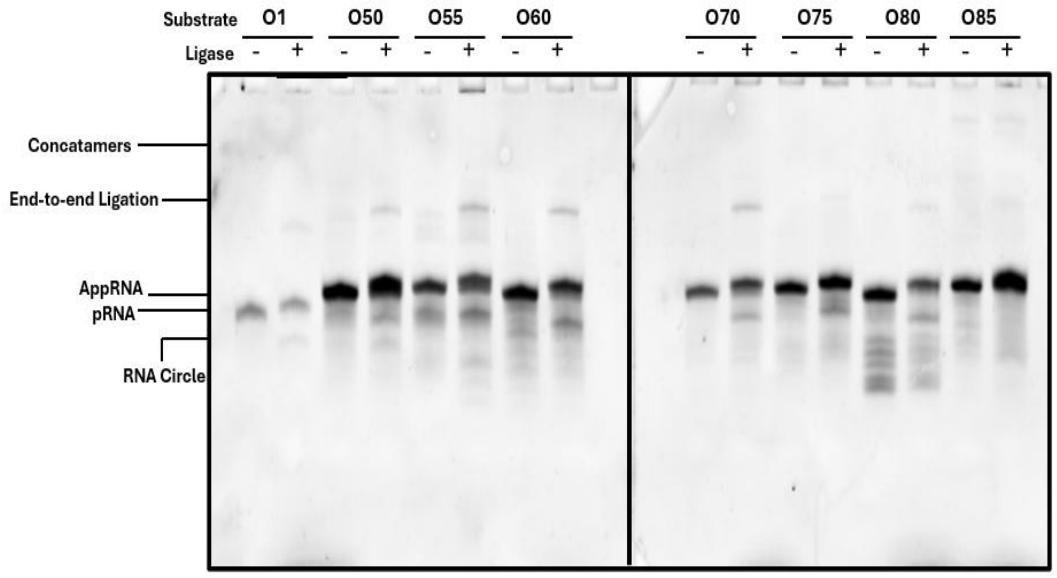

Figure 3.21 *Assay performed on UREA-PAGE gel using D248A with an incubation temperature at 65 ℃. Each reaction with each oligonucleotide, was set up with a negative, containing gel filtration buffer instead of enzyme prior to the reactions with the presence of enzyme (denoted by the "+").*

3.3.4 Thermostability of PpaRnl Mg^{2+} Binding Mutants

Temperature melts were performed to assess the thermal stability of the PpaRnl D94A and PpaRnl D248A mutants. Within these melts, folded protein samples were incubated with SYPRO Orange dye and subjected to a controlled, systematic increase in temperature. As the proteins began to unfold, hydrophobic regions became exposed, allowing the dye to bind, which resulted in an increase in fluorescence intensity. Fluorescence emissions were monitored throughout the temperature gradient, providing an outline of protein unfolding for each conditions tested. This enabled the determination of melting temperatures (T_m) for the mutant proteins.

The mutation of D94A and D248A were introduced into the WT of *P. pacificus* RNA ligase, which typically exhibits a melting temperature (T_m) of 86 °C at a pH of 6.5 (Rousseau et al.,2024). Based on these observations, it was anticipated that the generated mutants would display thermal stability within an equivalent range. For the D94A mutant however the melting curve did not align with expectations. At pH 5, the largest amount of fluorescence emission was displayed, with a midpoint of this peak at approximately 60 °C (Figure 3.23). At pH 5.5 and pH 6

slight curvature was observed near 80 °C, although the fluorescence emission suggests that this was not strongly displayed. At higher pH values such as pH 7 and pH 8, the results were inconsistent and exhibited greater variability, further complicating interpretation. Collectively, this data did not permit a reliable estimate of the T_m for this mutant, highlighting that this experiment would require repeating with possible optimisation of the protocol. The low fluorescence values are an indication of a high signal to noise ratio which could possibly be remedied with an increase in the protein concentration used, or possible adjustment of the dye concentration (Figure 3.22).

In contrast, the D248A mutant displayed melting curves that more closely resembled those of the WT protein (Figure 3.22). Across pH 5.5 to pH 8, unfolding transitions were consistently observed between 80 °C and 98 °C (Figure 3.23), with the major midpoint of the peak occurring at approximately 85 °C. The reproducibility of these curves across multiple pH conditions suggested that the D248A substitution did not significantly compromise protein stability. In comparison to the D94A mutant the thermal shift profile of D248A more closely reflects the results of the wild type, while also increasing the activity of the enzyme.

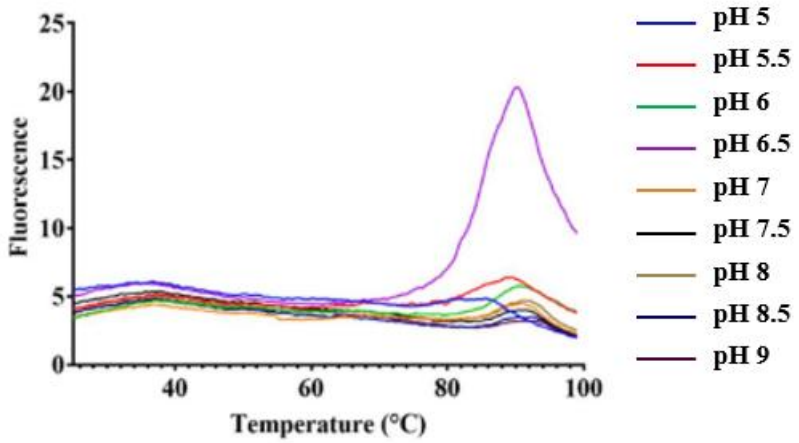

Figure 3.22 *Thermal melt curves for PpaRnl WT (Rousseau et al.,2024) SYPRO Orange fluorescence (y-axis) was measured from 25 °C-98 °C (x-axis) using buffer ranging in pH from 5-9.*

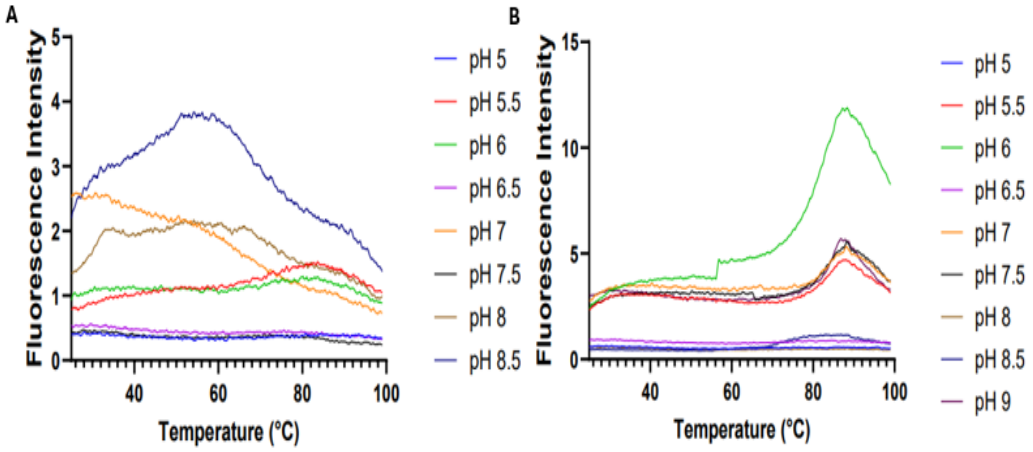

Figure 3.23 *Temperature melt curves for PpaRnl D94A (A) and PpaRnl D248A (B) mutants. SYPRO Orange fluorescence (y-axis) was measured from 25 °C-98 °C (x-axis) using buffer ranging in pH from 5-9.*

3.3.5 Structural Characterisation of D94A

To determine the structure of the D94A mutant via X-ray crystallography, a series of robot crystallisation screens and hanging-drop fine screens were conducted at varying conditions to obtain crystals suitable for X-ray diffraction at the Australian Synchrotron in Melbourne.

3.3.5.1 Robot Screens

Robot screens were set up using a Mosquito robot which dispensed protein into 96 wells varying in conditions (2.7.1) For the D94A mutant, robot screens were conducted on two separate occasions under the same set of conditions across two plates. The first trial was carried out using protein at a concentration of approximately 4 mg/mL, while the second trial used a higher protein concentration of approximately 7mg/mL. Improved crystal formation was observed in the higher concentration screens, suggesting that the protein concentration played a role in promoting nucleation and crystal growth. Despite these results, no further optimisation via hanging drop fine screens was carried out, as screens around the original crystallisation condition for PpaRnl (Rousseau et al., 2024) that were carried out simultaneously generated crystals of suitable quality for further experiments.

3.3.5.2 Hanging-drop Fine Screens

Using the same crystallisation conditions that were previously established (Rousseau et al., 2024), fine screens were set up with PpaRnl D94A at 7 mg/mL the conditions increased in pH (6.5, 7, 7.5, 8) from top to bottom, while increasing in precipitant across the screen. After five days, well-formed crystal growth was observed in eight out of 24 conditions. Most crystals exhibited a diamond morphology, with a minority formed as plates and rods. Crystals varying in morphology was collected (looped) and transferred into cryoprotectant solution composed of the original crystallisation condition, supplemented with 15% (v/v) glycerol. This additive was included to protect the crystals from damage and to minimise degradation during the flash freezing in liquid nitrogen, in preparation for transport to the Synchrotron. The rod and plate shaped crystals produced significantly better diffraction compared to the diamond shaped crystals, potentially due to twinning present within the latter.

3.3.6 Structural Characterisation of PpaRnl D248A

As with the D94A mutant hanging-drop fine screens were conducted at varying conditions to obtain crystals suitable for X-ray diffraction at the Australian Synchrotron in Melbourne. Due to the struggle of obtaining large amounts of purified protein with this mutant, robot screens were not set up with the D248A mutant. Instead, the same crystallisation conditions that were used in the hang-drop fine screens that resulted in crystallisation of PpaRnl D94A were used. These were the same fine screen conditions that were implemented in the crystallisation of PpaRnl K238G (Rousseau et al., 2024).

3.3.6.1 Hang-drop Fine Screens

In contrast to the D94A mutant, D248A only required approximately 3 mg/mL of protein to produce well-formed crystals. These crystals began to appear within three days, with continued growth observed over the course of a week. Unlike the crystals formed with D94A, which exhibited relatively uniform morphology, the crystals produced under these conditions displayed a highly diverse range of shapes and sizes, with crystals present in all 24 of the screened conditions. In addition, these crystals lacked the smooth, even surface observed previously; instead, they exhibited cracks and irregularities, which may have been a consequence of their rapid formation. Due to this, crystals were only looped from six of the 24 conditions.

There was a noticeably higher proportion of plate and rod-shaped crystals present. Based on the previous diffraction patterns obtained from different morphologies, plates and rods, were prioritised over diamond-shaped crystals. Nevertheless, a combination of morphologies was still transported to the Synchrotron for analysis, with the same differences in diffraction observed across different morphologies.

3.4 Structural Analysis of PpaRnl Mutants D94A and D248A

The crystal structures of the wild-type (WT) protein (8UCE) and the K238G mutant (8UCF and 8UCI with no AMP) had previously been solved (Rousseau et al., 2024), allowing for a reference point for comparison with the newly generated Mg^{2+} B site mutants. Structural assessment of these variants was initially carried out through the evaluation of the data collection statistics, and furthermore through the refinement statistics obtained after the final round of model building. The space group was changed in CCP4 during scaling through the AIMLESS interface using XDS output files. Symmetry mates were generated from one of the PpaRnl monomers, using that as an input for the Autobuild, to ensure the full dimer was built in the correct orientation.

For the single mutants D94A and D248A, many of the parameters fell within acceptable ranges; however, additional refinement was required to improve the overall accuracy and reliability of the models. This need for further refinement was most clearly reflected in the completeness values of the datasets. The D94A mutant exhibited statistics comparable to those of the WT structure (Table 4), indicating a reasonable degree of data quality. In contrast, the D248A mutant displayed noticeably poorer completeness, suggesting that in initial data processing in CCP4i, too much data had been removed and that further model optimisation would have been necessary to achieve a structure of comparable quality with that complete assurance of accuracy and reliability.

Unfortunately, due to the time constraints associated with the duration of this study, it was not possible to perform the additional refinement steps required to improve the final models for these mutants.

Table 4 *Statistics of the collected data and final models of PpaRnl Wild-type and K234G (Rousseau et al., 2024) and the PpaRnl mutants D94A and D248A. The highest resolution shell is shown in brackets. Data statistics were generated through phenix.tableone and AIMLESS*

Data Collection	**WT**	**K238G ATP/Mg2+**	**D94A**	**D248A**
Wavelength (Å)	0.953725	0.953732	0.953735	0.953734
Space Group	$P22_12_1$	$P22_12_1$	$P22_12_1$	$P22_12_1$
a/b/c/ (Å)	81.216/98.799/ 142.977	80.849/99.722/ 143.073	80.972/97.481/1 42.96	81.23/98.516/ 142.772
α/β/γ (º)	90/90/90	90/90/90	90/90/90	90/90/90
Resolution (Å)	2.12 (2.196-2.12)	2.14 (2.216-2.14)	2.28(2.33-2.28)	2.34(2.39-2.34)
R-merge	0.1462 (1.275	0.2477 (1.325)	0.2580(0.899)	0.982(1.033)
Completeness (%)	93.20 (84.44)	99.88 (99.95)	92.03(85.18)	88.81(73.91)
Redundancy	13.6 (13)	13.7 (14)	12.4(12.6)	11.4(11.3)
No. of observations	894 046 (85022)	884 932 (89121)	648127(56567)	561500(50257)
No. of unique reflections	65 953 (6536)	64 482 (6353)	48 184(2587)	43 578 (2388)
Mean $I/\sigma I$	14.53 (3.36)	11.12 (2.64)	5.9(2.3)	16.6(2.2)
Refinement				
R-work	0.1891 (0.2738)	0.2054 (0.2758)	0.2383(0.2868)	0.2119(0.3564)
R-free	0.2363 (0.3294)	0.2358 (0.3267)	0.2792(0.3274)	0.2629(0.4698)
No. of protein residues	761	762	762	760
Average *B* value (Å)	27.64	18.79	28.26	31.26
RMSD				
Bonds lengths (Å)	0.008	0.003	0.008	0.008
Angles (º)	0.87	0.59	0.88	0.87
Model contents				
Protomers/ASU	2	2	2	2
Ligands	21	21	14	0
Water	478	477	457	402

During the structural analysis of the Mg^{2+} B site mutants (D94A and D248A), several clear differences were observed when compared to the wild-type enzyme. While the overall domain architecture of the mutants was preserved; notable changes were detected in the coordination of the active site ligands. In both the WT and K238G mutant structures (Figure 4.2), two Mg^{2+} ions were consistently present: one bound within the active site and a second bound in a separate location.

3.4.1 PpaRnl D94A Structural Analysis

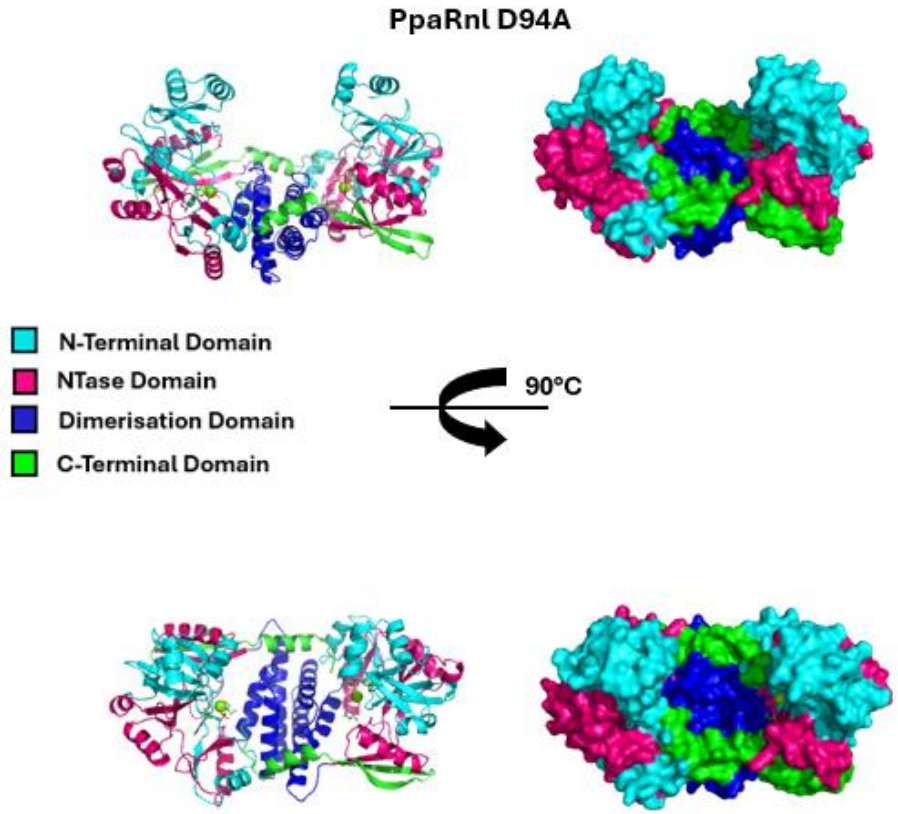

Figure 3.24 *The architecture of the domains and the arrangement of the α-helices and β-sheets that are present in the Mg^{2+} B mutant PpaRnl D94A. The domains and their corresponding colour are outlined in the legend.*

Like other thermophilic RNA ligases PpaRnl D94A forms a homodimer. Dimerisation plays a critical role in protein stability, regulation, and function of the enzyme. Each monomer within the homodimer is made up of four domains: an N-terminal (residues 1-66), an Ntase domain (66-240), the dimerisation domain (residues 240-312) and lastly the C-terminal domain (residues 312-382). The dimerisation interface of PpaRnl is centred on the AMP and Mg^{2+} ions binding

pocket, where the residues from both the monomers contribute to substrate stabilisation. The interface is stabilised by complementary hydrophobic contacts (Phe169) and electrostatic interactions, particularly seen between Lys92 and the phosphate of AMP (Figure 3.25). This cooperative interface creates a symmetrical catalytic pocket, highlighting the dimerisation that is essential for enzymatic activity. There is a central helical bundle, comprising several long α-helices which forms the dimerisation domain (Figure 3.24). This region was bordered by extended β-sheets, which bridged the two lobes of the structure. The peripheral domains contained alternating α-helices and β-strands, generating compact α/β folds. Together these elements created a stable scaffold that positioned the functional regions of the protein appropriately within the overall fold. The way that the secondary structural elements had been arranged, created well defined binding pocket for the ligands. The central α-helices had formed a scaffold that helped to position the residues within the active site, while adjacent β-sheets provided a rigid framework that helped to stabilise the ligand interactions. The amphipathic nature of several of the helices allowed hydrophobic side chains to orient toward the ligands, whereas polar charged residues located on the β-sheet surface contributed to specific hydrogen bonding and electrostatic interactions. All up the coordinated positioning of α-helices and β-sheets ensured precise ligand binding and maintained the geometry of the active site (Figure 3.25).

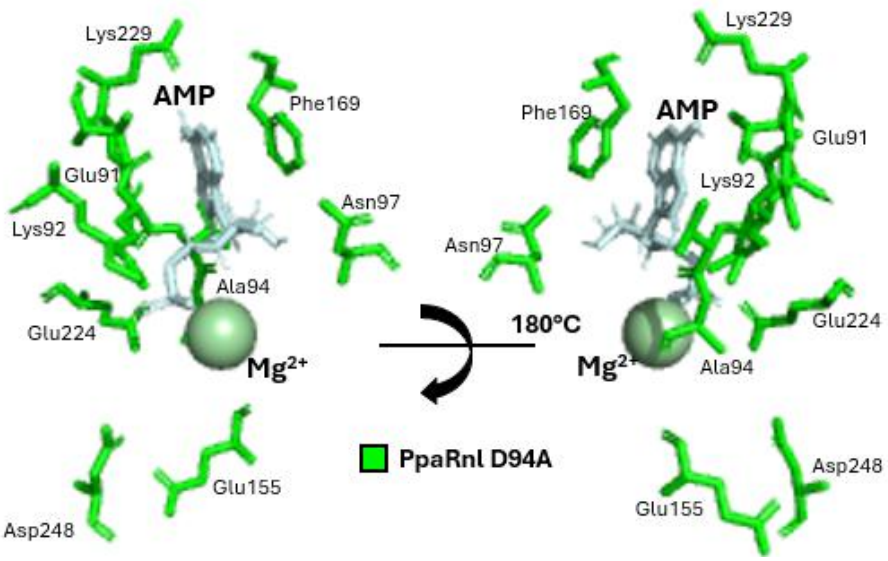

Figure 3.25 *Close up of the active site of PpaRnl D94A with the labelled amino acids. The AMP and Mg^{2+} ion are coloured in a lighter colour in comparison to the rest of structure $Mg^{2+}B$ was not present in this structure. AMP is covalently bound to Lys92 indicating the enzyme is purified in an adenylated state.*

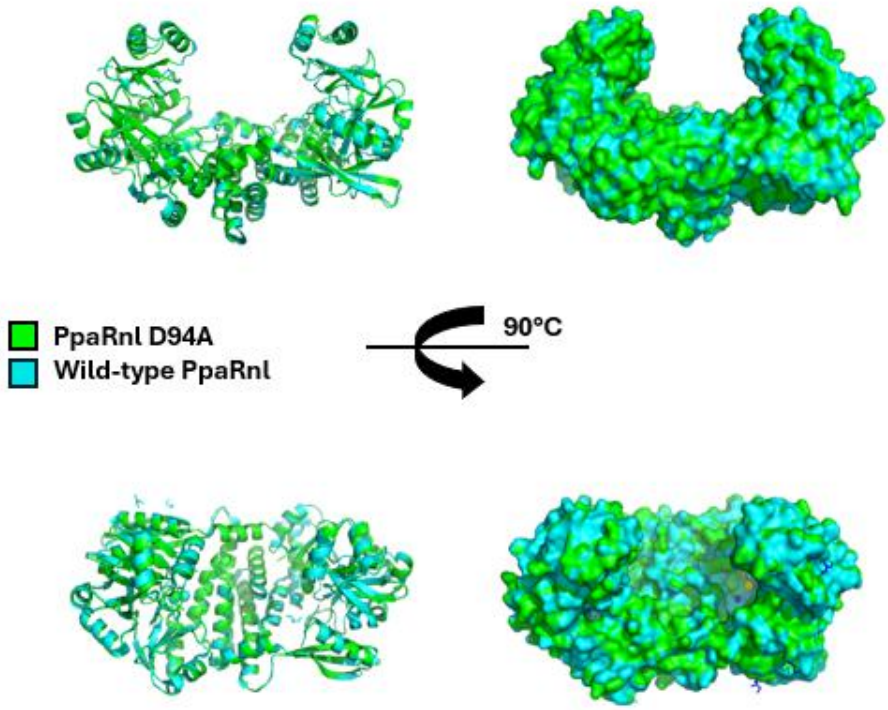

Figure 3.26 *Alignment of PpaRnl D94A mutant with the PpaRnl wild type. The ligands present within the active site are coloured yellow if present in D94A and blue if found within the wild type.*

Structural alignment of the WT protein and the D94A mutant (Figure 3.26) revealed they are both relatively similar in structure, with an RMSD value of 0.225, which was expected as only a point mutation differs between the two. In the WT protein, the aspartic acid residue at position 94 was positioned to coordinate the Mg^{2+} B ion. In contrast, substitution of this residue with alanine in the D94A mutant introduced a neutral side chain in place of the negatively charged one, resulting in a clear misalignment among the surrounding residues (Figure 3.27). The single point mutation had a direct effect on metal ion coordination. Comparatively, WT structure contained well-defined Mg^{2+} ions in the active site, the D94A mutant lacked the Mg^{2+}B ion. Instead, electron density consistent with a glycerol molecule was observed in the same position, likely introduced during crystallisation, due to its presence within the conditions. The replacement of the Mg^{2+} B with glycerol in the mutant structure suggested the absence of the negatively charged aspartic acid disrupted ion binding capacity, supporting the conclusion that Asp94 plays a critical role in stabilising Mg^{2+} coordination. Surprisingly, despite the absence of the

second Mg^{2+} and the disruption caused by this substitution, the enzymatic activity of the D94A mutant remained comparable to that of the WT, with the only activity observed being adenylation. It is also of note the change in position of Glu155 in D94A compared to that of the WT. This may be since the second Mg^{2+}B ion was not present within the mutant, but it could also be an indication that the point mutation had resulted in a conformational change within the Glu155 residue. This indicated that while the Asp94 is essential for proper Mg^{2+} B binding, loss of this ion does not significantly impair or improve catalytic function under the conditions tested, though the coordination of surrounding residues, Glu155, had been altered.

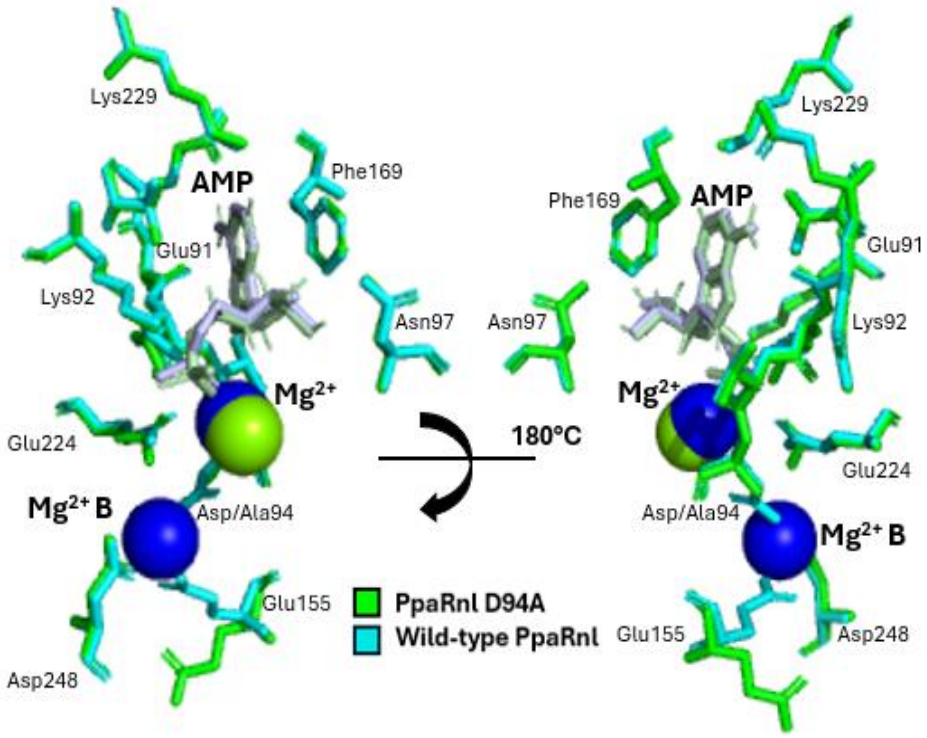

Figure 3.27 *Alignment of the active site of both PpaRnl wild-type and PpaRnl D94A. AMP of PpaRnl WT can be seen in a pale green. It is important to note the change from Asp94 to Ala94 which is represented by in the PpaRnl D94A enzyme.*

Further structural analysis was performed by aligning the D94A mutant with the K238G mutant (Figure 3.28). The comparison revealed that the residues located within the active site did not align precisely between the two structures, though the RMSD value of 0.231 still indicated they were highly similar in overall structure. Although the Mg^{2+}A ions were present within the active site of both proteins (Figure 3.28), the D94A mutant lacked the Mg^{2+}B ion. The positions of the Mg^{2+}A ions present, differed slightly, consistent with observations from the earlier comparison

between D94A and the WT enzyme (Figure 3.26). Analysis of the active site of the two structures (Figure 3.28) revealed that the coordination of the Mg^{2+}B ion was dependent on the side chain of Asp94. Substitution of this residue with alanine resulted in the loss of the carboxylate side chain, thereby eliminating a key coordinating group for the Mg^{2+}B. Examination of the surrounding active site residues further demonstrated that this substitution not only disrupted the direct coordination of Mg^{2+}A ion and a lack of the Mg^{2+}B ion in the mutant, but also altered the positioning and interactions of upstream coordinating residues, in particular Glu155, with a completely different orientation in the mutant compared to that observed in the WT. These structural changes indicated that mutation of Asp94 to alanine substantially modified the metal binding environment of the active site. Given that the point mutation of D94A resulted in no improvement of activity, like what was observed in K238G, these finding suggested that the coordination of the second Mg^{2+}, Mg^{2+} B, contributed to the catalytic efficiency of the enzyme and that disruption of this interaction impaired activity.

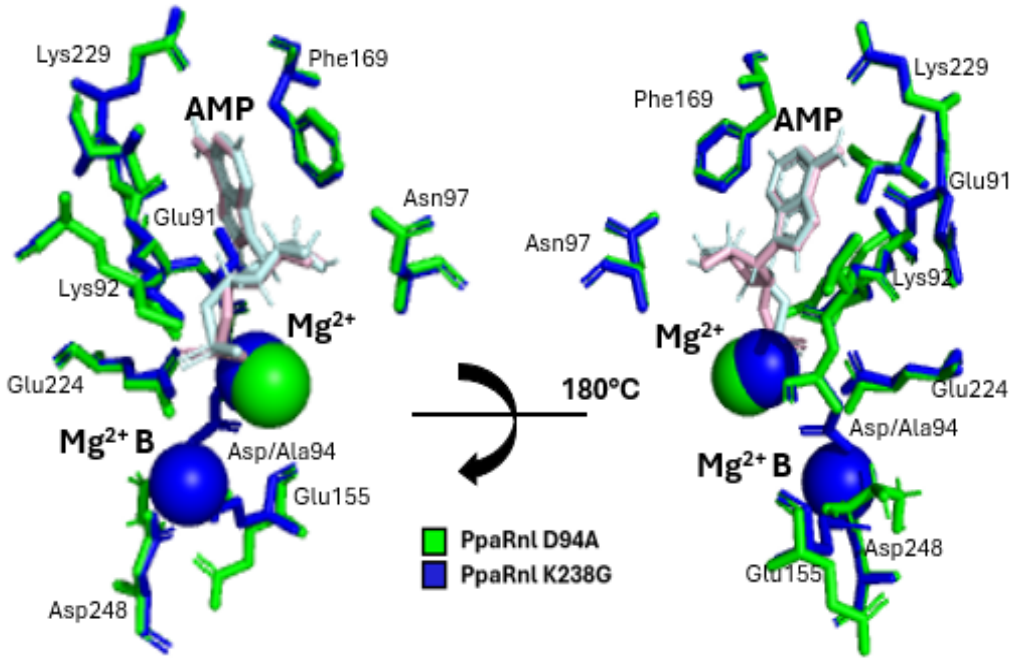

Figure 3.28 *Alignment of the active site of both PpaRnl K238G and PpaRnl D94A. AMP of PpaRnl can be seen in a light pink, compared to the pale green AMP of PpaRnl D94A.It is important to note the change from Asp94 to Ala94 which is represented by in the PpaRnl D94A enzyme.*

3.4.2 PpaRnl D248A Structural Analysis

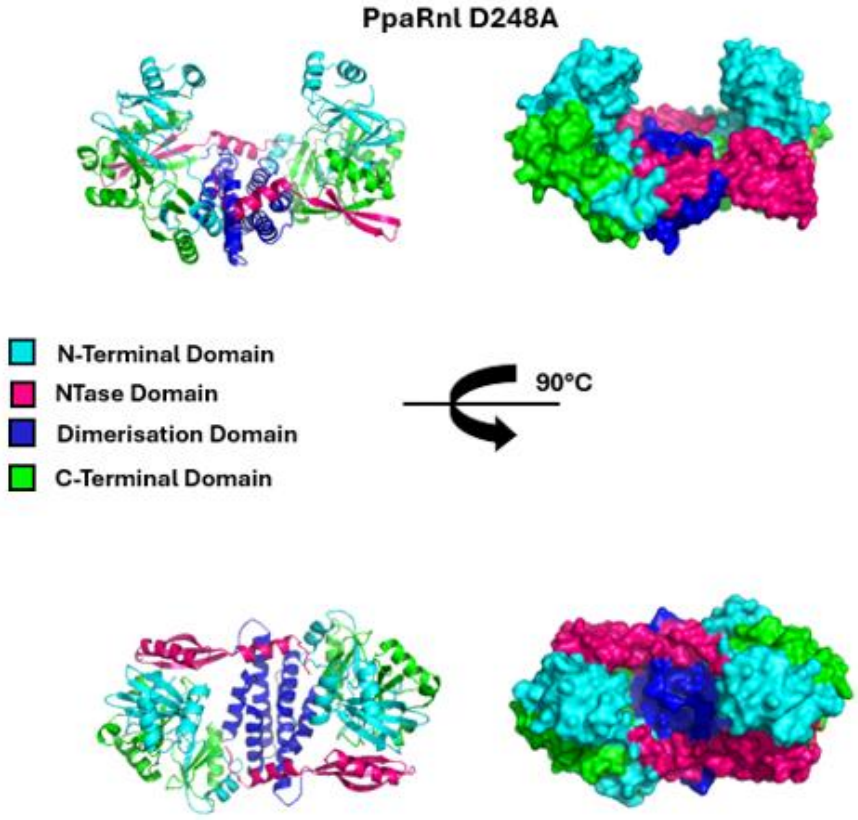

Figure 3.29 *The architecture of the domains and the arrangement of the α-helices and β-sheets that are present in the Mg^{2+} B mutant PpaRnl D248A. The domains and their corresponding colour are outlined in the legend.*

The overall secondary structure of the mutant D248A mutant is made up of the same four domains as seen in D94A (Figure 3.29); an N-terminal (residues 1-66), an Ntase domain (66-240), the dimerisation domain (residues 240-312) and lastly the C-terminal domain (residues 312-382). It is also a homodimer, which was expected across the mutants as the only change was a single point mutation. D248A consisted of alternating α-helices and β-sheets. Like seen for PpaRnl D94A (Figure 3.24) the central region of D248A was also dominated by a cluster of long α-helices, which packed together to form the stable dimerisation domain. This domain was also bordered by β-strands which were arranged into extended β-sheets, bridging the two lobes of the protein together into its dimeric form. The outer regions displayed a mixture of shorter helices and β-sheets, contributing to the compact α/β fold that stabilised the global architecture. The arrangement of these secondary structural elements created a distinct place for ligand accommodation. The helical bundle provided a rigid scaffold that orientated key residues toward the active and

secondary binding site. Adjacent β-sheets had contributed structural rigidity and helped define the outline of these binding sites. Compared with the wild-type, K238G, and D94A mutants, the absence of both the Mg^{2+} ion in the D248A structure appeared to show the helices and β-sheets surrounding the site maintained the overall shape of the pocket but lacked the precise metal-mediated coordination observed previously.

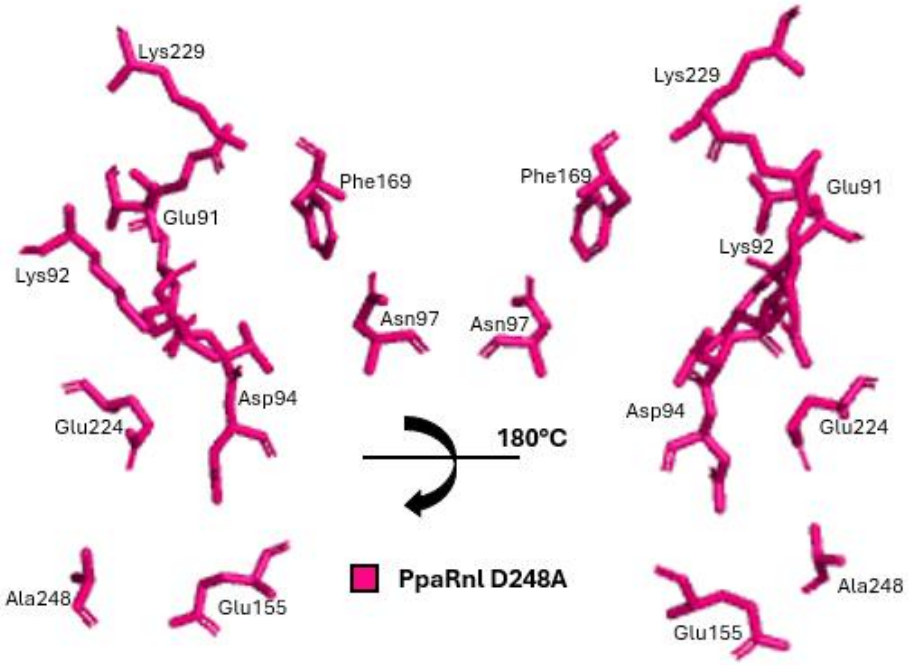

Figure 3.30 *A close view of the active site of PpaRnl D248A with any visible amino acids labelled. No Mg^{2+} ions were present in this structure with no additional ligands present also.*

As with the D94A mutant, structural comparisons of the D248A mutant were carried out against both the wild-type and the K238G mutant to assess the effect of the substitution on the architecture of the protein. The initial alignment was performed between D248A and the WT protein (Figure 3.31). When the structure of the D248A mutant was solved, it was found that there was no AMP molecule bound within the active site; it was determined through the RMSD value (0.216) that the structures were still highly similar. The absence of AMP likely prevented a conformational change that is normally required for proper binding of the Mg^{2+} ions. To address this restraint, re-crystallisation in the presence of ATP, or alternatively an ATP soak of pre-formed crystals would have been necessary to ensure the presence of AMP in the active site. Although within the D248A structure there appeared to be space for the AMP and associated residues appeared to be present,

electron density in these regions was ambiguous, repeated rounds of refinement in *phenix.refine* placed water molecules into this density rather than nucleotide ligands, despite multiple attempts, the size of the density and the coordination of the side chain present within those regions would not accommodate the expected ligands.

Like with D94A, the structural difference between the WT protein and D248A is a single point mutation of Asp248 to an alanine, again making the change from a negatively charged residue to a neutral one. D248A may have exhibited a disruption like that observed in the WT versus D94A alignment (Figure 3.26), where subtle but notable differences in the positioning of coordinating residue were detected. The change from the negatively charged aspartic acid 248 was done as it has been observed (Figure 3.32) to play a dominant role in the coordination of the Mg^{2+}B ion. With the change to an alanine, it was expected that this would affect the coordination as there would no longer be that pull of the negative residue to attract the positive charge of the Mg^{2+}. It is important to note that this point mutation consistently resulted in an improvement in enzymatic activity when compared with the activity observed for the WT protein. This enhancement was noted within completed activity assays (Figure 3.16), indicating that the substitution granted a measurable functional advantage relative to the native enzyme.

Further structural analysis was conducted with the alignment of the WT protein and the D248A mutant (Figure 3.31). As previously mentioned it was seen that the structures were similar but with the lack of Mg^{2+} ions or AMP within the D248A mutant, it was hard to make an accurate comparison; a comparison of the two models' active sites (Figure 3.32) revealed differences in the residues surrounding the Mg^{2+} ions present in the WT structure. Alterations were observed in the coordination geometry of several side chains. Most notable, the side chain of Glu224 in the D248A mutant which adopted a conformation that appeared suboptimal for the coordination of the Mg^{2+} B ion, though this may have simply been due to its absence within the D248A structure. Although the D248A mutation was located directly within the Mg^{2+}B binding region, additional differences were identified in residues distanced from the substitution. Specifically, Asn97 displayed positional shifts when compared to the WT structure, though this amino acid is not directly interacting with either the AMP or Mg^{2+} ions, its proximity to the AMP binding site, suggests that it may contribute to the correct orientation of the AMP and its offset could directly correlate with the lack of AMP present in the active site

of D248A. The findings suggest that the mutation could have potentially introduced broader conformational changes extending beyond the immediate active site.

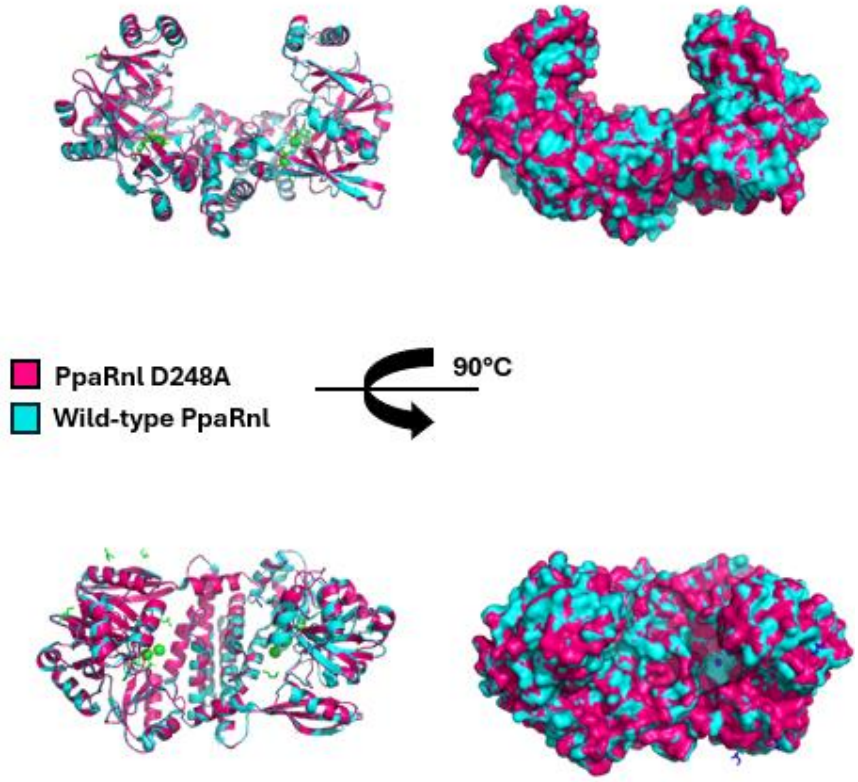

Figure 3.31 *Alignment of PpaRnl D248A mutant with the PpaRnl wild-type. The ligands present within the active site are coloured green and blue and were only present with the WT protein.*

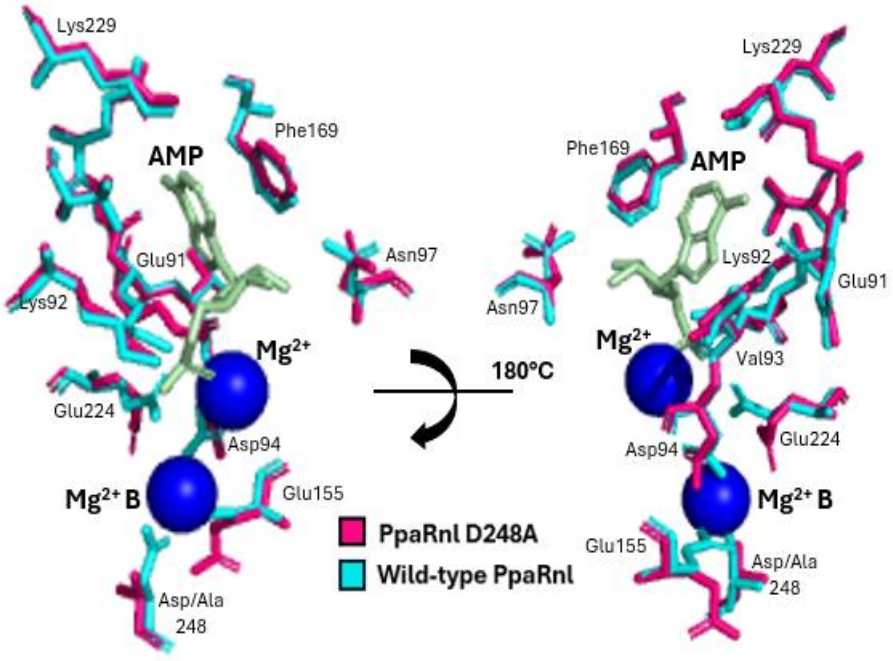

Figure 3.32 *Alignment of the active site of both PpaRnl wild-type and PpaRnl D248A. AMP of PpaRnl WT can be seen in a pale green. It is important to note the change from Asp248 to Ala248 which is represented by in the PpaRnl D248A enzyme.*

The functional and structural analysis of D248A versus K238G was successively carried out (Figure 3.33). The enzymatic activity of D248A more closely resembled that of K238G than that of the WT enzyme with the RMSD of 0.256 indicating that this was the case. With the mutation producing comparable improvement in activity to that observed in K238G (Figure 3.16). However, as with the comparison of D248A and the WT, due to the additional ligands not being present within the structure, making a direct and accurate comparison of D248A versus K238G was not possible. Like the comparison between the WT enzyme and the D248A mutant, the structural model of D248A suggested that presence of sufficient space within the structure to accommodate the additional ligands.

Analysis of the active site alignment of the mutants D248A and K238G (Figure 3.33) highlighted the key differences in the coordination of the residues surrounding the two Mg^{2+} ions. The primary issue in the D248A mutant appeared to be the altered positioning and coordination of the side chains in this region. Although the substitution of Asp248 with alanine removed one of the coordinating residues, the side chain of Glu155 remained positioned in such a way that it may have contributed to partial coordination of the Mg^{2+} B ion. However, the absence of AMP in the

active site of the PpaRnl D248A structure, as noted previously, likely hindered the conformational change required to position the surrounding side chains appropriately for ion binding. It is theorised that the enzyme works through a process of abortive ligation, which is thought to be why only adenylation is observed within the WT (Rousseau et al., 2024). With the improvement in activity, it is suggesting that the mutation had slowed down the enzyme's reaction time, allowing it to complete all three ligation steps. It is interesting to note that only within enzymes where the activity was improved, there was a lack of AMP in the active site, and additional soaks with ATP were required. Without this conformational change, the residues were unable to adopt orientations necessary for the effective coordination, ultimately preventing Mg^{2+} ions from stably occupying the active site.

In contrast, the K238G mutant maintained a more favourable configuration of active site residues; though the RMSD value had slightly increased compared to what was seen in the WT (0.216). Despite the lysine to glycine substitution, the overall geometry of the coordinating network around the Mg^{2+} ions remained intact, and AMP was present within the active site, The presence of AMP likely stabilised the local confirmation and facilitated the arrangement of surrounding residues, enabling effective Mg^{2+} coordination. This comparison emphasised that the inability of D248A to bind Mg^{2+} ions was not solely due to the loss of Asp248, but also to the absence of AMP and the subsequent failure to induce the conformational state required for proper metal ion coordination. Nevertheless, further investigation would be required to determine whether these ligands could indeed be accommodated in this space in a similar fashion to what was observed in the K238G mutant, or whether structural discrepancies would emerge, as was observed in the comparison between K238G and D94A (Figure 3.28).

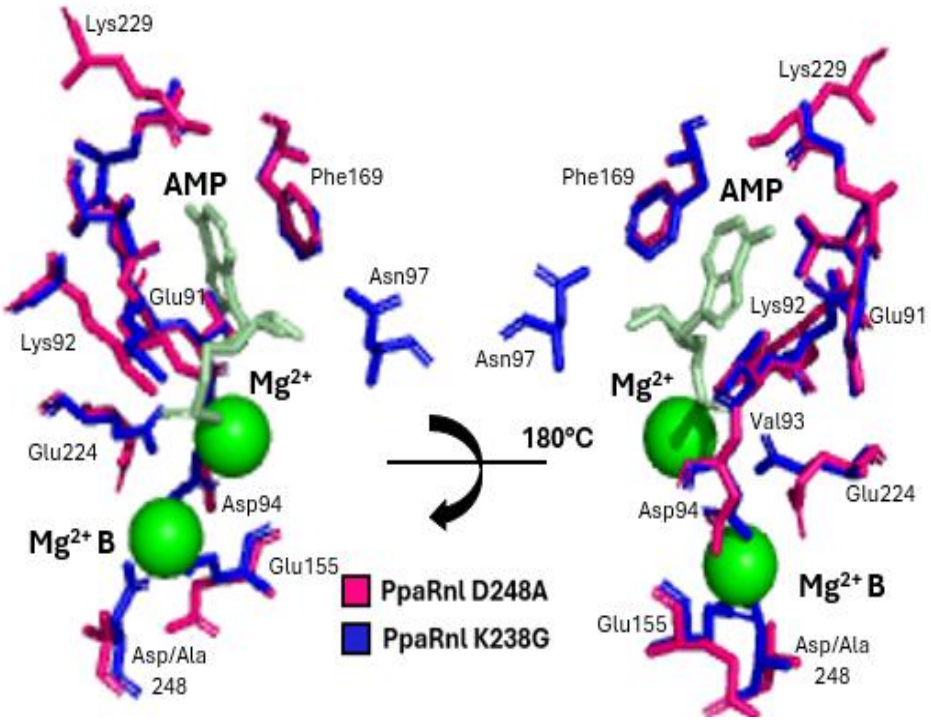

Figure 3.33 *Alignment of the active site of both PpaRnl K238G and PpaRnl D248A. AMP of PpaRnl can be seen in a pale green. It is important to note the change from Asp248 to Ala248 which is represented by in the PpaRnl D248A enzyme.*

Structures of K238G with no AMP present in the active site were compared with D248A, to compare the conformation of active site residues in the absence of AMP (Figure 3.34). It was found that the lack of AMP in both mutants didn't account for the changes in the side chain of the amino acids orienting the Mg^{2+} ions in PpaRnl D248A, although the differences were not as dramatic as previously observed which was reflected in the slight decrease in the RMSD value from 0.256 to 0.247. It was clear that the orientation of Asn97 was again twisted in the D248A structure, with the lack of AMP present in both structures it was expected that this discrepancy would not be present. The other notable change was seen within Glu155 and Glu224; in the K238G structure it could be observed that the side chains are orientated in a way the ideally coordinated the Mg^{2+}A ion (Glu224) and the Mg^{2+}B ion (Glu155), though were are no Mg^{2+} ions present in the D248A mutant so this may be the cause, as there was no positive charge of the Mg^{2+} ions to help coordinate the negative charge of the glutamic acid side chains (Figure 3.34).

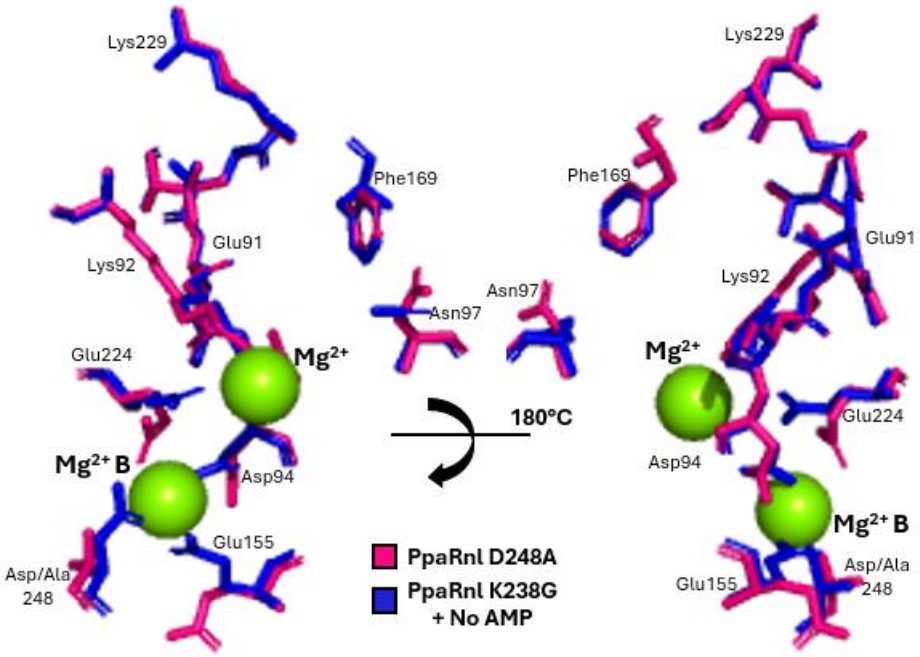

Figure 3.34 *Alignment of the active site of both PpaRnl K238G with no AMP present and PpaRnl D248A. It is important to note the change from Asp248 to Ala248 which is represented by in the PpaRnl D248A enzyme.*

A final structural comparison was conducted between the single mutants D94A and D248A. Based on their positions within the protein, it was anticipated that these substitutions would not introduce substantial alterations to the overall structural architecture. This expectation was confirmed through structural alignment, which revealed the mutants exhibited highly similar conformations with no major deviations observed within the secondary of tertiary structure (Figure 3.35), apart from the missing residues within the active site of D248A. These findings indicated that the integrity of the two mutants was preserved. The differences in the two structures was better observed in the comparison of the active sites of the two mutants (Figure 3.36); The most notable being that there was nothing present within the active site of D248A, and AMP and Mg^{2+} A, present in the D94A mutant, this difference was most clear in Glu155 and Glu 224 of D248A with the orientation of the residues side changes angled in differently compared to what was observed in the D94A mutant. This could again be due to the lack of the Mg^{2+} ions in D248A. The same twist in Asn97 was also seen in contrast to D94A, as seen in the WT (Figure 3.32), K238G (Figure 3.33) and the K238G mutant with no AMP (Figure 3.34).

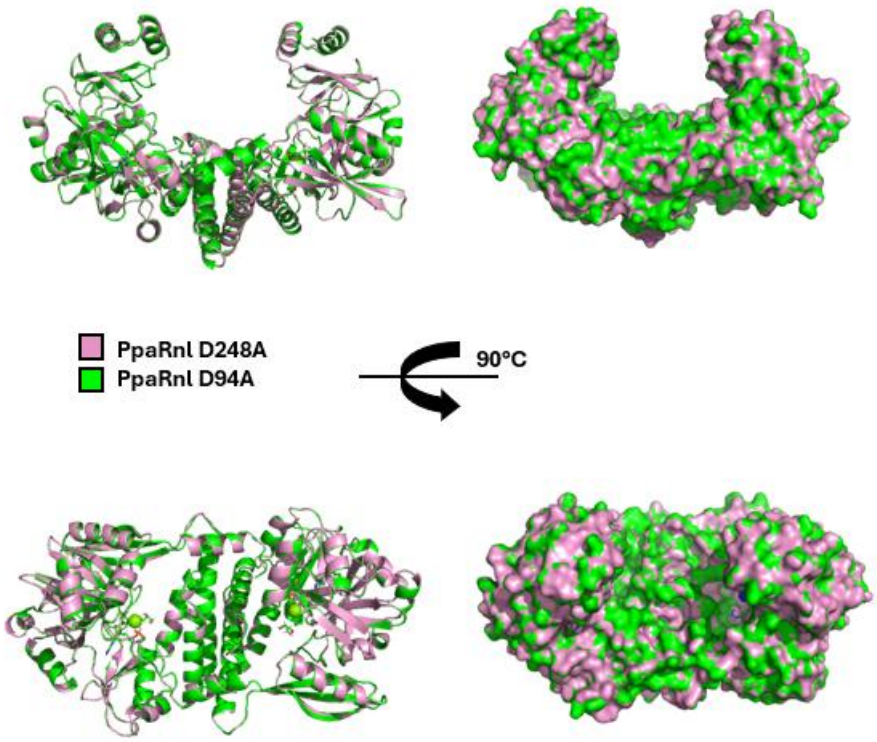

Figure 3.35 *Alignment of PpaRnl D248A mutant with the PpaRnl D94A. The residues present within the active site are coloured in contrast to the surrounding protein and found within the D94A mutant.*

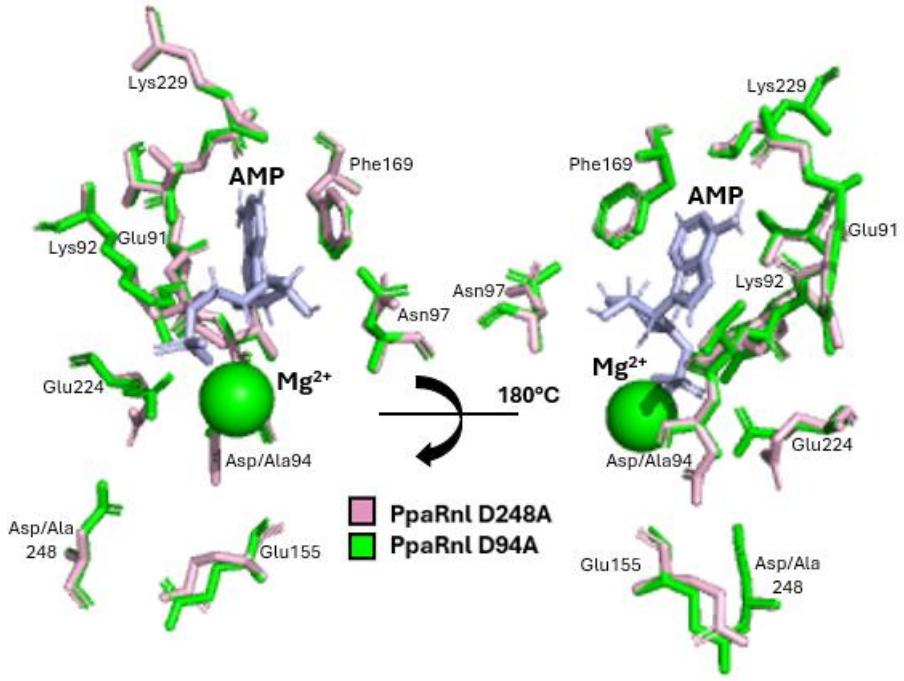

Figure 3.36 *Alignment of the active site of both PpaRnl D94A and PpaRnl D248A. AMP of PpaRnl D94A can be seen in a pale green. It is important to note the change from Asp248 to Ala248 which is represented by in the PpaRnl D248A enzyme and the change of Asp94 to Ala94 represented by PpaRnl D94A.*

Chapter Four: Conclusion and Future Directions

The biological world operates according to the central dogma, in which genetic information flows from DNA to RNA and ultimately to protein, through transcription and translation (Koonin, 2015). Over the last decade it has become clear that non-coding RNAs such as long non-coding RNAs (lncRNAs), small interfering RNAs (siRNAs), and micro RNAs (miRNAs) play essential roles in the regulation of gene expression, with a complex layer of post-translational and epigenetic control (Ratti et al., 2020), that is yet to be fully uncovered.

With the development of next generation sequencing (NGS) detection and understanding of the role of miRNAs is becoming clearer. In contrast to the traditional methods such as hybridisation-based miRNA microarrays, and RT-qPCR, NGS opens the door to understanding miRNA biology with its high sensitivity and the identification of novel miRNAs, though it does come with its disadvantages. During the small RNA library preparation, (Hafner et al., 2011). Ligation efficiency and accuracy can be compromised by unintended side reactions and enzyme-specific biases. AMP transfer reactions may remove the 5'-adenylation from adaptors, and transfer it to RNA substrates, promoting self-ligation, circularisation, or concentration of RNA molecules (Zhuang et al., 2012). Secondary structures such as hairpins or base paired termini, can hinder ligase accessibility, reducing ligation efficiency (Jayaprakash et al., 2011). In addition, RNA ligases exhibit sequence and structure dependent substrate preferences, introducing systematic biases in miRNA quantification, as observed for T4 RNA ligase 1 (T4Rnl1) and T4 RNA ligase 2 (T4Rnl2) which show preferences based on the nucleotide composition, secondary structure, and terminal chemistry of their RNA substrates (Hafner et al., 2011; Sorefan et al., 2012).

With the recent advancements in small RNA sequencing technologies, understanding the structural and mechanistic characteristics of the RNA ligase enzyme is fundamental for the desired accuracy of miRNA sequencing. With a diverse pool of archaeal RNA ligases, the structural and functional knowledge of these is ever growing, resulting in the potential for the exploration of different enzymes outside of the T4 RNA ligases with unique mechanistic functions that are

ideal for small RNA sequencing protocols. With recent studies looking into the thermophilic RNA ligases such as *Methanobacterium thermoautotrophicum* (MthRnl) where the overall, structural, kinetic, and mutagenesis data, show that MthRnl's dimeric architecture, surface-exposed residues, and active site configuration, act in harmony to direct substrate recognition, binding, and the multi-step catalysis required for RNA circularisation (Gu et al., 2016). *Pyrococcus abysii* (PabRnl) RNA ligase has strong thermal stability, multi-turnover efficiency, structural adaptability, and a broad substrate range, making it a promising model for ancestral ligases and a valuable tool for biotechnological applications requiring thermostable nucleic acid circularisation (Brooks et al., 2008). The focus of this thesis was the RNA ligase from the thermophilic archaeon *P. pacificus*. This enzyme was further characterised through the introduction of new substrates, and point mutations, to explore the role of a second Mg^{2+} ion binding site identified in previous research.

The PpaRnl WT enzyme displays only adenylation of the RNA substrate in a process that is hypothesised as abortive ligation where step 1 and step 2 of the ligation reaction occur quickly leading to the adenylated RNA leaving the active site before step 3 forms a phosphodiester bond, either between the ends of the substrate (intramolecular ligation/circularisation), or two substrate molecules (intermolecular ligation/end to end joining). Mutation of Lys238 in the active site to glycine improved activity with both intra- and intermolecular ligation of RNA substrates observed. Lys238 plays a key role in coordination of the AMP in the active site which in turn helps to coordinate the Mg^{2+}A ion. Previous research established that the PpaRnl K238G enzyme will display a substrate presence with no solid determination on the origination of this preference; was it sequence, or secondary structure? To answer this question a series of gel-based activity assays with PpaRnl K238G were conducted. A panel of previously tested oligonucleotides was expanded with newly designed variants to probe the influence of sequence and structural features based on the established preference of PpaRnl K238G for oligonucleotide 65 (Rousseau et al., 2024). Overall, the data confirmed that PpaRnl K238G has a distinct substrate preference, with both sequence and structural elements influencing activity. The assays completed with the new panel of oligonucleotide revealed that the structure plays a crucial role with reduced activity observed in oligos with blunt ends though the sequence was the same. This suggests

ligation is improved when the substrates have either a 5' or 3' overhang i.e. there is a specific structural feature underlying this bias. However, although closer to an answer, a firm conclusion cannot be made without further investigation. Additionally, temperature-dependent differences highlighted the importance of RNA folding in modulating ligation efficiency.

To investigate the role of the second magnesium binding site in the activity of PpaRnl, activity assays were performed with the PpaRnl mutants D94A and D248A. Since the presence of two Mg^{2+} ions within the enzyme active site is unique to PpaRnl, the investigation of point mutations of the amino acids that coordinate these ions, allows for a better understanding of enzyme mechanism, taking one step closer to being able to utilise these thermophilic ligases in NGS technologies. This research initially also aimed to characterise the E155A mutant and the double mutant D94A-D248A; despite repeated attempts, these proteins could not successfully be expressed in high enough yields for characterisation. Further investigation is warranted to explore whether these mutations rendered the protein non-viable or whether disruption of the Mg^{2+}B binding site was detrimental to protein purification. Such findings could help reveal unique mechanistic features of PpaRnl with potential significance for small RNA sequencing applications like the identification of biomarkers found in cancer and viral diseases (Cui et al., 2025; Girardi et al., 2018; Zhang et al., 2025).

In the case of the D94A mutant, activity assays across multiple oligonucleotides and temperatures showed only adenylation of the RNA substrates, indicating that this mutation neither enhanced nor impaired the enzymatic activity compared to the WT enzyme. The wild-type enzyme also displayed only adenylation of the RNA substrate. Structural characterisation of this mutant provided further insight: disruption of the Mg^{2+} B site altered the coordination of the surrounding side chains. Although AMP and the other Mg^{2+} ion were present, their alignment compared to both the WT and K238G structures was slightly offset due to the change in orientation of the side chains and the Mg^{2+} B's absence. These findings highlight the importance of residue Asp94 in maintaining proper coordination of Mg^{2+} B within the active site.

The D248A mutation, by contrast, produced an unexpected improvement in activity, like that observed for the K238G mutant with both intra- and intermolecular ligation of RNA substrates observed. This variant also exhibited comparable substrate bias,

confirming that both the sequence and structural features contribute to substrate preference. However, its activity was slightly reduced at the higher temperature of 75 °C. Further assays with the original panel of oligonucleotides demonstrated that, consistent with previous results with PpaRnl K238G (Rousseau et al., 2024), oligo 65 remained the preferred substrate. From a structural perspective the D248A mutant lacked both AMP and Mg^{2+} ions in the active site. This absence strongly suggests that the residues within this region are key determinants of cofactor coordination and may drive conformational changes essential for enzymatic function. Both mutants diffracted to a high resolution, enabling comparison with the previously solved wild-type and K238G structures. Although the preliminary models fell within acceptable statistical ranges, further refinement will be required for definitive interpretation.

Taken together, these results highlight how point mutations within the active site can profoundly influence enzyme mechanism. Future work is required to determine whether the discrepancies observed were due directly to the mutations themselves or to secondary effects such as impaired cofactor binding. Seeing improvement in activity in the D248A enzyme despite any Mg^{2+} ions present poses more questions than answers about the role of the Mg^{2+} B in this enzyme, although other ligases utilise a Mg^{2+} ion there is only one present in the active site. Therefore, the presence and position of the two Mg^{2+} ions in PpaRnl is unique and further investigation into the "why" is required. This could be achieved through the soaking of the crystalised protein with ATP to ensure that AMP would be present within the protein. It would also be valuable to be able to analyse the structure of these proteins with RNA bound within the active site; being able to observe the interactions of a protein in real time is not feasible, but seeing the way the RNA binds and the conformational changes that occur within the protein would be invaluable. Ultimately, gaining a clearer understanding of these structural dynamics could provide valuable insight into substrate specificity, the functional contribution of the Mg^{2+} B ion, and the potential to exploit these mechanisms in reducing enzymatic bias during small RNA sequencing.

As substrate preference is central to the utility of RNA ligases in small RNA sequencing, a broader understanding of this bias is critical. Future studies should expand the diversity of tested oligonucleotides, as the limited panel used here does not fully represent the range of naturally occurring species. Incorporating

oligonucleotides from additional organisms, such as human derived miRNAs, could provide a more representative dataset to be utilised in a sequencing library, as a library is only as good as the data set it is exposed to. Thermophilic RNA ligases such as PpaRnl, and its mutants, show considerable promise for reducing bias in miRNA sequencing, and this thesis provides an early foundation for uncovering their potential. Although the preliminary models of both D94A and D248A fell within acceptable statistical ranges, further refinement will be required for definitive interpretation. The combination of biochemical and structural results presented here provides important advances in clarifying the role of residue coordination within PpaRnl.

This research has advanced our understanding of thermophilic RNA ligases by combining mechanistic, biochemical, and structural analyses, to elucidate both substrate preferences and active-site dynamics. In doing so it has identified new potential candidates for use in small RNA library preparation, where reducing bias remains a significant challenge in sequencing technologies. Beyond substrate specificity, this work has also uncovered previously uncharacterised structural mechanisms that govern cofactor coordination and enzymatic conformational change.

By integrating activity assays with structural characterisation of key mutants, this thesis has provided framework for understanding how subtle differences in sequence or active-site residues can profoundly influence enzyme behaviour. These findings not only expand the fundamental knowledge of RNA ligase function but also establish a foundation for rational engineering of enzymes with tailored properties. In particular, the structural insights gained here highlight avenues for modifying cofactor interactions, or substrate preferences, with potential applications in the design of next-generation sequencing tools.

Ultimately, this research has laid important groundwork for the development of RNA ligases as powerful biotechnological tools. Future studies that build on these discoveries may enable the creation of engineered ligases with reduced substrate bias, improved efficiency, and broader applicability, thereby driving innovation in sequencing-based technologies and expanding the toolkit available for RNA biology.

References

Becker, H. F., L'Hermitte-Stead, C., & Myllykallio, H. (2019). Diversity of circular RNAs and RNA ligases in archaeal cells. Biochimie, 164, 37–44. https://doi.org/10.1016/j.biochi.2019.06.011

Benesova, S., Kubista, M., & Valihrach, L. (2021). Small RNA-Sequencing: Approaches and Considerations for miRNA Analysis. Diagnostics (Basel, Switzerland), 11(6), 964. https://doi.org/10.3390/diagnostics11060964

Brooks, M. A., Meslet-Cladiére, L., Graille, M., Kuhn, J., Blondeau, K., Myllykallio, H., & van Tilbeurgh, H. (2008). The structure of an archaeal homodimeric ligase which has RNA circularization activity. Protein Science, 17, 1336–1345. https://doi.org/10.1110/ps.035493.108

Cui, Q., Guo, H., Cai, Y., Fei, Y., & Wang, S. (2025). Interval-shared information integration and false-negative association reduction in Multi-source miRNA-disease association prediction. IEEE Journal of Biomedical and Health Informatics, 1–13. https://doi.org/10.1109/jbhi.2025.3562617

Girardi, E., López, P., & Pfeffer, S. (2018). On the importance of host microRNAs during viral infection. Frontiers in Genetics, 9, 439. https://doi.org/10.3389/fgene.2018.00439

Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., ... & Caldas, C. (2010). Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA, 16(5), 991–1006. https://doi.org/10.1261/rna.1947110

Gu, H., Yoshinari, S., Ghosh, R., Ignatochkina, A. V., Gollnick, P. D., Murakami, K. S., & Ho, C. K. (2016). Structural and mutational analysis of archaeal ATP-dependent RNA ligase identifies amino acids required for RNA binding and catalysis. Nucleic Acids Research, 44(5), 2337–2347. https://doi.org/10.1093/nar/gkw094

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J. T. G., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., & Tuschl, T. (2011). RNA-ligase-dependent biases in miRNA

representation in deep-sequenced small RNA cDNA libraries. RNA, 17(9), 1697–1712. https://doi.org/10.1261/rna.2799511

Hammond, S. M., Lee, R. C., Wightman, B., Zamore, P. D., Grishok, A., Bartel, D. P., Ventura, A., Woods, K., Marson, A., Chang, T. C., Raver-Shapira, N., Gregory, R. I., Vickers, K. C., Tabara, H., Matranga, C., Shin, C., Liu, Y., Tomari, Y., Meister, G., Nykanen, A., … O'Donnell, K. A. (2015). An overview of microRNAs. Advanced Drug Delivery Reviews, 87, 3–14. https://doi.org/10.1016/j.addr.2015.05.001

Ho, C. K., Wang, L. K., Lima, C. D., & Shuman, S. (2004). Structure and mechanism of RNA ligase. Structure, 12(2), 327–339. https://doi.org/10.1016/j.str.2004.01.011

Jayaprakash, A. D., Jabado, O., Brown, B. D., & Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Research, 39(21), e141. https://doi.org/10.1093/nar/gkr693

Kabsch, W. (2010). XDS. Acta Crystallographica Section D: Biological Crystallography, 66(2), 125–132. https://doi.org/10.1107/S0907444909047337

Koonin, E. V. (2015). Why the central dogma: On the nature of the Great Biological Exclusion Principle. Biology Direct, 10, 52. https://doi.org/10.1186/s13062-015-0084-3

Lohman, G. J. S., Zhang, Y., Zhelkovsky, A. M., Cantor, E. J., & Evans, T. C. (2014). Efficient DNA and RNA ligation with thermostable 5′ App DNA/RNA ligase. PLOS ONE, 9(7), e101847. https://doi.org/10.1371/journal.pone.0101847

Matthews, B. W. (1968). Solvent content of protein crystals. Journal of Molecular Biology, 33(2), 491–497. https://doi.org/10.1016/0022-2836(68)90205-2

McGeary, S. E., Lin, K. S., Shi, C. Y., Pham, T. M., Bisaria, N., Kelley, G. M., & Bartel, D. P. (2019). The biochemical basis of microRNA targeting efficacy. Science, 366(6472), eaav1741. https://doi.org/10.1126/science.aav1741

Motameny, S., Wolters, S., Nürnberg, P., & Schumacher, B. (2010). Next generation sequencing of miRNAs – strategies, resources and methods. Genes, 1(1), 70–84. https://doi.org/10.3390/genes1010070

Nandakumar, J., Shuman, S., & Lima, C. D. (2006). RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. Cell, 127(1), 71–84. https://doi.org/10.1016/j.cell.2006.08.039

Ratti, M., Lampis, A., Ghidini, M., Salati, M., Mirchev, M. B., Valeri, N., & Hahne, J. C. (2020). MicroRNAs (miRNAs) and long non-coding RNAs (lncRNAs) as new tools for cancer therapy: First steps from bench to bedside. Targeted Oncology, 15(3), 261–278. https://doi.org/10.1007/s11523-020-00717-x

Raymond, A., & Shuman, S. (2007). Nick sealing and ligase adenylylation by RNA ligase. Nucleic Acids Research, 35(3), 839–849. https://doi.org/10.1093/nar/gkl1075

Rousseau, M. D., Hicks, J. L., Oulavallickal, T., Williamson, A., Arcus, V. L., & Patrick, W. M. (2024). 8UCE: Thermophilic RNA ligase from Palaeococcus pacificus + AMP. RCSB Protein Data Bank. https://www.rcsb.org/structure/8UCE

Rousseau, M., Oulavallickal, T., Williamson, A., Arcus, V., Patrick, W. M., & Hicks, J. (2024). Characterisation and engineering of a thermophilic RNA ligase from Palaeococcus pacificus. Nucleic Acids Research, 52(7), 3924–3937. https://doi.org/10.1093/nar/gkae149

Shuman, S., & Lima, C. D. (2004). The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyl transferases. Current Opinion in Structural Biology, 14(6), 757–764. https://doi.org/10.1016/j.sbi.2004.10.005

Sorefan, K., Pais, H., Hall, A. E., Kozomara, A., Griffiths-Jones, S., Moulton, V., & Dalmay, T. (2012). Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence, 3(1), 4. https://doi.org/10.1186/1758-907X-3-4

Uciuleac, M.-C., Goldgur, Y., & Shuman, S. (2015). Structure and two-metal mechanism of a eukaryal nick-sealing RNA ligase. Proceedings of the

National Academy of Sciences, 112(45), 13574–13579. https://doi.org/10.1073/pnas.1516536112

Wang, N., Zheng, J., Chen, Z., Liu, Y., Dura, B., Kwak, M., Xavier-Ferrucio, J., Lu, Y.-C., Zhang, M., Roden, C., Cheng, J., Krause, D. S., Ding, Y., Fan, R., & Lu, J. (2019). Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. Nature Communications, 10, 95. https://doi.org/10.1038/s41467-018-07981-6

Yang, Z., Zhang, C., Lian, G., Dong, S., Song, M., Shao, H., Wang, J., Zhong, T., Luo, Z., Jin, S., Ding, C., & others. (2022). Direct adenylation from 5′-OH-terminated oligonucleotides by a fusion enzyme containing Pfu RNA ligase and T4 polynucleotide kinase. Nucleic Acids Research, 50(13), 7560–7569. https://doi.org/10.1093/nar/gkac604

Yuan, Y., Stumpf, F. M., Schlor, L. A., Schmidt, O. P., Saumer, P., Huber, L. B., Frese, M., Höllmüller, E., Scheffner, M., Stengel, F., Diederichs, K., & Marx, A. (2023). Chemoproteomic discovery of a human RNA ligase. Nature Communications, 14, 943. https://doi.org/10.1038/s41467-023-36451-x

Zhang, B., & Matsumoto, Y. (2025). Integrated network analysis decipher ZNF384-related miR-20b-5p and miR-424-5p in colon adenocarcinoma. Cancer Reports, 8(5), e70233. https://doi.org/10.1002/cnr2.70233

Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., & Robb, G. B. (2012). Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. Nucleic Acids Research, 40(7), e54. https://doi.org/10.1093/nar/gkr1269

Appendices

Appendix A: Cloning Information for Thermophilic RNA Ligases

A1: PpaRnl Wild-Type Sequence

PpaRnl peptide sequence. 401 residues including the N-terminal Hexahistidine-tag (underlined).

Asp94, Glu155, Lys238, Asp248 are in bold.

MGSSHHHHHHMVSSYFKGILLNLGLDEERIEVLENKGGIVEDEFEGMRYLRLKD
SARSLRRGTVVFDEHNIILGFPHIKRVVQLENGIRRAFKRKPFYVEEKV**D**GYNVR
VAKIGEKILVFTRGGFVCPFTTERIEDFITLDFFKDYPNMVLCG**E**MAGPESPYLVE
GPPYVKEDIQFFLFDIQEKKTGRSLPVEERLKLAEEYGIPSVEVFGLYDLSRIDELH
ALIDRLTKEKREGIVMKSPDMKKIV**K**YVTPYANIN**D**IKIGARIFFDLPHGYFMQRI
KRLAFYLAERKIRGEEFDEYARALGKVLEPFVESIWDISSGDDEIAELFTVRVKKL
ETAHKMVTHFERLRLKIHIDDIEVLDNGYWRITFKRVYPDATKEMRELWNGHAF
VD

Molecular Weight: 46.89 kDa (monomer)

Isoelectric point: 6.5

Extinction Coefficient: 40,465

A[280] of 1mg/ml: 0.86AU

A2: PpaRnl K238G Sequence

PpaRnl K238G peptide sequence. 401 residues including the N-terminal Hexahistidine-tag (underlined).

Gly238, in bold.

MGSSHHHHHHMVSSYFKGILLNLGLDEERIEVLENKGGIVEDEFEGMRYLRLKD
SARSLRRGTVVFDEHNIILGFPHIKRVVQLENGIRRAFKRKPFYVEEKVDGYNVR
VAKIGEKILVFTRGGFVCPFTTERIEDFITLDFFKDYPNMVLCGEMAGPESPYLVE
GPPYVKEDIQFFLFDIQEKKTGRSLPVEERLKLAEEYGIPSVEVFGLYDLSRIDELH
ALIDRLTKEKREGIVMKSPDMKKIV**G**YVTPYANINDIKIGARIFFDLPHGYFMQRI
KRLAFYLAERKIRGEEFDEYARALGKVLEPFVESIWDISSGDDEIAELFTVRVKKL
ETAHKMVTHFERLRLKIHIDDIEVLDNGYWRITFKRVYPDATKEMRELWNGHAF
VD

Molecular Weight: 46.89 kDa (monomer)

Isoelectric point: 6.5

Extinction Coefficient: 40,465

A[280] of 1mg/ml: 0.86AU

A3: PpaRnl D94A Sequence

PpaRnl D94A peptide sequence. 401 residues including the N-terminal Hexahistidine-tag (underlined). Ala94 in bold.

MGSSHHHHHHGLVPRGSHMMVSSYFKGILLNLGLDEERIEVLENKGGIVEDEFE

GMRYLRLKDSARSLRRGTVVFDEHNIILGFPHIKRVVQLENGIRRAFKRKPFYVE

EKV**A**GYNVRVAKIGEKILVFTRGGFVCPFTTERIEDFITLDFFKDYPNMVLCGEM

AGPESPYLVEGPPYVKEDIQFFLFDIQEKKTGRSLPVEERLKLAEEYGIPSVEVFGL

YDLSRIDELHALIDRLTKEKREGIVMKSPDMKKIVKYVTPYANINDIKIGARIFFDL

PHGYFMQRIKRLAFYLAERKIRGEEFDEYARALGKVLLEPFVESIWDISSGDDEIA

ELFTVRVKKLETAHKMVTHFERLRLKIHIDDIEVLDNGYWRITFKRVYPDATKE

MRELWNGHAFVD

Molecular Weight: 46.85 kDa (monomer)

Isoelectric point: 6.6

Extinction Coefficient: 40,465

A[280] of 1mg/ml: 0.86AU

A4: PpaRnl D248A Sequence

PpaRnl D248A peptide sequence. 401 residues including the N-terminal Hexahistidine-tag (underlined). Ala248 in bold.

MGSSHHHHHHGLVPRGSHMMVSSYFKGILLNLGLDEERIEVLENKGGIVEDEFE

GMRYLRLKDSARSLRRGTVVFDEHNIILGFPHIKRVVQLENGIRRAFKRKPFYVE

EKVDGYNVRVAKIGEKILVFTRGGFVCPFTTERIEDFITLDFFKDYPNMVLCGEM

AGPESPYLVEGPPYVKEDIQFFLFDIQEKKTGRSLPVEERLKLAEEYGIPSVEVFGL

YDLSRIDELHALIDRLTKEKREGIVMKSPDMKKIVKYVTPYANIN**A**IKIGARIFFDL

PHGYFMQRIKRLAFYLAERKIRGEEFAEYARALGKVLLEPFVESIWDISSGDDEIA

ELFTVRVKKLETAHKMVTHFERLRLKIHIDDIEVLDNGYWRITFKRVYPDATKE

MRELWNGHAFVD

Molecular Weight: 46.85 kDa (monomer)

Isoelectric point: 6.6

Extinction Coefficient: 40,465

A[280] of 1mg/ml: 0.86AU

A5: PpaRnl E155A sequence

PpaRnl E155A peptide sequence. 401 residues including the N-terminal Hexahistidine-tag (underlined). Ala155 in bold.

MGSSHHHHHHGLVPRGSHMMVSSYFKGILLNLGLDEERIEVLENKGGIVEDEFE

GMRYLRLKDSARSLRRGTVVFDEHNIILGFPHIKRVVQLENGIRRAFKRKPFYVE

EKVDGYNVRVAKIGEKILVFTRGGFVCPFTTERIEDFITLDFFKDYPNMVLCGAM

AGPESPYLVEGPPYVKEDIQFFLFDIQEKKTGRSLPVEERLKLAEEYGIPSVEVFGL

YDLSRIDELHALIDRLTKEKREGIVMKSPDMKKIVKYVTPYANINAIKIGARIFFDL

PHGYFMQRIKRLAFYLAERKIRGEEFAEYARALGKVLLEPFVESIWDISSGDDEIA

ELFTVRVKKLETAHKMVTHFERLRLKIHIDDIEVLDNGYWRITFKRVYPDATKE

MRELWNGHAFVD

Molecular Weight: 45.90 kDa (monomer)

Isoelectric point: 6.42

Extinction Coefficient: 40,465

A[280] of 1mg/ml: 0.88AU

GMO approval number:

E. coli: GMD101146

A6. Luria Bertani (LB) Agar:

Combine the following:

5 g Peptone

10 g NaCl

5 g Yeast Extract

5 g Agar

1 L of mQ H_2O

Autoclave at 121 °C before use.

A7. Luria Bertani (LB) Broth:

Same methods as LB agar, but no addition of agar.

A8. Terrific Broth (TB):

12 g Tryptone

24 g Yeast

4 mL Glycerol

900 mL of mQ H_2O

Autoclave at 121 °C before use.

2.13 g KH_2PO_4

12.54 g K_2HPO_4

100 mL of mQ H_2O

Autoclave at 121 °C before use and combine with above.

Appendix B: Components for Protein Purification

B1: Purification Buffers for Thermophilic RNA Ligases

Table 5 *Purification buffers for RNA ligase purifications*

Buffer	Composition
Lysis Buffer	50 mM Tris pH 7.4, 200 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol
High Salt Wash Buffer	50 mM Tris pH 7.4, 4 M NaCl, 20 mM imidazole, 10% (v/v) glycerol
Elution Buffer	50 mM Tris pH 7.4, 200 mM NaCl, 1 M imidazole, 10% (v/v) glycerol
Gel Filtration Buffer	50 mM Tris pH 7.4, 200 mM NaCl, 10% (v/v) glycerol

B2: SDS-Page Gel Components

SDS-Loading Dye

250 mM Tris (pH 6.8)

20% glycerol (v/v)

4% SDS (w/v)

10% beta-2-mercaptoethanol (w/v)

0.025% bromophenol blue (w/v)

Tris-glycine SDS Buffer

25 mM Tris (pH 8.5)

250 mM glycine

0.1% SDS (w/v)

Fairbanks Staining Solution

0.05% Coomassie blue

25% isopropanol

10% acetic acid

Destaining Solution

10% acetic acid

EDTA Buffer (1x)

93.05 g EDTA disodium salt

400 mL MQ H_2O

pH to 8.0 with NaOH

Make up to 500 mL with MQ H_2O

DNA Loading Buffer (6x)

3 mL 30% (v/v) glycerol

25 mg 0.25% (w/v) Bromophenol Blue

25 mg 0.25% (w/v) Xylene Cyanol

Make up to 10 mL with MQ H_2O

Appendix C: Thermal Shift Assay Data

C1: Thermal Shift Assay components

50x SYPRO Orange

1 µL SYPRO™ Orange Dye (5,000X Concentrate in DMSO)

99 µL Gel Filtration Buffer (Appendix B1)

5x SYPRO Orange

40 µL 50x SYPRO™ Orange Dye

36 µL Gel Filtration Buffer (Appendix B1)

Reaction Buffer (per buffer)

0.124g Boric Acid

0.2g Phosphoric Acid

0.12g Acetic Acid

pH with NaOH

Make up to 50 mL with MQ H_2O

Appendix D: Oligonucleotide and Assay Components

Table 6 *Oligonucleotide substrates used in ligation assays. P = 5' phosphate: /5Phos. Oligonucleotides were resuspended in Ultrapure DNase/RNase-free H_2O.*

Substrate	Sequence	T_m (°C)
Oligonucleotide 1	P-GAGCUAGCAUUAACUUGG	40.8
Oligonucleotide 50	P-AUGUAUGCACUCUUGAUACAGU	51.6
Oligonucleotide 55	P-UUAUUUGGUCAUGGGGUGAGC	54.9
Oligonucleotide 60	P-AUGUAUGCACUCUUGAUACAGU	61
Oligonucleotide 65	P-GGACGGAGGUUAUUGGCCGCGU	63.7
Oligonucleotide 70	P-UGAGUCUUGAUGCCCCGGCUCC	69.2
Oligonucleotide 75	P-GGGCAAGUUGUGAAGCCGCUAU	77.3
Oligonucleotide 80	P-GUUUGUUCUCCGCACCCUGCGA	80.2
Oligonucleotide 85	P-AGGGGCGUGGGACAACGUACCG	85.4
Oligonucleotide V2	P-ACGGGGUUAUUGGCCGCGUCA	63.7
Oligonucleotide V3	P-UACGGAGGUUAUUGGCCGCGUA	63.7
Oligonucleotide N1	P-ACGCUAAGCUUAGCUAGUA	67.1
Oligonucleotide N3	P-ACGCUAAGCUUAGCUAACG	67.3
Oligonucleotide N4	P-ACGCUAAGCUUAGAUAGCA	68.3

D2: Assay Dye and Buffer Components

NEBuffer 1.1 (New England BioLabs)

10 mM Bis-Tris-Propane-HCl

10 mM $MgCl_2$

1 mM DTT

pH 7 at 25 °C

T4 RNA Ligase Reaction Buffer (New England BioLabs)

10 mM Tris-HCl

10 mM $MgCl_2$

1 mM DTT

pH 7.5 at 25 °C

2x Formamide Loading Dye

Deionised Formamide 9.5 ml

Bromophenol Blue 2.5 mg

Xylene Cyanol FF 2.5 mg

EDTA (0.5 M, pH 8) 100 µL

Make up to 10 mL with Ultrapure DNase/RNase-free H_2O

D3. Urea-PAGE Gel Buffer Components

10 x TBE Buffer, pH 8.3

121.1 g Tris

61.8 g Boric Acid

7.4 g EDTA

Make up to 1 L with MQ H_2O

5 x TBE Buffer, pH 8.3

200 mL of 10 x TBE Buffer

Make up to 1 L with MQ H_2O

1 x TBE Buffer, pH 8.3

100 mL of 10 x TBE Buffer

Make up to 1 L with MQ H_2O

Appendix E: Crystallisation and Structure Solving Information

E1. Structural images

Final refinement settings for PpaRnl D94A and PpaRnl D248A

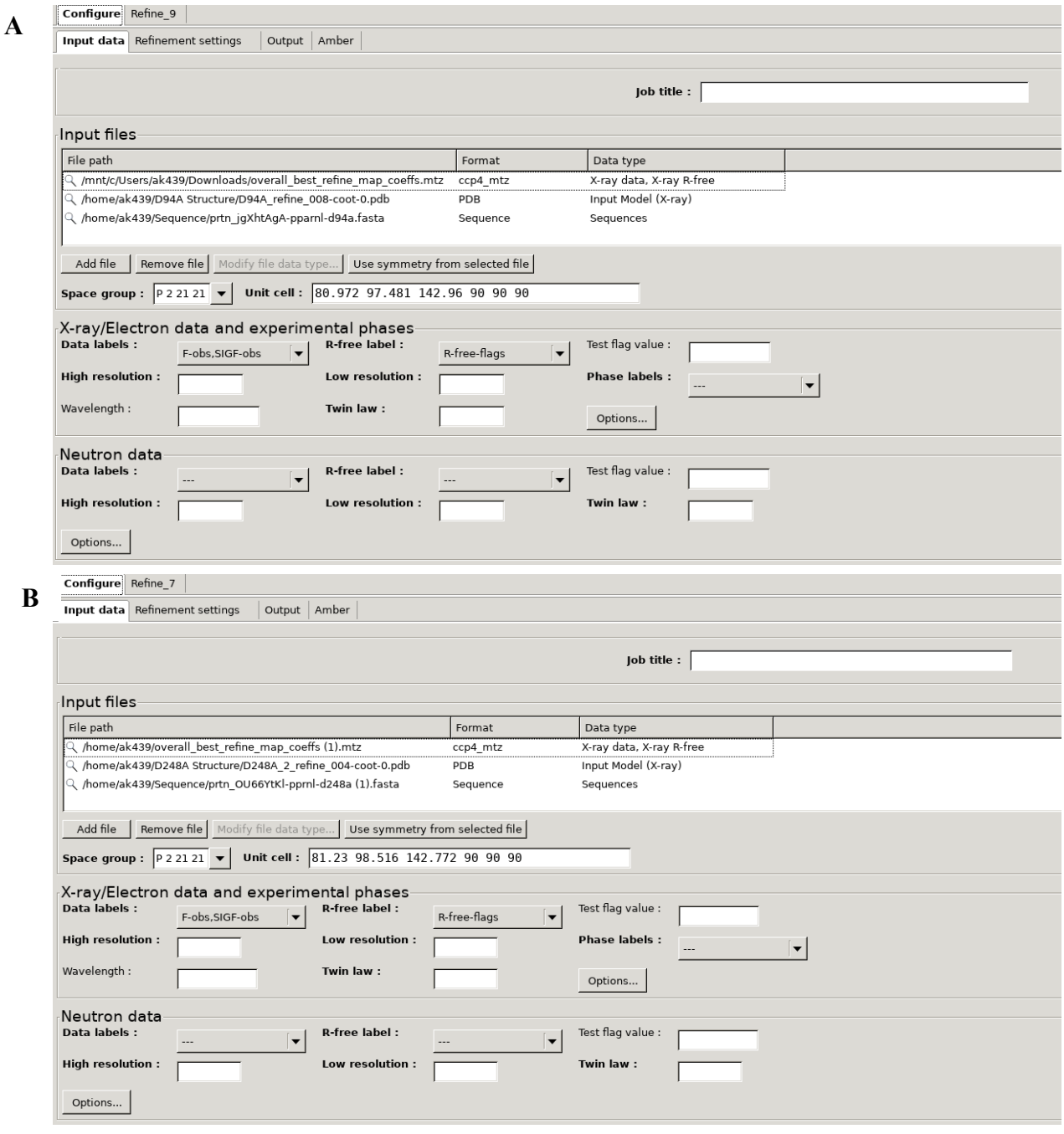

Figure 4.1 *Screenshots of the final input in phenix.refine used for building and refining the modes of PpaRnl D94A (A) and PpaRnl D248A (B).*

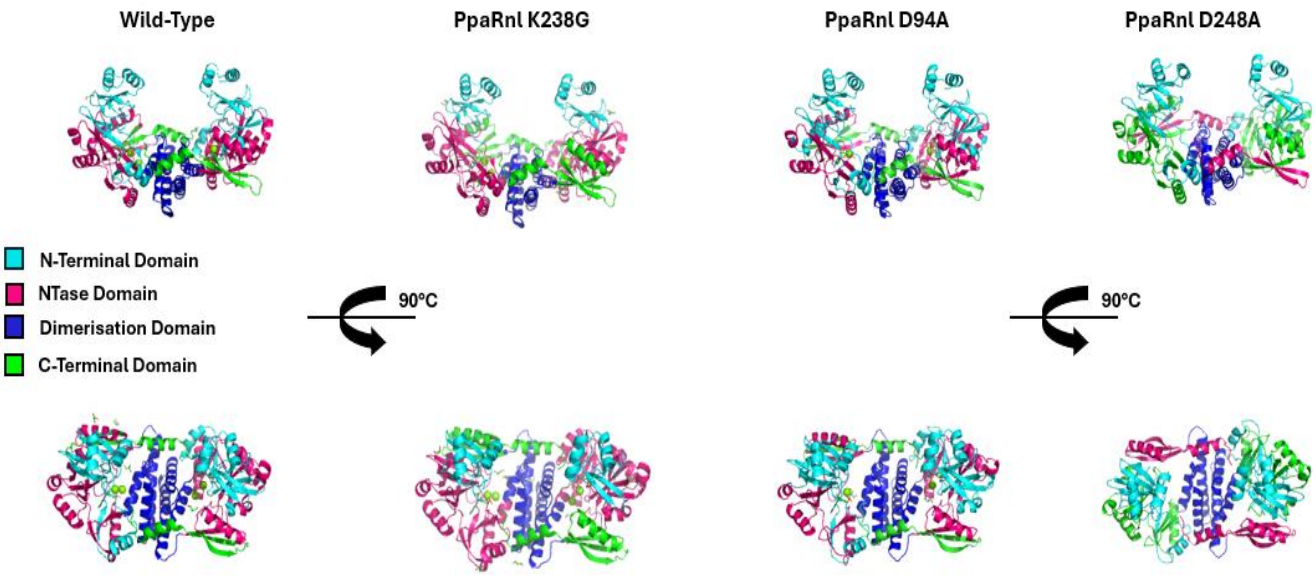

Figure 4.2 *Individual secondary structures of the PpaRnl wild-type and mutants K238G, D94A, and D248A with each domain highlighted according to the figure legend.*

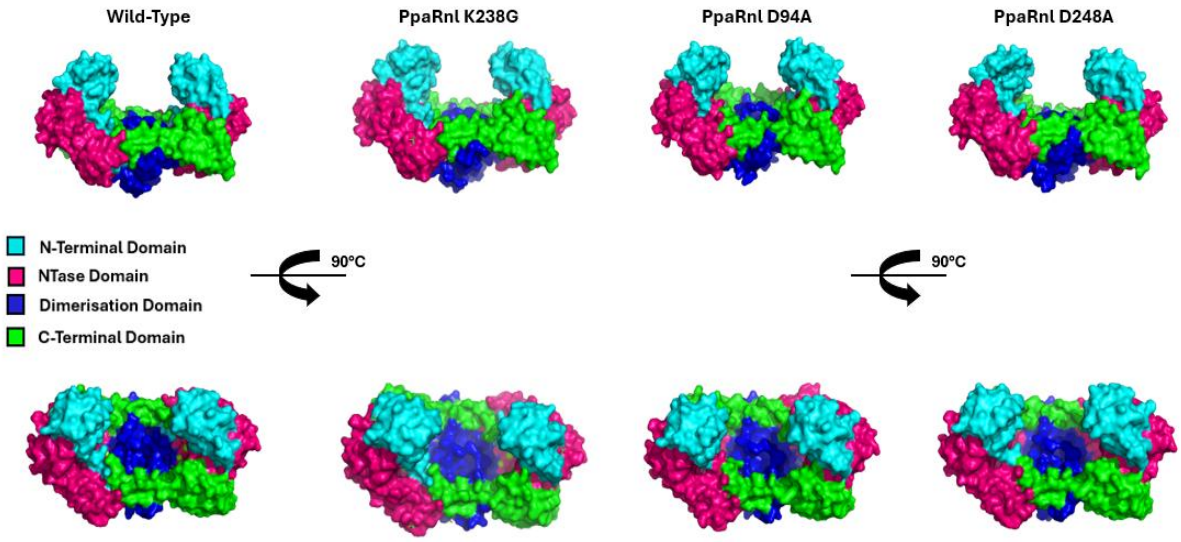

Figure 4.3 *Surface views of tertiary structures of the PpaRnl wild-type and mutants K238G, D94A, and D248A with each domain highlighted according to the figure legend.*

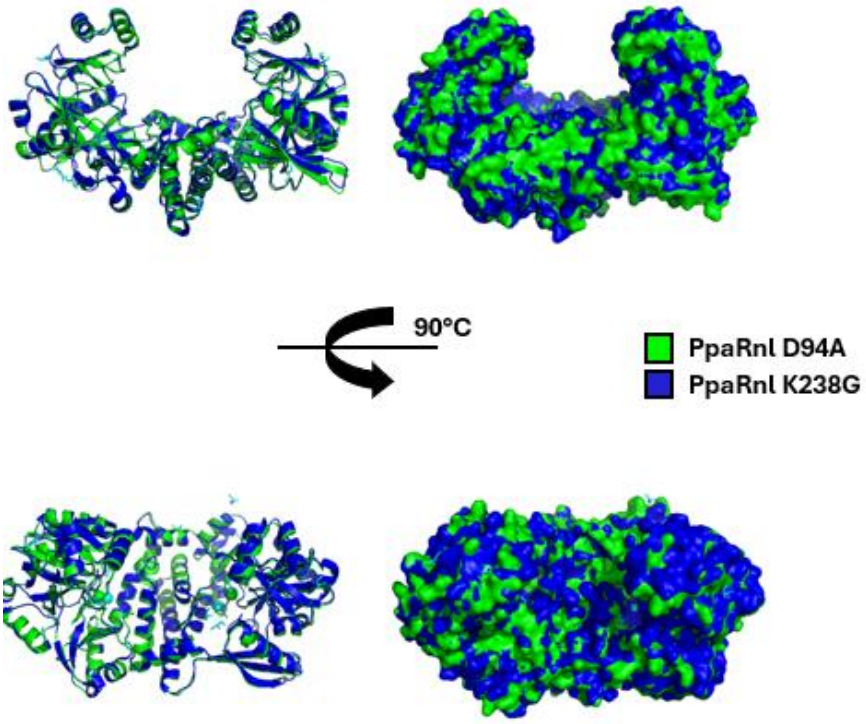

Figure 4.4 *Alignment of PpaRnl K238G and PpaRnl D94A with active site ligands in pale blue in K238G and green in D94A, note the lack of the Mg^{2+} B ion in D94A*

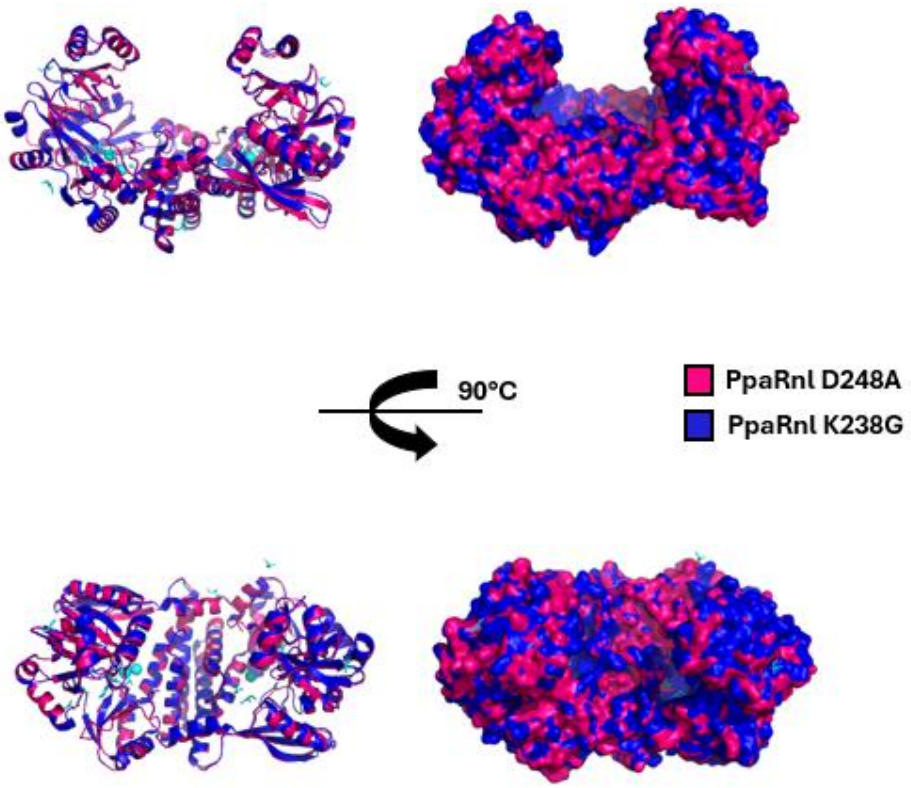

Figure 4.5 *Alignment of PpaRnl K238G and PpaRnl D248A with active site ligands of K238G in pale blue, note the lack of any ligands present in the active site of D248A.*