ENGINEERING A THERMOSTABLE RNA LIGASE FOR USE IN AN UNBIASED MICRORNA SEQUENCING PROTOCOL

BY

Hamish John Francis Trlin

2021

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Biotechnology

Abstract

MicroRNAs are small, non-coding RNA sequences that act to regulate gene expression at a post translational level. In humans, irregular expression of microRNAs has been associated with the development of numerous diseases. The ability to determine alterations in microRNA expression may, therefore, provide a novel method for predicting the onset, severity, and outcomes of these diseases. Identifying changes in microRNA expression relies on accurately sequencing and profiling microRNAs. Current microRNA sequencing protocols, however, are plagued by bias. To successfully sequence microRNAs, adapters must be ligated to both ends of the molecule. Current protocols utilise two separate reactions to achieve this. In the first, a pre-adenylated, 3′ amino modified DNA adapter is ligated to the 3′ end of the microRNA. In the second, a ligase must adenylate the 5′ phosphate of the microRNA, utilising ATP, before ligating it to the 3′ end of a 5′ dephosphorylated RNA. It is in the ligation of these adapters that the bias originates. Different microRNA secondary and tertiary structures have been shown to differentially alter the ability of ligase enzymes to interact with the microRNA. By running the adapter ligation steps of the sequencing protocol at temperatures high enough to melt microRNA structures, this bias should be eliminated.

The RNA ligase from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pfu*) was identified as a target of interest for use in our sequencing protocol due to its remarkable thermostability. To optimise the activity of the ligase for ligation of a pre-adenylated 3′ adapter to microRNA, its adenylation activity needed to be removed. This activity is associated with the generation of undesirable ligation products. Preliminary research, carried out by Dr Tifany Oulavallickal, identified residues K92 and K238 as targets of interest for mutagenesis. A K92A variant of the *Pfu* RNA ligase was generated as a proof of concept, displaying significantly reduced adenylation activity while retaining ligation activity at the desired temperatures. The research conducted for this thesis built upon this work, characterising all other possible K92 substitutions, and utilising that information to inform separate, and co-substitution of K238 (i.e., single, and double mutants).

All the possible K92 variants were successfully generated. Ligation activities with both DNA and RNA sequences were then assessed by endpoint TBE-urea gel assays, with K92A, K92G, K92S, K92T, and K92Y being identified as substitutions of interest. These same amino acids (A, G, S, T, and Y) were then used to replace the second active site lysine, K238. Double

mutants were also generated by substituting K238 of the K92A variant for A and Y. All K238 single mutants, and both double mutants were successfully characterised.

All K92 variants displayed significant decreases in adenylation activity, with the amino acid substitutions A, G, S, T, and Y resulting in increased ligation activities. All K238 variants, bar K238Y, displayed significant ligation activities, but continued to display adenylation activity.

As such, K92A was identified as the best candidate for the 3′ DNA adapter ligation. This variant displayed the most promising ligation activity of all K92 variants, while displaying low enough adenylation activity to minimise the production of undesirable ligation products. The adenylation activity observed in the K238 variants made them ideal for the 5′ RNA adapter ligation, as this requires the microRNA to be adenylated. As such, the most active K238 variant, K238G, was identified as the best candidate for this adapter ligation.

Acknowledgements

Constructing this thesis was always going to be a difficult task, filled with trials and tribulations. Unfortunately, COVID-19 only made it harder. I consider myself extremely fortunate to have been able to complete this thesis despite the raging pandemic, and that is in part due to the incredible effort of New Zealanders. As such, I would like to begin by thanking everyone who did their part to keep New Zealand COVID-free. Without all your efforts, this year would have been unimaginably difficult.

To my supervisor, Assoc. Prof. Wayne Patrick, thank you for all the guidance you have given me over the past year, and for allowing me to be part of such an incredible lab group. Your passion for science is infectious and serves as inspiration for all your students.

A huge thank you to the wider MME lab, not only for the support you gave me, but also the help and support you provide to all your members. It can be easy to forget how fortunate we are to be part of such a cool group of people who are all so willing to help each other out. You all made the past year an absolute pleasure. A special thanks to team engineering for the additional support, and to Monica S., Mike, and Josie for helping with the final editing of this thesis.

A special acknowledgement is required for the incredible Dr Tifany Oulavallickal. You are the heart and soul of this project. The time and effort you put into helping me throughout the duration of my research and writing is beyond what I could have ever asked for. You are an absolute inspiration! Thank you for being there for me every step of the way.

Finally, I would like to thank my friends and family. You guys are my world. I couldn't have done any of this without you and count my lucky stars that you are all part of my life. To my mum, Katrina, and my dad, Matthew, thank you for everything. The past 24 years have been an incredible journey, and I have you to thank for every minute of it. All my greatest achievements are thanks to you, and never would have been possible without your undying support. To my brothers, Josh and Nick, thank you for being such incredible friends, and for always keeping me grounded. I can't put into words how much I appreciate everything you have done for me. L + 1.

Table of Contents

List of Figures

List of Tables

Abbreviations

3′ AmMO	3′ amino modification
AIM	Auto-induction media
AMP	adenosine monophosphate
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	base pair
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
His_6-tag	hexa-histidine tag
IPTG	isopropyl β-D-1-thiogalactopytanoside
LB	Luria-Bertani
NAD^+	nicotinamide adenine dinucleotide
nt	nucleotide
OD_{600}	optical density at the 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
PPi	pyrophosphate
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
T_a	annealing temperature
TAE	tris-acetic acid-EDTA
TBE	tris-borate-EDTA
TE	tris-EDTA
TEMED	tetramethylethylenediamine
T_m	melting temperature

Chapter 1 Introduction

1.1 Ligase enzymes

Ligases are enzymes which catalyse the formation of chemical bonds between two molecules, resulting in the generation of a larger product. Of particular interest to the research conducted for this thesis are the ligases which repair phosphodiester bonds between deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) sequences.

Damage to DNA and RNA is defined as any alteration to the structure of the sequence that may cause cellular injury or reduce the viability or reproductive fitness of an organism (Kaufmann & Paules, 1996). There are two primary subdivisions of damage: endogenous damage, caused by reactive oxygen species or by-products of metabolic processes; and exogenous damage, caused by radiation, hydrolysis, plant toxins and viruses (Cervelli *et al.*, 2012). Alterations to DNA or RNA can range from base alterations to breakage of phosphodiester bonds. It is important to note that all forms of damage will eventually require a ligase enzyme to seal strands of DNA or RNA back together. DNA and RNA ligases therefore allow cells to efficiently repair damage as it occurs, slowing its accumulation and preserving cell functionality.

1.1.1 DNA ligases

Two distinct classes of DNA ligases have been defined according to their preferred cofactor. NAD^+-dependent DNA ligases are specific to prokaryotes (Wilkinson *et al.*, 2001), and entomopox viruses (Sriskanda & Shuman, 2001). ATP-dependent DNA ligases on the other hand have been found across all phylogenetic kingdoms, with all known archaebacteria and eukaryotes possessing at least one such ligase (Ellenberger & Tomkinson, 2008; Martin & MacNeill, 2002). To date, all bacteria have been found to encode at least one NAD^+-dependent DNA ligase, however, may also contain an additional ATP-dependent DNA ligase (Cheng & Shuman, 1997; Wilkinson *et al.*, 2001; Williamson *et al.*, 2016; Williamson *et al.*, 2014).

1.1.2 RNA ligases

According to current literature, there are five families of RNA ligase enzymes (Rnl1 -Rnl5), each of which can be classified by specific structural features and polynucleotide substrate

specificities (Gu *et al.*, 2016). Crystal structures have been solved for at least one member of each family of enzymes. These ligases are typically utilised during RNA processing events, such as sequence editing, splicing, and repair (Bellacosa & Moss, 2003; Gu *et al.*, 2016; Ho & Stewart, 2002; Zhelkovsky & McReynolds, 2014). All RNA ligases are believed to be ATP dependent, catalysing the joining of RNA molecules by means of three nucleotidyltransfer steps like those utilised by ATP-dependent DNA ligases (Chauleau & Shuman, 2013; Ho *et al.*, 2004; Zhelkovsky & McReynolds, 2012). Rnl1 ligases, such as T4 RNA ligase1, are typically associated with repairing the anticodon loop of tRNAs. Rnl2 ligases, such as T4 RNA ligase2, and Rnl5 ligases are involved in double stranded RNA (dsRNA) nick repair. Rnl2 and Rnl5 ligases can be differentiated by the enzymatic domain associated with substrate recognition. Rnl3 ligases can be found in archaeal species and have been shown to circularise both single stranded RNA and DNA (ssRNA and ssDNA), as well as convert 3′-phosphorylated RNA termini into 2′, 3-cyclic phosphate. The biological function of these processes is currently unknown, and other functions have yet to be identified. Rnl4 ligases interact with protein Hen1 to repair ribotoxin damaged RNA molecules.

1.1.3 Mechanisms

All ligations share a three-step mechanism, shown for RNA ligase in Figure 1-1 (Shuman, 2009). The hydrolysis of one high energy molecule is required per enzymatic event (Weiss *et al.*, 1968). Firstly, a lysine residue in the active site of the ligase performs a nucleophilic attack on the α-phosphate of ATP or NAD^+ forming a ligase-AMP intermediate and releasing pyrophosphate (PP_i) or nicotinamide mononucleotide (Figure 1-1, step 1). The adenylation of the ligase yields a relatively stable and highly reactive conformation that has increased specificity for binding to oligonucleotide substrates (Sriskanda & Shuman, 1998). The AMP is then transferred to the 5′ phosphate terminating end of the oligonucleotide strand, forming an oligonucleotide-adenylate (Figure 1-1, step 2). This occurs by the oxygen from the 5′ phosphate attacking the ligase-adenylate complex, allowing the lysine residue to act as a leaving group. The reaction is resolved with an attack by a 3′ hydroxyl group from the second oligonucleotide substrate, on the 5′ phosphate, resulting in the formation of a new phosphodiester bond and the liberation of AMP.

According to the 2011 paper published by Viollet *et al.*, step 2 of this mechanism is reversible, allowing ligases to remove AMP from an RNA-adenylate and re-form the ligase-AMP intermediate. Re-adenylated ligases can transfer the AMP to new oligonucleotide.

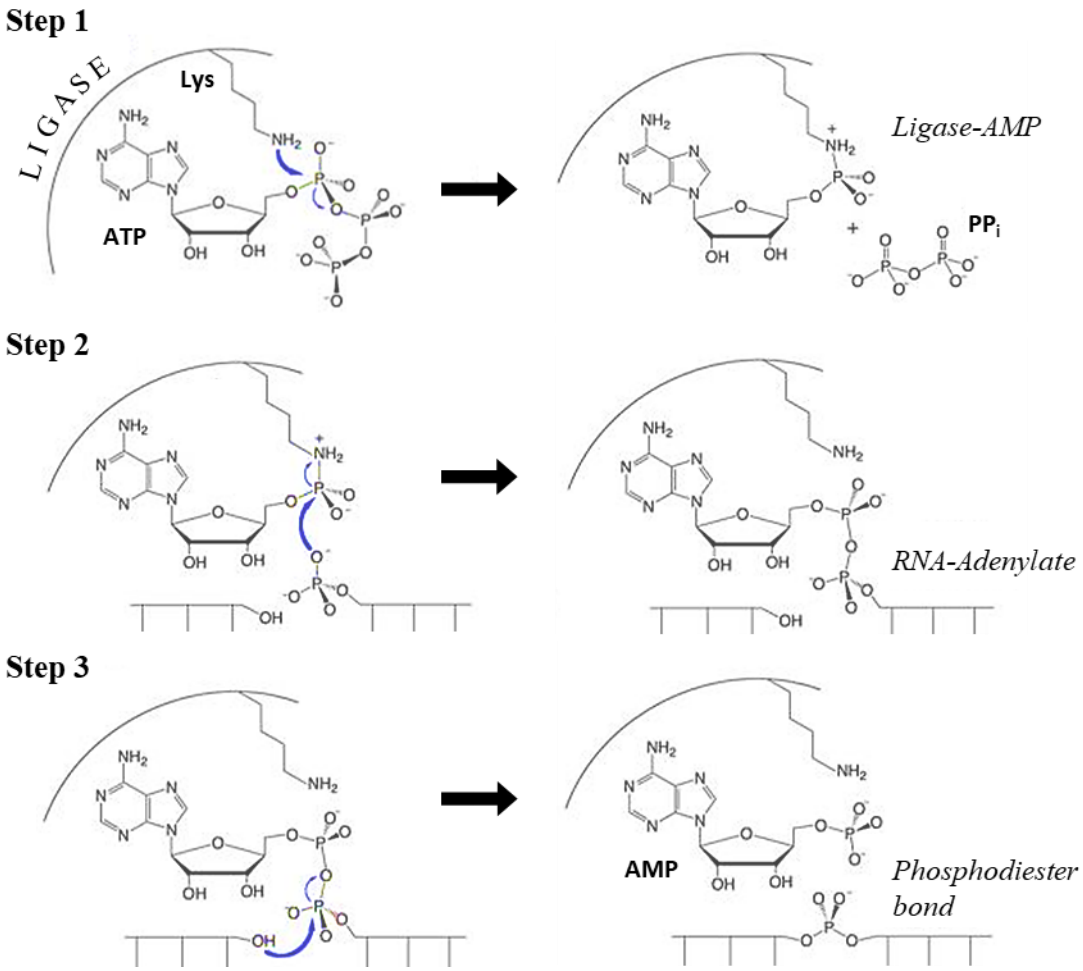

Figure 1-1 Mechanism of ligation

This figure depicts the three steps of the ligation mechanism. Step 1 involves adenylation of a lysine in the active site of the ligase, resulting in the formation of the ligase-AMP intermediate. In step 2, the AMP is transferred from the ligase-AMP intermediate to the target RNA molecule, generating an RNA-Adenylate. Step 3 sees the ligase facilitate the formation of a phosphodiester bond between the RNA-adenylate and a second RNA molecule, resulting in the AMP being released into solution.

1.1.4 Structure

DNA and RNA ligases have a relatively high level of sequence and structure similarity in their active sites. At the core of every ATP-dependant ligase lie six conserved nucleotidyl-transferase motifs (I, Ia, III, IIIa, IV, and V) that form a common adenylyl-transferase domain (Ho *et al.*, 2004; Martin & MacNeill, 2002; Sriskanda & Shuman, 1998). These motifs contain the key amino acid residues involved in the enzymatic activity of the ligase. The adenylyl-transferase domain can be separated into two separate sub-domains: the adenylation domain; and a C-terminal oligonucleotide-binding domain. The adenylation domain contains five of the six previously mentioned motifs, and thus most of the key active site residues, as well as the ATP binding site. The oligonucleotide-binding domain contains the sixth motif and has two key functions. Its primary function is to increase substrate binding affinity,

stabilising and orienting DNA or RNA substrates such that ligation can be catalysed. This also means it dictates the substrate specificity of the enzyme (Wang *et al.*, 2007). Its secondary function is to assist with the transfer of the AMP to the enzymes active site, allowing the conformational change that prepares the enzyme for substrate orientation and ligation (Nandakumar *et al.*, 2006). Some ligases have an additional N-terminal binding domain that interacts with the adenylation and oligonucleotide-binding domains to encircle the target substrate. The one exception to this is the Rnl5 family of RNA ligases, which lack a C-terminal domain, relying on a distinct N-terminal domain to recognise the target substrate and facilitate AMP transfer (Gu *et al.*, 2016).

Outside of the six key catalytic sequence motifs, ligases can vary quite dramatically in sequence and overall structure. It is these other structural features that determines the optimal conditions for enzymatic function, as well as the overall stability of the specific ligase (Brooks *et al.*, 2008; Pascal, 2008; Torchia *et al.*, 2008).

1.1.5 Applications

1.1.5.1 Molecular cloning technology

The first DNA ligase was purified in 1967, marking the beginning of the modern era of biotechnology (Weiss *et al.*, 1968). These enzymes play an essential role in molecular cloning technologies, which allow recombinant DNA sequences to be generated through the insertion of DNA fragments into target vectors. This is achieved by "cutting" target vectors with restriction enzymes and ligating a DNA fragment into the resultant gap (Benz, 1990; Carroll, 1993; David *et al.*, 1972). The impact this technique has had on all biological science disciplines is immeasurable, receiving recognition for its importance in 1980 with a Nobel Prize. Paul Berg, the first person to construct a recombinant DNA molecule, was the recipient of this award.

Several next generation cloning technologies, such as Gibson, and Golden Gate assembly, have been developed since, expanding upon the capabilities of traditional molecular cloning. All these techniques still require DNA ligase enzymes. Gibson assembly, for example, is a single isothermal overlap recombination system that utilises Taq DNA polymerase, exonuclease III, and Taq DNA ligase (Gibson *et al.*, 2009). This system allows multiple DNA inserts to be assembled in a single reaction using complementary overhangs, while also extending the maximum length of individual inserts to several hundred kb. These types of developments are accelerating the rate at which genetic pathways are discovered and re-

engineered, improving library construction, site-directed mutagenesis, and the production of various compounds and pharmaceuticals.

1.1.5.2 High throughput sequencing

The first method of polynucleotide sequencing was established by Ray Wu at Cornell University in 1970 (Wu, 1970). The method utilised a location-specific primer extension strategy, and successfully generated several sequence reads between 1970 and 1973. This method laid the foundations for the development of Sanger sequencing, which was developed in 1977 by Fredrick Sanger and colleagues (Sanger & Coulson, 1975). For nearly 20 years, Sanger sequencing was the method of choice for all manner of sequencing experiments, capable of generating reads of up to 1000 bp with relatively high levels of accuracy. This came to an end in the mid to late 90s as genomics drove the need for a markedly higher throughput method of sequencing. Second-generation sequencing technologies provided the desired improvement to throughput, introducing methods for in-parallel sequencing, as well as increasing maximum read length.

Ion Torrent, Illumina, and 454 pyrophosphate sequencing (Margulies *et al.*, 2005) are all second-generation sequencing technologies that utilise sequencing by synthesis techniques. Sample preparation and workflows are similar for all three, relying on DNA ligase enzymes to ligate target sequences to immobilised adapters (Linnarsson, 2010). Target sequences of dsDNA are fragmented by either mechanical or enzymatic means, such that the fragments are of appropriate length. Fragments can then be ligated to adapters which are immobilised on emulsion beads or in microfluidic channels, allowing for millions of sequencing reactions to be carried out in parallel. Modified nucleotide bases are then incorporated a single base pair at a time to the target sequences. The identity of the specific base incorporated can then be determined by fluorescence or a change in current depending on the sequencing technique, ultimately allowing for the determination of the full sequence. The adapters utilised have two primary functions. Firstly, they act as a primer sequence for DNA polymerase binding, allowing for synthesis to occur. Secondly, they act as a barcode. This allows for the results from multiple individual sequencing experiments to be pooled for analysis, with separate reactions being identified by their barcode. This process results in the generation of a highly accurate consensus sequence, which can then be evaluated by bioinformatic tools.

MicroRNA sequencing protocols are of particular interest to this research. Using RNA ligase enzymes, it is possible to ligate adapter sequences to a pool of microRNAs, allowing for

primer binding, reverse transcription, and then amplification of the sequences. Massively parallel sequencing techniques, such as those described above, can then be used to analyse the resultant pool of cDNA sequences.

1.2 MicroRNA

Believed to be implicated in the regulation of approximately one third of all human genes (Hammond, 2015), microRNAs are one of the most significant regulators of human gene expression. Acting post-transcriptionally, these small, non-coding RNA sequences hybridise to the 3′-untranslated regions (3′UTR) of specific mRNAs, leading to degradation or translational repression of the sequence. Malfunction of this system can result in the down regulation or silencing of genes required for regular cellular function. To date, the mis-regulation of microRNA expression has been correlated with the development of over 100 human diseases (Chen & Zhang, 2013; Chen *et al.*, 2018; Chen & Yan, 2014; Jiang *et al.*, 2009; Salta & De Strooper, 2017). Developing accurate microRNA sequencing protocols will improve microRNA expression pattern analysis, expanding the repertoire of tools available for the prediction of onset, severity, and outcomes of disease (Lu *et al.*, 2005; Rupaimoole & Slack, 2017; Takamizawa *et al.*, 2004; Wang *et al.*, 2016; Yanaihara *et al.*, 2006). Accurate profiling of microRNA expression also has the potential to greatly broaden our understanding of other biological systems. Many multicellular organisms utilise microRNAs in the development of both innate, and adaptive immune responses, as well as for regulating numerous other biological processes.

1.2.1 Sequencing

Existing microRNA sequencing protocols utilise two different RNA ligase enzymes to ligate flanking adapter sequences to the microRNAs within a sample (Newman *et al.*, 2011), such that a reverse transcriptase can be utilised to generate cDNA for sequencing (Figure 1-2). First, an engineered RNA ligase catalyses the ligation of pre-adenylated, 3′ modified ssDNA adapter sequences to the 3′ ends of each microRNA (step 1). Using pre-adenylated adapters bypasses steps 1 and 2 of the ligation mechanism (Figure 1-1), allowing ATP to be omitted from the reaction. The 3′ modification on the adapter ensures the sequence can only be ligated to the 3′ end of microRNAs, forcing the desired ligation to occur. To ensure the ligase is unable to facilitate the transfer of AMP from pre-adenylated adapters to itself (Section 1.1.3), resulting in the generation of unwanted ligation products, a ligase lacking the lysine residue required to catalyse steps 1 and 2 of the ligation mechanism is utilised. The

combination of modified ligase and pre-adenylated 3′ adapter forces microRNA-adapter ligations reaction to occur, minimising the generation of unwanted ligation products.

T4 RNA ligase1 is then utilised to ligate a 5′ modified RNA adapter to the 5′ end of the microRNA (step 2). This is done without pre-adenylation and in the presence of ATP. The 5′ modification on the RNA adapter prevents its adenylation, meaning only the product of the initial ligation reaction can be adenylated. This microRNA-adapter product retains the 3′ modification of the DNA adapter, preventing its ligation to anything other than the 3′ end of the 5′ modified adapter. As a result, only the desired ligation reaction can occur, meaning ATP can be utilised without generating off-target products.

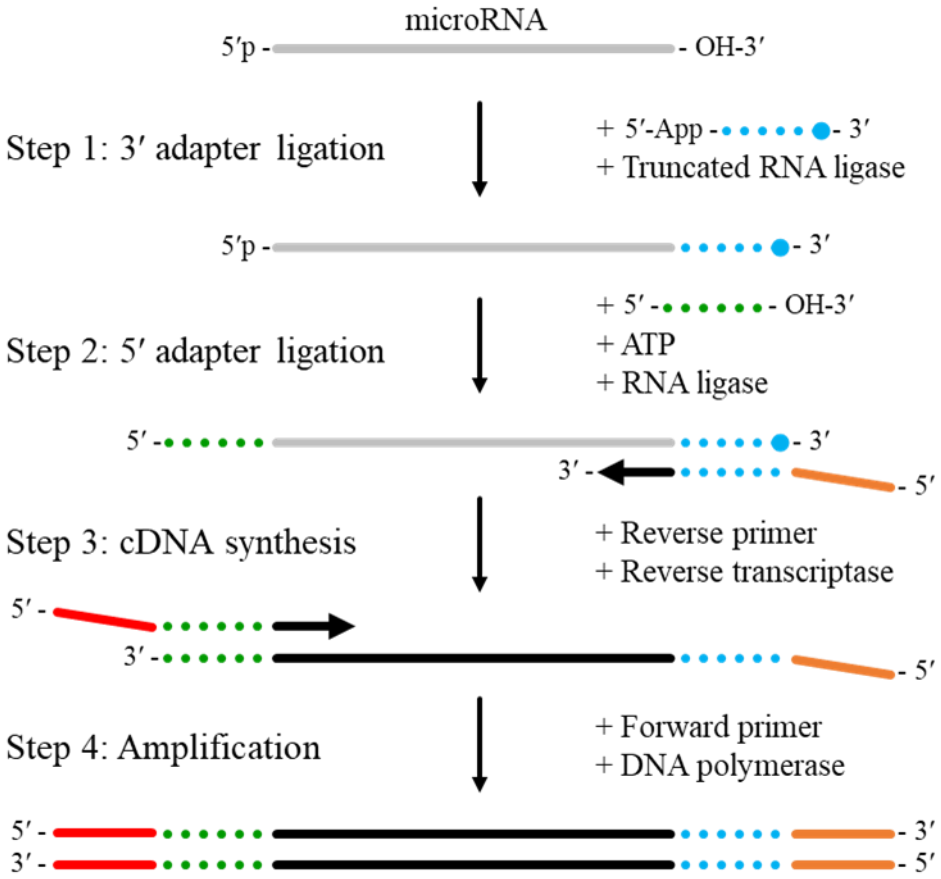

Figure 1-2 Adapter ligations

This figure displays the four steps involved in generating cDNA from microRNAs. The 3′ adapter ligation utilises a truncated T4-RNA Ligase II to ligate 5′ pre-adenylated, 3′ blocked adapters to the 3′ end of the microRNA. For the 5′ adapter ligation, the 5′ phosphate of the microRNA needs to be adenylated before it can be ligated to the 3′ end of 5′ de-phosphorylated adapters. Dephosphorylation of the adapters prevents their adenylation. T4-RNA ligase I is used for this as it can carry out the adenylation reaction in the presence of ATP. For cDNA synthesis, a primer anneals to the 3′ adapter, allowing a reverse transcriptase to generate a cDNA copy of the sequence. PCR amplification is then utilised to amplify the cDNA sequence.

Adapted from RNA-seqlopedia, Figure 3.4 (Cresko Lab, University of Oregon; https://rnaseq.uoregon.edu/#figure3.4)

A primer of complimentary sequence to the 3′ DNA adapter is then hybridised to the molecule (step 3), allowing reverse transcriptase enzymes to synthesise the complementary DNA strand of the adapter ligated RNA. cDNA product can then be amplified and enriched by means of PCR, then sequenced by next-generation methods (Patrick *et al.*, 2008; Raabe *et al.*, 2014; Yanaihara *et al.*, 2006).

1.2.2 MicroRNA expression profile analysis

MicroRNA sequences must be complementary to the target mRNA sequence for hybridisation to occur. As a result, it is possible to identify the target mRNA, and therefore the gene being regulated by the microRNA. The extent to which a gene is regulated is proportional to the quantity of microRNA present. Generating an accurate profile of microRNAs within a sample can therefore be used to quantify the extent to which the target genes are being regulated, providing a unique opportunity to have near real-time readout of transcriptional regulation. This has many medical applications, ranging from the design of personalised medicine to the identification of malfunctioning regulatory systems; the latter of these has shown promise in the identification of a range of diseases, particularly cancers (Carbonell *et al.*, 2012; Esquela-Kerscher & Slack, 2006; Liz & Esteller, 2016). For the most part, this provides an alternate method for diagnosing many currently diagnosable diseases, while potentially providing a more accurate prognosis. The real benefit, however, is apparent in a smaller group of diseases, which currently have no simple or reliable method of diagnosis (Bjaanaes *et al.*, 2014; Cuk *et al.*, 2013; Patrick *et al.*, 2008; Phua *et al.*, 2014; Xu *et al.*, 2016).

1.2.2.1 Example: expression profiles in peripheral blood and faecal samples

Pancreatic or biliary-tract cancers are renowned for their elusive nature, often evading detection until late-stage development (Perkhofer *et al.*, 2019). This results in poor prognosis for patients, with a significant proportion of cases becoming terminal before their discovery. Identifying microRNA markers, which exist stably in peripheral blood, is a promising new technique for identifying these cancers. Recent studies have identified numerous microRNA sequences, 66 and 81 in biliary and pancreatic cancers respectively, that show a statistically significant difference in expression between patients with and without these cancers (Kojima *et al.*, 2015). This provides a substantial pool of biomarkers for which diagnosis may be based. The observation of statistically significant differences shows promise for the use of microRNAs in identifying specific cancers, while also being able to determine malignancy.

The highly invasive procedures required to diagnose colorectal cancers often discourage people from seeking medical advice. This compounded with the often-asymptomatic nature of colorectal cancer, results in a large portion of cases going undiagnosed for significant lengths of time, reducing the chance of a positive prognosis. Numerous research articles have detailed the potential of faecal microRNA sequencing to provide a non-invasive replacement for current diagnostic techniques (Kanaoka *et al.*, 2013; Link *et al.*, 2010; Phua *et al.*, 2014; Tepus & Yau, 2020).

1.2.3 Sequencing bias

In 2014, a paper published by Raabe et al. highlighted the bias issue observed in current methods of microRNA sequencing. It was noted that the expression levels of some microRNAs appeared to be artificially enhanced, while others were diminished, or even absent from deep sequencing datasets. This was backed up by a study in 2015 by Fuchs et al. in which a pool of randomised ligation substrates, defined mixtures of microRNAs, and several combinations of adapters were used in the construction of high throughput screening (HTS) libraries. The use of a defined mixture of RNAs allowed for a direct comparison of different library preparation conditions to an expected result. The first library was generated simply to observe the extent of the bias in sequencing protocols. Sequencing reads were annotated with zero mismatch tolerance, counted, then normalised, meaning each RNA was expected to have a read value of 1. A minority of the microRNAs were within 2-fold of the expected value, with nearly a 10,000-fold spread between the sequences which had the highest and lowest abundances. Sequencing and library construction replicates were used to prove replicability of the result. qPCR was used to confirm the bias. The paper goes on to show that the discrepancy observed in sequencing is caused by the folded structures of microRNAs, with irregular conformations making it difficult for the ligase enzyme to function consistently from one structure to another. Due to this, the ligase preferentially interacts with microRNAs of certain structures (Argyropoulos *et al.*, 2017; Fuchs *et al.*, 2015; Hafner *et al.*, 2011; Raabe *et al.*, 2014; Zhuang *et al.*, 2012).

MicroRNA secondary and tertiary structures, as with any RNA sequence, are highly dependent on the sequence of the molecule (Mathews *et al.*, 1999; Rogers & Heitsch, 2016; Singh *et al.*, 2019; Tan *et al.*, 2017; Zuker, 2003). Where possible, microRNA sequences will hydrogen bond, or hybridise to themselves, folding into unique shapes and patterns with differing stabilities (Figure 1-3). The stability of the structure assumed by the sequence depends on two factors: the sequence complementarity; and the guanine-cytosine (GC)

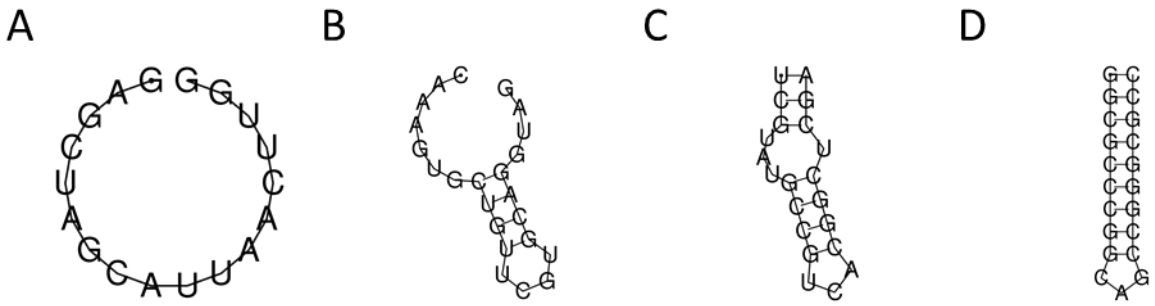

Figure 1-3 Examples of microRNA structures

This figure displays four different microRNA structures. Low enough levels of sequence complementarity may prevent the formation of a secondary structure, as displayed in A. At the other extreme, high enough levels of sequence complementarity may result in all bar one or two bases pairing, as displayed in D. Sequences displayed in B and C fall somewhere between these two extremes. While the level of base pairing within an oligonucleotide sequence is directly proportional to the thermotolerance of the secondary structure, so is the level of G-C pairing. As such, A and D respectively represent the least and most thermostable oligonucleotide secondary structures, with A displaying no secondary structure due to a lack of complementarity, and D displaying the highest level of complementarity possible with only G-C pairing.

composition. Hydrogen bonding between adenine and uridine bases is weaker than that between guanine and cytosine, meaning high GC content is typically associated with greater stability. Formation of these hydrogen bonds requires complementarity between sections of the sequence, however. The lower the complementarity, the less perfectly the sequence can hybridise to itself. As a result, some GC poor sequences may be more stable than some GC rich sequences. The greater the stability of a microRNA structure, the higher its melting temperature, or T_m.

Several issues arise from the folding of microRNAs and the bias it introduces into adapter ligation during sequencing. The primary issue is inconsistency, making it difficult for microRNA sequences to be accurately quantified. Fortunately, it is still possible to determine whether levels of expression have changed. This is due to the bias being relative to the microRNAs conformation. As such, if two samples are sequenced under the same conditions, the relative rates of sequencing for a specific microRNA should remain constant, meaning a change in expression should be identifiable (Linsen *et al.*, 2009). Unfortunately, this type of expression comparison can only identify proportional change. The bias problem makes it impossible to determine the exact concentration of a microRNA in solution, thus making quantitative change impossible to determine. Not only does this make comparing rates of expression between microRNAs futile, it also means exact levels of gene regulation are incalculable. Until there is a method for removing bias, these methods of analysis will remain unreliable. There have been some attempts to remove the observed bias, but none have

managed to eradicate the problem (Baran-Gale *et al.*, 2015; Jayaprakash *et al.*, 2011; Sorefan *et al.*, 2012).

1.3 High temperature microRNA sequencing

The simplest solution to removing the bias in microRNA sequencing protocols appears to be increasing the temperature at which the ligation steps take place (Zhelkovsky & McReynolds, 2012). Due to the bias being a function of structural differences between microRNAs, carrying out adapter ligations at temperatures high enough to unfold the pool of microRNAs could remove the bias. While this appears to be a straightforward solution, high temperatures can also cause proteins to destabilise, resulting in a loss of enzymatic function. The target temperature for the ligation reaction is approximately 75 ^{o}C, at which most, if not all microRNA sequences should unfold. The ligase enzyme used must therefore be able to withstand temperatures exceeding 70 ^{o}C. The T4 RNA ligase enzymes used in most current sequencing protocols functions optimally at 25 ^{o}C and are inactivated by a 20-minute incubation at 65 ^{o}C, rendering them functionless at the temperatures we desire.

1.3.1 Thermophiles

Thermophiles are a class of organisms that thrive in temperatures between 41 and 122 ^{o}C. While they are typically found in hot springs, or around deep-sea hydrothermal vents, they can also be found in decaying plant matter. The temperatures these organisms thrive in are not normally conducive to life, with most proteins becoming unstable, or even denaturing, resulting in a loss of function that can cause cell damage or death. This has forced these organisms to evolve proteins that remain stable and functional when others would denature. These organisms are therefore of particular interest to those looking for proteins that retain their function at higher-than-normal temperatures (Dalmaso *et al.*, 2015; Egamberdieva *et al.*, 2018; Holden, 2019; Stetter, 1999; Verma *et al.*, 2015). The following sections illustrate the impact that thermophiles have had on molecular biology, and the potential they have in microRNA sequencing protocols.

1.3.1.1 Thermus aquaticus

Thermus aquaticus, *Taq*, was first isolated from the Lower Geyser Basin of Yellowstone National Park (Brock & Freeze, 1969) and can survive in temperatures as high as 80 ^{o}C. It has since been found in similar thermal habitats around the world. The DNA polymerase from this bacterium was first isolated in 1976 (Chien *et al.*, 1976), and would go on to revolutionise the way we conducted PCR reactions. In the original protocols for PCR,

Escherichia coli (*E. coli*) DNA polymerase was utilised to amplify target sequences. The low melting temperature of this enzyme resulted in its denaturation every PCR cycle, meaning fresh polymerase had to be added at the beginning of each extension phase. The stability of the *Taq* polymerase at temperatures above the melting temperature of DNA allowed for PCR reactions to be carried out with a single initial addition of polymerase enzyme (Saiki *et al.*, 1988). While this reduced the total amount of enzyme required to carry out PCR, it more importantly allowed thermocyclers to be used to automate PCR cycles. This resulted in a massive efficiency gain, revolutionising the way in which PCRs were conducted. Thus, *Taq* polymerase became one of the first commercially produced thermophilic proteins.

1.3.1.2 *Methanobacterium thermoautotrophicum*

Methanobacterium thermoautotrophicum, *Mth*, is a thermostable, anaerobic, methane-producing archaeon which may be found in the anaerobic sediment under ponds and marshes, thriving in temperatures as high as 65 ^{o}C. This archaeon was first isolated in 1972 (Zeikus & Wolee, 1972). The RNA ligase it produces is currently utilised in the adenylation of oligonucleotide sequences. While this thermostable ligase has been shown to function at higher temperatures than those normally used in microRNA sequencing protocols, it does not appear to have been utilised for this purpose. While the exact reason for this is uncertain, it is may be due to the functional temperature of the enzyme not being high enough to ensure the complete removal of microRNA tertiary structures.

1.3.1.3 *Pyrococcus furiosus*

Pyrococcus furiosus, *Pfu*, is an aquatic, anaerobic, hyperthermophile archaeon first isolated from a hydrothermal vent near Vulcano Island, Italy, in 1986 (Fiala & Stetter, 1986). Sporting an optimal growth temperature of 100 ^{o}C, this organism produces remarkably stable proteins, capable of withstanding temperatures well above that which is required to denature almost all nucleotide and amino acid sequences. It is this property that has drawn the attention of our research team. Our hope was that the RNA ligase this organism produces could be modified such that it could be utilised in a much higher temperature microRNA sequencing protocol than any current RNA ligases. During the initial research leading into this study, Dr Tifany Oulavallickal optimised the protocol for expressing recombinant *Pfu* RNA ligase in *E. coli* cells, and purifying the protein using hexahistidine-tag (His_6-tag) metal affinity chromatography. Protein successfully purified by Dr Oulavallickal was utilised in several activity assays using a variety of RNA and DNA substrates. The results of these assays provided the base upon which my research was carried out.

1.4 Protein engineering

The aim of this thesis has been to build upon the work of Dr Oulavallickal, engineering the *Pfu* RNA ligase such that it is better suited for use in microRNA sequencing protocols. Protein engineering was originally inspired by the process of evolution. Driven by natural selection, evolution has produced countless highly specialised enzymes, each capable of catalysing unique cellular reactions with incredible precision and efficiency. The activity of these enzymes, however, is often dependent on the internal environment of the organism in which it evolved, with small variations from these conditions often resulting in a loss of specificity, precision, or efficiency. Enzymes are often only desirable when they exclusively, and efficiently, catalyse the production of a specific product. Producing low concentrations of product, or unnecessary by-products, may effect the viability of an enzyme for use in any extracellular application. By manipulating (engineering) protein sequences at the gene level, scientists are often able to alter an enzyme's properties. This process of enzyme engineering can be used to produce superior or desirable enzymes for any given number of industrial purposes. Engineered enzymes can boast numerous improved functional characteristics over their wild type counterparts, including catalytic activity, pH optimum, and protein stability.

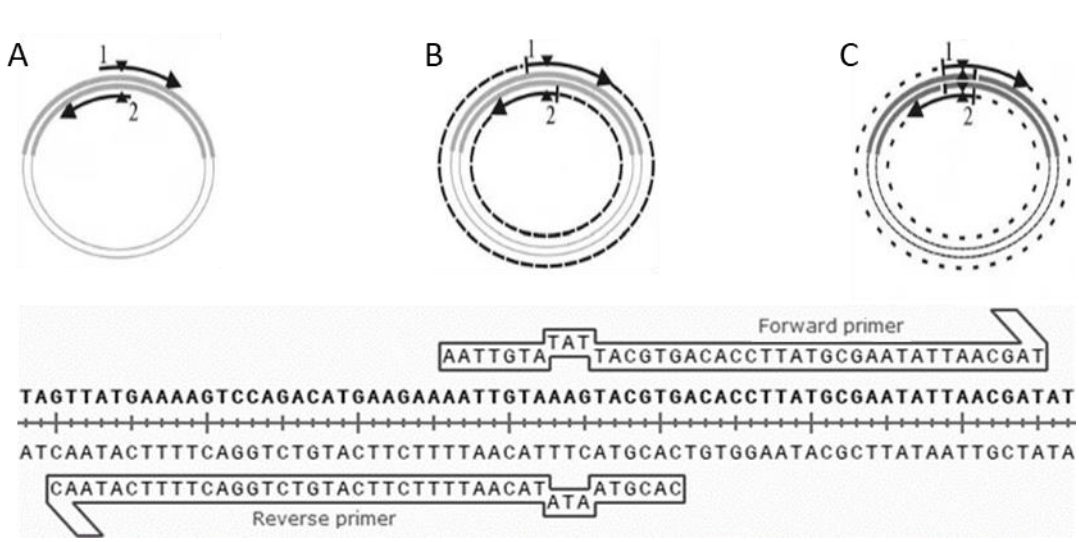

Figure 1-4 Site directed mutagenesis

This figure displays a method by which site-directed mutagenesis can be utilised to introduce mutations at specific locations within a gene. First, a primer containing the desired site-specific mutation is annealed to the target sequence (A). As this mutation needs to be introduced to both strands of the DNA sequence, primers are required for each strand. The two primers therefore overlap, sharing some sequence complementarity. A high-fidelity polymerase is then utilised to amplify the whole plasmid using the designed primers (B). This results in the polymerase incorporating the mutation into the amplified sequence. The mutant amplified product then becomes the template for concurrent rounds of amplification (C). The sequence displayed at the bottom of the figure provides an example of the primers annealed to opposite strands of DNA, with the small section of sequence complementarity containing the desired mutant sequence.

Directed evolution and rational design are the two main strategies used to conduct enzyme engineering, however, semi rational design, a hybrid strategy, can also be utilised.

1.4.1 Rational design

Rational design utilises pre-existing knowledge of protein structure and function to identify specific amino acid deletions, insertions, or substitutions which are likely to improve protein function. Despite decades of research focused on understanding the link between protein structure and function, it is still impossible to accurately predict the effect that a deletion, insertion, or substitution will have on the function of the protein. While this may currently limit the potential applications and impact of rational design, it is still an incredibly useful tool for identifying targets of interest and manipulating protein function (Barrozo *et al.*, 2012; Dombkowski, 2003; Gu *et al.*, 2016; Lutz, 2010).

1.4.1.1 Site directed mutagenesis

Modern site directed mutagenesis protocols can introduce specific mutations to target proteins with incredible efficiency and accuracy. Utilising primers with a small sequence overlap, in which the desired mutation lies (Figure 1-4), it is possible to introduce the desired mutation during sequence amplification, be it a deletion, insertion, or substitution (Liu & Naismith, 2008).

1.4.1.2 Modified RNA ligases

Site directed mutagenesis has been utilised to develop improved versions of both T4 RNA ligase2 (Viollet *et al.*, 2011) and the *Mth* RNA ligase utilised in current microRNA sequencing protocols (Zhelkovsky & McReynolds, 2012). As previously mentioned in Sections 1.1.2 and 1.1.4, the catalytic domains of DNA and RNA ligases contain highly conserved sequence motifs. As such, the biochemical mechanism of the enzymes was well understood.

Using pre-existing knowledge of the enzymes structure, researchers were able to identify the specific lysine residue in T4 RNA ligase2 responsible for catalysing the adenylation reaction (Shenmin *et al.*, 2003). Substitution of this lysine for an alanine residue proved to eliminate the adenylation activity of the enzyme without inhibiting ligation activity. This work was used to inform the rational design of the *Mth* RNA ligase. Due to the similarity between the conserved motifs in the active site, structural overlays of T4-RNA Ligase II and *Mth* RNA ligase were utilised to identify K97 as the lysine of interest in the *Mth* RNA ligase. A second

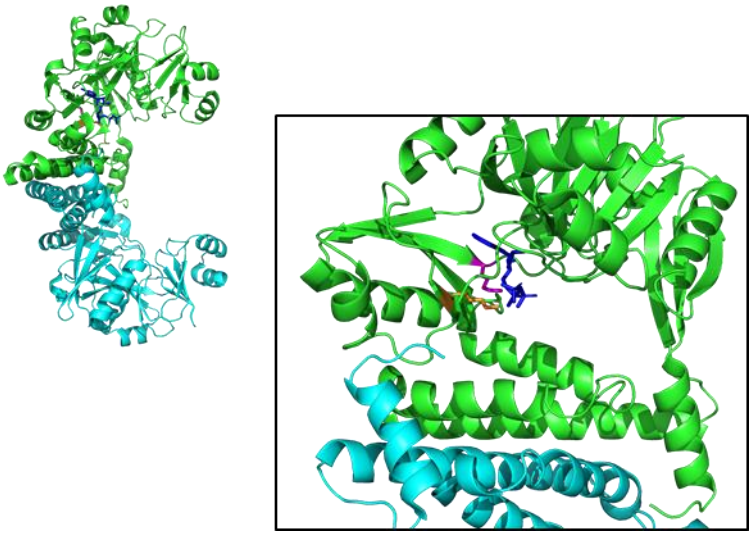

Figure 1-5 Active site of Mth RNA ligase

Displayed in the above figure is the ribbon structure of Mth RNA ligase. Mth RNA ligase is composed of two chains (indicated by green and blue) of identical sequence and structure. Both contain an active site. A close-up of one of the active sites is displayed within the box. The two lysine residues of interest, K97 (pink) and K246 (orange), can be seen near an ATP molecule (dark blue), highlighting that both residues are in positions from which they can interact with the molecule.

```
CLUSTAL 2.1 multiple sequence alignment

Pfu_RNA_ligase      ----MVSSKFKELLYTLGIPEDKVEILEARGGIMEDEFEGIRYLRFKNSVGKLRRGTVLF
Mth_RNA_ligase      MNSMNSDIPFDLIQERTGVPSSRLKVAFARGSLRLLESAGMQALLFKKPLGDLEAGTVIY
                          .  *. :    *:*..::::  ***.:   *  *:: * **:.:*.*. ***::

Pfu_RNA_ligase      EDGT--TVFGFPHIKRIVNLSAGVRKIFKSSEFYVEEKVDGYNVRVVKFKDR-ILGITRG
Mth_RNA_ligase      LGDETEVIRGFPKIRRTLLLSPTIQEHFR-DRVAVEEKMNGYNVRIACLSSGETVALTRG
                     ..   .: ***:*:* : **. ::: *: ... ****::*****:. :..   :.:***

Pfu_RNA_ligase      GFICPYTTERIAEFVP-EEFFKDHKDLVLVGEMAGPESPYLVEGPPYVKEDIQFFLFDIQ
Mth_RNA_ligase      GHVCPFTTRKAQELLDLSEFFREHPDLVICGEMIGRDNPYVSQDYPEVGP-LGFRVFDLR
                    *.:**:**.:  *::  .***::* ***: *** * :.**: :. * *   : * :**::

Pfu_RNA_ligase      DIKTGSSLPVEERLKLAEEYGINHVEVFGRYSYKDID-DLYELIERLSREGREGIVMKSP
Mth_RNA_ligase      EKNTNRPLPVEERRALLDSYGLPNVRLFGVYPIEEAASEVADIIRALGMAGREGVVMKDP
                    : :*. .******  * :.**: :*.:** *. ::   :: ::*. *.  ****:***.*

Pfu_RNA_ligase      DMKK-IVKYVTPYANINDIKIGARVFYELPGGYFTSRISRLAFYIAEKKIRGEELHNLAL
Mth_RNA_ligase      SMEVPPLKYTSSQAHARELAYAFSYPFDFGRPFFFSRVIREGFQAYELDESDDETRERAR
                    .*:   :**.:. *: .::  .    :::   :* **: * .*   * .  .:* :: *

Pfu_RNA_ligase      QLGKALLQPLVEAIHDVTQGDVIAERFRVRVRKIETAYKMVTHFEKLGLEIEIEDIEEIE
Mth_RNA_ligase      RLGEAIIYPMLERIKSISAGEAAYEDTVIDVEDREAAEEFIRHLVRLGVSATLADYRDGR
                    :**:*:: *::* *:.:: *:.  *   : *.. *:* ::: *: :**:.  : * .: .

Pfu_RNA_ligase      GGWRVTFKRVYPEATREIRDLIGGKAFVD
Mth_RNA_ligase      ----ATIRRFYQSTTDRINNYLKGGLY--
                        .*::*.* .:* .*.: : *  :
```

Figure 1-6 RNA ligase sequence alignment

Displayed in the above figure is a sequence alignment between the Pfu RNA ligase and the Mth RNA ligase. Conserved motifs are contained within the grey boxes, while the lysine residues of interest are indicated by the arrows. Both residues are located within conserved motifs. The lysine residues indicated here are the same as those that are highlighted in the Mth RNA ligase active site displayed in Figure 1-5.

lysine, at position 246 in the *Mth* RNA ligase, was also identified as a residue of interest during this research. These residues are displayed in the active site of the Mth RNA ligase in Figure 1-5.

Due to step two of the ligation mechanism being reversible, both Viollet et al. and Zhelkovsky & McReynolds were able to show that substituting the key lysine involved in the adenylation mechanism of the RNA ligase for alanine not only eliminated adenylation activity, but also sequestered de-adenylation activity. This prevented the ligase enzymes from transferring the AMP from adenylated substrates to itself, then to off-target substrate. As the first adapter ligation reaction utilises pre-adenylated substrate in a bid to remove off target activity, this lack of de-adenylation activity is desirable.

Using the structural overlay method utilised by Zhelkovsky and McReynolds, the *Pfu* RNA ligase structure was compared to the *Mth* RNA ligase, identifying two lysine residues that correspond to K97 and K246 in the *Mth* RNA ligase. K92 was identified as the K97 equivalent, while K238 was correlated to K246. A sequence alignment between the Pfu and Mth RNA ligases is displayed in figure 1-6, showing the alignment of these lysine residues. Dr Oulavallickal utilised site-directed mutagenesis to substitute the primary lysine, K92, for alanine. This resulted in the generation of the original K92A variant. While this substitution effectively eliminated adenylation activity, confirming this as the primary residue involved in adenylation, some activity persisted. The remaining activity can be explained in one of two ways. Either K238 can substitute for K92 in its absence, albeit at a significantly reduced efficiency, or the alanine substitution is still allowing limited activity through an unknown mechanism.

1.4.2 Directed evolution

By intentionally introducing random mutations into genes of interest, libraries of random protein variants can be screened for improved activity. Through application of artificial selection pressures, the process of natural selection can be replicated experimentally, albeit on a significantly shorter time scale. As a result, enzymes can be “evolved” to have improved functionality without needing to understand the link between structural and functional changes (Arnold, 2018; Kuchner & Arnold, 1997). This process is known as directed evolution. Typically, directed evolution experiments begin with the creation of a library of mutant gene variants. There are numerous methods by which one can construct a library of

randomly mutated sequences, however error prone PCR is the most common technique utilised for rapid library construction. By amplifying the target gene with low fidelity polymerases, random mutations can be introduced into the gene at a significantly faster than natural rate. The more rounds of PCR conducted, the greater the number of mutations introduced into the gene. Once the library has reached the desired size, a selection pressure or screen is utilised to identify desirable traits, ranging from overall activity to stability of the protein. This process can be repeated numerous times, using the best variants from previous rounds as the template for consecutive rounds of mutagenesis. This allows for multiple desirable traits to be selected and improved upon.

1.4.3 Semi-rational design

Site-saturation mutagenesis is a form of semi-rational design, combining aspects of site-directed mutagenesis and directed evolution. Once a residue of interest has been identified, as per regular site-directed mutagenesis, the residue of interest is substituted for all other amino acids, generating a library of mutants. This combines the benefits of targeting specific residues involved in specific activities associated with site-directed mutagenesis, with the benefit of the selection pressure associated with directed evolution. This allows residues to be optimised.

1.5 Hypothesis and specific aims

The research conducted for this thesis aimed to optimise the mutagenesis of K92 by generating and characterising a K92 site-saturation mutagenesis library. The information gathered during the screening of this library was to be used to inform the rational mutagenesis of K238, such that its role in adenylation could be better characterised. By manipulating these two residues, I hoped to produce a variant of the *Pfu* RNA ligase gene with a more desirable activity profile than the wild type or K92A variant. The variant/s with the most desirable activity would be selected for use in a high temperature microRNA sequencing protocol, with the hope of overcoming the bias observed in current sequencing protocols.

Chapter 2 Methods and Materials

2.1 Bacterial strains and plasmids

2.1.1 Bacterial strain

The *E. coli* strain E. cloni® 10G (Lucigen® Corporation) was used for all experiments. Its genotype is F¯ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *end*A1 *rec*A1 φ80*dlac*ZΔM15 Δ*lac*X74 *ara*D139 Δ(*ara,leu*)7697 *gal*U *gal*K *rps*L (Str^R) *nup*G λ− *ton*A

2.1.2 Plasmids

The plasmid used in this study was pCA24N (Kitagawa *et al*., 2005). This is a high copy number, expression vector in which the gene of interest is under the control of an IPTG inducible T5 promoter. pCA24N can also be used for high levels of protein expression and purification, encoding an N-terminal His_6-tag connected by a short linker. The antibiotic resistance selection marker is chloramphenicol.

2.2 Primers

All primers were purchased from Macrogen Inc. (Seoul, South Korea) and supplied as a lyophilised powder. All stocks were made up to 100 µM using sterile TE buffer. Primers were further diluted to 10 µM with sterile molecular biology grade water to form working stocks. All solutions were stored at -20 °C. (refer to appendix A4 for primer sequences)

2.3 Bacterial cell growth and storage

E. cloni10G cultures were grown at 37 °C and aerated at 200 rpm for liquid culture, or static for solid culture growth, unless otherwise stated. The short-term maintenance and storage of bacteria on agar plates was at 4 °C, whereas long-term storage of liquid bacterial cultures was at a 1:1 ratio with autoclaved 50% (v/v) glycerol and stored at -80 °C indefinitely.

2.3.1 Liquid media

LB broth and AIM were prepared using pre-made powders supplied by ForMedium. LB broth was prepared by dissolving 25 g of LB Miller broth in 1 L of distilled water. AIM was prepared by dissolving 55.85 g of AIM terrific broth base including trace elements, supplemented with 10 mL of glycerol, in 1 L of distilled water.

2.3.2 Solid media

Prior to autoclaving, 1.5% (w/v) agar was added to LB media. Any necessary supplements were added after the agar had cooled to <50 °C post-autoclave. Agar was poured in a laminar flow cabinet as ~20 mL aliquots into sterile 90 mm Petri dishes and left to solidify at room temperature. Storage of agar plates containing supplements was at 4 °C.

2.3.3 Media supplements

The only media supplement used in this study was the antibiotic chloramphenicol. This was prepared by dissolving chloramphenicol powder in 100% analytical grade ethanol to a stock concentration of 34 $mg.mL^{-1}$. Stocks were filter-sterilised with a 0.22 µM syringe filter and stored at -20 °C until required. Antibiotic stock was added to media as a 1:1000 ratio, resulting in a final concentration of 34 $\mu g.mL^{-1}$.

2.4 Site directed mutagenesis

A modified QuikChange® Site-Directed Mutagenesis protocol, described in Section 2.4.3, was used to introduce site specific mutations into the *Pfu* RNA ligase gene – specifically, variations of the lysine's at positions 92 and 238. The plasmid used as a template for these reactions was a recombinant pCA24N plasmid, containing a copy of the *Pfu* RNA ligase gene. This was isolated from pre-existing stored cultures that contained the recombinant plasmid.

2.4.1 Plasmid DNA isolation

Plasmid DNA was prepared from 1-2 mL overnight cultures, using the New England Biolabs inc. Monarch Mini-Prep Kit according to the manufacturer's instructions.

2.4.2 DNA quantification

The purity and concentration of DNA was measured using a NanoPhotometer® NP80 (Implen Inc.; Westlake Village, CA, USA), according to the manufacturer's instructions.

2.4.3 Whole plasmid PCR (WP-PCR)

All WP-PCR reactions were performed in a Labnet Thermal Cycler (ThermoFisher; Waltham, MA, USA). WP-PCR amplifications utilised Q5® High-Fidelity DNA polymerase in a final volume of 25 µL. QuikChange® Site-Directed Mutagenesis, modified according to the protocol of Liu & Naismith et al. as outlined in Table 2-1, was utilised to introduce the desired mutations into the target gene. All reactions were set up using a PCR master-mix

(excluding template DNA), and aliquoted proportionately into PCR tubes. Standard PCR parameters were optimised as required for each combination of primers and genes.

The pCA24N plasmid encoding the K92A variant of the *Pfu* ligase gene, provided by Dr Tifany Oulavallickal, was used as a template for the generation of 17 of 18 other K92 variants, while K92Q was used as the template for the 18th. For generation of the K238 variants, the wild type gene sequence was used, while K92A and K92Y were used as templates for generation of double mutants.

2.4.3.1 Agarose gel electrophoresis

WP-PCR products were analysed using 1% agarose gels made with 1× TAE buffer (40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA, pH 8.0). 5 µL of each PCR reaction product was added to 1 µL of 6× KAPA DNA loading dye before being loaded onto the gel. The gel was then covered in 1× TAE buffer and run using a Cleaver Scientific Power Pro Power Supply at 100-110 V for 25-30 minutes to resolve a 5.6 kb band indicative of the amplified, linear plasmid.

Table 2-1 Composition and cycling parameters for QuikChange® PCR reactions

Component	Volume per 25 µL	Final Concentration
5× Q5 Reaction Buffer	5 µL	1×
5× Q5 High GC Enhancer	5 µL	1×
10 mM dNTPs	0.5 µL	200 µM
10 µM Forward Primer	1.25 µL	0.5 µM
10 µM Reverse Primer	1.25 µL	0.5 µM
Q5 High-Fidelity DNA Polymerase	0.25 µL	~0.02 $U.\mu L^{-1}$
Template DNA (10 $ng.\mu L^{-1}$)	0.5 µL	0.2 $ng.\mu L^{-1}$
Sterile molecular biology grade H_2O	11.25 µL or to 25 µL	

Step		Temp (°C)	Time
Denaturation		98	30 seconds
Amplification (×16)	*Melt*	98	10 seconds
	Anneal	T_a*	30 seconds
	Extend	72	30 $sec.kb^{-1}$
Final extension		72	10 minutes
Hold		4	∞

* The annealing temperature (Ta) was altered depending on the melting temperature (T_m) of the primers. The T_a used was 5 °C lower than the lowest T_m value of the primer(s)

For staining, the gel was placed in an ethidium bromide bath (concentration 0.1 µg.mL^{-1}) for 15-20 minutes. KAPA universal DNA ladder was used as a DNA standard and the bands were visualised and imaged under UV light on a UVITEC Cambridge Gel Documentation System.

Successfully amplified product was then restriction enzyme digest treated with DpnI, before being cleaned up by means of a PCR product clean up kit, as described in the following sections.

2.4.3.2 Restriction enzyme digests

All WP-PCR products were treated with 0.5 µL (500 units) of DpnI (New England BioLabs inc.) in restriction digest buffer to degrade leftover template DNA. This was incubated at 37 °C for a minimum of 4 hours, before heat inactivation at 80 °C for 20 min. DpnI treated product was typically cleaned by means of a PCR DNA clean up kit, before being stored at 4 °C.

2.4.3.3 PCR product clean up

Product clean-up was done using the Monarch PCR DNA clean up kit, supplied by New England BioLabs inc., according to the manufacturer's instructions.

2.5 Bacterial cell transformation and electroporation

2.5.1 Preparation of electrocompetent cells

E. cloni® 10G cultures were inoculated from a glycerol stock and grown in LB media overnight at 37°C, with shaking at 200 rpm. A 1:100 dilution of overnight culture was used to inoculate fresh LB media and grown at 37°C, 200 rpm until an optical density at 600 nm (OD_{600}) of 0.4-0.5 was reached. Cells were transferred to 50 mL conical tubes and harvested by centrifugation at 3190*g* for 15 min at 4°C. All subsequent steps were performed on ice and centrifuged at 3190*g*, 4°C. Cells were gently resuspended in a half culture volume of ice-cold, sterile de-ionised water and pelleted by centrifugation for 10 min. The cell pellet was resuspended in a quarter culture volume of ice-cold, sterile 10% glycerol (v/v) and pelleted again for 10 min. The resulting pellet was resuspended in 50 mL of 10% glycerol and collected by centrifugation for 15 min. The supernatant was completely discarded, and the pellet resuspended in the minimum volume of 10% glycerol required to form a liquid suspension. 50 µL aliquots were distributed into pre-chilled 1.5 mL microcentrifuge tubes on

ice and snap-frozen using a metal tube block cooled to -80°C. Cells were stored at -80 °C until needed.

2.5.2 Transformation of electrocompetent cells

Stored electrocompetent cell aliquots were thawed on ice. Once fully thawed, 1 µL of the desired plasmid DNA was added and gently mixed. Samples were then transferred to 2 mm gap electroporation cuvettes and electroporated at 2.5 kV, 25 µF, 100 Ω. Immediately following electroporation, 500 µL of fresh LB broth was added to the cuvette with the resulting contents being transferred to a sterile 15 mL conical tube. Cell recovery was carried out at 37 °C, 200 rpm for approximately 60 minutes. Between 50 µL and 100 µL of undiluted cells were plated on LB agar, supplemented with 34 $\mu g.mL^{-1}$ chloramphenicol, and incubated at 37 °C overnight.

2.6 Mutant sequence confirmation

Colonies growing on the antibiotic supplemented LB agar plates (mentioned in Section 2.3.2) were picked and resuspended in 10 µL of sterile molecular biology grade water, with a fraction of this being used for the DNA template in a colony PCR. Successfully amplified product was sent for sequencing to confirm the mutant sequence. What remained of the resuspension was utilised in the generation of a glycerol stock for long term storage.

2.6.1 Colony PCR

All colony PCRs were performed in a Labnet Thermal Cycler (ThermoFisher; Waltham, MA, USA). The Quick-Load® Taq 2× Master Mix Protocol (New England BioLabs inc.) was utilised for all colony PCRs and was carried out according to the manufacturers protocol, with minor modifications, as outlined in Table 2-2.

An aliquot of undiluted colony suspension (1 µL) was used for the template DNA. Primers pCA24N.for and pCA24N_rev2 were designed to flank the gene inserted into a recombinant pCA24N plasmid. As a result, amplified product should only be the gene and a small amount of flanking sequence, a total sequence length of 1.3kb.

Of the remaining 9 µL, 4 µL was used to seed a 400 µL overnight culture in LB broth, supplemented with 34 $\mu g.mL^{-1}$ chloramphenicol. To this culture, 400 µL of autoclaved 50% glycerol was added, such that it could be stored at -80 °C until the sequence could be confirmed.

Table 2-2 Composition and cycling parameters for colony PCRs

Component	***Volume per 20 µL***	***Final Concentration***
2× Quick-Load® PCR Buffer	10 µL	1×
10 µM pCA24N.for Primer	1 µL	0.5 µM
10 µM pCA24N.rev_2 Primer	1 µL	0.5 µM
Template DNA	1 µL	
Sterile molecular biology grade H_2O	7 µL or up to 20 µL	

Step		***Temp (°C)***	***Time***
Denaturation		95	5 minutes
Amplification (×35)	*Melt*	95	20 seconds
	Anneal	T_a*	30 seconds
	Extend	72	30 sec.kb^{-1}
Final extension		72	5 minutes
Hold		4	∞

2.6.2 Agarose gel electrophoresis

Colony PCR products were visualised using the same 1% agarose gel protocol described in Section 2.4.3.1. Due to the composition of the Quick-Load® Master Mix, addition of loading dye was not necessary. As such, 3 µL of each reaction product was loaded directly onto the gel. The gel was then covered in 1× TAE buffer and run using a Bio-rad PowerPac™ Basic Power Supply at 100-110 V for 25-30 minutes to resolve a 1.3 kb band indicative of the amplified product. PCR tubes containing the successfully amplified product were sealed and prepared for transport to the sequencing facility of choice.

2.6.3 DNA sequencing

Colony PCR product was sent to Macrogen Inc. (Seoul, South Korea) for both purification and sequencing. Primers pCA24N.for and pCA24N.rev_2 were utilised for these sequencing reactions.

Sequencing results were visualised using the software program Snap Gene (version 5.2.4), allowing comparisons to be easily made against a reference sequence. This simplified the process of identifying the presence of mutations.

2.6.4 Long term storage (glycerol stock)

Once the sequence of a colony was confirmed, temporary glycerol stocks (refer to Section 2.3) were used to seed an identical 400 µL overnight culture. To this culture, 400 µL of autoclaved 50% glycerol was added before being transferred to a cryotube, such that it could be stored at -80 °C indefinitely.

2.7 Protein purification

All proteins to be expressed and assayed were cloned into pCA24N vector containing an N-terminal His_6-tag and transformed into *E. coli* strain E. cloni® 10G for purification via a batch purification protocol (Figure 2-1) utilising Talon® metal affinity resin (Sigma-Aldrich). All centrifugation steps throughout the purification protocol were done at 4 °C unless stated otherwise.

2.7.1 Protein expression

A glycerol stock of E. cloni® 10G cells with the ligase construct of interest was used to inoculate a 5 mL starter culture of chloramphenicol supplemented LB media and incubated overnight at 37 °C, 200 rpm. The starter culture was added to 500 mL Autoinduction Media (AIM) supplemented with chloramphenicol and incubated at 37 °C, 200 rpm. OD_{600} of the culture was periodically measured in a 1 mL cuvette to monitor cell growth, until a reading of between 0.5 - 0.6 was observed. Protein expression was then induced by an overnight incubation at 18 °C, 200 rpm. The cells were harvested by centrifugation in 50 mL conical tubes at 7068 *g*, 4 °C, for 30 min. The supernatant was discarded, and cell pellets were stored at -20 °C until required.

2.7.2 Lysate preparation

Cell culture pellets were resuspended in chilled lysis buffer (50 mM Tris, 200 mM NaCl, pH 7.4) up to 15 mL, followed by the addition of lysozyme to a final concentration of 0.5 $mg.mL^{-1}$, and 100 µL of Protease Inhibitor Cocktail (Sigma-Aldrich). The cell suspension was lysed by sonication on ice with 10 rounds of 10 seconds on, 30 seconds off at a 30% amplitude (Sonics Vibra-cell, 3 mm tip). To remove insoluble cellular debris, samples were transferred to high-speed centrifuge tubes and spun for 20 minutes at 30,000*g* in a Sorvall LYNX 4000 Superspeed Centrifuge (thermo ScientificTM; Waltham, MA, USA). The soluble fraction was collected and filtered through a 0.45 µM syringe-driven filter.

2.7.3 Talon resin batch purification

2.7.3.1 Column preparation

2 mL of Talon® metal affinity resin was washed by addition of 8 mL of Milli-Q® water, followed by centrifugation for 2 minutes at 700*g*. The pellet was resuspended in lysis buffer (50 mM Tris, 200 mM NaCl, pH 7.4) up to 15 mL, and centrifuged for 2 minutes at 700*g*. This step was repeated twice to properly equilibrate the resin.

Sample His_6-tag protein was bound to equilibrated resin by combining the resin and the clarified cell lysate and incubating for 30-90 minutes on a shaker at room temperature. Protein bound resin was isolated by centrifugation for 5 minutes at 700*g*. A 10 µl sample of the supernatant was taken for analyse by means of SDS-PAGE.

The protein bound resin was resuspended in lysis buffer up to 10 mL and incubated at 4 °C with agitation for 10 minutes to wash off unbound protein. The washed resin was isolated by centrifugation for 5 minutes at 700*g*. This step was repeated twice to remove as much unbound protein as possible. A 10 µL sample of the supernatant was taken after the first wash for analysis.

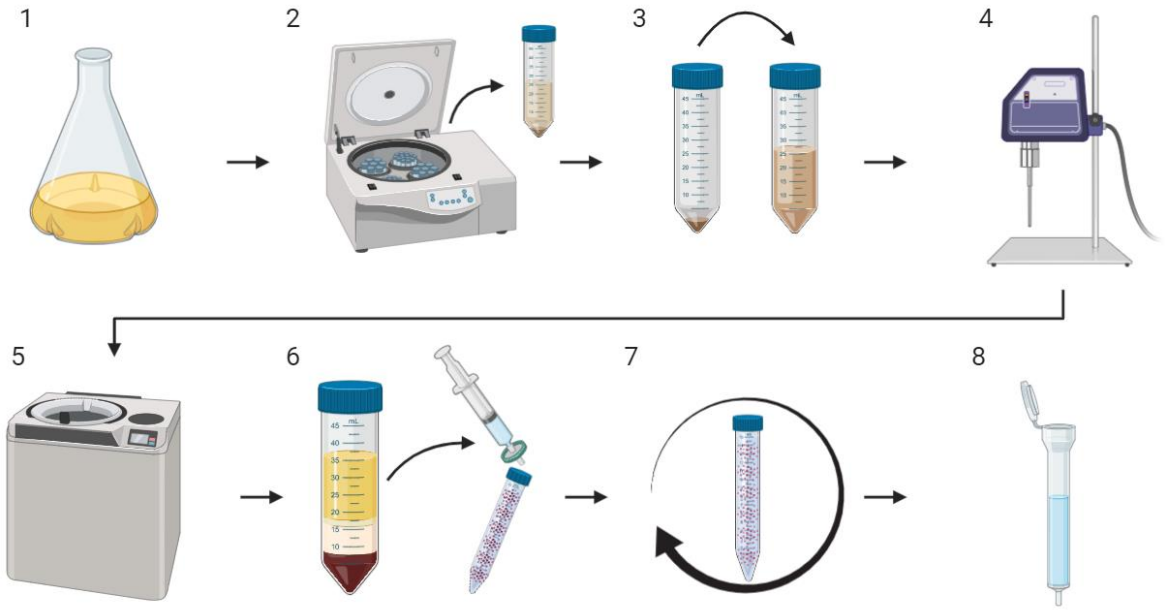

Figure 2-1 Batch purification

Culture was grown in 500 mL of antibiotic supplemented AIM (1). Cells were pelleted by centrifugation (2) with the supernatant being discarded. Collected pellet was resuspended in lysis buffer (3) before sonication (4). Lysate was centrifuged at 15000g (5) to separate soluble protein from insoluble protein and cellular debris. The clarified lysate was syringe filtered, then added to Talon resin (6). Protein was bound to the resin by a 90-minute incubation with agitation at room temperature. Unbound protein was removed, and resin washed before addition to a gravity flow column for protein elution (8)

Washed beads were resuspended in 1 ml lysis buffer and added to a Poly-Prep® Chromatography gravity flow column (Bio-Rad), draining the buffer until it just covered resin bed. The column was then washed a final time with 5 mL of lysis buffer, followed by 5 mL of wash buffer (50 mM Tris, 200 mM NaCl, 5 mM imidazole, pH 7.4).

2.7.3.2 Protein elution

Ligase protein was eluted from the column using 5 mL of elution buffer (50 mM Tris, 200 mM NaCl, 300 mM Imidazole, pH 7.4) and collected as 500 µL samples.

2.7.3.3 Column cleaning / regeneration

Following elution, resin was washed with 10 mL of resin cleaning buffer (20 mM MES, 100 mM NaCl, pH 5.0), followed by 10 mL of Milli-Q® water. This process cleaned and recharged the resin, which could then be stored as a 50% slurry in 20% ethanol. Properly cleaned and recharged resin was used a maximum of three times, but only for the same protein. This ensured that improperly cleaned resin would not contaminate the purified protein with undesirable protein. Resin was therefore utilised for an average of two purifications.

2.7.3.4 Bradford assay

To quickly test for the presence of protein, 2 µL aliquots of fractions eluted from the column were combined with 50 µL of 1× Bradford Protein Dye (Bio-rad; Hercules, CA, USA) in PCR tubes. If protein were present, the solution would turn blue. This was used to determine which samples should be analysed by means of SDS-PAGE.

2.7.3.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels were made according to the recipe outlined in Table 2-3. Protein samples were mixed at a 1:1 ratio with 2× SDS loading dye (150 mM Tris-Cl, pH 6.8; 6% SDS (w/v), 30% glycerol (v/v), 0.03% bromophenol blue (w/v), 300 mM β-mercaptoethanol) and heated for 5 min at 95 °C before being loaded on to the gel. All protein samples were run alongside 5 µL Bio-Rad Protein Ladder 10 – 180 kDa to estimate molecular weight. Gels were run in 1× running buffer (24.9 mM Trisbase, 192 mM glycine, 0.1% SDS (w/v)) at 200 V for 40 min. For visualisation of protein bands, gels were stained using Coomassie Blue Stain (2.5 $g.L^{-1}$ Coomassie Blue Brilliant, 50% methanol (w/v), 10% acetic acid (w/v)) for 20 minutes and destained using destain buffer (10% (w/v) acetic acid, 40% (w/v) methanol) until bands could be visualised and imaged.

Total protein was analysed by running a fraction collected from the sonicated lysate (Section 2.7.2), pre-centrifugation. The composition of soluble and insoluble protein was analysed by collecting fractions of the post-centrifugation supernatant (soluble), and pellet (insoluble). Due to the viscosity of the pellets, a 100× dilution was made such that the sample could be run on a gel. Protein that failed to bind to the column was analysed by taking a fraction from the column flow through, while a fraction of the subsequent column wash was also analysed to identify the protein washed from the column before elution. Protein was eluted in approximately 1000 µL fractions, of which samples were also analysed.

Table 2-3 Recipe for 10× SDS-PAGE gels

Component	*Resolving*	*Stacking*
Sterile de-ionized H_20	23.9 mL	12.7 mL
Tris [a]	13.75 mL	5 mL
10% w/v SDS	550 µL	200 µL
Acrylamide [b]	16.5 mL	2 mL
Aps	275 µL	100 µL
TEMED	55 µL	20 µL
Total	55 mL	20 mL

2.7.3.6 Buffer exchange / protein concentration

To ensure protein stability, the protein was transferred into storage buffer (50 mM Tris, 200 mM NaCl, pH 7.4). This was done using 10 kDa Amicon® Ultra-15 Centrifugal protein concentrators (Millipore).

2.7.3.7 Measurement of protein concentration

Protein concentrations were initially evaluated using a NanoPhotometer® NP80 (Implen Inc.; Westlake Village, CA, USA), with A_1 and $mg.mL^{-1}$ values being recorded. When a more precise measurement was required, a 10 µL aliquot of protein was diluted in 90 µL storage buffer. This was then transferred to a 10 mm rectangular quartz cuvette (Agilent Technologies; Santa Clara; CA, USA) where the absorbance spectrum was scanned from 200 nm to 300 nm at a scan rate of 600 $nm.min^{-1}$ using a Cary 100 UV-Vis Spectrophotometer (Agilent Technologies; Santa Clara; CA, USA). 90 µL storage buffer was utilised as a blank. Protein concentrations were calculated using the measured absorbance (A), a molecular extinction coefficient (ε) of 32320 $m^2.mol^{-1}$, a molar mass (c) of 45934.84 $g.mol^{-1}$, and a pathlength (l) of 1 cm, according to the following equation.

$$mg.mL^{-1} = A / \varepsilon cl$$

The molarity of protein samples was calculated using protein concentration ($mg.mL^{-1}$), as determined by either the NanoPhotometer®, or a Cary 100 UV-Vis Spectrophotometer (Agilent Technologies; Santa Clara; CA, USA), and molecular weight ($g.mol^{-1}$), according to the following equation.

$$\mu M = mg.\, mL^{-1} \div g.\, mol^{-1}$$

2.7.3.8 Storage of ligase proteins

All protein aliquots were stored at the eluted concentration or diluted to 50 µM using storage buffer before being stored at 4 °C. Due to the extreme thermostability of the protein, lower storage temperatures were not required.

2.8 Protein activity experiments

Endpoint TBE-urea PAGE assays were employed to determine the adenylation and ligation activity patterns of each variant. These assays involved analysing the products generated during ligation reactions by means of Urea-PAGE, a technique frequently used to analyse RNA and DNA fragments. All the following reactions were carried out with strict sterile technique to minimise potential RNAse contamination.

2.8.1 Ligation reaction

Ligation reactions were carried out using 200 µM of the desired ligase variant, and 5 µM of each substrate. This meant single substrate assays contained 5 µM total substrate, while dual substrate assays contained 10 µM. All reactions were carried at 70 °C. Reactions were stopped after 90 minutes by addition of 2× Novex® TBE-Urea Sample Buffer (Invitogen) at a 1:1 ratio, then heating at 95 °C for 5 minutes.

2.8.1.1 Pre-adenylation reaction

Reagents and buffers used for this reaction can be seen in Table 2-4. Pre-adenylation reactions were carried out at 65 °C, for 60 minutes. Reactions were stopped by heating at 85 °C for 5 minutes. Adenylated substrate was precipitated out of solution by addition of 2.5 reaction volumes of 100% ethanol and 1/3 reaction volumes of 10 M ammonium acetate, before being left on ice for 30 minutes. Precipitate was pelleted by centrifugation at 15,000g, 4 °C for 20 minutes. The pellet was then washed thoroughly with 70% ethanol, before being centrifuged for another 20 minutes. A pipette was used to remove as much ethanol as possible before leaving the pellet to air dry. Pellets were stored at -80 °C until needed, then resuspended in sterile molecular biology grade water to a final concentration of 20 µM.

Table 2-4 Pre-adenylation reaction

Component	Volume	Final Concentration
RNAse Inhibitor (40 U.µL^{-1})	0.5 µL	20 U
NEBuffer™ 1 (10×)	2 µL	1×
ATP (10 mM / 2 mM)*	2 µL	500 / 100 µM
Substrate (100 pmol)	1 µL	5 µM
Mth RNA ligase (80 µM.µL^{-1})	2.5 µL	200 µM
Sterile molecular biology grade H_20	Up to 20 µL (12 µL)	-

2.8.1.2 Initial ligation activity screen

A basic RNA ligation reaction, utilising pre-adenylated Oligo1 (refer to appendix A5) as the sole substrate, was employed to assess ligation activity (see Table 2-5). A second RNA ligation reaction, done in the presence of cofactor ATP while utilising non-adenylated Oligo1 as the substrate, was employed to determine whether a variant was capable of both adenylation and ligation activity. Variants of interest were those that lacked adenylation activity but retained ligation activity.

Table 2-5 Single substrate assay

Component	Ligation activity	Adenylation activity	Final Concentration
RNAse Inhibitor (40 U.µL^{-1})	0.5 µL	0.5 µL	20 U
NEBuffer™ 1 (10×)	2 µL	2 µL	1×
ATP (300 µM)	0 µL	1 µL	0 / 15 µM
Oligo1 (100 µM)	0 µL	1 µL	0 / 5 µM
Pre-adenylated Oligo1 (20 µM)	5 µL	0 µL	5 / 0 µM
Enzyme (50 µM.µL^{-1})	4 µL	4 µL	200 µM
Sterile molecular biology grade H_20	Up to 20 µL (8.5 µL)	Up to 20µL (11.5 µL)	-

2.8.1.3 SR1 RNA ligation assay

To determine the ability of a ligase to join a pre-adenylated 3′ modified RNA sequence to a microRNA substrate, ligation reactions were carried out utilising pre-adenylated 3′ amino modified SR1 RNA (refer to appendix A5) and non-adenylated Oligo1 RNA substrates (refer to Table 2-6). No cofactor was added. The 3′ amino modifier (3AmMO) acted to block the 3′ end of the SR1 RNA oligonucleotides, forcing the desired ligation to occur. Variants of interest were those that were able to produce the most microRNA-SR1 RNA product.

Table 2-6 Dual substrate assays

Component	Volume	Final Concentration
RNAse Inhibitor (40 U.µL^{-1})	0.5 µL	20 U
NEBuffer™ 1 (10×)	2 µL	1×
ATP (300 µM)	1 µL	15 µM
Oligo1 (100 µM)	1 µL	5 µM
SR1 oligonucleotide (100 or 20 µM)*	1 µL / 5 µL	5 µM
Enzyme (50 µM.µL^{-1})	4 µL	200 µM
Sterile Molecular Biology Grade H_20	Up to 20 µL	-

* 100 µM untreated or 20 µM adenylated

2.8.1.4 SR1 DNA ligation assay

To determine the ability of each ligase to join a pre-adenylated, 3′ blocked DNA sequence to a microRNA substrate, ligation reactions were carried out utilising pre-adenylated 3′ amino modified SR1 DNA (refer to appendix A5) and non-adenylated Oligo1 RNA substrates (refer to Table 2-5). No cofactor was added. The 3′ amino modifier (3AmMO) acted to block the 3′ end of the SR1 DNA oligonucleotides, forcing the desired ligation to occur. Variants of interest were those that were able to produce a microRNA-SR1 DNA product.

2.8.1.5 5′ blocked microRNA assays

To ensure that the ligated product observed in both the SR1 RNA and DNA ligation reactions was in fact Oligo1-SR1, the reactions described in Sections 2.8.1.3 and 2.8.1.4 were repeated with a modified Oligo1 RNA, lacking a 5′ phosphate.

2.8.2 Urea-PAGE

Urea-based acrylamide gels (see Table 2-7) were utilised to analyse ligation reaction results. Gels were pre-run in 1× TBE buffer (0.13 M Tris-base, 45 mM Boric acid, 2.5 mM EDTA, pH ~8.3) at 150 V for 30 minutes before flushing the wells with running buffer and loading the samples. Gels were then run at 200 V for ~120 minutes, then soaked in a 50 mL bath of 1× TBE buffer with 5 µL of Sybr® Gold to stain. Gels were then visualised under UV light.

Table 2-7 Recipe for 1× TBE urea PAGE gel

Component	Volume / Weight
5× TBE	0.75 mL
Urea	3.15 g
40% polyacrylamide	3.75 mL
Sterile de-ionised H_20	Up to 7.2 mL
-	-
*10% APS**	30 µL
*TEMED**	6 µL

*Other components were combined and heated to 60 °C to dissolve Urea prior to addition of APS and TEMED

Chapter 3 Characterisation of K92 *Pfu* RNA Ligase Variants

3.1 Introduction

To overcome the bias present in modern microRNA sequencing protocols, our aim has been to develop a method for high temperature microRNA sequencing. To do this, an RNA ligase capable of functioning in temperatures exceeding 70 ^{o}C is required. As mentioned in Section 1.3.1.3, the RNA ligase isolated from *Pfu* was identified as a promising candidate. As with all other RNA ligases utilised in microRNA sequencing, it is necessary that the adenylation, and therefore de-adenylation activity of this enzyme be inhibited to ensure unwanted ligation products are minimised.

Previous research (Zhelkovsky & McReynolds, 2012) identified that substitution of K97 in the *Mth* RNA ligase for an alanine effectively eliminated the adenylation activity of the enzyme. Due to high levels of similarity between the *Mth* and *Pfu* RNA ligase sequences, sequence alignments were able to identify K92 as the probable *Pfu* equivalent. K92 was subsequently substituted for an alanine (K92A) by Post-Doctoral Fellow Tifany Oulavallickal. While initial testing showed that the K92A variant retained ligation activity at temperatures exceeding 70 ^{o}C, it also identified the variant had retained limited adenylation activity. While the observed decrease in adenylation was promising, it was not the total elimination of adenylation activity that was hoped for. To determine whether other amino acid substitutions at position 92 would aid in further reducing de-adenylation activity, while potentially improving ligation activity, I carried out site saturation mutagenesis on position 92 of the *Pfu* RNA ligase. Over 2019, in preparation for this thesis, I was able to successfully mutate K92 to 17 of the remaining 18 amino acids, failing only to produce K92L.

This chapter aimed to complete and characterise the library of K92 variants, subsequently identifying the amino acid substitutions conferring the greatest benefit to protein activity. This information would be used to inform future mutations.

3.2 Results

Characterising the ligation activity of an enzyme is achieved using relevant activity assays. In these assays, the purified enzyme is combined with various substrates under a range of conditions, allowing the ability of the enzyme to generate desired product to be determined.

At the beginning of my research year, the only K92 variant missing from my library was K92L. As such, generating K92L, and therefore completing the library of variants, was the first milestone of this thesis. This was followed by a massively parallel batch purification of the 20 proteins, such that all activity assays could be carried out simultaneously with freshly purified proteins.

3.2.1 Generation of *Pfu*Rnl(K92L)

A sample of the plasmid encoding the K92M variant was successfully prepared and used as the template for introducing the K92L mutation. PCR reactions were then carried out according to a modified QuikChange® mutagenesis protocol (refer to Section 2.4.3), using K92L-specific primers to amplify the K92M template. Plasmid miniprep and QuikChange® PCR products were analysed on 1% agarose gels by means of electrophoresis. (Figure 3-1A and B)

Successfully amplified product was transformed, prepared for culture, and plated according to the protocols described in Section 2.6. Successful uptake of the desired plasmid was indicated by the growth of numerous colonies. Three colonies were picked and resuspended for use in colony PCRs. All three reactions successfully amplified the target sequence (Figure 3-1C), with two of three samples being identified as the K92L variant. This concluded generation of the K92 site saturation mutagenesis library.

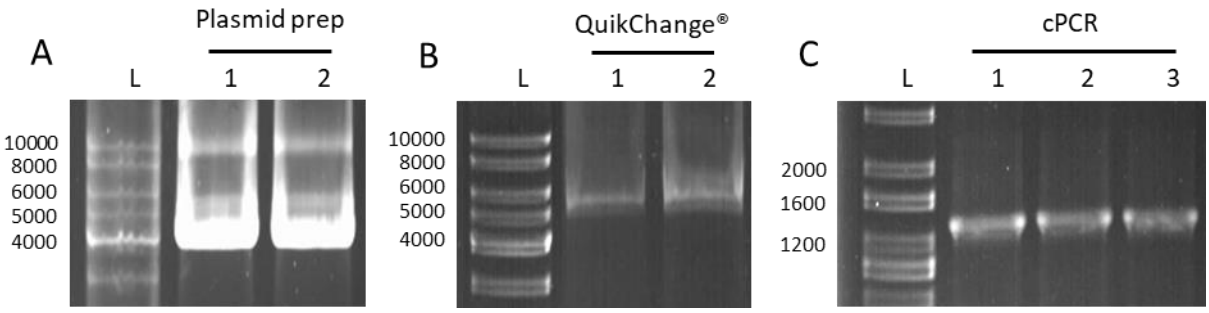

Figure 3-1 Generation of K92L

Lanes identified by an L contain the ladder. Numbered lanes contain separate reactions. Circular DNA sequences travel more freely through agarose gels than their linear counterparts, making them appear smaller in size. As plasmid isolated during the miniprep (A) is circular, the 5633 bp template sequence appears closer in size to 4000 bp. QuikChange® mutagenesis PCR amplifies the whole plasmid, generating a linear product from the circular template. In B, two product bands of around 5500 bp can be seen, indicating the plasmid sequence was successfully amplified. To confirm the sequence of mutants, colony PCR only needs to amplify the target gene (1146 bp), and a small amount of flanking DNA. Amplifying the full plasmid would be unnecessary. In C, presence of an approximately 1300 bp product indicates the desired sequence was successfully amplified.

3.2.2 Protein purification

To ensure comparisons between variants were accurate, all twenty proteins were purified within the same 36-hour window to minimise potential effects of protein degradation. A batch purification protocol (Figure 2-1) was utilised so that multiple purifications could be carried out simultaneously (refer to Section 2.7.3) Fractions collected at each stage of the purification protocol were analysed by means of SDS-PAGE (Figure 3-2). While the target protein is approximately 44 kDa, previous purifications identified the desired band appearing closer to the 37 kDa refence band than expected.

Two significant bands of approximately 40 kDa (seen in Figure 3-2) were observed in the elution fractions of all variants. This indicates that target protein, as well as a potential contaminant, was eluted from the column. Faint bands of similar sizes can be observed in all pre-elution fractions, indicating limited protein loss during purification. Eluted fractions were not as pure as they could have been, presumably due to the batch purification protocol. The larger of the two bands observed in the elution fractions is likely a contaminant. While the presence of this contaminant is not ideal, standardised samples (Figure 3-3) do not appear to contain significant proportions of the same contaminant. As the contaminant band appears to be of little significance in the standardised samples, it is unlikely to have a significant effect on the activity of the final protein sample.

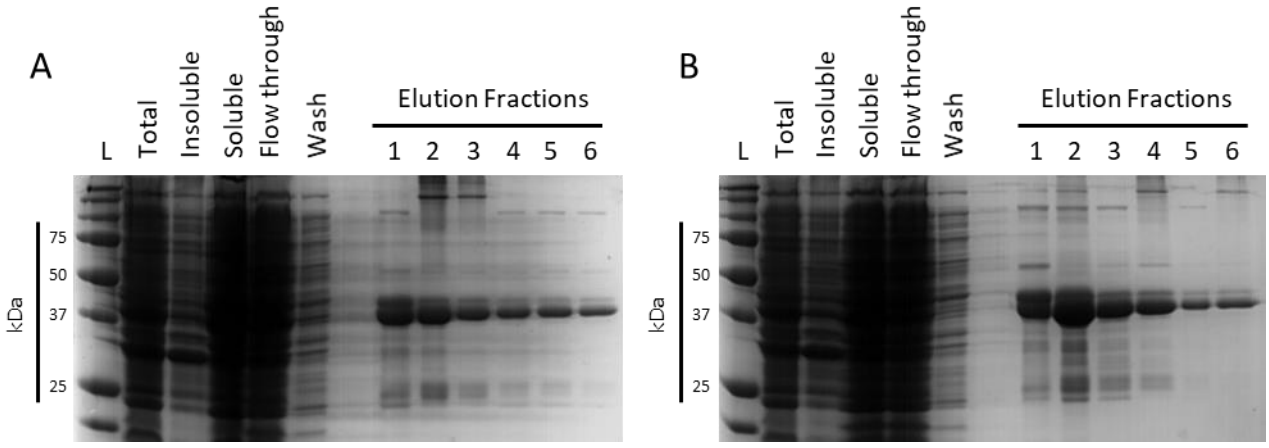

Figure 3-2 K92 variant purifications

Analysed samples include fractions collected of pre-centrifugation lysate (total protein), as well as of post-centrifugation supernatant (soluble protein) and pellet (insoluble protein). Also included are fractions of protein that failed to bind to the column (flow through) and protein washed from the column pre-elution (wash). Protein elution fractions were also analysed to ensure only fractions with reasonable purity were combined for buffer exchanging. Despite target protein being approximately 44 kDa, it is regularly identified around the 40 kDa mark. The second band of significance identified in these elution fractions is likely a contaminant.

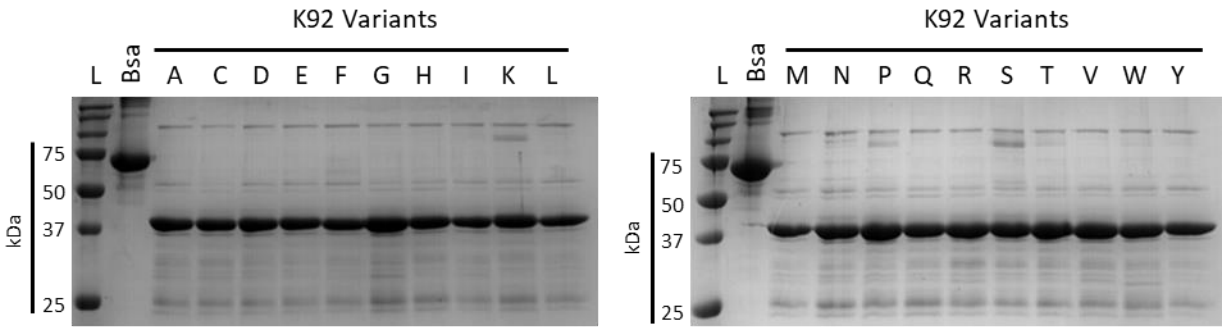

Figure 3-3 Standardised protein sample comparison

Aliquots of all proteins were diluted to 50 µM stock concentration and run side by side on the above gels. A 50 µM Bovine serum albumin (Bsa) control was utilised for reference. The bands produced by each variant were determined to be of similar enough intensity for variations to be attributed to minor dilution and loading errors. While all variants display a less intense band than the Bsa control, it was not large enough to be of concern.

Purifications of all 20 proteins yielded between 10 and 25 grams of protein per litre of inoculated media (10-25 $mg.mL^{-1}$). Further optimisations to the batch purification protocol, or use of a different protocol (i.e., HisTrap™ chromatography using a Cytiva ÄKTA pure purification system), may produce a higher yield of more pure protein.

3.2.3 Activity characterisation

Numerous assays were run with a variety of substrate combinations, such that a comprehensive ligation activity profile could be generated. Single substrate assays were utilised as a basic screen to identify variants with the most promising adenylation and ligation activities. Variants expressing the most promising activity profiles were then used in directional, dual substrate assays, designed to represent the ligation reaction in the microRNA sequencing protocol more accurately.

3.2.3.1 Initial ligation activity screen

The basic ligation activity of each variant was tested by means of a single substrate activity assay, with each variant being added to a reaction containing only pre-adenylated Oligo1 substrate (an 18-nt RNA sequence - refer to appendix A5). Variants able to ligate the provided Oligo1 molecules together, therefore creating dimer-sized products, were determined to have retained their viability. To directly compare the activity of all twenty ligase variants, all activity assays were carried out in parallel. The results of these assays were run on three identical gels under the same conditions to ensure compatibility (Figure 3-4A).

While most variants displayed some level of ligation activity, variants K92 A, G, S, T and Y all stood out as having the most significant ligated product bands. Variant K92C also expressed an interesting pattern of activity. While its ligated product band was not of particular significance, it displayed significant de-adenylation activity, with almost no identifiable pre-adenylated Oligo1 band. This result is reminiscent of that of the wild type enzyme, resulting in the variant being identified as one of possible interest. While the gels displayed in Figure 3-4A show the activity patterns typically displayed by each of the 20 proteins, some replicates indicated variants K92 D and E were of potential interest. To ensure the best variants were selected, the eight variants identified above, K92 A, C, D, E, G, S, T, and Y, plus the wild type, were used in a further round of replicate assays and run on a single gel to remove any gel variability from the equation. This result can be seen in Figure 3-4B. A result of particular note is that of K92Y. While expressing a similar level of ligation activity to that of the other frontrunners, it displayed little to no de-adenylation or circularisation activity, both of which are undesirable due to reducing the pool of available adenylated substrate.

A second assay was done to determine the adenylation activity of the eight variants. This reaction was carried out in the presence of cofactor ATP with a non-adenylated Oligo1 substrate. For ligation to occur, variants needed to adenylate the Oligo1 substrate. Without an adenylated substrate, a ligase cannot catalyse formation of a phosphodiester bond to join the two substrates. Figure 3-4C displays the results of these assays. As can be seen, all variants display a reduction in ligated product, indicating that the adenylation activity has been successfully diminished by the substitution. While this result is positive, substitution of K92 should have resulted in a total inhibition of adenylation activity as this lysine residue is adenylated in a key step of the reaction. The generation of ligated product by variants K92 A, C, S and T therefore indicates the potential involvement of a second active-site lysine in adenylation, and potentially ligation.

These assays reaffirmed the best variants to carry forward were K92 A, G, S, T and Y, each of which displayed a significant improvement in activity over that of the wild type (Figure 3-4B), while displaying a markedly decreased adenylation activity. The total lack of adenylation activity by variants K92 G and Y, alongside strong adenylation activity, highlight these variants as potential frontrunners.

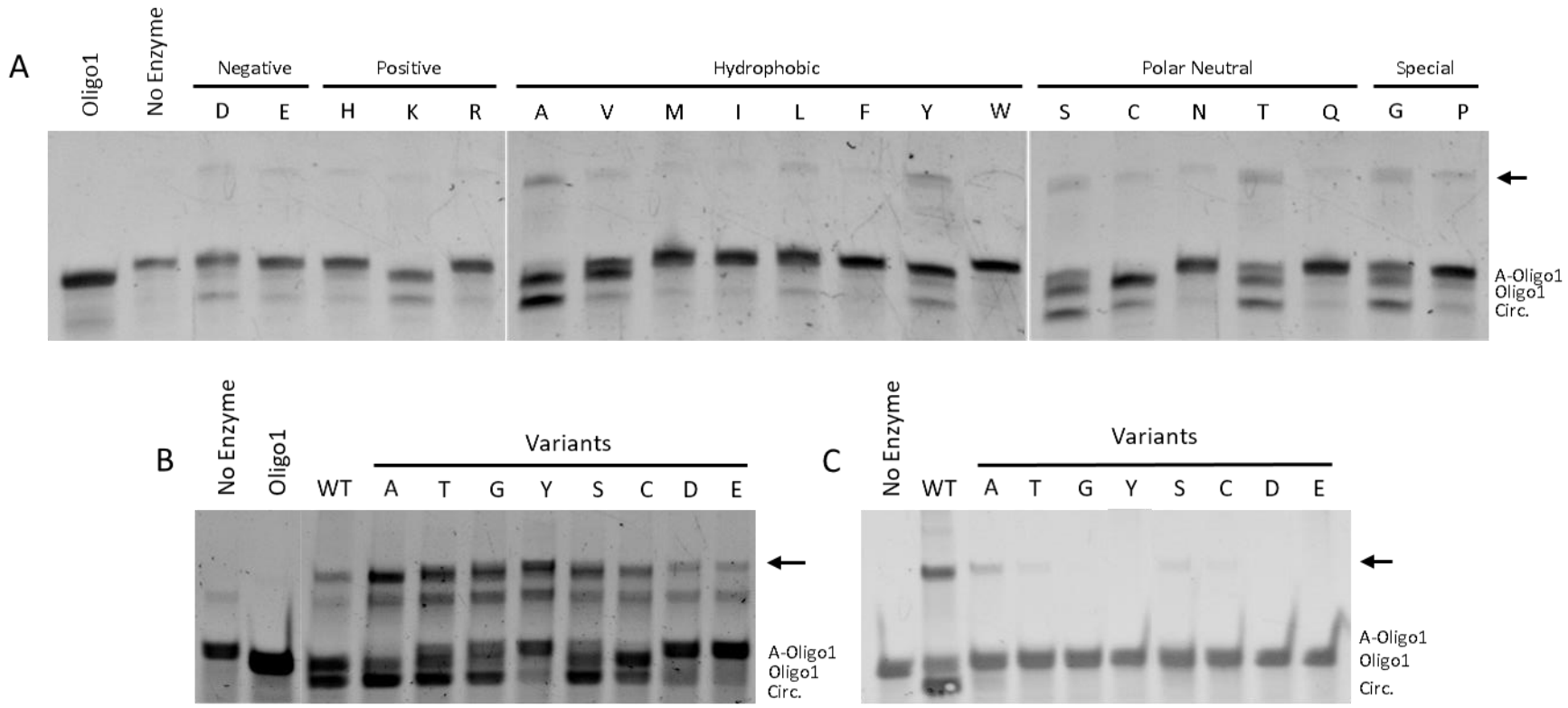

Figure 3-4 Initial activity assays

No enzyme controls contain all required reagents and substrates (pre-adenylated Oligo1, or non-adenylated Oligo1), excluding only the enzyme. Oligo1 controls lack pre-adenylation, and enzyme. Arrows indicate ligated product.

To compare the viability and activity of each variant, each of the twenty variants were assessed on their ability to dimerise pre-adenylated Oligo1 substrate. Utilising a master mix of substrate and buffer, all assays were run in parallel and run on three identical gels, allowing for direct comparisons to be made (A). Variants were grouped according to basic side chain characteristics (negative, positive, hydrophobic, polar neutral, and special case) then arranged from smallest to largest. This allowed any correlation between activity, and side chain characteristics to be easily identified. A minimum of two biological replicates, each carried out in duplicate, were used to ensure generation of an accurate activity profile. The variants identified as most promising were then re-assayed and run side by side on a singular gel (B), allowing for more accurate comparisons to be drawn between them. Presence of two substrate bands is explained by circularisation of dimerised product. These variants were also utilised in an adenylation trial, where variants had to utilise cofactor ATP to adenylate substrate before it could be ligated. A lack of adenylation activity makes production of a ligated product impossible.

3.2.3.2 SR1 RNA ligation assay

By running assays with variants of interest (identified in Section 3.2.3.1) that employed both pre-adenylated RNA oligonucleotide (SR1) and untreated Oligo1 substrates, the 3′ adapter ligation reaction from the sequencing protocol could be more accurately replicated. This provided more accurate information regarding the likely behaviour of the selected variants during the desired reactions. K92A appeared to generate the largest amount of product, identifying it as the likely variant of choice, closely followed by K92 G, S, and T (refer to Figure 3-5-A). Surprisingly, K92Y displayed the smallest amount of ligated product following its promising activity with a single substrate. All mutant variants displayed significantly stronger ligated product bands than the wild type, which produced very little of the desired Oligo1-SR1 RNA product, and significant levels of the off-target ligation and circularisation activity. This activity is associated with the functional adenylation mechanism of the wild type, as it allows the enzyme to transfer the AMP molecule from the pre-adenylated RNA oligonucleotide substrate to the 5′ phosphate of the Oligo1. The diminished ability of the mutant variants to adenylate the 5′ phosphate of the Oligo1 substrate prevents these off-target activities from occurring. The smaller size of the ligated product can be attributed to Oligo1 sequences (18-nt) being marginally smaller than the SR1 RNA oligonucleotide substrate (21-nt), resulting in dimerised Oligo1 being a smaller product than the Oligo1-SR1 RNA product. The only mutant variant investigated potentially displaying any of this activity is K92A, which appears to have a faint indication of a band in line with the dimerised Oligo1 product band and the circularised substrate band displayed by the wild type. While these were a constant presence in K92A assays, they were always very faint. None of the other mutant variants displayed signs of this activity, indicating that the adenylation activity of all mutant variants has diminished enough to adequately minimise any associated off-target activities. While K92A does appear to display some adenylation activity, the insignificance of the associated product bands indicates it is unlikely to have a significant impact on the desired activity.

3.2.3.3 SR1 DNA ligation assay

All twenty variants were screened for potential activity with DNA adapters. Pre-adenylated SR1 DNA and non-adenylated Oligo1 RNA were utilised as the substrates for this assay (Figure 3-5B). Due to similar assays, carried out before the commencement of this research, generating negative results with the *Pfu* RNA ligase wild type, these assays were expected to return with negative results. However, several variants displayed ligation activity,

successfully producing an Oligo1-DNA product. Variants K92 A, C, D, G, S, and T, as well as the wild type (i.e., the "K" variant in Figure 3-5B), were all able to produce product. Off target activity is once again displayed by the wild type, with a circularised substrate band being observed. While ligation activity does not appear to have been affected, the SR1 DNA band is close enough in size to Oligo1 for homo and heterodimer products to appear the same size. Even though the SR1 DNA sequence has the same number of bases as the SR1 RNA sequence, the DNA sequence migrates through the gel faster than the RNA sequence, making it appear smaller.

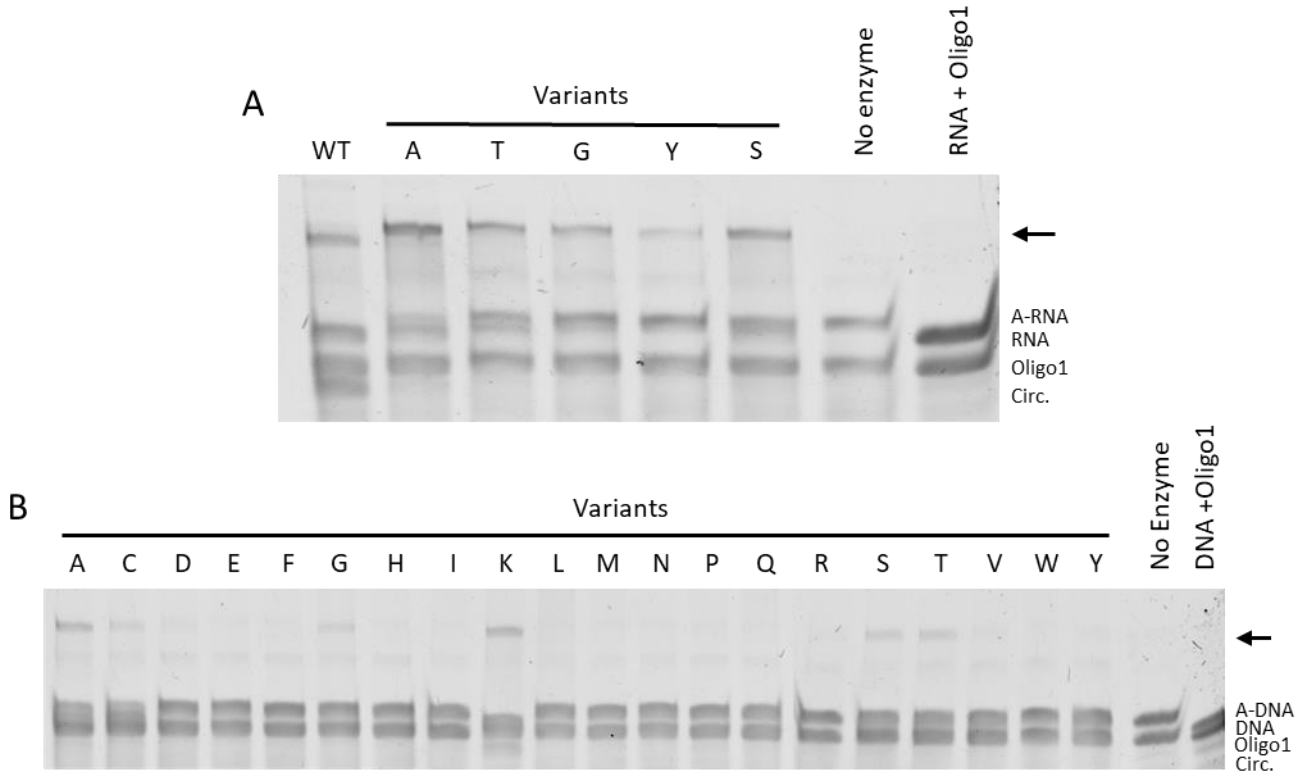

Figure 3-5 SR1 RNA and DNA ligation assays

No enzyme controls contain all required reagents and substrates (pre-adenylated RNA and DNA + Oligo1), excluding only the enzyme. RNA and DNA + Oligo1 controls lack pre-adenylation, and enzyme. Arrows indicate ligated product.

Figure A displays the results of the RNA adapter ligation assays. Mutant variants were arranged from most promising to least. K92A stands out as the best performing variant, while K92Y performed the poorest. The wild type produced a ligated product smaller in size than the mutant variants, as well as displaying circularisation activity. This is indicative of the off-target activity associated with a functional adenylation mechanism. Figure B displays the results of the DNA adapter ligation assay. Variants were arranged alphabetically. Most variants displayed a lack of activity, while variants A, C, G, K, S, and T displayed relatively strong ligation activity. Other variants, such as D, appear to have very low levels of activity, producing very faint ligated product bands. wild type, or variant K, again displaying apparent circularisation activity.

3.3 Discussion

The aim of this chapter was to characterise alterations to the catalytic activity of *Pfu* RNA ligase when the lysine residue at position 92 was substituted for any of the nineteen other amino acids. Completing this library of 20 variants allowed a side-by-side comparison of the effects each substitution had, allowing for the selection of the variant with the most desirable activity. The knowledge gained from this site saturation mutagenesis will aid in future rational engineering of ligase enzymes.

3.3.1 K92L

As the final piece of the puzzle, significant effort was put into generating this variant. A common issue experienced during the generation of all the mutants was an apparent abundance of unaltered wild type template DNA being transformed into cells, with between a quarter and one half of all samples sent for sequencing containing this unaltered sequence. This issue was significantly more troublesome while generating K92L, with over 20 potential clones containing the wild type *Pfu* RNA ligase gene. This issue was initially put down to a failure to digest all remaining template DNA before transformation, resulting in some template DNA being transformed. Alterations to the restriction digest protocol, however, did not improve results. Increasing the duration of the digest, or the concentration of DpnI still resulted in samples returning as template DNA. While the cause of this issue remains unknown, a positive result was eventually obtained after changing the template sequence. The K92A template, which was used in the generation of all other K92 variants, was substituted for a K92M template. This template was selected due to its sequence differing from that of K92L by a single base. K92P could also have been utilised for this same reason.

The new template sequence was successfully amplified utilising the K92L specific designer primers, with colonies containing the amplified product being generated following the normal transformation protocol. Of the three colonies sent for sequencing, two encoded the K92L variation of the *Pfu* ligase gene. The speed and ease of generating the variant using the new template sequence raised numerous questions regarding the failed attempts with K92A. However, as the required mutant had been successfully generated, the cause of the issue was not investigated further.

3.3.2 Purification

The decision to go with parallel batch purification of all 20 proteins saved significant time and effort. Ensuring that proteins were purified at a similar time allowed for comparisons to

be made between all proteins with a greater level of accuracy than if each protein had been purified at a different time. While this method of purification had its benefits, it sacrificed the purity that can be achieved by other means. To ensure efforts were not spread too thin, I opted for four rounds of five purifications over the span of two days. Attempting to purify more proteins than this was impractical. All purifications went according to plan.

Fortunately, all active variants were stable, with most retaining activity well beyond a year after purification. As a result, only one round of purifications was necessary to carry out the desired comparisons. Due to the apparent stability of the proteins, it is likely that carrying out the purifications in parallel was not necessary. This, however, was not known at the time, justifying the precautionary approach taken. While all direct comparisons were carried out from this single round of purifications, all proteins were purified on at least one separate occasion to ensure a minimum of duplicate results. This was done to ensure each variant was accurately characterised. Variants of interest (i.e., K92 A, G, S, T, Y, and wild type) were purified on multiple occasions to ensure that results were reproducible.

3.3.3 Activity assays

Utilising a master mix during the preparation of assays resulted in a uniform concentration of substrate across all assays. This also kept the concentration of RNAse inhibitor, buffer, and ATP constant between reactions, reducing variability. This was important due to the inability for accurate quantification of ligation activity. Attempts to analyse the intensity of product and substrate bands by densitometry often returned confusing results. Substrate control lanes, which contained the same quantity of substrate as each of the reactions, were in almost all instances identified as containing less total substrate than lanes containing assay product. As it is impossible for substrate to have increased, it was determined that either the stain used to visualise the results of the assay must bind to the contents of each product band with a different affinity, resulting in disproportionate representations, or the gel analysis software must not be able to accurately determine changes in band intensity past a certain point. The possibility of the latter being the issue was ruled out as the sole cause of this issue due to gels with lower overall intensity experiencing the same issue. This highlighted the limitations of the assays. However, by limiting variation between reactions as much as possible with the master mix, and running them on the same gels, it was possible to directly compare the intensity of each reaction’s product band, circumventing the need to identify exact rates of ligation. While this resulted in a more binary result than an accurate quantification, it still

served to identify variants with improved ligation activity. Had more specific results been necessary a different means of analysis would have been required.

The initial round of assays was intended to be a basic qualitative assessment of each variant's activity, identifying ligases that were rendered inactive by the substitution at K92. Utilising adenylated Oligo1 as a sole substrate allowed the functionality of each variant to be determined independent of adenylation activity. Despite being unable to quantify the exact rates of ligation, this round of assays drew attention to several variants. K92A, G, S, T, and Y each expressed rates of ligation greater than that of the wild type in each replicate. K92C was identified as a sixth variant of interest due to displaying a similar activity profile to that of the wild type, while variants K92D and K92E were identified as variants of interest due to displaying occasional positive results. Other points of interest were differences in de-adenylation and circularisation activity. While neither of these activities is desirable in the context of a microRNA sequencing workflow, circularisation still requires the ligase to ligate the sequence to itself. When provided with substrates unable to be circularised, it was thought that some of this intra-molecular ligation activity may be translated to inter-molecular ligation activity.

The nine mutants identified above, plus the wild type, were subjected to a further round of assays with pre-adenylated oligonucleotides to provide a direct, side by side comparison of the nine. Alongside this, they were also utilised in an assay containing ATP and non-adenylated Oligo1 substrate. Variants lacking adenylation activity were unable to utilise ATP to adenylate Oligo1 substrate and were therefore unable to generate a ligated product. These two sets of assays provided enough information on each mutant to narrow the selection of mutants to just K92 A, G, S, T and Y, eliminating K92 C, D, and E from contention. As can be noted above, these five variants were identified in the initial round of assays as expressing rates of ligation clearly greater than the wild type.

While the assays carried out up to this point provided a quick and easy means of identifying trends in activity, they were not truly representative of the activity required in the sequencing workflow. The reaction we are selecting a ligase for requires the selective ligation of a 3′ blocked, 5′ adenylated adapter sequence to unknown microRNA sequences. This reaction is significantly more directional than the sole substrate ligations, requiring both the adapter and microRNA to be correctly oriented within the active site of the ligase. A 3′ chemical block on the adapter prevents the ligase from joining adapter sequences to one another, forcing the

microRNA-adapter ligation to occur. While this does not vary drastically from the sole substrate assays, it is an extra level of complication that may impact the overall efficiency of the ligase.

To mimic the required ligation reaction more accurately, the best variants from the first round of assays, including the wild type, were tasked with ligating pre-adenylated SR1 RNA, acting as our adapter substrate, to Oligo1. The results of this assay were in line with the predicted results, bar K92Y. This variant expressed some of the most promising ligation activity during the sole substrate assays, while also exhibiting no apparent adenylation or de-adenylation activity. This had placed it as one of the frontrunners leading into this round of assays. Unexpectedly, this variant had the weakest result of all mutants, producing the least ligated product. It is important to note that all mutants, including K92Y, appear to outperform the wild type, which generated negligible Oligo1-SR1 RNA product, but significant amounts of circularised substrate (indicated by the band residing beneath the Oligo1 substrate band in Figure 3-4A) and Oligo1 dimer (indicated by the smaller ligated product band). This activity was expected due to the off-target activities associated with an active adenylation mechanism (Zhelkovsky & McReynolds, 2012). All other variants displayed little to none of this activity, with K92A being the only one to display any potential signs of it. The shadow bands produced by K92A indicating these activities are faint enough to be mistaken as background contamination and can be easily overlooked. If this were a one-off result, it would have been attributed to random background noise, however, all three replicates displayed the same shadow bands. K92A has been identified as displaying relatively significant levels of de-adenylation activity compared to other mutant variants, meaning it may have retained a limited ability to re-adenylate itself. As such, it is fair to assume that limited off-target activity may be occurring. While the potential off-target activity displayed by K92A is not ideal, this variant also produced the most significant product band out of any variant, followed closely by K92S.

While the results of the SR1 RNA ligation assays were promising, RNA sequences are more difficult to use in sequencing protocols than their DNA counterparts. This is primarily due to the ease at which RNA is degraded, making RNA adapter ligation protocols significantly more delicate than those that use DNA adapters. Alongside this, RNA sequences are also significantly more expensive to synthesise, with Macrogen inc., a reputable and commonly utilised company for oligonucleotide synthesis, charging more than 10 times more for each Uridine (U) base incorporated into the sequence than that of the A, T, C, and G bases. Due to

this, current microRNA sequencing protocols all utilise DNA adapters, meaning a *Pfu* RNA ligase capable of functioning with DNA adapters would not only be easier and cheaper to work with, but would also be a more viable drop-in replacement for current ligase enzymes than one that requires a change of adapter. Some attempts had been made by Dr Oulavallickal to use DNA adapters with *Pfu* RNA ligase before commencement of this thesis, however they had mostly been unsuccessful. Surprisingly, several of the variants I tested managed to successfully produce a ligated product. This result proved to be replicable, with K92A, C, G, S, and T, as well as the wild type producing ligated product in all replicates. This further highlighted the benefit of the K92 A, G, S, and T variants, as they were also the variants that conferred the biggest potential benefit to RNA adapter ligations. The apparent front runner was once again K92A, highlighting it as the variant most likely to be carried forward into future sequencing experiments. The ability of K92C to ligate SR1 DNA to Oligo1 was also a point of interest. While this variant had not been identified as particularly promising in the SR1 RNA ligation assays, it had been noted to display a very similar activity profile, bar adenylation activity, to that of the wild type.

Chapter 4 K238 and Double Mutants

4.1 Introduction

According to the results published by Zhelkovsky and McReynolds (2012), there are two lysine residues in the active site of *Mth* RNA ligase capable of being adenylated, and therefore catalysing adenylation activity: K97 and K246. While substitution of K97 effectively eliminated adenylation activity, indicating that K246 was not able to catalyse adenylation, the substitution of K246 had been shown to alter the adenylation activity of the enzyme. The same sequence alignment utilised to identify K92 in the *Pfu* RNA ligase was also able to identify K238 as the equivalent of K246.

The initial K92A *Pfu* RNA ligase variant produced by Dr Oulavallickal experienced a significant reduction in adenylation activity compared to the wild type, however, it was not eliminated. As such, it was determined that the secondary lysine in *Pfu* RNA ligase, K238, must be capable of catalysing limited adenylation activity. Utilising the information gained from the site saturation mutagenesis of K92, site directed mutagenesis was utilised to substitute K238 for the specific amino acids likely to confer the greatest benefit. This process was carried out on both the wild type (K92, K238), and the K92A variant (K92A, K238), generating a library of both single and double mutant ligase enzymes. In making and characterising these variants, I hoped to definitively prove this residue was involved in the adenylation activity observed in K92 variants. Generation of these variants also provided the opportunity to generate new, novel mutations, with the potential to uncover further improvements to the activity profile of the ligase.

It is important to note that K238 single mutants still had the wild type lysine residue at position 92. This meant that we were expecting the single mutants to continue displaying adenylation activity, as we were confident in this being the primary residue involved in adenylation. We were, however, hoping to observe alterations to the activity pattern that could be used to identify the role of this lysine more accurately.

One area of uncertainty was the impact of removing both lysine residues on the overall activity of the enzymes. While we were confident that the double mutants would lack adenylation activity, therefore preventing the generation of undesirable ligation products, we were uncertain of the effect it would have on ligation activity.

4.2 Results

4.2.1 K238 variant generation

To generate single mutants, the wild type *Pfu* RNA ligase gene was used as the template for QuikChange® mutagenesis. Primers designed to introduce the desired mutations were used to successfully amplify the template sequence (Figure 4-1). As all reactions were done in duplicate, and all amplifications were successful, duplicate products were combined before proceeding with the same PCR product clean up, restriction digest, transformation, plating, and colony PCR protocols as those used to generate K92L. Twenty colony PCR reactions were run in total (four for each variant), 16 of which successfully produced amplified product. These 16 PCR products were all sequenced, with nine being identified as the desired variants. These nine positive results covered all desired mutants. As such, K238 variants A, G, S, T and Y were all successfully generated.

Generation of double mutants was achieved utilising the K92A variant as the template for mutagenic PCRs, as this was identified as the best of the K92 variants in Chapter 3. The primers used for these reactions were designed to substitute K238 for alanine or tyrosine. Alanine was selected due to the success of the K92A variant, while tyrosine was selected due to the interesting results produced by the K92Y variant. Amplified product was treated as above to generate colonies that could be sent for sequencing. Eight colonies were picked (four of each), with seven being successfully amplified. Four of the seven were identified as double mutants. The resulting variants had alanine residues at position 92, with either an alanine or tyrosine residue at position 238.

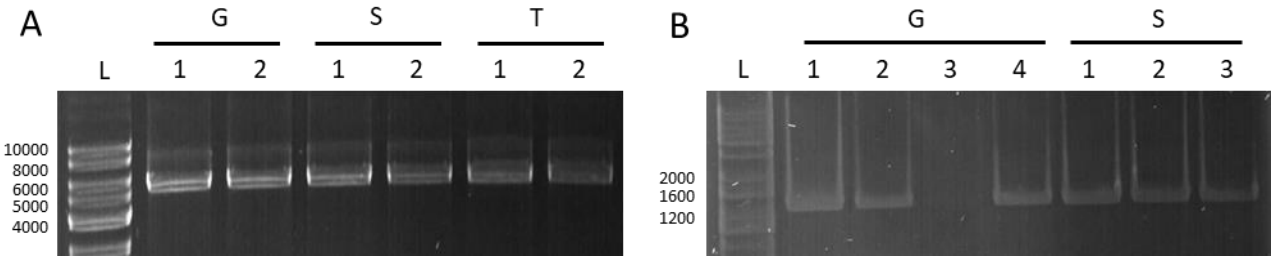

Figure 4-1 Examples of K238 QuikChange® mutagenesis and colony PCR

Figure A provides an example of the successful K238 mutagenesis reactions. These were carried out in duplicate for redundancy. As the 5633 bp product was successfully amplified in all reactions, duplicate reactions were combined before any further steps were taken. Figure B displays several colony PCR reactions. As can be seen, most reactions successfully amplified the target sequence, with only a limited number failing. Four colonies were picked for each variant, with twenty being picked in total. One A and one G colony failed to be amplified, while two T colonies failed.

4.2.2 Protein purification

As with the K92 variants, batch purification protocols were utilised to purify each of the new variants. Fractions were once again taken throughout the purification process and analysed by SDS-PAGE. As a result, it was determined that substituting the K238 of both the wild type and K92A variant did not significantly alter protein expression. While the image displayed in Figure 4-2 does a poor job of distinguishing individual bands in the pre-elution fractions, analysis of the physical gel identified similar protein profiles to each of the pre-elution K92 variant fractions. The protein elution fractions displayed the same band at approximately 44 kDa as the K92 elution fractions, indicating the presence of the target protein, while lacking the same contaminant band. As with the K92 variants, alteration of the purification protocol is likely to improve the purity and concentration of eluted product.

4.2.3 Activity assays

The activity assays used to characterise each of the variants in this chapter are the same as those used in Chapter 3, beginning with single substrate assays before moving onto dual substrate assays.

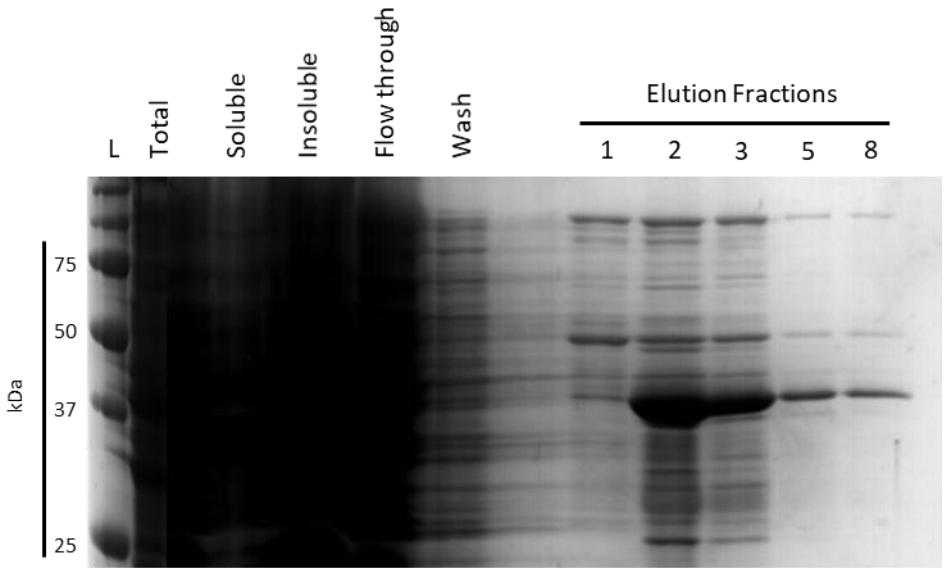

Figure 4-2 K238Y purification

Pre-elution fractions analysed include total protein, soluble protein, insoluble protein, column flow through, and column wash. Protein elution fractions were also analysed to ensure only fractions with reasonable purity were combined for buffer exchanging. Protein was eluted in smaller fractions than the K92 variants. As a result, the first elution fraction contained almost exclusively contaminant protein, and very little target protein. The second elution fraction contains the first significant portion of protein, without some of the contaminants found in the first elution fractions.

4.2.3.1 Single substrate assays

Utilising pre-adenylated Oligo1 as the single substrate, as discussed in Section 3.2.3.1, the baseline activity of each protein was quickly determined. Variants K238 A, G, S, T, and Y all exhibited ligation activity, with only variants A and G displaying rates of ligation noticeably greater than that of the wild type (Figures 4-3A and C). Notably, variants A, G, S, and T left little to no adenylated Oligo1 behind, either ligating, circularising, or de-adenylating the supplied substrate. K238Y on the other hand displayed significant levels of remaining adenylated substrate, while also lacking any notable circularisation activity. Both the K92A, K238A (AA) and K92A, K238Y (AY) double mutants displayed limited ligation activity, with no sign of de-adenylation or circularisation activity. The ligation activity of these double mutants was significantly lower than that of the wild type.

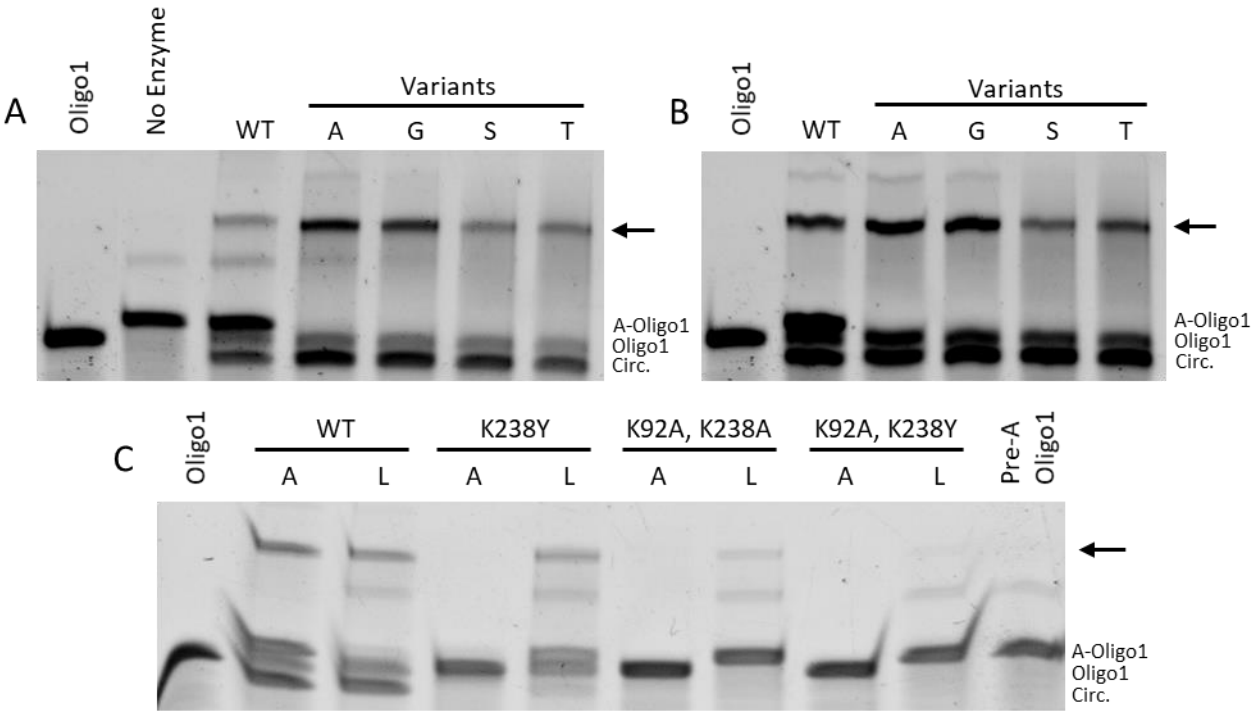

Figure 4-3 K238 initial activity assays

Ligated product is indicated by the arrow. Oligo1 lanes contain non-adenylated Oligo1 substrate. No enzyme lanes contain all substrates used in the respective reactions. In figure A, the ligation activity of K238 variants A, G, S, and T is assessed by comparing ability of each variant to join pre-adenylated Oligo1 substrates compared to that of the wild type. All variants display ligation activity, with K238 A and G appearing to be the most active. In figure B, the adenylation activity of K238 variants A, G, S, and T is assessed by determining the ability of variants to adenylate then ligate Oligo1 substrates in the presence of ATP, compared to that of the wild type. Presence of bands other than Oligo1, be they circularised (circ.), adenylated (A-Oligo1), or ligated, indicates an active adenylation mechanism. In figure C, both the ligation (as per figure A) and adenylation (as per figure B) activities of K238Y and the generated double mutants are assessed. Adenylation assays are indicated by an A, while ligation assays are indicated by L.

Assays carried out in the presence of cofactor ATP, utilising non-adenylated Oligo1 as the sole substrate, identified adenylation activity in all variants bar the AA and AY double mutants. While K238Y exhibited adenylation activity, it was extremely limited, with only a faint adenylated substrate band being displayed. K238Y also lacked circularised substrate, and ligated product bands. Variants A, G, S, and T had no adenylated substrate band, but very intense ligated product and circularised substrate, indicating that the totality of adenylated substrate was either ligated or circularised.

4.2.3.2 SR1 RNA ligation assays

Following proof of variant activity, K238 A, G, S, and T were utilised in the same SR1 RNA ligation protocol described in Section 3.2.3.2, tasked with ligating pre-adenylated 3′ amino modified SR1 RNA to Oligo1 sequences. The wild type ligase (K92, K238) was utilised as a control. Use of the pre-adenylated, 3′ modified SR1 RNA, and unmodified Oligo1 substrates should have prevented the generation of any product other than the desired Oligo1-SR1 RNA product. The 3′ modification on the SR1 RNA prevents its circularisation, while ensuring it can only be ligated to the 3′ end of the Oligo1. Use of an unmodified Oligo1 should have prevented it from being circularised or ligated to other Oligo1 sequences, as it lacks the required adenylation of the 5′ phosphate. Despite this, each of the variants, including the wild type, produced circularised substrate bands, Oligo1-Oligo1 homodimer (smaller) and Oligo1-SR1 RNA heterodimer (larger) ligated products, and concatemer product bands (Figure 4-4). This is likely possible due to the high rates of adenylation (Figure 4-3B) and de-adenylation activity (Figure 4-4) displayed by the K238 variants, as de-adenylation of SR1 RNA by a K238 ligase will result in the AMP group being transferred from SR1 RNA to the ligase, allowing it to then adenylated Oligo1.

Only variants G and T visibly failed to de-adenylate all remaining pre-adenylated SR1 RNA substrate. The biggest differences between variants can be seen in the Oligo1-SR1 RNA heterodimer product bands (Figure 4-4). Wild type produces a notably fainter band than the mutant variants, with K238G appearing to display the most intense band, followed closely by K238A. The results presented by these variants differ dramatically from those displayed by the K92 variants – mostly due to the off-target ligation and circularisation activity displayed by all assayed variants.

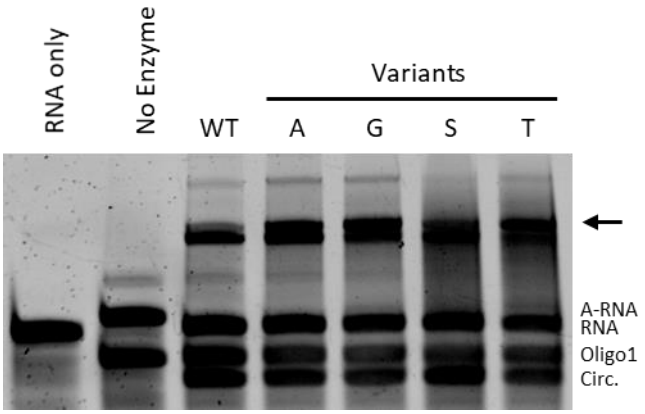

Figure 4-4 SR1 RNA ligations

Ligated product band is indicated by the arrow. Any bands above the arrow are likely concatemer products. RNA only lanes contain non-adenylated SR1 RNA substrate, while the no enzyme lanes contain both pre-adenylated SR1 RNA (A-RNA) and non-adenylated Oligo1 substrates. While all variants, including the wild type, display significant levels of ligation activity, they also display significant levels of unwanted ligation and circularisation activity. These activities are indicated by the presence of two product bands and a circularised substrate band (circ.). The active adenylation mechanism of the variants allows the enzymes to transfer the AMP molecule from pre-adenylated SR1 RNA to the Oligo1 substrates, which can then be ligated and circularised. As the 3′ Amino modification on the SR1 RNA prevents it from being circularised, this transferal explains the presence of circularised substrate, while also suggesting the second product band is an Oligo1-Oligo1 product.

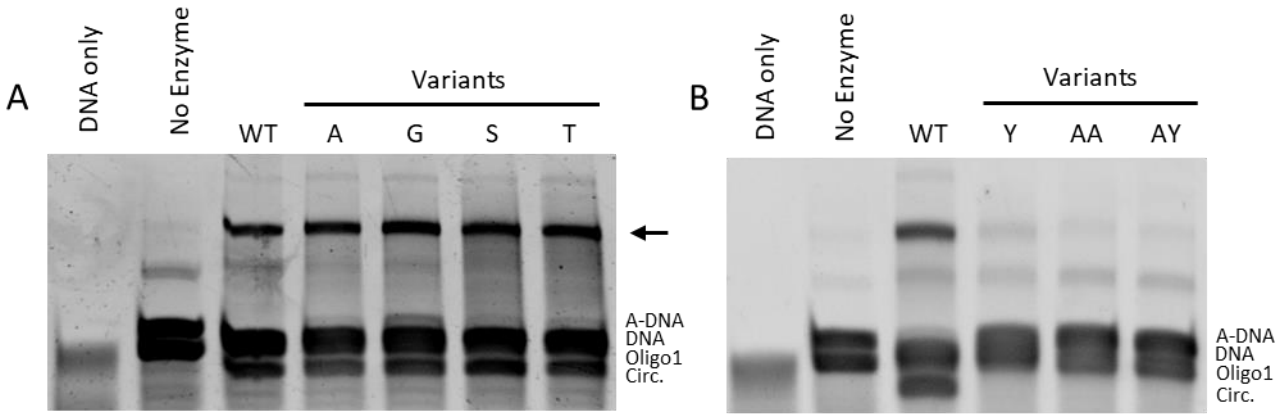

Figure 4-5 SR1 DNA ligation

Ligated product band is indicated by the arrow. Any bands above the arrow are likely concatemer products. DNA only lanes contain non-adenylated SR1 DNA substrate, while the no enzyme lanes contain both pre-adenylated SR1 DNA (A-DNA) and non-adenylated Oligo1 substrates. In gel A, K238 A, G, S, and T single mutants display significant levels of activity. Unfortunately, off target activity is indicated by the presence of circularised substrate band (circ.). The active adenylation mechanism of the variants allows the enzymes to transfer the AMP molecule from the pre-adenylated substrate to the 5′ phosphate of Oligo1 sequences, allowing the ligase to circularise and ligate them. As the 3′ Amino modification on the SR1 DNA prevents it from being circularised, this transferal explains the presence of circularised substrate. This also suggests that an Oligo1-Oligo1 product is likely present, raising doubt regarding the ligation efficiency of the variants with DNA. In gel B, both the K92A, K238A (AA) and K92A, K238Y (AY) double mutants, and the K238Y single mutant display no significant activity, producing little-to-no ligated product, and displaying no apparent off target activities.

While the 2012 paper published by Zhelkovsky and McReynolds indicates the observed off-target activity is likely a result of enzymes transferring the AMP molecule from pre-adenylated substrates to non-adenylated Oligo1 substrate, the adenylation of Oligo1 could also be explained by the presence of ATP in the reaction. The adenylation reaction used to generate the pre-adenylated substrate contains high concentrations of ATP, meaning a failure to properly remove the residual ATP from the product could result in ATP being carried over into the ligation.

4.2.3.3 SR1 DNA ligation assays

All K238 variants were screened for activity with DNA substrate using the SR1 DNA substrate. Again, wild type was used as a control. Both the AA and AY double mutants displayed an apparent lack of ligation activity with the SR1 DNA sequence, failing to produce a ligated product band with greater intensity than the background contaminants (Figure 4-5B). K238Y did not fare much better, displaying a ligated product band only marginally stronger than the background contaminant. None of these three variants displayed a result strong enough to be of particular interest for use in the sequencing protocol. In contrast, variants K238 A, G, S, and T, as well as the wild type, all produced a single substantial ligated product band, indicating substantial ligation activity (Figure 4-5A). K238G appears to display a slightly more significant ligated product band that the other variants. Each of these variants also displayed circularisation activity, indicating the presence of the same off-target activity observed in Section 4.2.3.2.

4.2.3.4 Mock adenylation reaction

The potential for ATP carryover from the adenylation reaction was investigated using mock adenylation reactions. By following the pre-adenylation reaction protocol with a substrate lacking a 5′ phosphate, we could ensure that the substrate was not adenylated. The product of this reaction could then be used in ligation reactions alongside standard Oligo1 substrate. As neither substrate was adenylated, and no ATP was added to the reaction, production of circularised substrate or ligated product should not have been possible. The only way these products could be produced was if ATP were carried over from the mock adenylation reaction. As such, the presence of any bands other than the Oligo1 and mock adenylation substrates (i.e., adenylated substrate, circularised substrate, or ligated product) would indicate

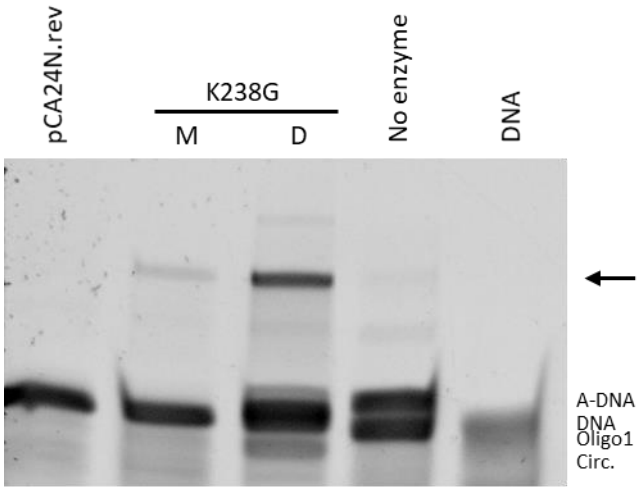

Figure 4-6 Mock adenylation

All reactions were carried out using K238G. Ligated product band is indicated by the arrow. Any bands above the arrow are likely concatemer products. Lane labelled pCA24N.rev is the control for the mock adenylation reaction, containing a sample of the substrate used in the mock adenylation reaction. The mock ligation reaction (M) displays a limited amount of ligated product, and no apparent circularised substrate band (circ). In contrast, the SR1 DNA ligation (D) displays significant circularised substrate and ligated product bands. This indicates that a significant majority of the off-target activity observed in the SR1 RNA and DNA ligation reactions is not a product of ATP carryover. No enzyme and DNA only controls were used as usual for controls of the regular reactions.

ATP carryover. The DNA oligonucleotide primer pCA24N.rev was utilised as the substrate for these reactions. Mock reactions were all treated as regular pre-adenylation reactions, including the precipitation wash steps. Assays run with these substrates utilised variant K238G as the ligase. This variant was selected due to K238G displaying the strongest activity of the K238 variants thus far. The results of these assays displayed no apparent circularisation of substrate, and a small amount of ligated product (Figure 4-6). This indicates that ATP carryover may explain some of the observed off target activities, not enough, however, to explain the extent of the circularised substrate and Oligo1 homodimer product bands observed in the SR1 RNA and DNA ligation assays. This reaction was run alongside a SR1 DNA ligation, as described in Section 4.2.3.3, providing a reference for the level of off target activity displayed in these reactions.

4.2.3.5 Pre-adenylation optimisation

To try and eliminate the limited effects of ATP carryover identified in Section 4.2.3.4, several adenylation reactions were carried out with varying ATP concentrations, such that the lowest concentration of ATP that could be used to successfully adenylate all desired sequences could be determined. This was designed to ensure that all off target activities observed in future ligation reactions would be a product of the mechanism identified by Zhelkovsky and

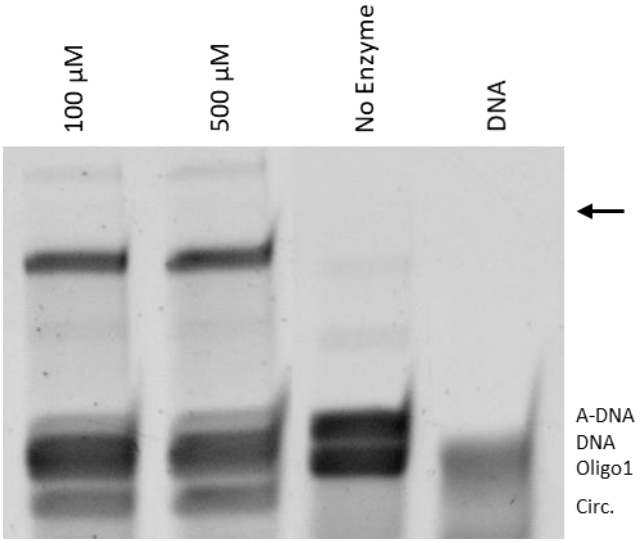

Figure 4-7 Optimisation of pre-adenylation ATP concentrations

Ligated product is in line with the arrow. Both ligation reactions are identical, only being separated by the concentration of ATP used in the pre-adenylation of their SR1 DNA substrate. As can be seen, changing the amount of ATP used in the adenylation reaction does not change the result in any obvious way, with the two lanes displaying near identical ligated product, adenylated DNA (A-DNA), DNA, Oligo1 and circularised substrate (circ.) bands. No enzyme and DNA only controls were used once again.

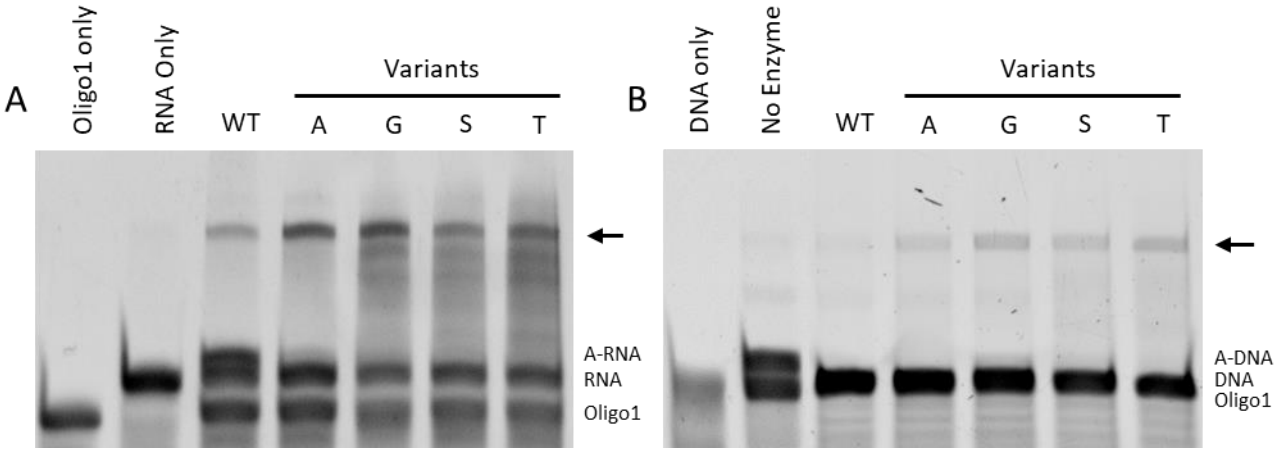

Figure 4-8 Dual blocked substrate assay

Ligated product band is indicated by the arrow. This figure displays the activity of variants when provided with modified Oligo1, and either SR1 RNA (A) or DNA (B) substrates. SR1 RNA ligation reaction was carried out in the presence of ATP, while SR1 DNA ligation reaction utilised pre-adenylated SR1 DNA. All unexpected bands observed in the regular adapter ligation assays are not present, indicating no off target activities occurred. As a result, the product bands observed are known to be the desired Oligo1-SR1 heterodimers, providing definitive results regarding activity with either type of substrate.

McReynolds (2012). Adenylation reactions carried out up to this point had been using a final ATP concentration of 500 µM. In these reactions a gradient of final concentrations ranging from 500 µM down to 100 µM was utilised to test rates of adenylation over the identified ATP concentrations. All tested concentrations resulted in a successful adenylation reaction. A comparison assay was then run using two different stocks of pre-adenylated SR1 DNA substrate, one prepared using the normal ATP concentration of 500 µM, and one prepared using the lower 100 µM final concentration. K238G was utilised as the ligase for these reactions. As can be seen in Figure 4-7, the results of these assays cannot be differentiated. Rates of circularisation were identical, with both reactions also displaying the same intensity ligated product band. The lack of variation between the two indicates one of two things, either ATP carryover did not decrease with the lower concentration, or that it plays an insignificant role in the observed activity.

4.2.3.6 Double block assays

A set of assays were carried out utilising two types of substrate, with modifications made to opposite ends of their sequence. This meant that one substrate contained a 5′ modification, and the other a 3′ modification. The purpose of this was to ensure that the only ligations that could occur would be between the unmodified ends of the substrates, forcing the desired ligation reaction to occur, and eliminating the possibility of any off-target activity. The same SR1 RNA and DNA sequences were used as in the assays described in Sections 4.2.3.2 and 4.2.3.3, with a modified, 5′ dephosphorylated Oligo1 being utilised as the second substrate. All assays utilising two modified substrates displayed a total lack of circularisation, alongside a noticeable reduction in ligated product. In the SR1 RNA ligation assays, the smaller homodimer product was entirely lost, with the slightly larger heterodimer product remaining a similar intensity (compare Figures 4-4 to 4-8A). In the SR1 DNA ligation assays, the singular ligated product band became significantly fainter (compare Figure 4-5A to 4-8B). In the SR1 RNA gel, variants A and G appear to display the most significant activities, while in the SR1 DNA gel, variants G and T appear to display the most significant activities. As such, K238G was identified as the most promising variant.

As all SR1 RNA and DNA double block assays display no circularised substrate, we can assume all off-target activities have effectively been eliminated. As such, comparisons can be made between the product bands displayed in Figures 4-4 and 4-8. The Oligo1-SR1 RNA product bands in Figure 4-4 and Figure 4-8A appear to be of similar intensity, suggesting that the off-target activities did not affect the rate at which the Oligo1-SR1 RNA product was

generated. As such, we can assume the same is true for generation of the Oligo1-SR1 DNA product. As there is a notable discrepancy between the Oligo1-SR1 DNA product bands displayed in Figures 4-5 and 4-8B, the difference can be attributed to a lack of off-target Oligo1-Oligo1 product. This indicates that the Oligo1-SR1 DNA and Oligo1-Oligo1 products overlap, exaggerating the extent of the activity observed in Figure 4-5A.

4.3 Discussion

Utilising the information generated in Chapter 3, informed mutagenesis of lysine 238 was carried out. Through this, I aimed to determine its role in adenylation and ligation activity, while potentially improving the desired activity of the enzyme. Utilising both wild type and K92A sequences as templates for mutagenesis, a small library of single and double mutants was generated, allowing the activity of each lysine residue to be analysed independently. This aided in developing our understanding of the role lysine residues play in the activity of this ligase.

4.3.1 Generation of variants

All the desired mutations at position K238 were achieved successfully. Compared to the issues encountered in generating the K92L variant (Section 3.3.1), mutagenesis was straightforward, with more than 50% of sequenced clones routinely containing the desired mutation.

4.3.2 Purification

The process of purifying these variants was comparable to that for the K92 variants. While the total number of variants needing purification was significantly less than that of the K92 mutants, batch purification protocols were utilised to ensure consistency with K92 purifications, and for the benefit of parallel purification.

4.3.3 Activity assays

Qualitative, gel-based activity assays were employed to ensure consistency with the results described in Chapter 3. Master mixes were utilised to reduce variation between reactions, while the reduced number of variants compared to Chapter 3 meant all desired reactions could be run on single gels.

4.3.3.1 Ligation and adenylation activity

The initial round of assays utilised the same single substrates as in Chapter 3. The assays with pre-adenylated substrates immediately highlighted issues with the AA and AY double mutants,

with neither displaying levels of activity comparable to the wild type (Figure 4-3C). While both de-adenylation and circularisation activities were absent from these variants, the intensity of product bands was low enough to rule these mutants out from subsequent characterisation. On the same gels as the double mutants, the K238Y single mutant was compared to the wild type (Figure 4-3C), displaying remarkably low levels of circularisation. De-adenylation activity and ligation activity on the other hand were comparable to that of the wild type, resulting in a relatively significant amount of leftover pre-adenylated substrate. While there is not any apparent benefit to ligation activity, the lack of circularisation alongside the remaining adenylated substrate leaves room for reaction optimisations to increase ligated product. The four other single variants, A, G, S, and T, all displayed similar patterns of activity to each other, but vastly different to that of the K238Y single variant (Figure 4-3). With no sign of leftover pre-adenylated substrate, the four variants either ligated, circularised or de-adenylated all the provided pre-adenylated substrate. Proportionately, de-adenylation activity appears to represent the smallest proportion of activity, with both the circularised and ligated product bands having an intensity significantly greater than the de-adenylated substrate band. Notably, variants A and G displayed the most promising ligation activity, with their ligated product bands displaying a significantly greater intensity than that of the wild type.

To build on these results, the adenylation assay was utilised to determine the impact of substituting K238 on adenylation activity. Both double mutants, as expected, displayed no apparent adenylation activity, while single K238 variants A, G, S, and T all displayed significant adenylation activity, as identified by the significant circularised substrate and ligated product bands. In contrast to this, the K238Y variant displayed an incredibly low level of adenylation activity, producing no identifiable circularised substrate or ligated product bands. K238Y also displayed limited adenylation activity, generating a band faint enough to miss at first glance. As this variant still has its primary lysine residue intact, it was assumed that for adenylation activity to have been impacted to this extent, the substitution would have had to have altered the conformation of the active site such that ligation would be impacted too. As the ligation activity of this ligase appears to have remained intact, the lack of adenylation activity is especially interesting.

4.3.3.2 RNA and DNA adapter activity

As with the K92 assays, only the most promising variants, in this case K238 A, G, S, and T, were carried forward for use in the SR1 RNA ligation assay. Each variant displayed significant levels of activity in these assays. Unfortunately, the activity displayed was not

ideal. The nature of the SR1 RNA sequence renders it impossible for anything to be ligated to its 3′ end due to an amino modification. As the SR1 RNA sequences were the only pre-adenylated substrate in the reaction, ligation of the adenylated 5′ end of the SR1 RNA substrate to the 3′ end of the Oligo1 substrate should have been the only possible ligation reaction. Dimerisation and circularisation of SR1 RNA substrates is impossible due to the 3′ modification, while the lack of ATP should have prevented off-target activities with Oligo1 substrates, such as circularisation and dimerisation. Despite this, circularised substrate bands, alongside a second ligated product band were observed (Figure 4-4). Due to the 3′ chemical modification on the SR1 RNA, the presence of circularised substrate can only be explained by the circularisation of Oligo1. For this to have occurred, Oligo1 sequences had to have been adenylated, also indicating that one of the two product bands is likely an Oligo1-Oligo1 product. As can be seen from the control lanes, the SR1 RNA sequence is slightly larger than the Oligo1 sequence. This means that an Oligo1 dimer product would be marginally smaller than an SR1 RNA-Oligo1 hybrid dimer, supporting the hypothesis that one of these bands is an Oligo1 homodimer. A third, much fainter product band can be seen above the two primary product bands, indicative of concatemerised substrate.

SR1 DNA ligation assays, utilising the same K238 A, G, S, and T variants, were run parallel to the SR1 RNA ligation assays. While these variants all appeared to display promising activity, the presence of circularised substrate indicated that some off-target ligation activity was also likely. As such, I was relatively confident that Oligo1 homodimer was also present, despite the lack of two separate ligated product bands. A single ligated product band therefore suggested that either no Oligo1-SR1 DNA heterodimer product was produced, or the two products were similar enough in size to overlap. Due to SR1 DNA migrating through urea gels at a similar rate to Oligo1 than the SR1 RNA, the potential for overlapping product bands was relatively high. As such, it was impossible to determine from this gel whether the product band was the desired Oligo1-SR1 DNA heterodimer, the undesirable Oligo1 homodimer, or a combination of the two.

K238Y and the double mutants were also utilised in SR1 DNA ligation assays, following their purification. Both double mutant variants appeared to lack activity with the SR1 DNA substrate, producing ligated product bands matching the intensity of the contaminant product observed in the no-enzyme control. This contaminant band became a regular issue from this point on, with all samples containing pre-adenylated SR1 DNA displaying it. The reason for this is unknown, but it was likely due to a contaminant being introduced into the stock. While

this was not desirable, it did not impact the ability to identify ligation activity. K238Y displayed limited activity with the SR1 DNA substrate, producing a product band marginally more intense than the contaminant. As expected, all three of these variants lacked a circularised substrate band. The lack of a functional adenylation mechanism prevented them from producing unwanted Oligo1 products regardless of the cause of activity.

There are two possible reasons for the observed adenylation activity. While no ATP was added directly to the reaction, it is possible that some was carried over from the pre-adenylation reaction. This reaction utilises a relatively high concentration of ATP, meaning a failure to properly remove any residual ATP could plausibly result in some carry over. However, as the product of the adenylation reaction is precipitated in ethanol, washed thoroughly twice to remove unused ATP, then resuspended in sterile RNAse free H2O, the risk of ATP carryover is marginal. The other possibility relates to the adenylation mechanism of the variants. Previous research has identified that the biochemical mechanism by which these ligase enzymes de-adenylate is the reverse of the adenylation mechanism, resulting in the transferal of the adenyl group back to the ligase (Zhelkovsky & McReynolds, 2012). If ATP carryover is not present, then this mechanism of de-adenylation may explain the observed results. A re-adenylated ligase could then transfer the adenyl group to a new substrate, explaining the adenylation and subsequent ligation of Oligo1.

4.3.3.3 Is ATP carryover to blame?

To determine the role of ATP carryover in the observed result, a mock adenylation reaction was carried out. This involved following the normal adenylation protocol with a substrate lacking a 5′ phosphate, such that adenylation could not actually take place. When the product of this adenylation reaction was used in an assay alongside Oligo1, without adding ATP, the presence of any circularisation or ligation activity would be attributed to enzymes utilising ATP carried over from the pre-adenylation reaction. Without this carryover, there would be no ATP to facilitate adenylation of the Oligo1 substrate. The reactions run with this substrate (Figure 4-5) produced a small amount of ligated product, but no circularised substrate. This amount was insignificant however, failing to replicate anything close to the rate of circularisation and Oligo1 dimerisation observed in the previous experiments. This indicated that, while ATP carryover may be responsible for some of the observed results, it does not appear to be the sole, or even primary cause.

The limited ATP carryover displayed by the mock adenylation could be the result of a failure to properly wash the ATP from samples during the precipitation step of the pre-adenylation protocol. Future work could involve investigating more thorough wash steps during the adenylation protocol.

To minimise the possible effect of ATP carryover, pre-adenylation optimisation reactions were carried out utilising a range of ATP concentrations, the lowest of which being one fifth of that used in the normal reaction. All reactions successfully adenylated all provided substrate, with the products of the highest and lowest concentrations being used in ligation activity assays. Due to the reduced concentration of ATP, carryover should have been minimised, resulting in a reduction of Oligo1 circularisation and dimerisation. The activities, however, appeared to remain constant across reactions, indicating that either ATP carryover was not reduced by the lower concentration of ATP used in the pre-adenylation reaction, or that ATP carryover is not the cause of the off-target activities.

Despite the lack of evidence that a lower concentration of ATP in pre-adenylation reactions reduces off target activity in the ligation reactions, the lower ATP concentration was used for all pre-adenylation reactions from this point forward to minimise the chance of ATP carryover impacting future experiments.

4.3.3.4 Activity with two blocked substrates

To eliminate the possibility of Oligo1 adenylation, a new set of SR1 RNA and DNA ligation reactions were carried out using a 5′ dephosphorylated Oligo1 substrate. By removing the 5′ phosphate of the Oligo1 substrate, which is required for sequence adenylation, off target adenylation of this substrate is prevented. The combination of this 5′ modification and the 3′ modification on the SR1 RNA and DNA substrates prevents the production of homodimer ligation products and circularised substrate. As such, the only possible ligation product is Oligo1-SR1 heterodimers. These “double-blocked” assays allowed me to isolate and characterise the desired ligation activity, providing a more accurate profile of specific ligation activities.

The 5′ adapter ligation involves ligating the 5′ end of the microRNA-adapter product (which has retained the 3′ modification of the 3′ adapter) to the 3′ end of a 5′ blocked RNA adapter. As such, an SR1 RNA “double-blocked” assay mimics this reaction. The 3′ modified SR1 RNA represents the product of the 3′ adapter ligation, while a 5′ blocked Oligo1 represents the 5′ blocked adapter. As this reaction would normally be carried out in the presence of

ATP, and the K238 variants have an active adenylation mechanism, ATP was used. The 5′ block on the Oligo1 ensured that 5′ of the SR1 RNA was adenylated, then ligated to the 3′ end of the 5′ blocked Oligo1.

These reactions displayed a single ligated product band, as well as a total lack of circularised substrate, as was expected. This meant that the use of oppositely modified substrates effectively eliminated the observed off target activity. All variants, including the wild type, managed to successfully adenylate SR1 RNA and generate a ligated product. Mutant variants all outperformed the wild type, with K238 A and G displaying the most prominent ligated product bands (Figure 4-8A).

As DNA adapters will only be utilised for the 3′ adapter ligation, which involves ligating pre-adenylated 3′ modified adapter sequences to the 3′ ends of an unmodified microRNAs, no SR1 DNA "double block" assay accurately mimics a real sequencing reaction. To keep the reaction as close as possible to a real DNA adapter ligation, a 5′ adenylated, 3′ modified SR1 DNA was used to represent the 3′ adapter, while 5′ dephosphorylated Oligo1 represented the unmodified microRNA.

Unfortunately, the ligated product band was significantly lighter than the original DNA adapter ligation assay (compare Figure 4-5 and 4-8B), indicating that the bulk of the product displayed in Figure 4-5 was in fact an Oligo1-Oligo1 homodimer. While this was not the desired result, each of the mutant variants displayed some level of ligation activity, outperforming the wild type which displayed a product band comparable to that of the contaminant in the no enzyme control. This result, however unimpressive compared to the original assay, still indicates that these variations confer a benefit to the enzyme. Variant K238G consistently displayed the most intense product band, solidifying its position as the frontrunner of the K238 variants.

Chapter 5 Towards a Sequencing Experiment

5.1 Introduction

The overarching aim of this thesis was to identify the *Pfu* RNA ligase variants best suited for use in a high temperature microRNA sequencing protocol. This protocol requires ligase enzymes to join adapter sequences to either end of target microRNAs, such that high throughput sequencing techniques can then be utilised to profile the microRNAs within the sample. Joining of these adapters is achieved through two separate ligation reactions, the first of which ligates an adapter sequence to the 3′ end of the microRNAs. Typically, this is done with a DNA adapter sequence. Not only are DNA sequences much more stable than their RNA counterparts, but also significantly cheaper to manufacture. This makes them a more desirable substrate than RNA adapters. While this is the case, some ligases enzymes display no activity with DNA adapters. For these ligase enzymes to be of use, RNA adapters are required for this step. The second ligation reaction involves ligating RNA adapter sequences to the 5′ ends of the microRNAs. DNA adapters cannot be used for this step, as they would prevent a reverse transcriptase from generating the cDNA sequences required for sequencing. Once both adapters have been joined to the target microRNA sequences, a primer is annealed to prepare the sequence for reverse transcription. The primer is a DNA sequence with complementarity to the 3′ adapter, capable of annealing to either RNA or DNA adapter sequence. A reverse transcriptase is utilised to generate a cDNA copy of the microRNA, which is then further amplified by PCR, to generate a double-stranded DNA product that is capable of being sequenced by means of standard high throughput sequencing techniques.

To ensure adapter sequences are ligated in the correct orientation, specific modifications are made to the sequences to ensure directionality. A 3′ amino modification (3AmMO) is made to the 3′ adapter, ensuring only the 5′ end is available for ligation. The 5′ adapter, on the other hand, has its 5′ phosphate removed, ensuring only the 3′ end of the adapter is available for ligation. To ensure adapters are not ligated to each other, two separate ligation reactions are required, the first ligating the 3′ adapter to the 3′ end of the microRNA, and the second ligating the 5′ adapter to the 5′ end of the microRNA. The product is cleaned in between ligations to remove all un-ligated substrate. This ensures that that only the desired ligations occur, resulting in the adapters being exclusively ligated to the desired end of the microRNA sequences.

A pre-adenylated 3′ adapter is utilised for the first ligation. Assuming the ligase utilised lacks a functional adenylation mechanism, preventing the undesirable re-activation of the ligase (Zhelkovsky & McReynolds, 2012), a pre-adenylated adapter forces the ligation of the 5′ adenylated adapter to the 3′ end of the microRNA. The resultant microRNA-adapter heterodimer product retains the 3′ amino modification of the adapter. On the other hand, the 5′ adapter ligation involves joining the 3′-OH of the adapter to the 5′-phosphate of the microRNA. As such, the second ligation reaction must be carried out in the presence of ATP with a ligase retaining its ability to adenylate the 5′-phosphate of the microRNA. The lack of a 5′ phosphate on the adapter ensures that only the 5′ end of the microRNA-adapter sequences can be adenylated. The 3′ modification on the microRNA-adapter sequence, and the 5′ modification on the 5′ adapter forces the desired ligation of the 5′ adapter to the microRNA-adapter sequence, preventing off-target activities such as the circularisation. The result is the target microRNA sequence flanked by the two adapter sequences.

Utilising the information gained from Chapters 3 and 4, this chapter aimed to determine the K92 and K238 variants best suited for each adapter ligation, such that they could be used in a trial sequencing experiment.

5.2 Results

Determining the variants best suited for use in the microRNA sequencing experiment was achieved by comparing the activity patterns of the best K92 and K238 variants to the activity required in each adapter ligation. As determined in Chapter 3, K92A was the top performing K92 variant, while Chapter 4 identified K238G as the most promising K238 variant. While both these variants show promise, the required activity for each of the ligation reactions favours different activity profiles.

5.2.1 3′ adapter ligation

To replicate the 3′ adapter ligation, pre-adenylated SR1 DNA and unmodified Oligo1 substrates were utilised in a reaction absent of ATP (Figure 5-1). Here, the DNA sequence acted as the adapter, while the Oligo1 sequence acted as the microRNA. K238G appears to display a greater level of ligation activity than K92A, however, it also displays evidence of circularisation activity. As SR1 DNA cannot be circularised, and no ATP is present for adenylation of Oligo1, this is only possible if de-adenylation of the SR1 DNA oligonucleotide by K238G allows it to re-form the ligase-AMP intermediate (with K92 adenylated), such that it can adenylate the 5' phosphate of Oligo1 then circularise it.

Adenylation of Oligo1 would also likely result in Oligo1 sequences being ligated to one another.

As the off-target activity of K238G had been observed in Chapter 4 (Figures 4-4 and 4-5A) a set of modified “double-blocked” assays, substituting the unmodified Oligo1 substrate for a 5′ dephosphorylated Oligo1, was also run. The lack of a 5′ phosphate prevents this Oligo1 substrate from being adenylated, ensuring that the activity displayed by the variants would be restricted to the desired ligation reaction. As expected, the K238G “double block” assay lacked circularised substrate band observed in the unmodified Oligo1 assay, while also displaying a reduction in ligated product. Despite this, the product band appeared to be of a similar intensity to the product band produced by K92A in the unmodified Oligo1 assay. K92A also appears to experience a minor decrease in ligation activity when provided with the 5′ dephosphorylated Oligo1, indicating that K92A may also exhibit some undesirable activities.

5.2.2 5′ adapter ligation

For the 5′ adapter ligation, 3′ blocked SR1 RNA and 5′ dephosphorylated Oligo1 substrates were utilised, in the presence of ATP. In this reaction, the 5′ blocked Oligo1 represented the 5′ blocked adapter, while the 3′ blocked SR1 RNA represented the 3′ blocked microRNA-adapter product of the first ligation. This allowed me to visualise the activities of both K92A and K238G on a single gel (Figure 5-2). As expected, K238G produced significant ligated product, while K92A produced no discernible product.

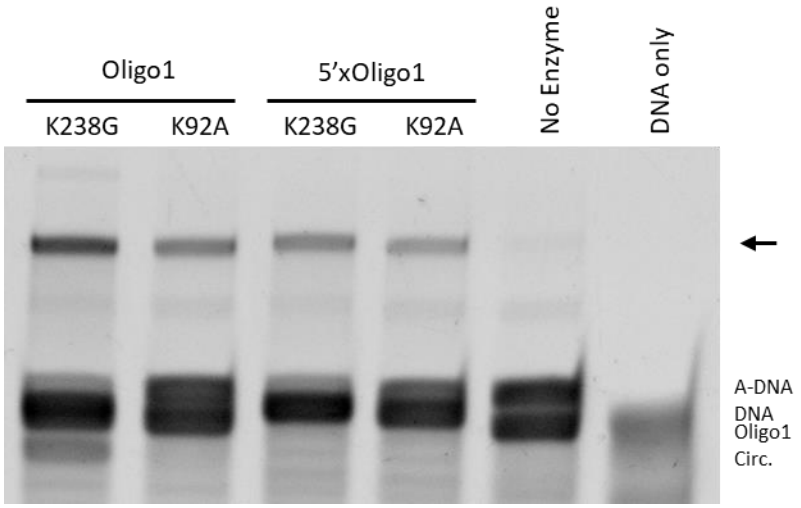

Figure 5-1 3′ ligation reaction

Ligated product is indicated by the arrow. This figure displays the results of both the regular (Oligo1) and modified (5′xOligo1) 3′ ligations with 5′ adenylated SR1 DNA (A-DNA). The results of the regular reaction show K238G generating a notably stronger ligated product band than K92A, while producing a notable circularised substrate band (circ.), and a faint concatemer product band (located at the very top of the lane). The results of the modified reaction show K238G producing less ligated product band, while lacking the circularise substrate and concatemer bands, and K92A producing a ligated product band of similar significance.

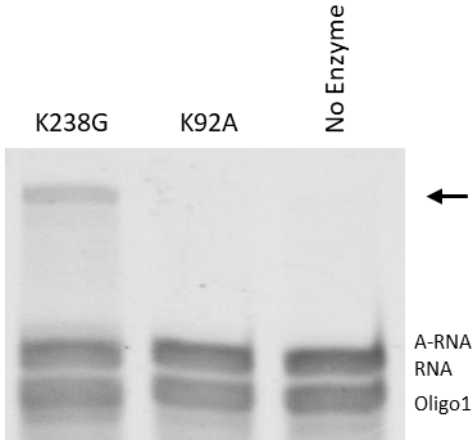

Figure 5-2 5′ adapter ligation

Ligated product is indicated by the arrow. This figure displays the results of the 5′ adapter ligation assay, utilising SR1 RNA and 5′ dephosphorylated Oligo1. K238G displays activity with the provided substrates, producing both adenylated RNA (A-RNA) and ligated product, while K92 does not.

5.3 Discussion

This chapter aimed to inform the selection of one or two variants for use in the final sequencing experiment. Unfortunately, the required activity for each adapter ligation is not the same, meaning a single enzyme solution was unlikely. By comparing the activity patterns of the most promising K92 and K238 variants, as identified in Chapters 3 and 4, I was able to determine the variant best suited to each adapter ligation for future sequencing experiments.

It is important to note that only microRNA sequences that have successfully had 3′ adapters ligated to them are carried forward into the second ligation reaction. All un-ligated microRNA and adapter sequences are removed, ensuring that the only thing the 5′ adapter can be ligated to is the microRNA-adapter product of the 3′ ligation. Only microRNA sequences that also have the 5′ adapter ligated to them can be reverse transcribed, and successfully amplified. This means that the pool of microRNAs gets progressively smaller after each step. For example, if 3′ adapters are successfully ligated to half of the available microRNAs during the 3′ adapter ligation, then the total number of microRNAs available for the 5′ adapter ligation will be half of the initial sample. As a result, half of all microRNAs within the sample have already been lost. If half of the microRNAs with 3′ adapters ligated then successfully have a 5′ adapter ligated to them, only one quarter of the initial microRNA sample will have had both adapters ligated and be ready for sequencing.

As neither the 3′ nor 5′ adapter ligations can successfully ligate all adapter sequences, a significant proportion of microRNA is expected to be lost in each ligation reaction. As accurate quantification of rates of ligation has not been achieved, the exact proportion of microRNAs likely to be lost is unknown. Selection of variants which produce the most significant product band are therefore going ensure that the maximal amount of microRNA is successfully prepared for sequencing.

5.3.1 3′ adapter ligation

Determining the variant for the 3′ adapter ligation was not straight forward. This reaction requires the ligase enzyme to ligate a pre-adenylated adapter sequence to an unmodified microRNA sequence in the absence of ATP, making ligation activity the primary criterion. As ligase adenylation activity is associated with the generation of unwanted ligation and circularisation products, a minimal adenylation, and therefore de-adenylation activity is also preferable.

While K238G appears to display greater levels of ligation activity with the desired substrates than K92A in the unmodified assay, the presence of circularised substrate indicates that some of the observed ligated product is not the desired product. As the 3′-amino modification of the SR1 DNA substrate prevents its circularisation, the observed circularised substrate must therefore be Oligo1, indicating off target adenylation. The absence of ATP in this reaction prevents K238G from using ATP to form the ligase-AMP intermediate (with K92 adenylated) required to adenylate, and therefore circularise Oligo1 (refer to Section 1.1.3 and Figure 1-1). Therefore, for K238G to produce circularised substrate, it must form the required ligase-AMP intermediate by de-adenylating SR1 DNA substrate. The fact that this happens means some of the ligation product displayed by K238G (indicated by the arrow in Figure 5-1) is likely to be two Oligo1 substrates ligated to one another.

As expected, the K238G "double block" assay lacked the circularised substrate band observed in the unmodified Oligo1 assay, while displaying a reduction in ligated product. This reduction was not as significant as expected, however, with the product band appearing to be a similar intensity to the product band produced in the unmodified K92A assay. K92A also displays a minor decrease in ligation activity in the modified "double block" assay. The SR1 RNA ligation reaction displayed in Chapter 3 (Figure 3-4A) is an exhibit of K92A displaying faint off-target ligation and circularisation activity. This faint off-target activity could explain the apparent reduction in activity.

While off-target activities are not desirable, their impact on the sequencing result may be negligible. Comparing Figures 4-4 (SR1 RNA ligation assays) and 4-8A (SR1 RNA "double block" ligation assays), it was determined that the amount of Oligo1-SR1 RNA product generated in the "double-blocked" assays was similar, if not identical to the amount generated in the equivalent standard ligation assays. This indicates that the undesirable activities are unlikely to alter the rate at which the desired microRNA-adapter products are generated by the ligase. As such, K238G may produce more microRNA-adapter product than K92A, despite the off-target activities. Assuming the undesirable ligation and circularisation activities exhibited by K238G are randomly distributed amongst all microRNA substrates (i.e., no microRNA sequences are favoured for adenylation and ligation), and therefore do not introduce a bias into the microRNA-adapter ligation, this enzyme may prove to be the better choice. However, as the distribution of these activities is unknown, K92A was selected for use in the sequencing experiment, avoiding any potential complications of K238G's off target activities.

While selecting the ligase for the first adapter ligation is important, a choice also needed to be made between RNA or DNA adapter sequences. While all variants appear to display a significantly greater level of activity with RNA adapters, the convenience of DNA adapters makes the decision significantly more complex than it appears. All current microRNA sequencing protocols utilise a DNA adapter for the 3′ ligation, meaning that a drop-in replacement for the ligases currently in use would need to be able to utilise the same DNA adapters. DNA is useful for several reasons, including its stability and affordability. Not only can it be stored at significantly higher temperatures (-20 °C vs -80 °C), it is also less susceptible to degradation (MeridianBioscience) and cheaper to purchase (Macrogen inc.). As we hope that our ligases can act as drop-in replacements for the ligases currently in use, DNA adapters would be the ideal choice over RNA.

Due to K92A displaying consistent activity with DNA adapters, these adapters will be used for the 3′ adapter ligation in future sequencing trials. This selection was made knowing that the use of an RNA adapter would likely result in a larger proportion of microRNAs being available for the 5′ adapter ligation reaction. Assuming the 3′ adapter, be it RNA or DNA, is ligated randomly and without bias to microRNA sequences, even limited production of adapter flanked microRNAs should still provide an unbiased profile of microRNAs. The overall benefit of DNA adapters made this decision relatively straight forward. If preparation of the microRNA sequences fails, changing to an RNA adapter will be one of the first alterations to the protocol considered.

5.3.2 5′ adapter ligation

The information gathered while constructing Chapters 3 and 4 of this thesis indicated that the best K238 variant was likely going to be the best option for the 5′ adapter ligation. This reaction requires adenylation activity, meaning that any variant having been identified as lacking adenylation could not be used. While this effectively ruled out the use of a K92 variant, no direct comparisons had yet been made between the K92 and K238 variants. To ensure all bases were covered, K92A and K238G were utilised in SR1 RNA “double block” assays and compared side by side on a single gel (Figure 5-2), where it was immediately obvious that K238G was the better choice. This confirmed K238G as the variant to use for the 5′ adapter ligation.

Chapter 6 Key Findings, Future Directions, and Conclusions

6.1 Research motivations

This thesis aimed to inform the selection of *Pfu* RNA ligase variants that could be utilised in high temperature microRNA sequencing protocols. Current methods of microRNA sequencing are plagued by bias, producing results that disproportionately represent nearly all microRNAs within a sample (Fuchs *et al.*, 2015; Raabe *et al.*, 2014). As a result, quantification of microRNAs is nearly impossible, limiting the potential applications and overall usefulness of the technique. This bias has been shown to be a product of structural differences between microRNAs, with RNA ligase enzymes expressing difficulty ligating the desired adapter sequences to all types of microRNA tertiary structures equally (Argyropoulos *et al.*, 2017; Fuchs *et al.*, 2015; Hafner *et al.*, 2011; Raabe *et al.*, 2014; Zhuang *et al.*, 2012). This results in the disproportionate ligation of adapter sequences to certain microRNA sequences, resulting in the misrepresentation of microRNAs within a sample. By carrying out the adapter ligation reactions at temperatures high enough to melt even the most stable microRNA secondary structures, structural bias can theoretically be eliminated. We believe this will allow for accurate microRNA profiles to be generated, broadening the scope of microRNA sequencing application, and increasing the overall usefulness of the technique. This has the potential to provide an avenue for novel diagnostics, amongst other applications. To achieve this, an RNA ligase is required that can function desirably at the required temperatures. Initial experiments with *Pfu* RNA ligase, carried out by Dr Oulavillickal, identified an optimal balance between temperature and ligation activity between 70 and 75 ^{o}C. While it is uncertain if this temperature range is high enough to ensure all microRNA secondary structures are eliminated, we believe it should be high enough to minimise structural bias, while also providing a platform from which future research can be continued.

My initial objective was to generate a site-saturation mutagenesis library of amino acid residue 92 in the *Pfu* RNA ligase enzyme. An initial K92A variant, inspired by the high temperature *Mth* RNA ligase engineered by Zhelkovsky and McReynolds, showed signs of promise in initial rounds of testing. Some undesirable activity was retained, however. By generating the site-saturation mutagenesis library, I hoped to optimise the substitution of K92, while generating information that could be used to inform the mutagenesis of K238.

6.2 Key findings

The required activities for the two adapter ligations in the sequencing protocol are not identical, meaning that a single enzyme solution was unlikely. By comparing the activity patterns of the most promising variants, as identified in Chapters 3 and 4, I was able to determine that a combination of variants K92A and K238G was the most likely to successfully ligate both adapter sequences to a random pool of microRNAs.

6.2.1 Activity assay analysis

Due to the nature of the endpoint TBE-urea gel-based assays, the limitations of the software available to analyse them, and potentially the limitations of the stain used to visualise product bands, quantification of assay results proved to be much more difficult than initially thought. While attempting to quantify the intensity of each band, such that rates of activity could be more accurately determined, it was determined that units of intensity could not be allocated a meaningful value with the information available.

For example, if the sum intensity of substrate bands within a no-enzyme control was assigned a value of 1, representing the total amount of substrate added to each reaction, then the sum intensity of substrate and product bands within any given lane should also equal 1. This was not the case, however, with the sum intensity of some reactions equalling 2 or 3, indicating one of two things: either a nonlinear relationship between the amount of substrate or product in a band, and band intensity exists, or the intensity is reaching a point of saturation. In certain scenarios, the intensity of Oligo1 substrate bands were observed to be exactly the same, or even marginally more intense in the activity assays than in the no-enzyme control. As the no-enzyme controls were an aliquot of the same master mix used for the reaction samples, and the reactions generated other visible products using the Oligo1 substrate, the intensity of this band should have been observed to decrease. Such similar intensities being observed in reaction samples identifies saturation as the likely explanation.

These issues drove the development of a new, semi-quantitative approach, in which all proteins were assayed in parallel using the same master mix. Ligation activities could then be directly compared, allowing the best variants to be identified.

6.2.2 *Pfu* RNA ligase K92

Substitution of K92 effectively reduced the adenylation and de-adenylation activities of all variants, but often failed to eliminate it. The adenylation activity observed by Dr

Oulavallickal in her initial screening of K92A persisted in many of the generated variants, with elimination of the activity typically coming at the cost of ligation activity.

All variants displaying the desired elimination of adenylation and de-adenylation activities lacked the desired ligation activity for use in the sequencing experiments. The one exception to this was variant K92Y, which displayed significant ligation activity, and no apparent adenylation or de-adenylation activities when provided with pre-adenylated Oligo1. Unfortunately, this variant did not display significant activity with pre-adenylated SR1 RNA or DNA sequences, ruling it out as a potential contender for use in the sequencing experiments. These results indicate that eliminating the remaining adenylation activity of K92A, while retaining significant ligation activity, cannot be achieved through optimisation of residue 92. Elimination of this activity is likely only achievable through further mutagenesis of the enzyme.

While K92A displayed the highest rate of adenylation and de-adenylation of the K92 variants, it also displayed the most significant ligation activity. While K92 G, S and T came close, displaying ligation activities with relative similarity to K92A, the significance of K92A's ligation activity was determined to outweigh the reduced adenylation and de-adenylation activities of these variants. The only similarity between these four amino acids (A, G, S, and T) is their small size, while their side-chain characteristics vary significantly. This indicates that alterations to activity, other than the loss of adenylation activity associated with the wild type amino acid, are likely due to minor conformational changes within the active site. Substituting the larger wild type amino acid, K, for something smaller, may allow the active site of the ligase to assume a structure more beneficial for catalysing ligation. It is important to note that none of these amino acids are negatively charged like the wild type. While this is almost certainly due to the larger size of the other negatively charged amino acids, it also provides evidence that the characteristic of amino acid side chain is likely unimportant for determining the best substitution.

6.2.3 *Pfu* RNA ligase K238

As identified by Zhelkovsky and McReynolds (2012) both K97 and K246 in the Mth RNA ligase (equivalent to K92 and K238 in the *Pfu* RNA ligase respectively) are capable of being adenylated, and therefore catalysing adenylation activity. Substitution of K97 eliminated the adenylation and de-adenylation activities of the ligase, identifying it as the key residue involved in adenylation, while substitution of K246 reduced de-adenylation activity but failed

to eliminate it. As substitution of K92 in the *Pfu* RNA ligase generated similar results to the substitution of K97 in *Mth* RNA ligase, substitution of K238 was therefore expected to generate similar results to the substitution of K246.

Surprisingly, K238 variants A, G, S, and T displayed an apparent increase in adenylation and de-adenylation activities relative to the wild type, accompanied by an apparent increase to other observed activities, including ligation and circularisation. Variants A and G displayed the most significant activity of the K238 variants. In assays containing SR1 oligonucleotide (RNA or DNA) and unmodified Oligo1 substrates (Chapter 4, Figures 4-4 and 4-5), these variants displayed equivalent levels of off-target ligation and circularisation activity to the wild type, while generating a notably larger amount of the desired ligation product.

As we believe both K92 and K238 are capable of being adenylated, and therefore catalysing the adenylation reaction, it is possible these residues "compete" for catalytic activity in the wild type. As K92 is the favoured substrate, capable of more efficiently catalysing the reaction than K238, it will have a higher affinity for the substrates involved. This, however, may not prevent K238 from occasionally being able to catalyse an adenylation reaction. As this residue is involved in a less efficient reaction, this would reduce the overall catalytic efficiency of the reaction. Removal of K238 may therefore eliminate this competition, allowing K92 to catalyse the reaction uninhibited.

K238 substitutions A and G are thought to be the best due to being the smallest of the amino acids. The non-disruptive nature of the substitutions minimises conformational changes to the active site, allowing the enzyme to retain an optimal conformation for ligation, while eliminating the K238.

6.2.4 Double mutants

Pfu RNA ligase variants lacking both lysine residues of interest were generated with the aim to profile the activity of the K238 residue more accurately. As several K92 variants retained some adenylation and de-adenylation activity, it was theorised that K238 must be catalysing the activity (albeit much less efficiently) in the absence of K92. Removal of both lysine residues should therefore eliminate all adenylation activity. Due to the K97A, K246A *Mth* RNA ligase double mutant generated by Zhelkovsky and McReynolds (2012) displaying a total loss of function, we were not expecting any of the double mutants generated to display significant activities.

Substitution of both lysine residues proved to eliminate adenylation and de-adenylation activities in both the K92A, K238A (AA) variant, and the K92A, K238Y (AY) variant, as expected. Assays run in the presence of ATP, with only non-adenylated Oligo1 as the substrate, failed to produce any sign of adenylated substrate, circularised substrate, or ligated product, indicating an absence of adenylation activity. Unexpectedly, when provided with only pre-adenylated Oligo1 substrate, both double mutants displayed signs of ligation activity (Figure 4-3C). While the ligated product bands produced by these variants were not particularly significant, relative to K92A or K238G, these results provided valuable information on the role of the lysine residues in ligation. The drastic decrease in ligation activity indicates that, while not necessary for catalysing ligation, at least one of these lysine residues is required for the enzyme to correctly orientate substrate within its active site. As such, absence of both residues may result in the ligase being unable to orientate substrate actively/efficiently within the active site, lowering the efficiency of ligation without inhibiting the ligation mechanism.

This result supports the hypothesis that K238 is responsible for the adenylation activity observed in the K92A single variant. Co-substituting the K238 residue of the K92A variant did reduce the ligation activity of the enzyme, however, the failure to eliminate it indicates the enzyme is still functional. Therefore, the elimination of the adenylation activity present in the "parent" K92A variant is most likely due to the lack of both lysine residues, not conformational changes.

This conclusion, while not identical to that reached by Zhelkovsky and McReynolds, does share some of its principles. Their paper also suggested that both lysine residues are involved in catalytic activity of the enzyme, and that interplay between the two allows for one to be removed without significantly altering ligation activity, however, they also suggest that one of the two lysine residues is required for ligation activity. While this may be true for *Mth*, my results show that this is not the case for *Pfu*.

6.2.5 Protein stability

The incredible stability of the *Pfu* RNA ligase is a characteristic that lends itself to use in a commercial sequencing kit. Most samples purified throughout this thesis retained functionality throughout the length of this project, being stored at 4 ^{o}C when not in use. The only variants displaying notable signs of degradation were those that were either already inactive, such as K92H, or those displaying signs of contamination. In the year of preparation

for this thesis, all generated protein samples had an aliquot stored at room temperature. Most of these continued to display signs of functionality six months post purification.

6.3 Critical evaluation of methodology

6.3.1 Protein purification protocol

The batch purification protocol employed to purify target proteins was used to great effect, however, the impact of the reduced protein purity on activity is uncertain. Ideally, all 20 proteins would have been purified using HisTrap™ high performance immobilised metal affinity columns on an ÄKTA automated liquid chromatography system. This would have reduced the variance introduced by manual gravity flow purifications, as well as producing a higher purity product. Due to time constraints, and the number of different individual proteins requiring purification, this was not achievable.

While the batch purification protocol allowed purifications to be carried out in parallel, reducing the total time spent purifying all desired proteins, this protocol primarily functioned to reduce the effect of protein degradation on comparative assays. By shrinking the timeframe within which all variants could be purified, the potential effects of protein degradation could be minimised. While this was successful, all proteins appeared to retain similar, if not identical functionality for several months following purification. This stability indicates that it may have been possible to pursue a less parallel approach of protein purification, allowing purer product to be compared without sacrificing the activity of the proteins. As such, better comparisons may have been possible. If I were to conduct this research again, I would likely utilise a batch purification technique for the initial screening of the entire K92 library, then HisTrap™ ÄKTA purifications to repurify the most promising variants, allowing for more accurate comparisons to be made.

6.3.2 Evaluating ligation using endpoint TBE urea-PAGE assays

The means by which activity was assessed throughout this thesis provided enough data to inform the final variant selection for use in future sequencing experiments. This, however, does not mean there is no better method. Should experiments like these be carried out in the future, a means by which activity can be accurately quantified would be preferable to the methods employed in this thesis. A possible solution would be the use of a fluorescently labelled adapter. This would allow an exact level of fluorescence to be determined, allowing for accurate quantification of ligation activity (Sharma *et al.*, 2020; Tang *et al.*, 2003; Williamson *et al.*, 2014). While this solution may provide an accurate assessment of ligation

activity, any activity with non-labelled substrate would not be identified. This would limit the ability of the assay to detect off-target activities associated with ligase enzyme re-adenylation.

6.4 Future directions

The results displayed in this thesis have been utilised to inform the selection of *Pfu* RNA ligase variants for use in high temperature microRNA sequencing experiments. The two variants of interest, K92A and K238G, are going to be utilised in a series of experiments designed to assess their effectiveness in real microRNA sequencing applications. Using combinations of our ligases and those used in the current microRNA sequencing protocols, we hope to show that both variants display the required activity to generate an unbiased pool of adapter flanked microRNA sequences which can then be amplified and sequenced.

Using a pool of designer oligonucleotides as microRNA equivalents, we will be able to ensure that the selected sequences have a broad range of T_ms, extending right up to the limit of microRNA T_ms. Assuming our ligase variants successfully generate a pool of adapter flanked "microRNAs" at 70 ^{o}C, the sequencing results should provide insight as to whether the proposed protocol will remove the structure related bias observed in current protocols.

If it becomes necessary to improve the overall activities of the K92A or K238G *Pfu* RNA ligase variants, a potential avenue is directed evolution. Discussed in Chapter 1, this technique could be used to screen for further improvements introduced through the random mutagenesis of the selected *Pfu* RNA ligase variant. While this technique does not guarantee a desired improvement in functionality, it may lead to improvements currently unachievable by means of semi-rational or rational design. It is possible that this avenue of mutagenesis may be required to achieve our eventual goal of identifying a ligase capable of displaying equivalent, or greater levels of activity than those used in current sequencing experiments, while at temperatures exceeding the identified 70 ^{o}C threshold.

Assuming our protocol effectively removes the structural bias observed in current microRNA sequencing protocols, we hope to generate an industry standard kit that can be utilised for all microRNA sequencing applications. Not only will this aid in the accurate generation of microRNA profiles, increasing our ability to understand the role they play in biological systems, but also provide the foundations upon which new bio-marker diagnostic techniques can be developed. This could provide the means for diagnosing numerous ailments and diseases currently without accurate means of detection.

6.5 Concluding remarks

This thesis effectively demonstrate three things: firstly, that K92A is the optimal substitution of *Pfu* RNA ligase residue 92 for limiting adenylation activity and maximising ligation activity; secondly, that K238 substitutions express higher levels of activity, including adenylation, ligation and circularisation, than the wild type; and thirdly, that substitution of both K92 and K238 eliminates any semblance of adenylation activity, while also reducing ligation activity of the *Pfu* RNA ligase to insignificant levels.

Assuming the limited adenylation activity displayed by variant K92A was catalysed by K238, it is possible that both lysine residues are capable of catalysing adenylation in the wild type, resulting in decreased catalytic efficiency due to a sharing of catalytic activity. Substitution of K238 for a non-disruptive amino acid, such as alanine or glycine, would therefore provide K92 with full catalytic control, improving the catalytic efficiency of the ligase. This could function to increase the overall activity of the ligase, supporting the hypothesis that K238 has minor adenylation capabilities.

While I was unable to generate a ligase variant better suited for the 3′ ligation reaction than K92A, confirmation of this as the best possible substitution was still an important result. Characterisation of the entire K92 site-saturation mutagenesis library also provided the information required for the rational mutagenesis of K238. Hopefully, this information will aid the rational mutagenesis of other ligase enzymes in the future.

Further work is likely going to be required to identify a ligase capable of functioning at temperatures upwards of 70 ^{o}C while matching the activity of the ligases used in current protocols, however, this research provides a solid foundation for the work to come.

References

Argyropoulos, C., Etheridge, A., Sakhanenko, N., & Galas, D. (2017). Modeling Bias and Variation in the Stochastic Processes of Small RNA Sequencing. *Nucleic Acids Research, 45*(11), e104-e104. doi:10.1093/nar/gkx199

Arnold, F. H. (2018). Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie (International ed.), 57*(16), 4143-4148. doi:10.1002/anie.201708408

Baran-Gale, J., Kurtz, C. L., Erdos, M. R., Sison, C., Young, A., Fannin, E. E., . . . Sethupathy, P. (2015). Addressing Bias in Small RNA Library Preparation for Sequencing: A New Protocol Recovers MicroRNAs That Evade Capture by Current Methods. *Frontiers in Genetics, 6*, 352-352. doi:10.3389/fgene.2015.00352

Barrozo, A., Borstnar, R., Marloie, G., & Kamerlin, S. C. L. (2012). Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution. *International Journal of Molecular Sciences, 13*(10), 12428-12460. doi:10.3390/ijms131012428

Bellacosa, A., & Moss, E. G. (2003). RNA Repair: Damage Control. *Current Biology, 13*(12), R482-R484. doi:10.1016/S0960-9822(03)00408-1

Benz, J. E. J. (1990). The Molecular Genetics of Cancer. Introduction to Principles of Recombinant DNA Technology. *Cancer, 65*(3 Suppl), 731-741.

Bjaanaes, M. M., Halvorsen, A. R., Solberg, S., Jørgensen, L., Dragani, T. A., Galvan, A., . . . Helland, Å. (2014). Unique MicroRNA-Profiles in Egfr -Mutated Lung Adenocarcinomas. *International Journal of Cancer, 135*(8), 1812-1821. doi:10.1002/ijc.28828

Brock, T. D., & Freeze, H. (1969). *Thermus Aquaticus* Gen. N. And Sp. N., a Nonsporulating Extreme Thermophile. *Journal of Bacteriology, 98*(1), 289-297. doi:10.1128/jb.98.1.289-297.1969

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*(8), 1336-1345. doi:10.1110/ps.035493.108

Carbonell, J., Alloza, E., Arce, P., Borrego, S., Santoyo, J., Ruiz-Ferrer, M., . . . Dopazo, J. (2012). A Map of Human MicroRNA Variation Uncovers Unexpectedly High Levels of Variability. *Genome Medicine, 4*(8), 62-62. doi:10.1186/gm363

Carroll, W. L. (1993). Introduction to Recombinant-DNA Technology. *American Society for Clinical Nutrition, 58*, 249-258.

Cervelli, T., Borghini, A., Galli, A., & Andreassi, M. (2012). DNA Damage and Repair in Atherosclerosis: Current Insights and Future Perspectives. *International Journal of Molecular Sciences, 13*(12), 16929-16944. doi:10.3390/ijms131216929

Chauleau, M., & Shuman, S. (2013). Kinetic Mechanism of Nick Sealing by T4 RNA Ligase 2 and Effects of 3'-Oh Base Mispairs and Damaged Base Lesions. *RNA (Cambridge), 19*(12), 1840-1847. doi:10.1261/rna.041731.113

Chen, H., & Zhang, Z. (2013). Similarity-Based Methods for Potential Human MicroRNA-Disease Association Prediction. *BMC Medical Genomics, 6*(1), 12-12. doi:10.1186/1755-8794-6-12

Chen, M., Liao, B., & Li, Z. (2018). Global Similarity Method Based on a Two-Tier Random Walk for the Prediction of MicroRNA–Disease Association. *Scientific Reports, 8*(1), 6481-6416. doi:10.1038/s41598-018-24532-7

Chen, X., & Yan, G.-Y. (2014). Semi-Supervised Learning for Potential Human MicroRNA-Disease Associations Inference. *Scientific Reports, 4*(1), 5501-5501. doi:10.1038/srep05501

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

Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic Acid Polymerase from the Extreme Thermophile *Thermus Aquaticus*. *Journal of Bacteriology, 127*(3), 1550-1557. doi:10.1128/jb.127.3.1550-1557.1976

Cuk, K., Zucknick, M., Heil, J., Madhavan, D., Schott, S., Turchinovich, A., . . . Burwinkel, B. (2013). Circulating MicroRNAs in Plasma as Early Detection Markers for Breast Cancer. *International Journal of Cancer, 132*(7), 1602-1612. doi:10.1002/ijc.27799

Dalmaso, G., Ferreira, D., & Vermelho, A. (2015). Marine Extremophiles: A Source of Hydrolases for Biotechnological Applications. *Marine Drugs, 13*(4), 1925-1965. doi:10.3390/md13041925

David, A. J., Robert, H. S., & Paul, B. (1972). Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular Sv40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia Coli. Proceedings of the National Academy of Sciences - PNAS, 69*(10), 2904-2909. doi:10.1073/pnas.69.10.2904

Dombkowski, A. A. (2003). Disulfide by Design: A Computational Method for the Rational Design of Disulfide Bonds in Proteins. *Bioinformatics (Oxford, England), 19*(14), 1852-1853. doi:10.1093/bioinformatics/btg231

Egamberdieva, D., Birkeland, N.-K., Panosyan, H., & Li, W.-J. (2018). *Extremophiles in Eurasian Ecosystems: Ecology, Diversity, and Applications*. Singapore: Springer Singapore.

Ellenberger, T., & Tomkinson, A. E. (2008). Eukaryotic DNA Ligases: Structural and Functional Insights. *Annual Review of Biochemistry, 77*(1), 313-338. doi:10.1146/annurev.biochem.77.061306.123941

Esquela-Kerscher, A., & Slack, F. J. (2006). Oncomirs — MicroRNAs with a Role in Cancer. *Nature Reviews. Cancer, 6*(4), 259-269. doi:10.1038/nrc1840

Fiala, G., & Stetter, K. O. (1986). *Pyrococcus Furiosus* Sp. Nov. Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100 C. *Archives of Microbiology, 145*(1), 56-61. doi:10.1007/bf00413027

Fuchs, R. T., Sun, Z., Zhuang, F., & Robb, G. B. (2015). Bias in Ligation-Based Small RNA Sequencing Library Construction Is Determined by Adaptor and RNA Structure. *PloS One, 10*(5), e0126049-e0126049. doi:10.1371/journal.pone.0126049

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature Methods, 6*(5), 343-345. doi:10.1038/nmeth.1318

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. doi:10.1093/nar/gkw094

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., . . . Tuschl, T. (2011). RNA-Ligase-Dependent Biases in MiRNA Representation in Deep-Sequenced Small RNA CDNA Libraries. *RNA (Cambridge), 17*(9), 1697-1712. doi:10.1261/rna.2799511

Hammond, S. M. (2015). An Overview of MicroRNAs. *Advanced Drug Delivery Reviews, 87*, 3-14. doi:10.1016/j.addr.2015.05.001

Ho, C. K., & Stewart, S. (2002). Bacteriophage T4 RNA Ligase 2 (Gp24.1) Exemplifies a Family of RNA Ligases Found in All Phylogenetic Domains. *Proceedings of the National Academy of Sciences - PNAS, 99*(20), 12709-12714. doi:10.1073/pnas.192184699

Ho, C. K., Wang, L. K., Lima, C. D., & Shuman, S. (2004). Structure and Mechanism of RNA Ligase. *Structure (London), 12*(2), 327-339. doi:10.1016/j.str.2004.01.011

Holden, J. F. (2019). Extremophiles: Hot Environments. In (pp. 127-146).

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-e141. doi:10.1093/nar/gkr693

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., . . . Liu, Y. (2009). Mir2disease: A Manually Curated Database for MicroRNA Deregulation in Human Disease. *Nucleic Acids Research, 37*(Database), D98-D104. doi:10.1093/nar/gkn714

Kanaoka, S., Kuriyama, S., Iwaizumi, M., Yamada, T., Sugimoto, M., Osawa, S., . . . Miyajima, H. (2013). Mo1179 Potential Usefulness of Fecal Immunochemical Test Plus Fecal MicroRNA Assay as a Marker for Colorectal Cancer Screening. *Gastroenterology (New York, N.Y. 1943), 144*(5), S-599-S-600. doi:10.1016/S0016-5085(13)62212-X

Kaufmann, W. K., & Paules, R. S. (1996). DNA Damage and Cell Cycle Checkpoints. *The FASEB Journal, 10*(2), 238-247. doi:10.1096/fasebj.10.2.8641557

Kojima, M., Sudo, H., Kawauchi, J., Takizawa, S., Kondou, S., Nobumasa, H., & Ochiai, A. (2015). MicroRNA Markers for the Diagnosis of Pancreatic and Biliary-Tract Cancers. *PloS One, 10*(2), e0118220-e0118220. doi:10.1371/journal.pone.0118220

Kuchner, O., & Arnold, F. H. (1997). Directed Evolution of Enzyme Catalysts. *Trends in Biotechnology (Regular ed.), 15*(12), 523-530. doi:10.1016/s0167-7799(97)01138-4

Link, A., Balaguer, F., Shen, Y., Nagasaka, T., Lozano, J. J., Boland, C. R., & Goel, A. (2010). Fecal MicroRNAs as Novel Biomarkers for Colon Cancer Screening. *Cancer Epidemiology, Biomarkers & Prevention, 19*(7), 1766-1774. doi:10.1158/1055-9965.epi-10-0027

Linnarsson, S. (2010). Recent Advances in DNA Sequencing Methods – General Principles of Sample Preparation. *Experimental Cell Research, 316*(8), 1339-1343. doi:10.1016/j.yexcr.2010.02.036

Linsen, S. E. V., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., . . . Cuppen, E. (2009). Limitations and Possibilities of Small RNA Digital Gene Expression Profiling. *Nature Methods, 6*(7), 474-476. doi:10.1038/nmeth0709-474

Liu, H., & Naismith, J. H. (2008). An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnology, 8*(1), 91-91. doi:10.1186/1472-6750-8-91

Liz, J., & Esteller, M. (2016). LncRNAs and MicroRNAs with a Role in Cancer Development. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms, 1859*(1), 169-176. doi:10.1016/j.bbagrm.2015.06.015

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., . . . Golub, T. R. (2005). MicroRNA Expression Profiles Classify Human Cancers. *Nature (London), 435*(7043), 834-838. doi:10.1038/nature03702

Lutz, S. (2010). Beyond Directed Evolution—Semi-Rational Protein Engineering and Design. *Current Opinion in Biotechnology, 21*(6), 734-743. doi:10.1016/j.copbio.2010.08.011

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., . . . Rothberg, J. M. (2005). Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature (London), 437*(7057), 376-380. doi:10.1038/nature03959

Martin, I. V., & MacNeill, S. A. (2002). Atp-Dependent DNA Ligases. *Genome Biology, 3*(4), REVIEWS3005-REVIEWS3005. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139351/pdf/gb-2002-3-4-reviews3005.pdf

Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *Journal of Molecular Biology, 288*(5), 911-940. doi:10.1006/jmbi.1999.2700

MeridianBioscience. Working with RNA. Retrieved from https://www.bioline.com/rna-hints-and-tips

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 (Cambridge), 127*(1), 71-84. doi:10.1016/j.cell.2006.08.038

Newman, M. A., Mani, V., & Hammond, S. M. (2011). Deep Sequencing of MicroRNA Precursors Reveals Extensive 3' End Modification. *RNA (Cambridge), 17*(10), 1795-1803. doi:10.1261/rna.2713611

Pascal, J. M. (2008). DNA and RNA Ligases: Structural Variations and Shared Mechanisms. *Current Opinion in Structural Biology, 18*(1), 96-105. doi:10.1016/j.sbi.2007.12.008

Patrick, S. M., Rachael, K. P., Evan, M. K., Brian, R. F., Stacia, K. W., Era, L. P.-A., . . . Muneesh, T. (2008). Circulating MicroRNAs as Stable Blood-Based Markers for Cancer Detection. *Proceedings of the National Academy of Sciences - PNAS, 105*(30), 10513-10518. doi:10.1073/pnas.0804549105

Perkhofer, L., Beutel, A. K., & Ettrich, T. J. (2019). Immunotherapy: Pancreatic Cancer and Extrahepatic Biliary Tract Cancer. *Visceral Medicine, 35*, 28-37. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6597904/pdf/vis-0035-0028.pdf

Phua, L. C., Chue, X. P., Koh, P. K., Cheah, P. Y., Chan, E. C. Y., & Ho, H. K. (2014). Global Fecal MicroRNA Profiling in the Identification of Biomarkers for Colorectal Cancer Screening among Asians. *Oncology Reports, 32*(1), 97-104. doi:10.3892/or.2014.3193

Raabe, C. A., Tang, T.-H., Brosius, J., & Rozhdestvensky, T. S. (2014). Biases in Small RNA Deep Sequencing Data. *Nucleic Acids Research, 42*(3), 1414-1426. doi:10.1093/nar/gkt1021

Rogers, E., & Heitsch, C. (2016). New Insights from Cluster Analysis Methods for RNA Secondary Structure Prediction. *Wiley Interdisciplinary Reviews. RNA, 7*(3), 278-294. doi:10.1002/wrna.1334

Rupaimoole, R., & Slack, F. J. (2017). MicroRNA Therapeutics: Towards a New Era for the Management of Cancer and Other Diseases. *Nature Reviews. Drug Discovery, 16*(3), 203-222. doi:10.1038/nrd.2016.246

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., . . . Erlich, H. A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science (American Association for the Advancement of Science), 239*(4839), 487-491. doi:10.1126/science.239.4839.487

Salta, E., & De Strooper, B. (2017). Noncoding RNAs in Neurodegeneration. *Nature Reviews. Neuroscience, 18*(10), 627-640. doi:10.1038/nrn.2017.90

Sanger, F., & Coulson, A. (1975). A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *Journal of Molecular Biology, 94*, 441-448.

Sharma, J. K., Stevenson, L. J., Robins, K. J., Williamson, A. K., Patrick, W. M., & Ackerley, D. F. (2020). Methods for Competitive Enrichment and Evaluation of Superior DNA Ligases. *Methods in Enzymology, 644*, 209-225. doi:10.1016/bs.mie.2020.04.061

Shenmin, Y., Ho, C. K., & Stewart, S. (2003). Structure-Function Analysis of T4 RNA Ligase 2. *Journal of Biological Chemistry, 278*(20), 17601-17608. doi:10.1074/jbc.M300817200

Shuman, S. (2009). DNA Ligases: Progress and Prospects. *The Journal of Biological Chemistry, 284*(26), 17365-17369. doi:10.1074/jbc.r900017200

Singh, J., Hanson, J., Paliwal, K., & Zhou, Y. (2019). RNA Secondary Structure Prediction Using an Ensemble of Two-Dimensional Deep Neural Networks and Transfer Learning. *Nature Communications, 10*(1), 5407-5413. doi:10.1038/s41467-019-13395-9

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), 4-4. doi:10.1186/1758-907X-3-4

Sriskanda, & Shuman, S. (1998). Mutational Analysis of Chlorella Virus DNA Ligase: Catalytic Roles of Domain I and Motif Vi. *Nucleic Acids Research, 26*(20), 4618-4625. doi:10.1093/nar/26.20.4618

Sriskanda, & Shuman, S. (2001). A Second Nad(+)-Dependent DNA Ligase (Ligb) in *Escherichia Coli. Nucleic Acids Research, 29*(24), 4930-4934. doi:10.1093/nar/29.24.4930

Stetter, K. O. (1999). Extremophiles and Their Adaptation to Hot Environments. *FEBS Letters, 452*(1-2), 22-25. doi:10.1016/S0014-5793(99)00663-8

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., . . . Takahashi, T. (2004). Reduced Expression of the Let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival. *Cancer Research (Chicago, Ill.), 64*(11), 3753-3756. doi:10.1158/0008-5472.can-04-0637

Tan, Z., Fu, Y., Sharma, G., & Mathews, D. H. (2017). Turbofold Ii: RNA Structural Alignment and Secondary Structure Prediction Informed by Multiple Homologs. *Nucleic Acids Research, 45*(20), 11570-11581. doi:10.1093/nar/gkx815

Tang, Z. W., Wang, K. M., Tan, W. H., Li, J., Liu, L. F., Guo, Q. P., . . . Huang, S. S. (2003). Real-Time Monitoring of Nucleic Acid Ligation in Homogenous Solutions Using Molecular Beacons. *Nucleic Acids Research, 31*(23), e148-148. doi:10.1093/nar/gng146

Tepus, M., & Yau, Tung O. (2020). Non-Invasive Colorectal Cancer Screening: An Overview. *Gastrointestinal Tumors, 7*(3), 62-73. doi:10.1159/000507701

Torchia, C., Takagi, Y., & Ho, C. K. (2008). Archaeal RNA Ligase Is a Homodimeric Protein That Catalyzes Intramolecular Ligation of Single-Stranded RNA and DNA. *Nucleic Acids Research, 36*(19), 6218-6227. doi:10.1093/nar/gkn602

Verma, A., Verma, A., Dhiman, K., Dhiman, K., Gupta, M., Gupta, M., . . . Shirkot, P. (2015). Bioprospecting of Thermotolerant Bacteria from Hot Water Springs of Himachal Pradesh for the Production of *Taq* DNA Polymerase. *Proceedings of the National Academy of Sciences, India. Section B: Biological sciences, 85*(3), 739-749. doi:10.1007/s40011-014-0412-x

Viollet, S., Fuchs, R. T., Munafo, D. B., Zhuang, F., & Robb, G. B. (2011). T4 RNA Ligase 2 Truncated Active Site Mutants: Improved Tools for RNA Analysis. *BMC Biotechnology, 11*(1), 72-72. doi:10.1186/1472-6750-11-72

Wang, J., Chen, J., & Sen, S. (2016). MicroRNA as Biomarkers and Diagnostics. *Journal of Cellular Physiology, 231*(1), 25-30. doi:10.1002/jcp.25056

Wang, L. K., Nandakumar, J., Schwer, B., & Shuman, S. (2007). The C-Terminal Domain of T4 RNA Ligase 1 Confers Specificity for TRNA Repair. *RNA (Cambridge), 13*(8), 1235-1244. doi:10.1261/rna.591807

Weiss, B., Thompson, A., & Richardson, C. C. (1968). Ezymatic Breakage and Joining of Deoxyribonucleic Acid. Vii. Properties of the Enzyme-Adenylate Intermediate in the Polynucleotide Ligase Reaction. *The Journal of Biological Chemistry, 243*(17), 4556-4563. Retrieved from https://www.jbc.org/content/243/17/4556.full.pdf

Wilkinson, A., Day, J., & Bowater, R. (2001). Bacterial DNA Ligases. *Molecular Microbiology, 40*(6), 1241-1248. doi:10.1046/j.1365-2958.2001.02479.x

Williamson, A., Hjerde, E., & Kahlke, T. (2016). Analysis of the Distribution and Evolution of the Atp-Dependent DNA Ligases of Bacteria Delineates a Distinct Phylogenetic Group 'Lig E'. *Molecular Microbiology, 99*(2), 274-290. doi:10.1111/mmi.13229

Williamson, A., Rothweiler, U., & Schrøder Leiros, H. K. (2014). Enzyme–Adenylate Structure of a Bacterial Atp-Dependent DNA Ligase with a Minimized DNA-Binding Surface. *Acta*

Crystallographica. Section D, Biological Crystallography., 70(11), 3043-3056. doi:10.1107/S1399004714021099

Wu, R. (1970). Nucleotide Sequence Analysis of DNA: I. Partial Sequence of the Cohesive Ends of Bacteriophage Λ and 186 DNA. *Journal of Molecular Biology, 51*(3), 501-521.

Xu, L.-h., Guo, Y., Zhang, X.-L., Chen, J.-j., & Hu, S.-y. (2016). Blood-Based Circulating MicroRNAs Are Potential Diagnostic Biomarkers for Leukemia: Result from a Meta-Analysis. *Cellular Physiology and Biochemistry, 38*(3), 939-949. doi:10.1159/000443046

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., . . . Harris, C. C. (2006). Unique MicroRNA Molecular Profiles in Lung Cancer Diagnosis and Prognosis. *Cancer Cell, 9*(3), 189-198. doi:10.1016/j.ccr.2006.01.025

Zeikus, J. G., & Wolee, R. S. (1972). *Methanobacterium Thermoautotrophicus* Sp. N., an Anaerobic, Autotrophic, Extreme Thermophile. *Journal of Bacteriology, 109*(2), 707-713. doi:10.1128/jb.109.2.707-713.1972

Zhelkovsky, A. M., & McReynolds, L. A. (2012). Structure-Function Analysis of *Methanobacterium Thermoautotrophicum* RNA Ligase - Engineering a Thermostable Atp Independent Enzyme. *BMC Molecular Biology, 13*(1), 24-24. doi:10.1186/1471-2199-13-24

Zhelkovsky, A. M., & McReynolds, L. A. (2014). Polynucleotide 3 '-Terminal Phosphate Modifications by RNA and DNA Ligases. *The Journal of Biological Chemistry, 289*(48), 33608-33616. doi:10.1074/jbc.M114.612929

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-e54. doi:10.1093/nar/gkr1263

Zuker, M. (2003). Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Research, 31*(13), 3406-3415. doi:10.1093/nar/gkg595

Appendix

A1 Pfu RNA ligase K92 site-saturation mutagenesis library

Variant	Sequence (mutations are bold and enlarged)
Wild type	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92A	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GCG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92C	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**TGC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG

	AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92D	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GAC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92E	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GAA**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92F	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**TTT**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92G	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC

	GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GGC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92H	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**CAC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92I	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**ATT**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92L	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**CTG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG

	TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92M	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**ATG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92N	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**AAC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92P	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**CCG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT

	CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92Q	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**CAA**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92R	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**CGC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92S	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**AGC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92T	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT

	AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**ACG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92V	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GTT**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92W	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**TGG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92Y	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**TAC**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT

	GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTAAAGTA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA

A2 Pfu RNA ligase K238 variants

Variant	Sequence (mutations are bold and enlarged)
K238A	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**GCG**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K238G	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**GGC**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K238S	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA

	CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**AGC**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K238T	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**ACG**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K238Y	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAGAAAGTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**TAC**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA

A3 Pfu RNA ligase double mutants

Variant	Sequence (mutations are bold and enlarged)
K92A, K238A	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GCG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**GCG**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA
K92A, K238Y	GAAAATATGGTAAGCTCGAAGTTTAAGGAGCTCCTTTATACCCTGGGAATTCCCGAAGAT AAAGTAGAGATTCTTGAAGCTAGGGGCGGAATTATGGAGGATGAGTTTGAAGGCATTAG ATATCTTAGGTTTAGAATTCAGTTGGAAAACTGAGGAGGGGAACTGTACTTTTTGAAGAC GGGACAACTGTTTTTGGCTTTCCTCCATCAAGAGGATTGTAAACCTTTCAGCTGGAGTTAG AAAAATCTTTAAGTCCTCTGAGTTCTACGTTGAGGAG**GCG**GTTGATGGGTACAATGTTAG AGTTGTGAAGTTCAAAGATAGGATTCTCGGAATAACTAGGGGAGGATTTATTTGTCCTTA CACTACCGAAAGAATTGCTGAATTCGTTCCAGAAGAATTCTTTAAAGATCATAAGGATTT GGTATTAGTTGGGGAAATGGCAGGCCCTGAAAGTCCCTATTTGGTGGAAGGACCTCCATA TGTAAAAGAGGACATACAATTCTTCCTATTTGATATACAGGACATCAAAACAGGTAGTAG TTTGCCTGTAGAGGAGAGGCTCAAGCTCGCCGAGGAGTATGGAATAAATCACGTGGAAGT CTTTGGGAGGTATTCCTATAAGGACATAGATGATCTTTATGAGCTAATTGAGAGGCTCAG CAGGGAAGGAGAGAAGGAATAGTTATGAAAAGTCCAGACATGAAGAAAATTGTA**TAC**TA CGTGACACCTTATGCGAATATTAACGATATAAAAATAGGGGCAAGAGTATTTTACGAGCT TCCTGGGGGGTACTTCACGAGCAGAATTTCGAGGTTAGCCTTCTACATAGCGGAAAAGAA AATTAGGGGGGAAGAACTCCATAACTTGGCATTGCAACTTGGGAAAGCTCTGCTTCAACCT CTTGTTGAGGCCATTCACGACGTTACCCAGGGAGATGTAATTGCAGAGAGATTCAGAGTTA GAGTTAGAAAAATTGAGACAGCATATAAGATGGTAACACACTTTGAAAAACTAGGGTTGG AAATAGAGATAGAGGATATTGAGGAGATAGAGGGAGGATGGAGAGTTACATTCAAGAGA GTTTATCCTGAAGCTACAAGAGAAATAAGAGACCTCATAGGGGGGAAGGCATTCGTTGATT AA

A4 Site-directed mutagenesis and colony PCR primers

Primer	Sequence
K92L.for	TGAGGAGCTGGTTGATGGGTACAATGTTAGAGTTGTGAAG
K92L.rev	ATCAACCAGCTCCTCAACGTAGAACTCAGAGGAC
K238A.for	AATTGTAGCGTACGTGACACCTTATGCGAATATTAACGAT
K238A.rev	CACGTACGCTACAATTTTCTTCATGTCTGGACTTTTCATAAC
K238G.for	AATTGTAGGCTACGTGACACCTTATGCGAATATTAACGAT
K238G.rev	CACGTAAGCTACAATTTTCTTCATGTCTGGACTTTTCATAAC
K238S.for	AATTGTAAGCTACGTGACACCTTATGCGAATATTAACGAT
K238S.rev	CACGTAGCTTACAATTTTCTTCATGTCTGGACTTTTCATAAC
K238T.for	AATTGTAACCTACGTGACACCTTATGCGAATATTAACGAT
K238T.rev	CACGTAGGTTACAATTTTCTTCATGTCTGGACTTTTCATAAC
K238Y.for	AATTGTATATTACGTGACACCTTATGCGAATATTAACGAT
K238Y.rev	CACGTAATATACAATTTTCTTCATGTCTGGACTTTTCATAAC
pCA24N.for	GATAACAATTTCACACAGAATTCATTAAAGAG
pCA24N.rev2	CAAATCCAGATGGAGTTCTGAGG

A5 Ligation reaction substrates

Oligo	Sequence
Oligo1	/5Phos/GAG CUA GCA UUA ACU UGG
5′xOligo1	GAG CUA GCA UUA ACU UGG
SR1 RNA	/5Phos/ UCG UAU GCC GUC UUC UGC UUG /3AmMO/
SR1 DNA	/5Phos/ TCG TAT GCC GTC TTC TGC TTG /3AmMO/

A6 Plasmid map of recombinant pCA24N