

**EVOLUTION AND EVALUATION OF ENGINEERED
DNA LIGASES FOR IMPROVED BLUNT-END
LIGATION**

BY

Janine Kelly Sharma

2020

**A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE IN CELL AND MOLECULAR BIOSCIENCE**



Abstract

DNA ligases are fundamental enzymes in molecular biology and biotechnology where they perform essential reactions, e.g. to create recombinant DNA and for adaptor attachment in next-generation sequencing. T4 DNA ligase is the most widely used commercial ligase owing to its ability to catalyse ligation of blunt-ended DNA termini. However, even for T4 DNA ligase, blunt-end ligation is an inefficient activity compared to cohesive-end ligation, or its evolved activity of sealing single-strand nicks in double-stranded DNA. Previous research from Dr Wayne Patrick showed that fusion of T4 DNA ligase to a DNA-binding domain increases the enzyme's affinity for DNA substrates, resulting in improved ligation efficiency. It was further shown that changes to the linker region between the ligase and DNA-binding domain resulted in altered ligation activity. To assist in optimising this relationship, we designed a competitive ligase selection protocol to enrich for engineered ligase variants with greater blunt-end ligation activity. This selection involves expressing a DNA ligase from its plasmid construct, and ligating a linear form of its plasmid, sealing a double-strand DNA break in the chloramphenicol resistance gene, permitting bacterial growth. Previous researcher Dr Katherine Robins created two linker libraries of 33 and 37 variants, from lead candidate ligase-cTF and (the less active form of p50-ligase variant) ligase-p50, respectively. Five rounds of selection were applied to each library. One linker variant, denoted ligase-CA3 showed the greatest improvement, comprising 42% of the final selected ligase-cTF population. In contrast, a lead linker variant from the ligase-p50 library was not obtained. In this study one additional round of selection was applied to the ligase-p50 library to test whether a lead variant would emerge. However, the linker variants selected at the end of Round 6 did not suggest a clear lead candidate, so one of the top variants (ligase-PPA17) was selected to represent this population in a fluorescence-based ligation assay that I optimised. Following identification of optimal reaction buffers to improve protein stability and DNA ligation, six engineered variants were compared for blunt-, cohesive-end, and nick sealing ligation activities. The five engineered variants exhibited improved blunt-end ligation activity over T4 DNA ligase. Ligase-PPA17 (1.9-fold improvement over T4 DNA ligase) was best performing for blunt-end ligation. This study found no evidence that ligase-CA3 was significantly improved over its predecessor, ligase-cTF in blunt-end ligation, however it was the best performing variant at cohesive-end ligation. Overall, we have evolved DNA ligase variants with improved blunt-end ligation activity over T4 DNA ligase which may be more advantageous in molecular biology and biotechnology for a variety of applications.

Acknowledgements

Many people told me that the Honour's year would be quite possibly one of the hardest year of your life, and they weren't wrong. However, with every struggle there is a silver lining, and the past 12 months have been some of the most memorable. With that, I'd like to thank my two amazing supervisors, Prof David Ackerley and Assoc. Prof. Wayne Patrick for the granting me the opportunity to work on the DNA ligase project and be apart of two incredibly lab groups.

To Dave, thank you for promoting a positive and supportive lab environment that has allowed me to grow not only as a researcher but also a person. Your endless support during the Covid lockdown has been outstanding. Writing a thesis is no easy feat, but to be isolated to my living room, you have been so encouraging and motivating, allowing me to stay on track and complete this body of work. Whilst you only beat me by 9 seconds at the Round the Bays 2019, you will continue to be miles ahead and I try continue to follow in your footsteps as a great leader and mentor.

A big thank you to Abby, Alistair, Kelsi, and Tif for being my go-to's for research advice and editing various exerts of my thesis (sometimes with little to no background or context). To Ed, Elsie, Ethan, Hannah, Jenni, Mark, Mitch, Sarah and the Metamax lab, thank you for allowing the year to fly by. From morning coffees, to Friday drinks, BYOs and to lab retreat, you have been a spectacular bunch of nerds who I look forward to seeing each day! Finally a special mention to my master's thesis partner in crime, Vincent. Thank for being one of my first friends at VUW. I am grateful to have gone through this journey with you, and have someone who understands the hardship of writing and submitting a thesis in the middle of a pandemic. I'm looking forward to getting out of lockdown to celebrate our success!

To Wayne, thank you for your support both inside and outside the lab environment. I am grateful that you have always believed in me even when I did not. I always enjoy talking running with you and your experience has been incredibly valueable. You have readily helped answer questions whenever I was in a 'sticky' situation, whether it be failing experiments or interpreting data. I'd like to acknowledge the MME lab, in particular the Protein Engineering Subgroup: Alex, Chris, Emma, Hamish, Lee, and Tif. Whilst we have often struggled with the Friday morning meeting, I am appreciative of the critical

analysis of my data which made problem solving much easier. Additionally, a massive thank you to Hamish for being my arms and completing my lab work when they decided to stop working properly from loading too many gels!

To the very dispersed Team Ligase: Adele, Katherine and Luke, thanks for making ends meet... and teaching me various aspects within the project over the last 12 months! I'd especially like to thank Luke. You have been a great help on many fronts, while you spent much of my Master's writing and handing your own PhD thesis, you still made time to answer my many questions, no matter how silly. You have been a great inspiration and someone who I very much look up to (both figuratively and literally).

A huge thanks to some of the greatest friends in the world: Aanchal, Jess, and Varun, you guys are absolute superstars! I love your ability to make me laugh when I am crying or make me laugh until I cry. To the BK lab members, I have appreciated your support and friendship, particularly during my Part 1 Masters. Furthermore, I'd like to acknowledge my running group (WoRM) for helping me retain a social life, stay fit and occasionally take a night off to destress. Lastly, a major thank you to my past and present flatmates come best mates: Mike, Tory and Grace, I appreciate your support, whether it be hugs, getting me food, or distracting me from stress.

Finally, a massive thank you to my parents, Robyn and Harry for the ongoing support and love, I'm very grateful to have you around the corner whenever I need it!

Table of Contents

ABSTRACT	I
ACKNOWLEDGEMENTS	II
TABLE OF CONTENTS	IV
LIST OF FIGURES	VII
LIST OF TABLES	VIII
ABBREVIATIONS	IX
CHAPTER 1: INTRODUCTION	1
1.1 DNA Ligases	1
1.1.1 Types of DNA Ligases	1
1.1.2 Structure of DNA Ligases	2
1.1.3 DNA Ligation Mechanism	3
1.1.4 <i>In Vitro</i> Applications of DNA Ligases	5
1.2 T4 DNA Ligase	6
1.2.1 <i>In vitro</i> Ligation Assay Conditions	7
1.3 Protein Engineering	7
1.3.1 Rational Protein Engineering	8
1.3.2 Directed Evolution	9
1.4 Engineering Improved DNA-Binding DNA Ligases	10
1.4.1 DNA-Binding Domains and Proteins	11
1.4.2 Ligation Activity Assays	13
1.4.3 Improving Ligation of PBCV-1 DNA Ligase with a DNA-Binding Domain	13
1.4.4 Small-Scale Linker Modification of Ligase-cTF	14
1.5 Competitive Ligase Selection	14
1.5.1 Protein Linkers	17
1.5.2 Creation of Ligase Linker Libraries	19
1.5.3 Enriched Engineered DNA Ligases	20
1.6 Aims and Hypothesis	22
MANUSCRIPTS IN PREPARATION	22
CHAPTER 2: GENERAL METHODS AND MATERIALS	23
2.1 Materials	23
2.2 Bacterial Strains and Plasmids	23
2.2.1 Bacterial Strains	23
2.2.2 Plasmids	24
2.3 Primers	24
2.4 Ligase Constructs Used in this Study	24
2.5 Bacterial Cell Growth and Storage	24
2.5.1 Liquid Media	25
2.5.2 Solid Media	25
2.5.3 Media Supplements	26

2.6	Routine Molecular Biology Protocols	26
2.6.1	Polymerase Chain Reaction	26
2.6.2	Agarose Gel Electrophoresis	27
2.6.3	Restriction Enzyme Digests	27
2.6.4	Isolation of Plasmid DNA	27
2.6.5	DNA Quantification	28
2.6.6	DNA Sequencing	28
2.7	Bacterial Cell Transformation and Electroporation	28
2.7.1	Preparation of Electrocompetent Cells	28
2.7.2	Transformation of Electrocompetent Cells	29
2.8	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	29
2.9	Competitive Ligase Selection Protocol	30
2.9.1	Whole Plasmid PCR	30
2.9.2	Ligase Protein Expression	30
2.9.3	Ligation Reaction	31
2.9.4	Proteinase K	31
2.9.5	Clean and Concentrate of DNA	31
2.9.6	<i>DpnI</i> and Alkaline Phosphatase Reaction	32
2.9.7	Isopropanol Precipitation of DNA	32
2.9.8	Transformation	32
2.9.9	Colony Screening	32
2.10	Protein Purification	33
2.10.1	Ligase Expression	33
2.10.2	Cell Lysis and Separation and Clean Up of Soluble Protein	33
2.10.3	Protein Purification via IMAC	34
2.10.4	ÄKTA start Purification	34
2.10.5	Purification with a Peristaltic Pump	34
2.10.6	Desalting of Eluted Protein	35
2.10.7	Measurement of Protein Concentration	35
2.10.8	Storage of Ligase Proteins	36
2.11	Fluorescence-Based Ligation Assay	36
2.11.1	Oligonucleotide/DNA Substrates	36
2.11.2	Ligation Substrate Master-Mix Preparation	37
2.11.3	Ligation Reaction	38
2.11.4	TBE-Urea Polyacrylamide Gel Electrophoresis (TBE-PAGE)	38
2.11.5	Ligation Quantification	39
CHAPTER 3: FURTHER EVOLUTION OF LIGASE-P50		41
3.1	Introduction	41
3.2	Results	41
3.2.1	DNA Substrates for Ligation Reaction	42
3.2.2	Diagnostic Linker PCR Confirmation	44
3.2.3	Determining the Linker Region Sequence	46
3.2.4	Analysis of Variants Enriched in Round 6 of the Ligase-P50 Linker Selection	48
3.3	Chapter 3 Discussion	49
3.3.1	Whole Plasmid PCR Amplification	49
3.3.2	Round 6 Ligase-p50 Linker Library Enrichment	50
CHAPTER 4: QUANTIFICATION OF LIGATION		52
4.1	Introduction	52
4.2	Purification and Confirmation of Ligation Activity for Engineered Ligases	52
4.2.1	Confirmation of Ligation Activity <i>in vivo</i>	52
4.2.2	Expression and Purification of Ligase Variants	53
4.3	Design and Optimisation of Fluorescence-Based Ligation Assays	56

4.4	Confirming Ligation Activity <i>in vitro</i>	61
4.5	Preliminary Time-Course Ligation Assays Revealed Extensive Batch-to-Batch Variation	62
4.6	Protein Buffer and Stability Optimisation	64
4.6.1	Protein Storage	64
4.6.2	Ligase Buffer Comparison	67
4.7	Protein Quantification	68
4.8	Evaluating the Ligation Activities of Engineered DNA Ligases	70
4.9	Discussion	74
4.9.1	Purification of Ligase Variants	74
4.9.2	Optimisation of Ligation Reactions	74
4.9.3	Quantification of Ligation Activity	76
	CHAPTER 5: KEY FINDINGS, FUTURE DIRECTIONS AND CONCLUSIONS	78
5.1	Research Motivations	78
5.2	Key Findings	78
5.2.1	Identifying Improved Ligase-p50 Linker Variants	78
5.2.2	Elucidating Protein Buffer Conditions for Improved Ligation Activity	79
5.2.3	Evaluating Ligation Activity of Engineered DNA Ligase Variants	79
5.3	Critical Evaluation of Methodology	80
5.3.1	Competitive Ligase Selection Protocol	80
5.3.2	Evaluating Ligation Activity using the Fluorescence-Based Ligation Assay	81
5.4	Future Directions	82
5.5	Concluding Remarks	84
	REFERENCES	86
	APPENDIX	92

List of Figures

Figure 1-1 Schematic diagram of the ligation reaction of a single-stranded DNA nick with ATP cofactor.	4
Figure 1-2 Engineered DNA ligases have increased surface area and potential to associate with dsDNA.	11
Figure 1-3 Schematic representation of the methodology for the competitive ligase selection protocol	16
Figure 1-4 Previous validation of the evolution of the optimised protocol to select for p50-ligase variants over less-active PprA-ligase variants over five rounds.	17
Figure 1-5 Schematic for the modification of the linker region in ligase libraries.	19
Figure 1-6 Evolution of ligase-cTF and ligase-p50 linker libraries following five rounds of selection.	21
Figure 3-1 Initial efforts to achieve whole plasmid amplicons were only successful from purified DNA, not diluted culture templates.	42
Figure 3-2 Small-scale amplification of whole plasmid amplicons.	43
Figure 3-3 Example of the amplification of whole plasmid amplicons for one biological plate used as the DNA substrates for the ligation reaction in Round 6 of the competitive ligase selection.	44
Figure 3-4 PCR amplification of ligase linker variants at the end of the competitive selection round.	45
Figure 3-5 Schematic representation of the work-flow to confirm the sequence of ligase variants using PCR reactions.	46
Figure 3-6 Validating the annealing temperature for the ligase-PPA17 linker with a gradient PCR amplification.	47
Figure 3-7 Enrichment of engineered ligase variants from the ligase-p50 linker library from Round 5 to Round 6.	48
Figure 4-1 Testing whether engineered DNA ligases relevant to this study can complement the temperature-sensitive ligase LigA in <i>E. coli</i> GR501.	53
Figure 4-2 Chromatograms (top) and SDS-PAGE gels assessing protein purification of T4 DNA ligase, ligase-CA3 and ligase-p50.	55
Figure 4-3 Schematic representation of the different fluorescence-based ligation assays.	56
Figure 4-4 Oligonucleotide separation of substrate and product on a TBE-PAGE gel.	58
Figure 4-5 Example of pre-optimised and optimised TBE-PAGE gels.	60
Figure 4-6 Evaluation of proteinase K concentration following fluorescence-based ligation reaction.	61
Figure 4-7 Engineered ligase variants maintain DNA nick-sealing ligation activity in the fluorescence-based ligation assay.	62
Figure 4-8 Initial efforts to plot a time-course of ligation activity for T4 DNA ligase, ligase-cTF, and ligase CA3.	63
Figure 4-9 Impact of storage on the activity of ligase-CA3 as measured by the fluorescence-based ligation assay.	65
Figure 4-10 Comparison of ligase-CA3 blunt-end ligation activity in two storage buffers.	66
Figure 4-11 Comparison of ligase-CA3 ligation activity in either the NEB or Williamson et al. 2014 reaction buffers.	67
Figure 4-12 Collation of blunt-end ligation data from Figures 4-8 to 4-11 in this study performed at 4 hr, 25°C.	68
Figure 4-13 SDS-PAGE to assess relative protein concentrations from a selection of the ligase variants assayed in Figure 4-8 and 4-10.	69
Figure 4-14 Ligation activity of engineered DNA ligase variants.	72

List of Tables

Table 1-1 DNA-Binding Proteins Fused to T4 DNA Ligase	11
Table 1-2 Ligation efficiencies of engineered DNA ligases with varied linker regions	14
Table 2-1 <i>Escherichia coli</i> strains used in this study	23
Table 2-2 Primers used in this study	24
Table 2-3 Ligase variant constructs assayed in this study	24
Table 2-4 LB Media	25
Table 2-5 Super-Optimal Broth (SOB-)* or SOB with Catalytic Repression (SOC)#	25
Table 2-6 Media supplements used in this study	26
Table 2-7 Composition and cycling parameters for PCR reactions containing Q5 DNA polymerase	26
Table 2-8 Composition and cycling parameters for PCR reactions containing Biomix™ Red	27
Table 2-9 Whole Plasmid PCR for Creation of Ligation Reaction Substrate	30
Table 2-10 Cycling profile for Whole Plasmid PCR	30
Table 2-11 Cycling profile for ligase colony screening	33
Table 2-12 Imidazole concentrations in Wash and Elution Buffer/s used in this study	34
Table 2-13 Storage Buffer Recipes for Protein Desalting	35
Table 2-14 Oligonucleotides used in the Fluorescent-Based Ligation Assay	37
Table 4-1 Commercial T4 DNA ligase storage buffers	66
Table 4-2 Blunt-end ligation activity over 4 hr at 25°C (from Figure 4-14, Panel A)	72
Table 4-3 Cohesive-end ligation activity over 60 min at 25°C (from Figure 4-14, Panel B)	72
Table 4-4 Nick-sealing ligation activity over 30 s at 25°C (from	73

Abbreviations

3' AmMO	3' amino acid blocker
AMP	adenosine monophosphate
ANOVA	analysis of variance
APS	ammonium persulfate
ATP	adenosine triphosphate
AUC	area under the curve
bp	base pair
<i>cam^R</i>	chloramphenicol resistance gene
CAT	chloramphenicol acetyltransferase
cTF	chimeric transcription factor
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
FAM	carboxy-fluorescein
FL	flexible linker
His-tag	hexa-histidine or poly-histidine tag
HL	helical linker
IDT	Integrated DNA Technologies
IMAC	immobilised metal ion affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
K_d	dissociation constant
K_M	concentration of substrate permitting a reaction to occur at half maximal velocity
KPhos	potassium phosphate
LB	Luria-Bertani
NAD ⁺	nicotinamide adenine dinucleotide
NEB	New England Biolabs
NF κ B	nuclear factor kappa light chain enhancer of activated B cells
NFAT	nuclear factor of activated T-cells
NGS	next generation sequencing

NL	natural linker
NLL	natural long linker
NLS	natural short linker
nt	nucleotide
OD ₆₀₀	optical density at the 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
PBCV-1	<i>Paramecium bursaria</i> Chlorella virus 1
PEG	polyethylene glycol
PPi	pyrophosphate
rpm	rotations per minute
rSAP	recombinant shrimp alkaline phosphatase
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SOB	super optimal broth
SOC	super optimal broth with catalytic repression
T _a	annealing temperature
TAE	tris-acetic acid-EDTA
TBE	tris-borate-EDTA
TBE-PAGE	TBE-urea polyacrylamide gel electrophoresis
TE	tris-EDTA
TEMED	tetramethylethylenediamine
T _m	melting temperature
UV-Vis	ultraviolet-visible
v/v	volume per volume
w/v	weight per volume

Chapter 1: Introduction

1.1 DNA LIGASES

Damage to DNA is a natural phenomenon that occurs continuously within cells and consequentially leads to cell death if not repaired. Damage can accrue due to external factors, such as UV radiation or chemical exposure; or by internal cellular processes that generate reactive compounds and by-products. DNA nicks and lesions may also arise through processing errors in DNA repair pathways or on the lagging strand during DNA replication (Ogawa and Okazaki 1980). The enzymes responsible for repairing and restoring the DNA backbone are DNA ligases. DNA ligases play an indispensable role in maintaining genomic integrity and ensuring the completion of the DNA helical structure during DNA replication, recombination, and repair (Shuman, 2009). Examples of their activities include the joining of Okazaki fragments at the replication fork on the lagging strand during DNA replication (Howes and Tomkinson 2012), and catalysing the final steps in base excision repair, or non-homologous end-joining repair pathways (Timson, Singleton, and Wigley 2000).

The discovery of *Escherichia coli* DNA ligases in 1967 was reported by several independent research groups, who showed that the enzyme circularised and joined breaks in DNA (Becker et al. 1967; Cozzarelli et al. 1967; Gefter, Becker, and Hurwitz 1967; Gellert 1967; Zimmerman et al. 1967). Subsequently, the mechanism for ligation was confirmed to be the formation of the phosphodiester bond between 5' phosphate and 3' hydroxyl termini in double-stranded DNA (dsDNA) (Lehman 1974).

1.1.1 Types of DNA Ligases

A plethora of DNA ligases have now been discovered from all domains of life. There are two distinct classes of DNA ligases defined according to their cofactor dependency. NAD⁺-dependent DNA ligases (EC: 6.5.1.2) are specific to prokaryotes (Wilkinson, Day, and Bowater 2001), and entomopox viruses (Verl Sriskanda, Moyer, and Shuman 2001). In contrast, ATP-dependent DNA ligases (EC: 6.5.1.1) have been found across all phylogenetic kingdoms, with all known archaeobacteria and eukaryotes possessing at least one such ligase (Ellenberger and Tomkinson 2008; Martin and Macneill 2002). All bacteria encode at least one conserved NAD⁺-dependent DNA ligase, however, may contain an additional ATP-dependent DNA ligase (Cheng and Shuman 1997; Kim and

Kwon 1998; Wilkinson, Day, and Bowater 2001; Williamson, Hjerde, and Kahlke 2016; Williamson, Rothweiler, and Leiros 2014).

1.1.2 Structure of DNA Ligases

DNA ligases are among a group of enzymes that make up the nucleotidyltransferase enzyme superfamily (Shuman and Schwer 1995). Sequence alignments of these enzymes has uncovered several regions of amino acid homology, with six sequence motifs of conserved residues forming a common tertiary structure and making up their catalytic core (Martin and Macneill 2002; Shuman and Lima 2004; Subramanya et al. 1996). The six motifs contain key active site residues that are critical for enzyme activity, as shown in mutagenesis studies (Verl Sriskanda and Shuman 2002a, 2002b).

The minimal structure of both ATP- and NAD⁺-dependent DNA ligases consists of an adenylation domain (or nucleotidyltransferase domain) fused to a C-terminal oligonucleotide-binding domain. The adenylation domain contains the ATP (or NAD⁺) binding site and five out of the six sequence motifs, which comprise key active site residues (Ellenberger and Tomkinson 2008). The oligonucleotide-binding domain contains the final sequence motif and is involved in two functions during catalysis (Nair et al. 2007). First, the oligonucleotide-binding domain assists with the transfer of AMP to the enzyme's active site before the domain undergoes a large conformational change to bind and correctly orientate the DNA for catalysis. Eukaryotic and archaic DNA ligases contain an additional N-terminus DNA-binding domain that interacts with the adenylation and oligonucleotide-binding domains, encircling the DNA. The main function of the DNA-binding domain is to increase DNA-binding affinity, and to stabilise and orientate the DNA in a conformation that allows for catalysis and ligation (Ellenberger and Tomkinson 2008). Mutations to this domain cause diminished or abolished activity, therefore it has been concluded the domain plays a crucial part in DNA ligation (Pascal et al. 2004; Rossi et al. 1997). Except for the catalytic sequence motifs, ATP- and NAD⁺-dependent DNA ligases do not exhibit high sequence or structural similarity. Owing to their structural exclusivity and substrate specificity, NAD⁺-dependent DNA ligases are being targeted for the discovery of new anti-microbial agents (Shuman 2009; Wilkinson, Day, and Bowater 2001).

1.1.3 DNA Ligation Mechanism

All ligations can be simplified into a core three-step mechanism (Figure 1-1, Shuman, 2009), requiring the hydrolysis of one high energy molecule per enzymatic event (Weiss, Thompson, and Richardson 1968). Firstly, a lysine residue in the active site of the DNA ligase performs a nucleophilic attack on the α -phosphate of ATP or NAD⁺ forming a ligase-AMP intermediate, and releasing pyrophosphate (PPi) or nicotinamide mononucleotide (Figure 1-1, step 1). The (pre-)adenylation of DNA ligases yields a relatively stable and highly reactive conformation that has increased specificity for binding to DNA substrates, particularly for single-strand DNA nicks (V Sriskanda and Shuman 1998). Next, the AMP is transferred to the 5' phosphate terminating end of the DNA strand, forming a DNA-adenylate (Figure 1-1, step 2). Concurrently, oxygen from the 5' phosphate attacks the ligase-adenylate complex, allowing the lysine residue to act as a leaving group. The reaction is resolved with an attack by the 3' hydroxyl on the 5' phosphate, resulting in the formation of a new phosphodiester bond and the liberation of AMP (Figure 1-1, step 3).

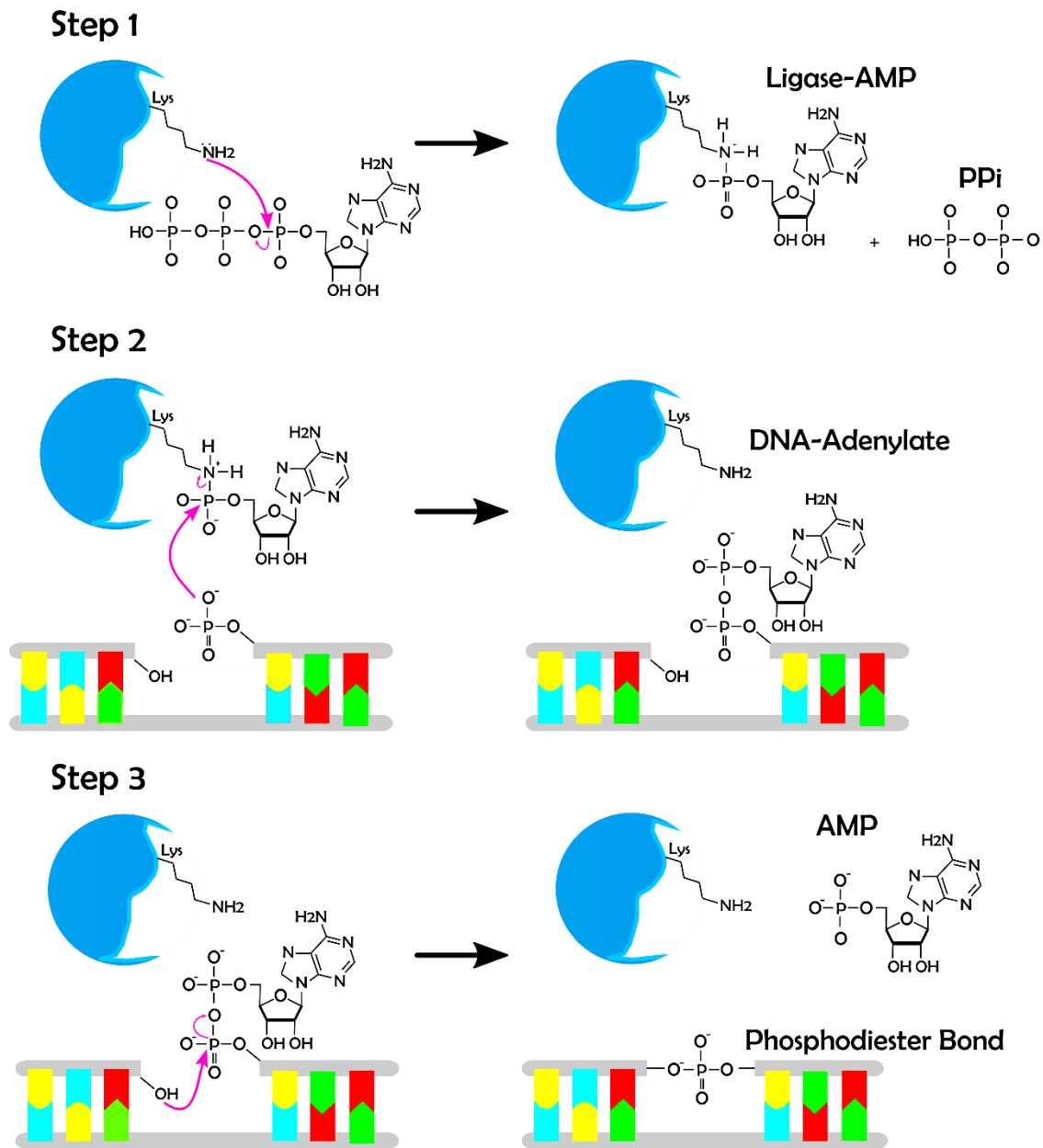


Figure 1-1 Schematic diagram of the ligation reaction of a single-stranded DNA nick with ATP cofactor.

The lysine residue within the active site of a DNA ligase (here, ATP-dependent) performs a nucleophilic attack on the α -phosphate of ATP, forming an adenylate-ligase intermediate. The adenylate-ligase complex reacts with the exposed 5' phosphate of a DNA strand, forming a DNA-adenylate intermediate. The 3' hydroxyl group reacts with the DNA-adenylate forming a phosphodiester linkage, completing the DNA backbone, and liberating AMP. Figure adapted from Shuman (2009).

1.1.4 *In Vitro* Applications of DNA Ligases

1.1.4.1 *Recombinant DNA Technology*

The discovery of DNA ligases marked a key milestone for molecular biology and biotechnology. DNA ligases have become a fundamental enzyme *in vitro* for numerous applications. In particular, the creation of recombinant DNA through the process known as molecular cloning, a technology recognised with the 1980 Nobel Prize, utilises a DNA ligase to ligate DNA fragments into a cut vector backbone (Carroll 1993; Jackson, Symons, and Berg 1972). It is difficult to overstate the impact that recombinant DNA technologies have had on all virtually all disciplines within the biological sciences.

More recent synthetic biology techniques such as Gibson assembly, a next-generation cloning technology, have also applied DNA ligases. Gibson assembly is a single isothermal overlap recombination system that utilises *Taq* DNA polymerase, exonuclease III, and *Taq* DNA ligase (Gibson et al. 2009). This technology is capable of cloning large DNA products (e.g., exceeding several hundred kb in length) or assembling multiple DNA inserts that have specific complementary overhangs in a single reaction. Techniques such as these are accelerating the discovery and re-engineering of genetic pathways, library construction, site-directed mutagenesis, and the production of various compounds and pharmaceuticals.

1.1.4.2 *Next-Generation Sequencing (NGS)*

Prior to the completion of the Human Genome Project in 2003, first-generation Sanger sequencing was the primary sequencing technology, used to characterise DNA sequences between 600-1000 bp with relatively high accuracy. Since 2003, several NGS technologies have evolved to improve the speed and accuracy of sequencing large amounts of DNA in a deep, in-parallel, high-throughput manner. Commonly used NGS technologies including 454 pyrophosphate sequencing (Margulies et al. 2005), Ion-Torrent, and Illumina (Kawashima, Farinelli, and Mayer 2005), use a sequencing by synthesis approach to determine sequence heterogeneity. These NGS technologies follow a similar sample preparation and workflow which relies on a DNA ligase for adaptor ligation to establish the array (Linnarsson 2010). Firstly, dsDNA is sheared by mechanical or enzymatic means to generate small blunt-end DNA breaks, (and any base overhangs are either complemented or removed). Individual blunt-end DNA fragments are then covalently linked to a solid surface, such as microfluidic channels or emulsion beads where millions of sequencing reactions are performed in parallel. The ligation of

adaptors is performed by a DNA ligase and a crucial step for NGS library creation. Adaptor sequences are short oligonucleotide sequences specific to each NGS platform which are ligated to both the 5' and 3' end of DNA fragments. These sequences together act as a barcode for the reaction, as well as a primer sequence for DNA polymerase binding and synthesis. Modified nucleotide bases are incorporated one by one basis, the fluorescence or current detected, and the nucleotide base determined. The synthesis process is repeated thousands to millions of times to obtain high sequence coverage and produce a consensus sequence that can be evaluated using bioinformatics tools. Due to the kinetic limitations and challenges of the enzyme (T4 DNA ligase) responsible for ligation of blunt-ended DNA and adaptors in NGS library preparation, this can often result in sub-optimal library creation, leading to unsatisfactory and incomplete results (Wilson et al. 2013). This identifies a clear need to improve and enhance the activity of a DNA ligase for greater blunt-end ligation activity.

1.2 T4 DNA LIGASE

The ATP-dependent, Mg²⁺-dependent bacteriophage T4 DNA ligase, discovered in 1967 (Weiss and Richardson 1967), is one of the most commonly used ligases in *in vitro* research (Tabor 2001) and is routinely used for blunt-end ligation. The enzyme is the product gene 30 of the *Enterobacteria* bacteriophage T4, which encodes an enzyme of 485 amino acids with a molecular weight of 55.2 kDa (Armstrong, Brown, and Tsugita 1983).

T4 DNA ligase readily accepts and catalyses the repair of single-stranded nicks in dsDNA (as shown in Figure 1-1). T4 DNA ligase achieves this reaction with high efficiency and a Michaelis constant (K_M) of 2.5 nM (Lohman, Chen, and Evans 2011). K_M is defined as the concentration of substrate which permits the reaction to occur at half-maximal velocity and is indicative of an enzyme's affinity for a substrate. Comparatively, the catalytic affinity of T4 DNA ligase decreases for cohesive-end dsDNA ligation ($K_M = 0.6 \mu\text{M}$).

Unlike many other characterised DNA ligases including those from bacteriophage T3, *Parmecium bursaria* chorella virus 1 (PBCV-1), and *E. coli*, T4 DNA ligase can successfully ligate blunt-end DNA substrates (Bauer et al. 2017; Sgaramella, Van De Sande, and Khorana 1970; Weiss and Richardson 1967), without the addition of macromolecular crowding agents (explained in detail in the Section 1.2.1). However, the catalytic affinity for blunt-end ligation is diminished by several orders of magnitude ($K_M = 50 \mu\text{M}$)

compared to cohesive-end and nick-sealing ligation (Sugino et al. 1977). To achieve comparable results, the amount of enzyme required to perform a blunt-end ligation is between 10- and 100-fold greater (Lohman, Tabor, and Nichols 2011). The sub-optimal activity of T4 DNA ligase for blunt-end ligation emphasises the opportunity to conduct enzyme engineering and improve its ligation activity.

1.2.1 *In vitro* Ligation Assay Conditions

There are several companies that sell T4 DNA ligase, including New England Biolabs (NEB), Thermo Fisher Scientific, and Roche. Ligation protocols and buffers for T4 DNA ligase have been optimised and use very similar reaction conditions. T4 DNA ligase reaction buffer contains 0.5-1 mM ATP as the AMP donor molecule for ATP-dependent ligase activity; 5-10 mM dithiothreitol (DTT) to maintain reduced cysteine residues and protein stability (Sevier and Kaiser 2002); and 5-10 mM Mg²⁺ to promote ligation by facilitating adenyl group transfer through stabilisation of the transition state of the DNA-adenylate intermediate (Cherepanov and De Vries 2003). Commercially available ligation protocols for T4 DNA ligase recommend that reactions are performed between 15 and 25°C, for anywhere between 10 min and 18 hr, depending on whether the DNA substrate has a blunt- or cohesive-end.

The addition of macromolecular exclusion molecules or condensing reagents such as Ficoll, hexamminecobalt chloride, or more commonly polyethylene glycol (PEG) can be used to enhance blunt-end ligation efficiency (Green and Sambrook 2019; Pheiffer and Zimmerman 1983). Condensing reagents act to increase the rate of blunt-end ligation. However, these macromolecular exclusion molecules alter the distribution of ligation products, exclusively favouring intermolecular ligation and the formation of linear products (Hayashi et al. 1986). Their diminished efficiency in promoting intramolecular ligation, or circularisation of DNA (Pheiffer and Zimmerman 1983) is problematic for some *in vitro* experiments, in particular for DNA cloning and transformation requires circular DNA to replicate within a host cell.

1.3 PROTEIN ENGINEERING

Enzymes have evolved to catalyse cellular reactions with incredible precision and efficiency, however their potential in industrial applications is often limited by their inability to catalyse key reactions effectively in an extracellular environment. Based on prior knowledge of protein structure and/or function, it is possible to manipulate the

protein sequence at the genomic level to alter an enzyme's properties. This process known as enzyme engineering can be used to produce superior or desirable enzymes for a given industrial purpose. Enzyme engineering can generate enzyme variants that have improved functional characteristics including catalytic activity, pH optimum, or protein stability. There are two main strategies that can be used to conduct enzyme engineering: rational design and non-rational design (also known as directed evolution).

1.3.1 Rational Protein Engineering

Rational protein engineering utilises biochemical and structural data to target protein residues where substitution is predicted to improve the activity of an enzyme. To achieve significant improvements using this method, a comprehensive understanding of the protein structure and function is required. An example of rational enzyme engineering of a DNA ligase was the structure-guided mutational analysis of *Pyrococcus furiosus*, which aimed to improve ligation efficiency, whilst retain thermostability (Tanabe, Ishino, and Nishida 2015). Comparing the enzyme's crystal structure to those of two other mesothermal DNA ligases, the team identified five ionic interactions between the adenylation and oligonucleotide-binding domains that were not present in the other enzymes (Tanabe et al. 2012). Site-saturation mutagenesis was then used to mutate these residues to non-ionic amino acids. The mutation D540A was found to generate a variant that displayed improved rate kinetics compared to the wild-type, by disrupting the interaction between the adenylation and oligonucleotide-binding domains proximal to the AMP-binding site. Further modifications to D540 were investigated to evaluate both the thermostability and ligation efficiency of variant enzymes (Tanabe et al. 2014). Ultimately, a D540R variant not only retained thermostability but had enhanced ligation efficiency caused by repulsion of PPi in the active site during the adenylation step of the reaction. Site-directed mutagenesis was also successfully applied to *Thermus thermophilus* DNA ligase to generate variants with improved ligation fidelity of single-stranded DNA nicks in the ligase chain reaction (Luo 1996). Furthermore, it is possible to use known *in vivo* interactions as a start point for rational engineering to create chimeric enzymes that may retain both enzymatic functions by the addition of a protein linker. For example, the genetic fusion of DNA-binding protein Sso7d was shown to increase the processivity of *Taq* DNA polymerase (Wang et al. 2004) and this fusion protein is now an established commercial enzyme sold by both Bio-Rad and Thermo Fisher Scientific.

1.3.2 Directed Evolution

Evolution is an ongoing process occurring over hundreds to millions of years that selects for catalytic excellence in key enzyme activities. Much of our present understanding linking protein structure to function is incomplete, making it difficult to predict the outcome of single residue substitutions. Directed evolution is a laboratory-based technique that mimics natural selection on a shorter time-scale, generating mutants of a given gene evolved towards a defined enzymatic activity or attribute, without the need for prior structural information (Arnold 2018; Kuchner and Arnold 1997). Conventional directed evolution experiments start with the construction of a mutant gene library by random mutagenesis, generally achieved by error-prone PCR or site-saturation mutagenesis. Next, an artificial selection pressure is applied to the library to obtain variants that bear a fitness advantage for the activity that is being screened or selected for. Several iterations of this process can be applied to enrich or enhance enzyme activity, with the previous round establishing the template or start point for the next round.

A critical component of this technique is an efficient screen or selection that isolates the beneficial mutations from a library. A screen refers to the survival-independent evaluation of the desired activity in recovered variants (Leemhuis, Kelly, and Dijkhuizen 2009; Olsen, Iverson, and Georgiou 2000). Numerous assay techniques, including cell growth, colourimetry, ligand binding, substrate consumption or product formation assays can be used to screen for improved enzymatic activity (Xiao, Bao, and Zhao 2015). In contrast, a selection directly couples the protein function to the survival of a gene, eliminating undesirable variants (Leemhuis et al. 2005). However, it is often difficult to link desired enzyme activity directly to enhanced survival. Exceptions are, for example, synthesis of essential nutrients or neutralisation of a toxin (Leemhuis, Kelly, and Dijkhuizen 2009). These scenarios offer scope for rapid and very high-throughput evaluation of evolved variants.

Generally directed evolution is a process that is uncoupled from the construction of the gene library. However, enzymes that interact with nucleic acids offer the opportunity to link the expression of an enzyme variant to an interaction with its own gene, thereby acting as their own selection pressure. Compartmentalised self-replication is one such directed evolution selection strategy that relies on an active DNA polymerase variant to replicate its own encoded gene under a subjected condition (Abil and Ellington 2018; Ghadessy, Ong, and Holliger 2001). Enhanced activity of a given polymerase variant

results in an overrepresentation of its encoded gene in the successive selection round/s. Conversely, inactive variants that fail to amplify their own gene disappear from the gene pool. Compartmentalised self-replication has been successfully used to evolve DNA polymerases for increased thermostability, as well as a 130-fold increased resistance to heparin (Ghadessy, Ong, and Holliger 2001). Furthermore, this strategy can be used to evolve DNA polymerases that have altered substrate specificities, for example enhancing ability to incorporate hydrophobic base analogues (Loakes et al. 2009), or preferentially incorporate dUTP over dTTP in PCR reactions (Tubeleveciute and Skirgaila 2010).

1.4 ENGINEERING IMPROVED DNA-BINDING DNA LIGASES

T4 DNA ligase has a favourable origin for enzyme engineering as it is capable of blunt-end substrate ligation without the addition of macromolecular exclusion molecules (Sgaramella, Van De Sande, and Khorana 1970). Inspired by the successful genetic fusion of *Taq* DNA polymerase to a DNA-binding protein (Wang et al. 2004), research led by Assoc. Prof. Wayne Patrick applied a similar rational engineering approach to T4 DNA ligase (Wilson et al. 2013). In a study initiated by Dr Robert Wilson, the team sought to determine whether fusion of T4 DNA ligase to a DNA-binding domain would increase ligation efficiency by improving the enzyme's affinity and interaction with DNA substrates (Wilson 2009). The key concept was that addition of a DNA-binding domain or protein would expand the possible DNA regions that the enzyme can interact with, permitting for binding at any part of the DNA strand, as opposed to binding at the small 5' phosphate group (Figure 1-2; Wilson et al. 2013).

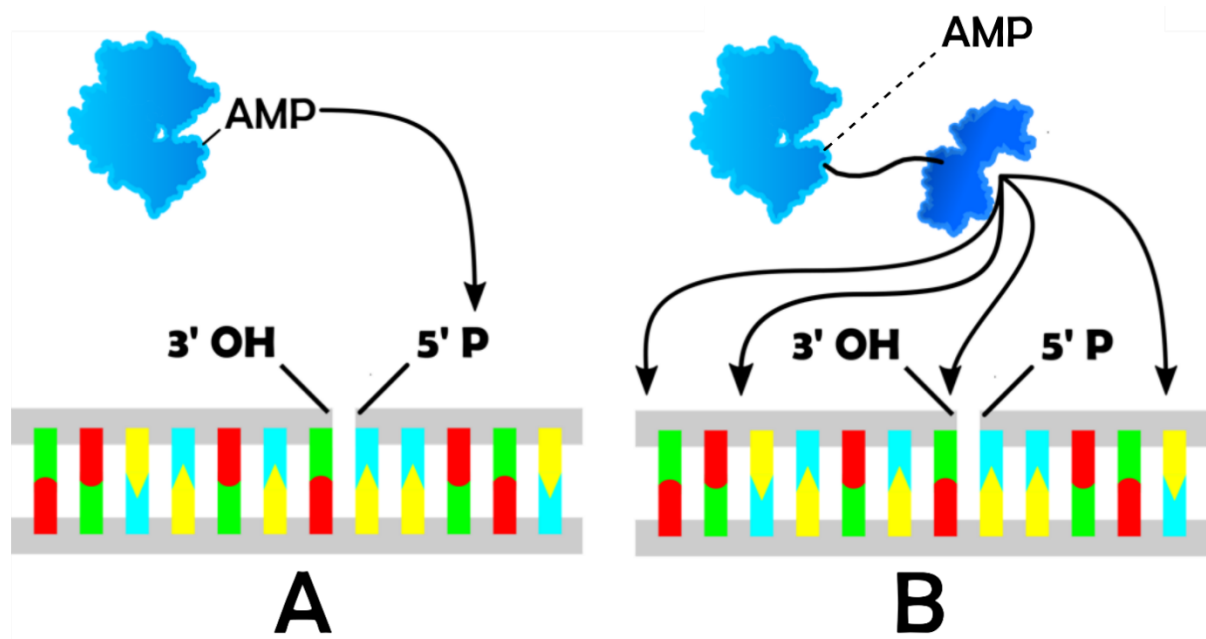


Figure 1-2 Engineered DNA ligases have increased surface area and potential to associate with dsDNA.

The addition of a DNA-binding domain (right; dark blue) permits the fusion ligase construct to interact and associate with a greater region of DNA and enabling the ligase-AMP intermediate to be formed at any stage of DNA association. In contrast, traditional DNA ligases, such as T4 DNA ligase (left) are only able to recognise the 5' phosphate on dsDNA which will only affiliate with DNA after the formation of the ligase-AMP intermediate.

1.4.1 DNA-Binding Domains and Proteins

Seven different DNA-binding proteins were fused to either the N- or C-terminus of T4 DNA ligase (Table 1-1; Wilson et al., 2013). The study reported that ligase variants displayed a diverse range of activity, for which p50-ligase and ligase-cTF showed superior ligation activity (as explained in detail in Section 1.4.2). (Note: throughout this work, if the DNA-binding protein was fused to the N-terminus of the ligase then it is mentioned first in the variant name, e.g. p50-ligase; whereas if it was fused to the C-terminus it is mentioned second, e.g. ligase-cTF).

Table 1-1 DNA-Binding Proteins Fused to T4 DNA Ligase

Protein	Source Organism	Source Reference
NF-κB (p50)	Homo sapiens	de Lumley <i>et al.</i> , (2004); Patrick and Blackburn, (2005)
NFAT	Mus musculus	de Lumley <i>et al.</i> , (2004)
cTF	H. sapiens/M. musculus	de Lumley <i>et al.</i> , (2004)
PprA	Deinococcus radiodurans	Narumi <i>et al.</i> , (2004)
Ku	Mycobacterium tuberculosis	Weller <i>et al.</i> , (2002)
Sso7d	Solfolobus solfataricus	Wang <i>et al.</i> , (2004)
[(HhH) ₂] ₂	Methanopyrus kandleri	de Vega <i>et al.</i> , (2010)

1.4.1.1 Nuclear Factor Kappa Light Chain Enhancer of Activated B Cells (NFκB)

The Rel/NFκB family of transcription factors are key regulators of gene expression for the activation of defensive immune mechanisms in response to injury or stress (Siebenlist, Franzoso, and Brown 1994). Rel proteins are distinguished by the Rel homology domain, a region of approximately 300 amino acids containing residues responsible for sequence-specific DNA recognition, and protein-protein interactions (Baeuerle and Henkel 1994; May and Ghosh 1997). Within the Rel protein family, active DNA binding is typically mediated by a dimeric complex comprised of two Rel subunits, with both homomeric and heterodimeric combinations possible (Chytil and Verdinef 1996; de Lumley et al. 2004).

NFκB1 or p50 is a DNA-binding subunit from mammalian transcription factor NFκB. This is a well characterised DNA-binding domain that has a strong non-specific association with DNA, denoted by a dissociation constant (K_d) of 5.7 nM (de Lumley et al. 2004). p50 commonly binds DNA as a dimeric complex with NFκB p65 (Fujita et al. 1992), however, it is also known to bind DNA as a homodimer, for which it recognises a palindromic sequence (5'-GGGAAATTCC-3'; Urban, Schreck, & Baeuerle', 1991) with a substantially higher affinity of $K_d = 8$ pM (de Lumley et al. 2004).

1.4.1.2 Nuclear Factor of Activated T Cells (NFAT)

NFAT is a family of transcription factors that share structural homology between their DNA-binding domain with the Rel-homology domain of NFκB and other Rel proteins (Jain et al. 1995; de Lumley et al. 2004). This class of proteins is capable of binding to DNA as a monomer (Chytil and Verdinef 1996) with a K_d of 39.5 nM (de Lumley et al., 2004). NFAT, like NFκB binds to sequence-specific DNA site (GGAAA; Macián, López-Rodríguez, & Rao, 2001), however this is only 4-fold greater affinity than the non-specific DNA binding ($K_d = 11.3$ nM; (de Lumley et al. 2004)).

To investigate the mechanisms and specificity of DNA binding of NFκB and NFAT proteins, de Lumley et al. (2004) used a structure-based approach comparing the wild-type proteins to chimeric fusions of different regions of the two proteins. NFκB and NFAT display low sequence homology, however their N- and C-terminals of the DNA-binding domains when superimposed occupied the same relative spatial positions and have a conserved tertiary structure. One chimeric transcription factor denoted cTF for the remainder of this study, contains the DNA-binding domain and linker region of NFAT,

and the dimerization domain of NF κ B. Variant cTF was chosen by Wilson et al. (2013) as a fusion partner for T4 DNA ligase.

1.4.2 Ligation Activity Assays

Ligase fusion constructs were screened to determine whether they exemplified superior catalytic activity compared to native T4 DNA ligase (Wilson et al. 2013). Agarose gel-based assays investigated both blunt- and cohesive-end fragment ligation abilities. The best performed variants were p50-ligase and ligase-cTF, which successfully ligated cohesive-end DNA substrates to near reaction completion, as shown by a near-total formation of product band on the gel. Additionally, Sso7d-ligase and p50-ligase were capable of ligating 90% of blunt-end DNA substrates, representing a 7-fold improvement over T4 DNA ligase. Ligase-cTF produced a less prominent increase with a 5-fold improvement compared to T4 DNA ligase in blunt-end ligation. Furthermore, it was evident by comparison with cTF-ligase and ligase-p50 that the specific termini for the DNA-binding protein was important, as the reciprocal fusion variants possessed weaker catalytic activity. To further quantify blunt-end ligation, the chimeric ligases were assessed for their abilities to ligate a small dsDNA fragment into a linearised vector that could be transformed and grown in *E. coli*. This assay provided a discrete method to determine the frequency of successful cloning events with a blue-white screen. Successful ligation was visually distinguished by the presence of white colonies. Ligase-cTF was again improved over T4 DNA ligase, with a 160% increase in the number of white colonies, or blunt-end fragments ligated. Finally, to investigate potential applications in the preparation of NGS libraries, a multiplex ligation assay was used to evaluate ligation of three random adaptor molecules to the 5' and 3' end of fragmented *E. coli* DNA. When the products were combined and sequenced, p50-ligase was found to be the most proficient at blunt-end adaptor ligation, contributing 49.1% of the total reads in the array. These assays each modelled relevant molecular biological experiments and proposed future commercial applications.

1.4.3 Improving Ligation of PBCV-1 DNA Ligase with a DNA-Binding Domain

Interestingly, studies by Bauer et al. (2017) investigated whether similar results could be observed for PBCV-1 DNA ligase. PBCV-1 DNA ligase is the smallest known DNA ligase, lacking a DNA-binding domain (Ho, Van Etten, and Shuman 1997). The ligase was fused to three different DNA-binding domains (Sso7d, the initial DNA-binding domain in Wang et al., 2004 and Wilson et al., 2013; Zinc finger from Human Ligase III; and the N-terminal

DNA-binding domain from T4 DNA ligase) and ligation activity for various DNA substrates was evaluated (Bauer et al. 2017). Similar to Wilson et al. (2013), the different DNA-binding domains resulted in various changes to DNA ligation. Although, DNA ligation was not majorly improved in this study, the researchers observed an increase in adenylated DNA in the presence of a DNA-binding domain, for which they reasoned increased the likelihood of a successful adenyl group transfer. The results from Bauer et al. (2017) support our initial hypothesis, whereby a DNA-binding domain increases ligation efficiency by improving the enzyme's affinity and interaction with DNA substrates.

1.4.4 Small-Scale Linker Modification of Ligase-cTF

An additional small-scale experiment in Wilson et al. (2013) investigated the linker region between T4 DNA ligase and the DNA-binding domain cTF, and whether this could influence the ligation efficiency. The original linker was compared to four different linkers including a natural short linker (NLS), natural long linker (NLL), flexible linker (FL), and a helical linker (HL). Changes to the linker region were found to result in varying blunt-end ligation activities (Table 1-2).

Table 1-2 Ligation efficiencies of engineered DNA ligases with varied linker regions

Ligase	Linker	Fraction of Substrate Ligated
Ligase-cTF	GTSGGGSGGG	0.27
Ligase-NLS-cTF	GTPTPTPTPTGEF	0.61
Ligase-NLL-cTF	GPTPTPTPTPTPTPTPTPTPTPTG	0.12
Ligase-FL-cTF	LSGGGGSGGGSGGGSGGGGSAAA	0.23
Ligase-HL-cTF	GPLAEAAAKEAAAKEAAAKEAAAKAAAPG	0.32

NLS - natural linker short
 NLL - natural linker long
 FL - flexible linker
 HL - helical linker

1.5 COMPETITIVE LIGASE SELECTION

To further investigate the relationship between T4 DNA ligase and the DNA-binding domain, a previous Ackerley lab Honours student implemented a selection protocol to enrich for ligase variants that have greater blunt-end ligation efficiency (Stevenson 2014). With assistance from Dr Katherine Robins, this competitive ligase selection protocol was optimised, and two libraries were subjected to five rounds of selection.

The initial cell culture (Figure 1-3, step 1) acts as the source of the plasmid DNA for a whole plasmid PCR amplification which is amplified from within a chloramphenicol resistance gene (*camR*) using 5' phosphorylated primers (Figure 1-3, step 2A). The

amplicon acts as the DNA substrate for the ligation reaction, with the 5' phosphate being essential because it is required for recognition and formation of the DNA-adenylate complex during ligation. The initial cell culture is also used as the source of expressed ligase variants (Figure 1-3, step 2B). The expressed DNA ligase protein and linear ligase plasmid amplicon are combined and ligated in a microwell, circularising the plasmid bearing the gene for that ligase variant (Figure 1-3, step 3). The ligation reactions from one or more microwell plates are then pooled, and proteinase K is added to remove any bound ligase (Figure 1-3, step 4). DNA is then purified using a clean and concentrator spin column to remove buffers, cellular debris, primers and proteins (Figure 1-3, step 5). To eliminate any residual plasmid DNA from the PCR reaction (which differs from the ligated PCR product in terms of having been methylated by its previous *E. coli* host), the mixture is digested with *DpnI*, a type II restriction enzyme that cuts as GATC wherever the adenosine residue is methylated (Mierzejewska et al. 2014). Simultaneously, recombinant shrimp alkaline phosphatase (rSAP) is used to dephosphorylate any residual linear DNA and thereby prevent any unwanted ligation of residual linear substrate *in vivo*, post-transformation (Figure 1-3, step 6). Finally, the circularised plasmid is used to transform *E. coli* host cells, which are plated on chloramphenicol supplemented LB agar to select for the most effective DNA ligase constructs (Figure 1-3, step 7). Single colonies are randomly selected and confirmed using a diagnostic PCR, after which they can be grown in a fresh microplate culture and used as the template for successive selection rounds.

This selection acts differently to typical directed evolution selections/screens. Rather than comparing the fitness of a cell, this method uses each ligase variant to ligate its own linearised plasmid, selecting for the ability to seal the double-stranded break in the *cam^R* gene. The ligase variants that are most active and successful at blunt-end ligation result in a greater number of chloramphenicol resistant colonies following the transformation of *E. coli*.

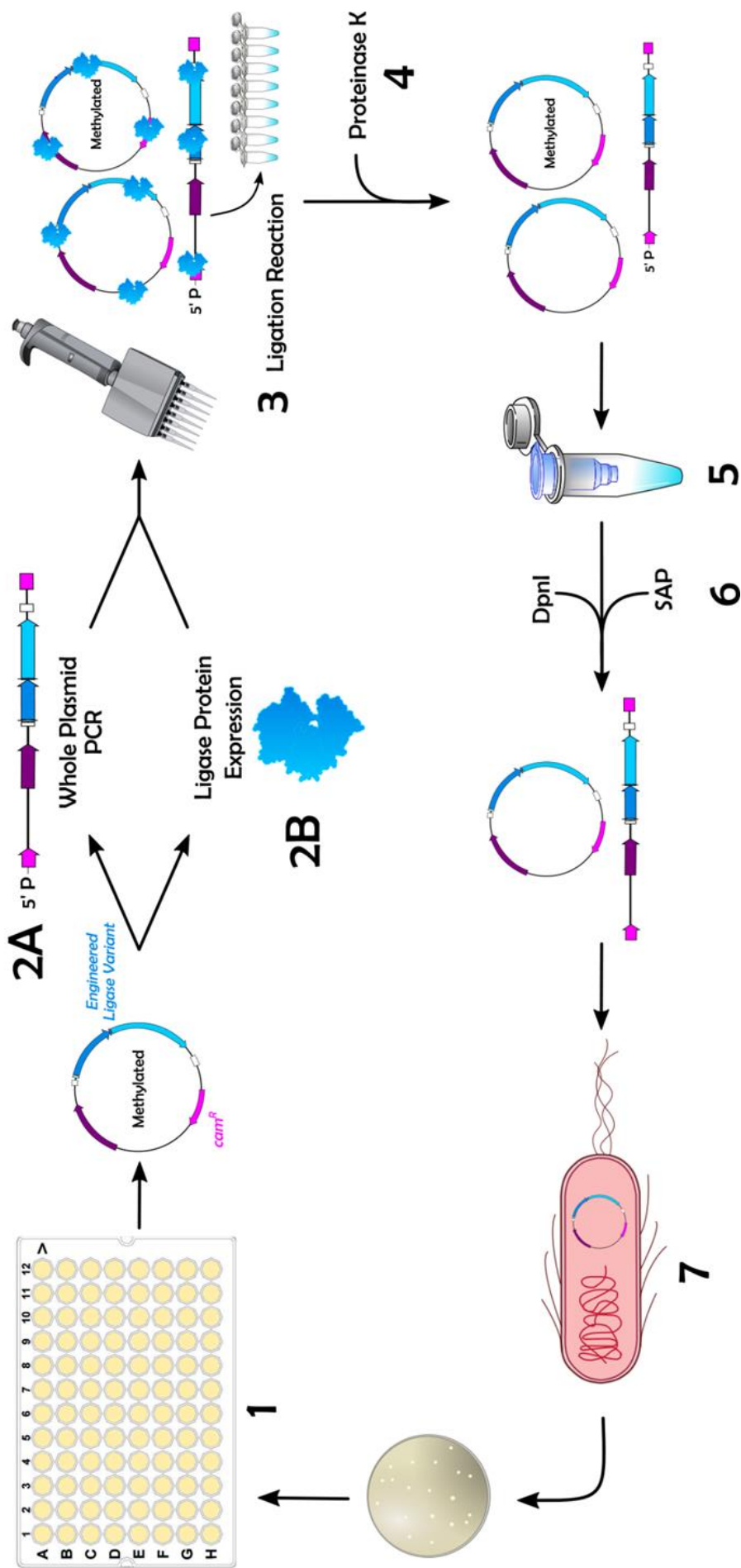


Figure 1-3 Schematic representation of the methodology for the competitive ligase selection protocol

To confirm that the competitive ligase selection was able to enrich for superior ligases exhibiting improved blunt-end ligation, p50-ligase, one of the ‘best’ ligases from Wilson et al. (2013) was challenged against one of the less active variants, PprA-ligase, across five rounds of selection. The initial pool of ligases contained a ratio of 1:5 of p50-ligase to PprA-ligase and after each round of selection, p50-ligase showed incremental increases between 10-20%. After five rounds of selection, 81% of the colonies screened expressed p50-ligase. The results validated the selection and paved the way for efforts to evolve for a superior DNA ligase variant for blunt-end ligation.

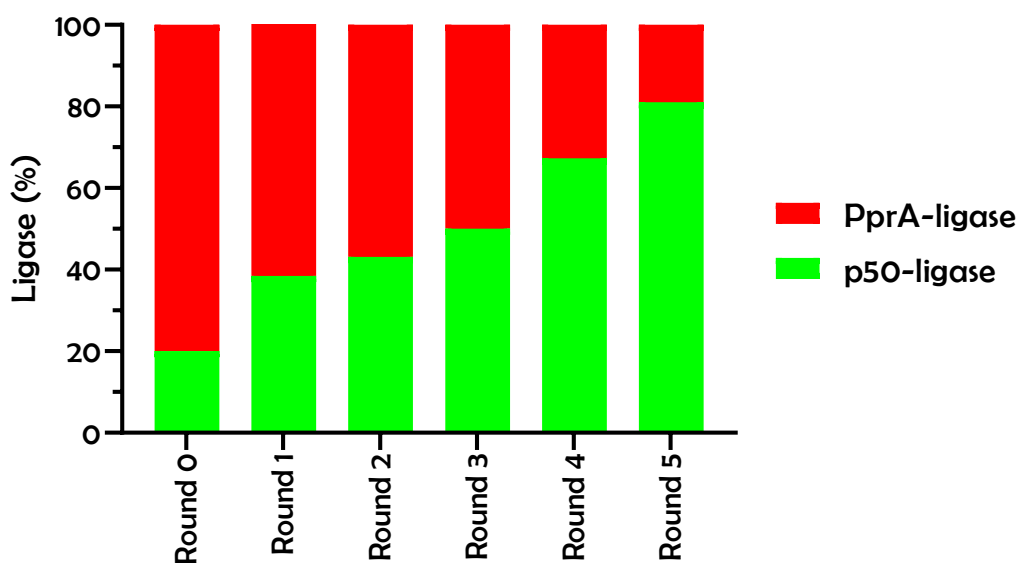


Figure 1-4 Previous validation of the evolution of the optimised protocol to select for p50-ligase variants over less-active PprA-ligase variants over five rounds.

The initial pool of ligase variants was comprised of 80% of the ‘bad’ ligase, PprA-ligase in red and 20% of the ‘good’ ligase, p50-ligase in green. Five rounds of competitive ligase selection were performed as shown in Figure 1-3. A colony PCR was used to assess the final percentage of each ligase at the end of each round. The forward primer bound to the DNA-binding domain of P50 or PprA and the reverse primer that bound to T4 DNA ligase. This gave a PCR product that was resolved by agarose-gel electrophoresis and could be distinguished by a size shift between the two variants. Data from Dr Katherine Robins (unpublished results).

1.5.1 Protein Linkers

Protein fusion provides a unique opportunity to improve a desired enzymatic activity through the introduction of short sequences of amino acids, or linker regions, between two (or more) desirable domains or proteins. Linkers are a well-established feature in biochemistry where they commonly assist in the purification process through the addition of affinity tags, such as the poly-histidine tag (His-tag), FLAG-tag, or calmodulin-binding peptide to name a few (Terpe 2003). In addition to connecting two proteins

together, linkers may provide added benefits to the fusion protein. These include increased protein stability or folding, increased protein expression, or improved biological activity. The identification of a suitable protein linker is not a simple process due to the complexity of their two- and three-dimensional structures. An effective linker must not only allow for the components to be connected but must also ensure they are compatible or do not interfere with their normal interactions or catalysis.

Although linkers must be empirically tested in each unique setting, a large number of naturally occurring multi-modular proteins have facilitated their design (Argos 1990; Chen, Zaro, and Shen 2013; George and Heringa 2003). There are three general subclasses of linkers: cleavable, flexible, and rigid linkers. As cleavability is not a desired quality, this study focuses on flexible and rigid linkers. Following the small-scale linker experiment, we aimed to create a larger ligase linker variant library that investigated varying the flexibility and length of the linker region between T4 DNA ligase and DNA-binding domains, cTF and p50. Dr Robins was the lead researcher who designed and created the ligase-linker libraries, considering several key criteria as described below.

Flexible linkers are commonly small, hydrophobic peptides containing glycine, serine, or threonine (Argos 1990). An advantage of flexible linkers is that allow for separation of two proteins, whilst at the same time typically do not adversely affect the function of the individual protein they attach to. Many glycine-rich or glycine-serine linkers have been identified from a range of proteins and their interacting partners. Glycine-rich linkers can influence a multitude of changes that are highly protein/s specific. These effects may include changing the oligomeric states of proteins, proteolytic resistance, and protein solubility or stability (Reddy Chichili, Kumar, and Sivaraman 2013). In addition, a general trend has been displayed with the incorporation of alanine or threonine residues to assist in maintaining flexibility. Conversely, glutamate or lysine have been incorporated to improve solubility (Chen, Zaro, and Shen 2013).

Alternatively, rigid linkers are used to retain a fixed distance between two domains and prevent unwanted interactions (George and Heringa 2003). This is important for proteins that undergo conformational changes or to prevent the introduction of steric clashes that may inhibit activity (Chen, Zaro, and Shen 2013). Rigid linkers either form α -helices with repetitive sequences containing glutamate, alanine, and lysine (EAAK), whereas non-helical linkers contain an iteration of proline residues with either alanine,

lysine, or glutamate. Poly-proline linkers have increased stiffness caused by its restricted conformation introduced by the secondary amine ring that effectively isolates the two proteins.

In addition to the composition of residues in linkers, altering the linker length has been found to influence enzymatic activity, protein stability, and domain orientations (George and Heringa 2003; Reddy Chichili, Kumar, and Sivaraman 2013). Typically, the length of linkers ranges anywhere from 2 to 31 amino acids and will often require optimisation for each specific protein fusion construct. The design and application of the competitive ligase selection allows us to speed up the optimisation of the linker for the promising candidate's ligase-cTF and ligase-p50 by varying compositions and lengths without the need for individual purification and testing of each variant.

1.5.2 Creation of Ligase Linker Libraries

To continue the work initiated by the small-scale linker modification of ligase-cTF, Dr Robins designed and generated a larger library than had been previously made by Wilson et al. (2013). Linkers were modified from the original ligase-cTF variant by whole plasmid PCR using primers that contained the varying linker sequence (Figure 1-5). The plasmid amplicon was phosphorylated and ligated together to produce the new modified linker variant. This method was used to create the ligase-cTF linker library and facilitated the investigation of a second DNA-binding domain, ligase-p50, by only requiring additional forward primers specific to p50. The notation for the ligase linker variants and their amino acid code can be found in Appendix S1 and S2.

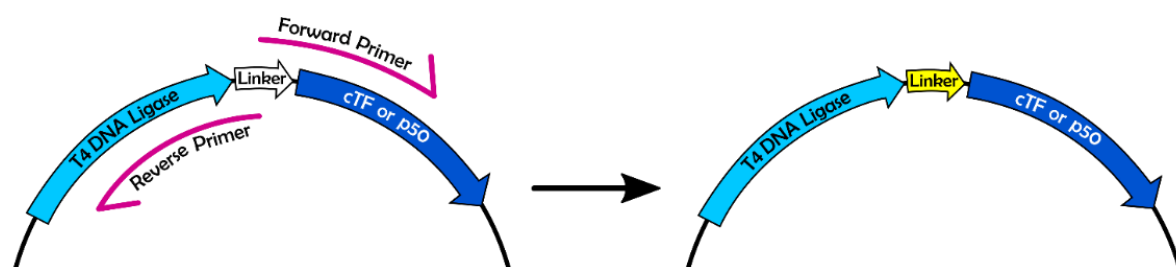


Figure 1-5 Schematic for the modification of the linker region in ligase libraries.

The linker region was modified by whole plasmid PCR with primers that contained the new sequence to be introduced. The forward primer bound to the DNA-binding domain, whereas, the reverse primer bound to T4 DNA ligase. The linearised amplicon was incubated with a kinase to create 5' phosphate ends, and the phosphorylated amplicon was ligated to form the new ligase-linker variant. A library of variants containing modified linkers was used to transform *E. coli* DH5 α and used in the competitive ligase selection protocol.

1.5.3 Enriched Engineered DNA Ligases

A total of 33 ligase-cTF linker variants and 37 ligase-p50 linker variants were created (defined in Appendix S1 and S2). Five rounds of selection were applied to each library with two biological replicates per library. In each case, the first round of selection was performed in triplicate (i.e. 99 starting microwells total for ligase-cTF, and 111 for ligase-p50) and each subsequent round used 95 picked colonies (with one negative control). The results of the changes are summarised in Figure 1-6. After five rounds of selection, the ligase-cTF linker library successfully selected for a single enzyme variant exhibiting improved blunt-end ligation. This linker variant, ligase-CA3, made up 42% of final variants in Figure 1-6, with approximately equal proportions arising in each of the two biological replicates performed (Figure 1-6, Ligase-cTF Linker Library, Replicate A and B). In comparison, the ligase-p50 linker library did not result in the enrichment of a dominant enzyme linker variant after five rounds. At the end of Round 5, the ligase-p50 library contained 21 out of 37 linker variants, whereas the ligase-cTF library contained only 13 out of the 33 starting linkers. Unlike, the two biological replicate plates from the ligase-cTF library, the two replicates from the ligase-p50 library resulted in the enrichment of two very divergent libraries (Figure 1-6, Ligase-p50 Linker Library, Replicate A and B). Across the two replicate plates from the ligase-p50 library, only one linker sequence, ligase-PHL10 displayed a notable enrichment after five rounds of selection and was the “top” candidate establishing 11.4% of the total variants recovered. The remainder of the linker variants showed varying degrees of enrichment with ligase-PPA17, ligase-PL1, ligase-PPSR3, and ligase-PG2G12 establishing a greater percentage of the variants in Replicate A; with ligase-PGSG15, ligase-PGSE18, ligase-PO2, and even original linker, ligase-p50 establishing a greater percentage of the variants in Replicate B.

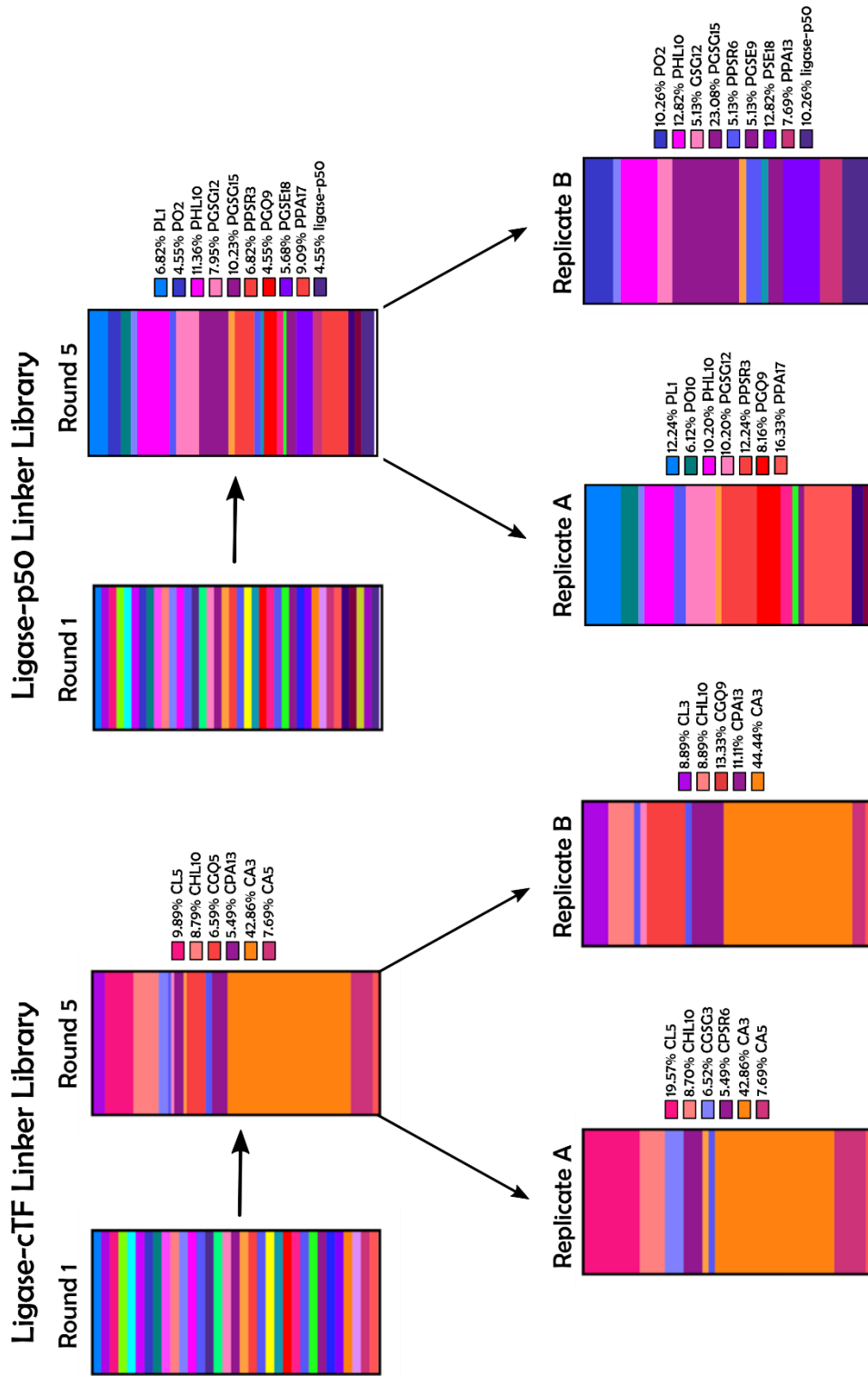


Figure 1-6 Evolution of ligase-cTF and ligase-p50 linker libraries following five rounds of selection.

The linker libraries contained three of each ligase linker variant at Round 1, each being represented by a different colour in the figure. Five rounds of selection were conducted and sequenced to determine the changes in expression/enrichment. The total summary of linker variants at the end of Round 5 for ligase-cTF and ligase-p50 libraries are displayed on the top row, whereas, the breakdown of the two replicate plates, denoted A and B for each library displayed on the bottom row. Ligase variants recovered from the Round 5 libraries are displayed if they contain 1 or more sequence identified linkers

1.6 AIMS AND HYPOTHESIS

After the departure of Dr Robins from the Ackerley lab, I took over the ligase project, at the point where five rounds of competitive ligase selection had just been completed. As no clear ‘winner’ was evident from the ligase-p50 library, the key goals of this thesis were to perform one more round of selection on the ligase-p50 library to determine if there was any evidence that a winning variant was starting to emerge. The second goal was to establish an improved method to quantify ligation activity in our lab and to evaluate the activity of the top selected DNA ligase variants.

The aims of this research were therefore to:

1. Perform one more round of ligase selection on the Round 5 ligase-p50 library.
2. Establish and optimise a fluorescence-based ligation assay.
3. Evaluate the ligation activity of our superior engineered DNA ligases from the selection.

Manuscripts in Preparation

Sharma, Janine K., Robins, Katherine J., Stevenson, Luke J., Williamson, Adele F., Patrick, Wayne M., Ackerley, David F. “Methods for Competitive Enrichment and Evaluation of Superior DNA Ligases.” *Submitted to a special edition of Methods in Enzymology for publication in October 2020.*

Robins, Katherine J., Stevenson, Luke J., Sharma, Janine K., Williamson, Adele F., Arcus, Vickery L., Patrick, Wayne M., Ackerley, David F. “A compartmentalised self-replication protocol enables selection of improved blunt-end DNA ligases.” *Manuscript in preparation.*

Chapter 2: General Methods and Materials

2.1 MATERIALS

Chemical and reagents used in this study were obtained from Sigma-Aldrich (St. Louis, MO, USA), Thermo Fisher Scientific (Waltham, MA, USA), or Total Lab Systems (Auckland, NZ), unless otherwise stated. Luria Bertani (LB) powder, chloramphenicol, isopropyl β -D-1-thiogalactopyranoside (IPTG), and micro-agar powder were purchased from Duchefa Biochemie (Haarlem, Netherlands). Agarose and ethidium bromide were purchased from Hydragene (Piscataway, NJ, USA). ATP disodium salt, and boric acid were purchased from Chem Impex (Wood Dale, IL, USA). Restriction enzyme *DpnI*, rSAP, and Q5® DNA polymerase and buffers were purchased from NEB (Ipswich, MA, USA). 2× Biomix™ Red, Hyperladder 1 kb, and proteinase K were purchased from Biorline (Memphis, TN, USA). PageRuler protein ladder 10-180 kDa, and yeast tRNA were purchased from Life Technologies (Carlsbad, CA, USA). Protein assay dye reagent concentrate and 40% acrylamide-bis, 19:1 were purchased from Bio-rad (Hercules, CA, USA). BugBuster® protein extraction reagent was purchased from Merck & Co. (Kenilworth, NJ, USA)

2.2 BACTERIAL STRAINS AND PLASMIDS

2.2.1 Bacterial Strains

The two bacterial strains used in this study are listed in Table 2-1.

Table 2-1 *Escherichia coli* strains used in this study

Strain	Genotype	Reference
DH5 α	F- Φ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1endA1</i> <i>hsdR17(rK-mK⁺) nupG</i> <i>supE44 (glnV44) λ-thi1</i> <i>gyrA96 relA1</i>	Hanahan (1983)
GR501	Hfr LAM- <i>ligA251 relA1 spoT1</i> <i>thi-1 ptsI⁺</i>	Dermody, Robinson, & Sternglanz, (1979)

2.2.2 Plasmids

The plasmid used in this study was pCA24N (Kitagawa et al. 2005). This is a high copy number, expression vector that is under the control of an IPTG inducible T5 promoter. pCA24N can also be used for high levels of protein expression and purification, containing a N-terminal hexahistadine tag (His-tag) connected by a short linker. The antibiotic resistance selection marker is chloramphenicol. Example plasmid maps can be found in Appendix S3.

2.3 PRIMERS

All primers were supplied as a lyophilised powder and purchased from Integrated DNA Technologies (IDT; Carlsbad, CA, USA). All stock solutions were made up to 100 μ M stock solution in 1 \times Tris-EDTA (TE; 10 mM Tris-Base pH 8.0, 0.1 mM EDTA). Primers were diluted to 10 μ M in dH₂O to form a working stock. All solutions were aliquoted to volumes that reduced freeze-thaw cycles and stored at -20 °C.

Table 2-2 Primers used in this study

Primer	Sequence (5' → 3')
CAT_Phos_Fwd ^a	5' Phosphate - GTTCAGCTGGATATTACGGCCT
CAT_Phos_Rev ^a	5' Phosphate - GGTCTGGTTATAGGTACATTGAGC
link_screen_fwd ^a	ACGTGAAGATAAAACTAAAGC
p50_int2_rev ^a	GCACCACGCTGAGGTCCATCTC
Fwd_p50_PA	CCGGCACCAGCCCCAGCAGATGGCCCATACCTT
Rev_PA12	GGCTGGCGCTGGTGCCGGGGCTGGTGCTGGCGCCGGTAGACCAGTTACCTCAT G

2.4 LIGASE CONSTRUCTS USED IN THIS STUDY

Ligase constructs were originally made by Dr. Robert Wilson with linker modifications made by Dr Katherine Robins via whole plasmid PCR. All resultant products were transformed into *E. coli* DH5 α .

Table 2-3 Ligase variant constructs assayed in this study

Ligase Variant	Linker	Molecular Weight (kDa)	Molecular Coefficient or ϵ (M ⁻¹ cm ⁻¹)
T4 DNA Ligase	N/A	57.14	57,300
Ligase-cTF	GTSGGGSGGG	91.49	81,710
Ligase-CAA3	AAA	91.02	81,710
Ligase-p50	GTSGGGSGGG	94.63	80,680
p50-ligase	GTSGGGSGGG	94.63	80,680
Ligase-PPA17	PAPAPAPAPAPAPAP	95.40	80,680

2.5 BACTERIAL CELL GROWTH AND STORAGE

E. coli DH5 α bacteria was grown at 37°C and aerated at 200 rpm for liquid culture, or static for solid culture growth, unless otherwise stated. *E. coli* GR501 bacteria was grown

at 28°C, 200 rpm for liquid culture growth. *E. coli* GR501 bacteria grown on solid agar plates was set up in duplicate plates and incubated at 28°C and 42°C to allow for the selection of ligase genes at the non-permissive temperature.

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 80% (v/v) glycerol and stored at -80°C indefinitely.

2.5.1 Liquid Media

Table 2-4 LB Media

Reagent	Final Concentration
NaCl	10 g.L ⁻¹
Tryptone	10 g.L ⁻¹
Yeast Extract	5 g.L ⁻¹

*Reagents were bought as a pre-made powder and 20 g.L⁻¹ was added per 1 L of dH₂O.

Table 2-5 Super-Optimal Broth (SOB-)* or SOB with Catalytic Repression (SOC)#

Reagent	Final Concentration
Tryptone	20 g.L ⁻¹
Yeast Extract	5 g.L ⁻¹
NaCl	0.5 g.L ⁻¹
KCl	2.5 mM
MgCl ₂ ^a	10 mM
D-Glucose ^a	20 mM

*Adjusted to pH 7.0 with NaOH before autoclaving

#Additional reagents were added post-autoclave and the media was renamed

^a Filter-sterilized using a 0.22 µM filter

2.5.2 Solid Media

1.5% (w/v) agar was added to LB media prior to autoclaving. Any necessary supplements were added after the agar had cooled to <50°C post-autoclave. Agar was poured in a laminar flow cabinet in ~20 mL aliquots into sterile 90 mm Petri dishes or ~10 mL aliquots into 30 mm Petri dishes and left to solidify at room temperature. Storage of agar plates containing supplements was at 4°C.

2.5.3 Media Supplements

All media supplements were filter-sterilized with a 0.22 μM syringe filter and stored at -20°C until required.

Table 2-6 Media supplements used in this study

Supplement	Buffer Preparation	Stock Concentration
Chloramphenicol ^a	100% EtOH	34 mg.mL ⁻¹
IPTG	dH ₂ O	100 mg.mL ⁻¹

^a antibiotics were made up at 1000× stock concentration

EtOH is ethanol

IPTG is isopropyl β -d-1-thiogalactopyranoside

2.6 ROUTINE MOLECULAR BIOLOGY PROTOCOLS

2.6.1 Polymerase Chain Reaction

All PCR reactions were performed in either a Super Cycler (Kyratec; Al Barsha South, Dubai, UAE), or a DNA Engine (Bio-rad; Hercules, CA, USA). Whole plasmid PCR amplification reactions were performed using Q5® High-Fidelity DNA Polymerase to a final volume of 10 μL . For other applications and colony screening, Biomix™ Red was employed. PCR reactions were set up according to the manufacturer's protocol and as outlined in Table 2-7 and Table 2-8. All experiments were set up with a PCR master-mix (excluding template DNA), and aliquoted to the desired volume in the PCR tubes. Standard PCR parameters were optimised as required for individual genes.

Table 2-7 Composition and cycling parameters for PCR reactions containing Q5 DNA polymerase

Component	Volume per 20 μL	Final Concentration
5× Q5 Reaction Buffer	4 μL	1×
5× Q5 High GC Enhancer	4 μL	1×
10 mM dNTPs	0.4 μL	200 μM
10 μM Forward Primer	1 μL	0.5 μM
10 μM Reverse Primer	1 μL	0.5 μM
Q5 High-Fidelity DNA Polymerase	0.2 μL	~0.02 U. μL^{-1}
Sterile Molecular Biology Grade Water	8.4 μL or to 20 μL	

Temperature (°C)	Time	Cycles
98	10 min	1
98	10 s	
T _a *	20 s	27-30
72	30 s.kb ⁻¹	
72	10 min	1

* The annealing temperature (T_a) 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).

Table 2-8 Composition and cycling parameters for PCR reactions containing Biomix™ Red

Component	Volume per 20 μ L	Final Concentration
2 \times Biomix™ Red	10 μ L	1 \times
10 μ M Forward Primer	1 μ L	0.5 μ M
10 μ M Reverse Primer	1 μ L	0.5 μ M
Template DNA ^a	1 μ L	2-10 ng
Sterile MilliQ Water	7 μ L or to 20 μ L	

Temperature ($^{\circ}$ C)	Time	Cycles
95	5 min	1
95	10 s	
T _a *	20 s	25-30
72	30 s.kb ⁻¹	
72	5 min	1

^a Template DNA was obtained from a colony or using 1 μ L of 10 ng. μ L⁻¹ plasmid prepped DNA

*The annealing temperature (T_a) was altered depending on the melting temperature (T_m) of the primers. The T_a used was 5 $^{\circ}$ C lower than the lowest T_m value of the primer(s).

2.6.2 Agarose Gel Electrophoresis

DNA samples and PCR products were analysed on agarose gels dissolved in 1 \times TAE buffer (40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA, pH 8.0) supplemented with 1 μ g.mL⁻¹ ethidium bromide. A 1% (w/v) agarose gel was typically employed, however, 2% (w/v) agarose gels were used when samples were between 0.2 kb and 3 kb. Gels were covered in 1 \times TAE buffer and run using a Bio-rad PowerPac™ Basic Power Supply at 110-130 V for 30-60 min. Hyperladder 1 kb marker was used as a DNA standard and the bands were visualised and imaged under UV light on a UVITEC Cambridge Gel Documentation System.

2.6.3 Restriction Enzyme Digests

The general reaction set up utilised NEB restriction digest buffers and enzymes to a final volume of 25 μ L. Restriction digests were incubated at 37 $^{\circ}$ C for 1-4 hr after which were heat inactivated at 80 $^{\circ}$ C for 20 min. Digested DNA was typically used immediately otherwise stored at -20 $^{\circ}$ C.

2.6.4 Isolation of Plasmid DNA

Overnight cultures of 1.5 mL to 5 mL were pelleted by centrifugation at 17,000 g for 1 min and supernatant discarded and repeated until all culture was pelleted. All centrifugations were performed at 17,000 g at room temperature. Cell pellets were fully resuspended in 200 μ L PD1 (100 μ g.mL⁻¹ Ribonuclease A, 50 mM Tris-HCl, pH 7.5, 10 mM EDTA). Cells were lysed by adding 200 μ L PD2 (0.2 M NaOH, 1% sodium dodecyl sulfate (SDS)) and inverted three times and left for 3 min at room temperature. The solution was

neutralised with the addition of 350 μL N3 (4.2 M GuHCl, 0.9 M potassium acetate) and quickly inverted to form a white precipitate. The precipitate was centrifuged for 4 min, and the supernatant was transferred to Plasmid DNA Mini-Prep Column (Epoch Life Science, Missouri, TX, USA). The DNA was bound to the column by centrifugation for 1 min and flow-through discarded. The column was then washed and centrifuged for 1 min with 400 μL of PB (5 M GuHCl, 30% isopropanol), followed by two washes with 500 μL PE (10 mM Tris-HCl, pH 7.5, 80% EtOH). The column was dried by centrifugation for 3 min and 40-60 μL of pre-heated (55°C) sterile MilliQ water was added to the column and eluted by centrifugation at for 1 min. Plasmid DNA was stored at -20°C.

When a high-throughput approach was required a 96 Well Batch DNA Binding Plate (Epoch Life Science, Missouri, TX, USA) was employed DNA binding and washing was conducted using a manifold vacuum regulator. DNA was eluted in 50 μL pre-heated (55°C) by centrifugation at 3190 g for 1-2 min.

2.6.5 DNA Quantification

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

2.6.6 DNA Sequencing

All Sanger sequencing was carried out by Macrogen Inc. (Seoul, South Korea) to view the changes between Round 5 and 6 in the competitive ligase selection protocol to determine the linker variant sequences. Where possible, at least 100 $\text{ng}\cdot\mu\text{L}^{-1}$ of plasmid DNA was sent as a sequence template. All reactions were performed with link_screen_fwd primer.

2.7 BACTERIAL CELL TRANSFORMATION AND ELECTROPORATION

2.7.1 Preparation of Electrocompetent Cells

The *E. coli* strain to be made competent was inoculated from a glycerol stock and grown in LB media overnight at 28°C or 37°C, 200 rpm for GR501 and DH5 α , respectively. A 1:100 dilution of overnight culture was used to inoculate fresh LB media and grown at 28°C or 37°C, 200 rpm for the respective strain until an OD₆₀₀ 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 MilliQ 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. 40 μ L 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.

2.7.2 Transformation of Electrocompetent Cells

Frozen cell aliquots were removed from storage and thawed on ice. Up to 10 ng of plasmid DNA (in 5 μ L maximal volume) was added to 40 μ L of cells and gently mixed. The cells were transferred to 2 mm gap electroporation cuvette and electroporated at 2.5 kV, 25 μ F, 100 Ω . Next, 800-1000 μ L of SOC media was immediately added to the cuvette and subsequently transferred to a sterile 15 mL conical tube. Cells were recovered for 60-80 min at 28°C, 200 rpm for GR501 cells or 37°C, 200 rpm for DH5 α . Typically, 100 μ L of undiluted cells was plated on LB agar, supplemented with necessary antibiotics, and incubated according to Section 2.5.2.

2.8 SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

ExpressPlus™ pre-cast PAGE gels (4-20%) were purchased from Genscript (Piscataway, NJ, USA). Protein samples were mixed at a 2:1 ratio with 3 \times 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 μ L PageRuler Protein Ladder 10 – 180 kDa (Life Technologies, Carlsbad, CA, USA) to estimate protein band sizes. Precast gels were run in 1 \times Tris-MOPS-SDS buffer supplied with the gels at 150 V for 60 min. For visualisation of protein expression, gels were stained in Coomassie Blue Stain (2.5 g.L⁻¹ Coomassie Blue Brilliant, 45% absolute ethanol (w/v), 10% acetic acid (w/v)) for 30+ min and soaked destain buffer (10% (w/v) acetic acid, 40% (w/v) methanol) until bands could be visualised and imaged.

2.9 COMPETITIVE LIGASE SELECTION PROTOCOL

2.9.1 Whole Plasmid PCR

A whole plasmid PCR was used to create a blunt-end linear PCR product that is the substrate for the ligation reaction in Section 2.9.3. The template was derived either from a 1:5 dilution with sterile molecular biology grade water of an overnight culture or 2-10 ng.µL⁻¹ of plasmid DNA. 5' phosphorylated primers (CAT_Phos_Fwd and CAT_Phos_Rev; Section 2.3) amplified the entire plasmid from within the *cam^R* gene, also known as the chloramphenicol acetyltransferase (CAT) selectable marker. A PCR master-mix was made up (as shown in Table 2-9), 9 µL of which was added to each well of a 96 well PCR plate.

Table 2-9 Whole Plasmid PCR for Creation of Ligation Reaction Substrate

Reagent	Volume
5× Q5 Reaction Buffer	190 µL
5× Q5 High GC Enhancer	190 µL
10 mM (each) dNTPs	19 µL
10 µM Forward Primer (CAT_Phos_Fwd)	47.5 µL
10 µM Reverse Primer (CAT_Phos_Rev)	47.5 µL
Q5 DNA Polymerase	9.5 µL
Sterile Molecular Biology Grade Water	450 µL

The final reaction volume of each PCR reaction was 10 µL following the addition of 1 µL template DNA. The PCR was performed using the following cycling profile:

Table 2-10 Cycling profile for Whole Plasmid PCR

Temperature (°C)	Time	Cycles
98	1 min	1
98	10 s	
61	20 s	27
72	3 min 30 s	
72	10 min	1
12	∞	

2.5 µL of each PCR reaction product was added to an equal volume of DNA loading dye and electrophoresed on a 1% agarose gel at 130 V for 35 min to resolve a 7 kb band indicative of the linearized plasmid. The PCR products were kept on ice (or stored at –20°C) until used as described in Section 2.9.3. For the ligation reaction to go ahead, 95% of successful PCR reactions products had to present with a 7 kb band on the gel.

2.9.2 Ligase Protein Expression

An overnight culture of the 96 well plate from the previous round of successful hits was set up in 200 µL chloramphenicol supplemented LB media and grown for >18 hr. 60 µL

of each well of the overnight culture was added to a 96 well plate containing 140 μL chloramphenicol supplemented LB media amended with 0.57 mM IPTG (i.e., 0.4 mM IPTG final concentration in 200 μL) and incubated at 30°C, 200 rpm, wrapped in aluminium foil for 3 hr. Next, 20 μL of the protein expression culture was lysed with 20 μL of BugBuster™ Protein Extraction Reagent (Merck; Darmstadt, Germany) and the mixture was incubated on an orbital shaker at room temperature, ~100 rpm for 40 min.

2.9.3 Ligation Reaction

Ligation reactions were set up in a 96 well plate to a final volume of 50 μL . Each reaction contained 7.5 μL PCR product, 20 μL lysed expression cells, 5 μL 10 \times T4 DNA ligase buffer, and 17.5 μL water. The PCR plate was sealed with a sealing film and the reaction was incubated at 25°C for 60 min, followed by a heat-inactivation step at 80°C for 20 min.

2.9.4 Proteinase K

All ligation reactions were pooled and thoroughly mixed by pipetting. To prevent any ligase interference in downstream reactions (e.g., ligase bound to the recircularised plasmid had previously been found to impair transformation efficiency (Wilson et al. 2013), 28 μL of 20 $\text{mg}\cdot\text{mL}^{-1}$ proteinase K was added to 100 μL of the pooled ligation reaction, to give a working concentration of 4.4 $\text{mg}\cdot\text{mL}^{-1}$ proteinase K. The reaction was carried out at 56°C for 60 min, followed by heat inactivation at 80°C for 20 min.

2.9.5 Clean and Concentrate of DNA

To remove cell debris, proteins, primers, and buffers, the total volume of the proteinase K reaction was combined into a 15 mL conical tube and 5 mL of PB buffer (5 M guanidinium hydrochloride, 30% isopropanol) was added. 1 mL of 0.5 M sodium acetate, pH 5.2 was added and mixed by inversion several times. 700 μL was added at a time into a silica spin column (Zymo Research; Irvine, CA, USA) and centrifuged at 17,000 g for 30 s until the total volume had been passed through. Next, 200 μL of PE buffer (10 mM Tris-HCl, pH 7.5; 80% ethanol) was added to the column and centrifuged at 17,000 g for 30 s and the flow-through discarded. The wash step was repeated once, and the column was dried by centrifugation at 17,000 g for 3 min. The spin column was transferred to a 1.5 mL microcentrifuge tube and 25 μL of sterile molecular biology grade water was added to the column and incubated for 5 min at 50°C. DNA was eluted by centrifuging for 30 s

at 17,000 g. The eluent was either stored overnight at -20°C or progressed through to the transformation step (Sections 2.9.6 to 2.9.8).

2.9.6 *DpnI* and Alkaline Phosphatase Reaction

The total volume of DNA elution from Section 2.9.5 was combined with 3 µL of 10× CutSmart buffer, 1 µL *DpnI*, and 1 µL rSAP and incubated at 37°C for 60 min, followed by a heat inactivation step at 80°C for 20 min.

2.9.7 Isopropanol Precipitation of DNA

In a 1.5 mL microcentrifuge tube, the DNA from Section 2.9.6 was combined with 1 µL 10 mg.mL⁻¹ yeast tRNA, 60 µL isopropanol, 24 µL 0.5 M sodium acetate, pH 5.2 before centrifugation at 13,000 g, 4°C for 10 min. The supernatant was carefully discarded, and the pellet was washed in 1 mL of ice-cold 100% ethanol before centrifugation at 13,000 g, 4°C for 10 min. Ethanol was removed and the DNA pellet was left for 10 min to dry. The DNA pellet was resuspended in 10 µL sterile H₂O and left at room temperature for a 2-3 hr to allow for complete resuspension. DNA was stored at -20°C if it was not transformed on the same day.

2.9.8 Transformation

2 µL of the resuspended DNA was added to 40 µL of electrocompetent DH5α cells. Cells were electroporated according to Section 2.7.2 and recovered in 800 µL SOC media for 75 min. 100 µL was spread over 8× LB agar plates amended with chloramphenicol and incubated overnight at 37°C.

2.9.9 Colony Screening

To determine whether each colony represented a genuine ligase variant and rule out the presence of false-positive colonies, a 96 well colony PCR screen was used to amplify across the linker region of T4 DNA ligase and the DNA-binding domain. A PCR master-mix consisting of 300 µL 2× Biomix Red, 15 µL of 10 µM link_screen_fwd primer and 15 µL of 10 µM p50_int_rev2 for ligase-p50 variants, plus 270 µL sterile MilliQ water was set up and 6 µL was aliquoted to each well of a 96 well PCR plate. A 96 well culture plate was also set up containing 200 µL LB supplemented with chloramphenicol. Using a sterile pipette tip, a single colony was picked from the agar plate. The pipette tip was dipped into one well in the PCR plate and briefly swirled to mix, followed by the pipette tip being dipped into the corresponding well in the culture plate and swirled to mix. The culture

plate was incubated at 37°C, 200 rpm overnight, and PCR was performed according to the following protocol:

Table 2-11 Cycling profile for ligase colony screening

Temperature (°C)	Time	Cycles
95	1 min	1
95	10 s	
52	20 s	27-35
72	30 s	
72	10 min	1

The entire volume from each of the PCR reactions was electrophoresed on a 2% agarose gel at 130 V for 35 min. A genuine ligase variant was confirmed by the presence of a product band at ~500 bp, indicative of a successful transformation. The correspondent culture wells from successful PCR amplifications were combined with an equal volume of 80% glycerol and stored at -80°C.

2.10 PROTEIN PURIFICATION

All proteins to be expressed and assayed were cloned into pCA24N vector containing a N-terminal His-tag and transformed into *E. coli* strain DH5 α for purification via immobilised metal affinity chromatography (IMAC).

2.10.1 Ligase Expression

A glycerol stock of *E. coli* DH5 α 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 400 mL LB media supplemented with chloramphenicol and incubated at 37°C, 200 rpm. Growth of cells was monitored by optical density at 600 nm (OD₆₀₀) in 1 mL cuvettes to an OD₆₀₀ of 0.5-0.7. The culture was transferred to ice for 15 min to cool before the addition of 0.4 mM (final concentration) IPTG and grown for an additional ~20-22 hr at 18°C, 200 rpm. The cells were harvested by centrifugation in 50 mL conical tubes at 3190 g, 4°C, for 30 min. The supernatant was discarded, and cell pellets were stored at -20°C until required.

2.10.2 Cell Lysis and Separation and Clean Up of Soluble Protein

For protein purification, cell pellets were thawed and resuspended in 15 mL of Lysis Buffer (50 mM potassium phosphate, pH 8.0, 300 mM NaCl, 10% glycerol (v/v)). The cell suspension was lysed by sonication on ice with 15 cycles of 10 s on, 20 s off at a 20% amplitude. The insoluble cellular debris was pelleted at 7,000 g in a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Scientific™; Waltham, MA, USA) at 4°C for 30 min. The

soluble fraction was collected, and syringe-filtered through a 0.45 μM pore filter, followed by a 0.22 μM pore filter to remove additional contaminants before loading on to a His-Trap column.

2.10.3 Protein Purification via IMAC

Purification of His-tag proteins was achieved either using an ÄKTA start (GE Healthcare; Chicago, IL, USA), or using the His·bind™ Ni-NTA chromatography kit (Novagen, Inc; Madison, WI, USA) and a peristaltic pump.

2.10.4 ÄKTA start Purification

Approximately 15 mL of the filtered soluble cell lysate was loaded via the sample pump on to a 5 mL HisTrap FF column at a flow rate of 0.5 mL.min⁻¹. The column was washed with 10 column volumes of Lysis Buffer to remove any unbound proteins. Proteins were eluted on a 0-100% gradient of Elution Buffer (50 mM potassium phosphate, pH 8.0, 300 mM NaCl, 10% glycerol (v/v), 500 mM imidazole), with protein typically eluting between 150-200 mM imidazole concentration depending on the ligase variant. The peak fractions from the trace were pooled and protein samples were concentrated using an Amicon® Ultra-15 Centrifugal Filter Unit, 30 kDa molecular weight cut-off (Merck, Kenilworth, NJ, USA). Concentrators were spun at 3190 g, 4°C to a final volume 1.5 mL.

2.10.5 Purification with a Peristaltic Pump

To maximise His-tag binding, the filtered soluble cell lysate was combined with ~ 2mL of pre-charged Ni²⁺ resin slurry, transferred to a 15 mL conical tube, and agitated for 1 hr on ice. The slurry was added to the column and the resin was left to settle. The soluble cell lysate was run through the column at a flow rate of 15-20 mL.min⁻¹ twice to increase His-tag binding before being discarded. 20 mL of Lysis Buffer was used to wash any unbound protein, followed by 10 mL of Wash Buffer 1 to discard any contaminant proteins. A second wash step with Wash Buffer 2 was used for ligase variants containing DNA-binding domain, p50. The ligase protein was eluted at a flow rate of ~10 mL.min⁻¹ with 8 mL of Elution Buffer 1 for ligase proteins fused to cTF or Elution Buffer 2 for ligases fused to p50.

Table 2-12 Imidazole concentrations in Wash and Elution Buffer/s used in this study

Wash Buffer 1	Wash Buffer 2	Elution Buffer 1 ^(a)	Elution Buffer 2 ^(b)
Lysis Buffer + 50 mM imidazole	Lysis Buffer + 125 mM imidazole	Lysis Buffer + 150 mM imidazole	Lysis Buffer + 200 mM imidazole

a Used to elute T4 DNA ligase, ligase-cTF, and ligase-CA3

b Used to elute ligase-p50, p50-ligase, and ligase-PPA17

2.10.6 Desalting of Eluted Protein

A 5 mL HiTrip® Desalting column (GE Healthcare, Chicago, IL, USA) was pre-equilibrated with 10 mL storage buffer (Table 2-13). 1.5 mL of eluted protein was loaded on to the column. 3 mL of Storage Buffer was added to the column and eluted protein was collected in 1.5 mL fractions. Ligase variants were initially desalted in potassium phosphate (KPhos) storage buffer until the buffer conditions were compared in Section 4.6.2. Thereafter, all subsequent ligase variants were eluted in 2× NEB storage buffer.

Table 2-13 Storage Buffer Recipes for Protein Desalting

KPhos storage buffer	(2×) NEB storage buffer
50 mM potassium phosphate, pH 7.0	20 mM Tris-HCl, pH 7.4
200 mM NaCl	100 mM KCl
10% glycerol (v/v)	2 mM DTT
	0.2 mM EDTA

2.10.7 Measurement of Protein Concentration

Initially, the protein concentration for ligase variants was evaluated from the nanodrop and purity assessed on an SDS-PAGE. For the the ligase variants used in the final ligation reaction, all ligase variants were run on an SDS-PAGE to compare the protein concentration to visualise equivalent concentrations next to a 1 mg.mL⁻¹ BSA standard. Multiple techniques were next employed to more accurately determine the protein concentration. The protein molarity was determined by using one of the following equations. The final concentration was calculated from the average of the three quantitative techniques (Sections 0, 2.10.7.2, and 2.10.7.3).

Protein Concentration Equations:

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

or

$$mg.mL^{-1} = A_{280} \div \text{Molecular Coefficient } (\epsilon; \text{Table 2.3})$$

2.10.7.1 Nanodrop Spectrophotometry

2 µL of each sample was measured using a NanoPhotometer® NP80 (Implen Inc.; Westlake Village, CA, USA) and the A₂₈₀ and mg.mL⁻¹ values were recorded.

2.10.7.2 Ultraviolet-Visible (UV-Vis) Spectrophotometry

In a 10 mm rectangular quartz cuvette (Agilent Technologies; Santa Clara; CA, USA), proteins were 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). All ligase variants were blanked with 90 μL 2 \times NEB storage buffer.

2.10.7.3 Bradford Assay

In 96 well plate, 10 μL of protein was added to 200 μL of 1 \times Bradford Protein Dye (Bio-rad; Hercules, CA, USA). The reaction was incubated at room temperature in the dark for 5 min and measured at an absorbance of 595 nm. All reactions were run alongside a BSA standard ranging from 0 – 0.5 $\text{mg}\cdot\text{mL}^{-1}$. Protein concentrations were calculated using the standard curve.

2.10.8 Storage of Ligase Proteins

All protein aliquots were stored at the eluted concentration or diluted to 5 μM in their respective storage buffer before being stored at -80°C .

2.11 FLUORESCENCE-BASED LIGATION ASSAY

To quantify DNA ligation activity, an endpoint TBE-urea gel-based ligation assay was used to assess ligation efficiency. This assay utilises a 5' carboxyfluorescein (FAM) modification to examine the ability to convert a 20 nt dsDNA into a 40 nt product. This assay was adapted from that previously reported by our collaborator, Dr Adele Williamson, and uses the same two template oligonucleotides (L1 and L2; Table 2-14) to compare ligation activity of a single-stranded DNA nick, blunt- and cohesive-end dsDNA substrates (Berg, Leiros, and Williamson 2019; Williamson, Grgic, and Leiros 2018; Williamson, Rothweiler, and Leiros 2014). Reactions were adapted to suit each condition by changing the combination of complementary oligonucleotides.

2.11.1 Oligonucleotide/DNA Substrates

All oligonucleotides were supplied as a lyophilised powder purchased from Macrogen Inc. (Seoul, South Korea) unless otherwise stated. Stock solutions were made up to 100 μM in 1 \times TE buffer and further diluted to 10 μM in 1 \times TE buffer.

Table 2-14 Oligonucleotides used in the Fluorescent-Based Ligation Assay

Oligonucleotide	Sequence (5' → 3')
L1 ^a	5' FAM – AGCCCATGGCTGATATCGGGA
L2 ^b	5' Phosphate – TCCGAATTCGAGCTCCGTCG – 3' AmMO ^c
L3 ^a	5' FAM – AGCCCATGGCTGATATCGGGATCCGAATTCGAGCTCCGTCG
L4	TCCCGATATCAGCCATGGCCT CGACGGAGCTCGAATTCGGA
L6	5' Phosphate - TCCCGATATCAGCCATGGCCT
L7	CGACGGAGCTCGAATTCGGA
L8	5' Phosphate - CCCGATATCAGCCATGG
L9	CCTCGACGGAGCTCGAATTCGGAT

5' FAM is 5' end of DNA modified with fluorescein

5' Phos is 5' end DNA modified with a 5' phosphate

a L1_Lig_Assay and L3_Lig_Assay were kept in light sensitive microcentrifuge tubes to protect against fluorophore quenching or degradation.

b Oligonucleotide was purchased from Integrated DNA Technologies (Coralville, IA, USA).

c 3' AmMO is a 3' amino blocker modification.

Working concentrations were diluted to their respective concentrations in 0.22 μ M filter sterilised MilliQ water. L1 and L3 were diluted to 0.5 μ M, L2 was diluted to 2.5 μ M, and all remaining complementary oligonucleotides were diluted to 0.7 μ M. All oligonucleotides were stored at -20°C.

2.11.2 Ligation Substrate Master-Mix Preparation

All ligation reactions contained oligonucleotides L1 and L2, which formed the template strand that was visualised with the at the end of the ligation reaction. Ligation reactions were performed in a final volume of 25 μ L, comprising 22.5 μ L ligation substrate master-mix and 2.5 μ L DNA ligase. Volumes described herein are for a single reaction, and were scaled according to the number of reactions performed. For a single reaction, a 6 μ L ligation buffer was created by combining 2.5 μ L 500 mM Tris-base, pH 8.0, 0.5 μ L 1 M MgCl₂, 2.5 μ L 100 mM DTT, and 0.5 μ L MilliQ water. For blunt- and cohesive-end substrates, the ligation buffer was split into two equal (3 μ L) fractions, however, single-stranded DNA nick-sealing substrate was kept in one tube. The blunt-end oligonucleotide-buffer master-mix was made by adding 4 μ L of oligonucleotides L1 and L6 to into one tube, and oligonucleotides L2 and L7 into the second tube. For cohesive-end ligation, 4 μ L of oligonucleotides L1 and L8 were added to one tube, and oligonucleotides L2 and L9 to the second tube. For the nicked DNA assay, 4 μ L oligonucleotides L1, L2, L4 and an additional 4 μ L MilliQ water were combined into the single tube. The oligonucleotide-buffer master-mix were vortexed to mix, incubated at 95°C in a heating block for 5 min and left to cool to room temperature away from light. The blunt- and cohesive-end oligonucleotide-buffer mixes were pooled into a single tube. 0.5 μ L of 50 mM ATP was added to 22 μ L. The solution was vortexed to mix to create the

final ligation substrate master-mix. The final concentrations in the ligation substrate master-mix were 50 mM Tris-base, pH 8.0, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP. The preliminary ligation buffer contained a final concentration of 50 mM Tris-base, pH 8.0, 50 mM NaCl, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP.

2.11.3 Ligation Reaction

On ice or cooling block, 2.5 µL of ligase enzyme was added to 22.5 µL ligation substrate master-mix in a PCR tube. The reaction was performed in a PCR machine with the desired temperature and timepoint with the lid heat turned off. Upon completion, all reactions were quenched and heat-inactivated with 5 µL of formamide loading dye (95% formamide, 0.5 M EDTA, 10 mg.mL⁻¹ bromophenol blue) at 95°C for 5 min. Samples were immediately run on the gel or stored in a light-proof container at -20°C.

2.11.4 TBE-Urea Polyacrylamide Gel Electrophoresis (TBE-PAGE)

A denaturing TBE-PAGE gel was employed to resolve the fluorescent oligonucleotide bands. A 20% TBE-PAGE stock solution was made up to a final concentration of 7 M urea, 1× TBE (100 mM tris-base, 100 mM boric acid, 20 mM EDTA), and 20% acrylamide/bis-acrylamide, 19:1, and stored at 4°C. For a single gel, 50 µL 10% ammonium persulfate (APS) and 1.5 µL TEMED was added to 7.5 mL 20% TBE-PAGE stock solution and immediately transferred to fill a Mini-PROTEAN® Tetra Handcast apparatus (Bio-Rad; Hercules, CA, USA). The comb was added and left for 20-30 min to set. Cast gels were either used immediately or stored at 4°C and used the following day. All TBE gels were hand-cast in 1.0 mm MINI PROTEAN spacer plates (Bio-rad; Hercules, CA, USA) with 10 or 15 well combs.

1× TBE buffer was pre-heated for >1 hr in a 55°C water bath prior to being added to the MINI PROTEAN Tetra cell gel apparatus (Bio-rad; Hercules, CA, USA) and pre-run at a constant voltage of 200 V for ~40 min. All wells were flushed to remove dissipated urea from the wells prior to sample loading. 10 µL of each reaction was carefully loaded into each well. Gels were run in a dark polystyrene box to protect against fluorophore degradation. Bands were resolved by running at a constant 200 V for ~75 min or until the loading dye had reached the bottom of the gel. Gels were imaged using a Typhoon FLA 5000 (GE Healthcare; Chicago, IL, USA) which detected the fluorophore probe by excitation at 473 nm and emission at 518 nm.

2.11.5 Ligation Quantification

Images generated in Section 2.11.4 were analysed using ImageJ software (Schneider, Rasband, and Eliceiri 2012). To visualise fluorescent bands, the contrast on all images was increased by 0.3% to allow for direct comparison of band location and intensity to the area under the curve (AUC) plot generated. Ligation activity was taken as a ratio between the upper 40 nt product band and the lower 20 nt substrate band and expressed as a percentage. The final ligation activity data was analysed using a one-way analysis of variance (ANOVA) with multiple comparisons and graphed using GraphPad Prism 8.0 (GraphPad Software Inc.; La Jolla, CA, USA).

Chapter 3: Further Evolution of Ligase-p50

3.1 INTRODUCTION

As noted in Section 1.2, T4 DNA ligase is far less efficient at blunt-end ligation compared to the repair of single-stranded DNA nicks and cohesive-end ligation (Sugino et al. 1977). To further improve the engineered DNA ligases developed by Assoc. Prof. Patrick and his team (Wilson et al. 2013), former members of the Ackerley lab designed and optimised a competitive selection protocol that enriches for DNA ligase variants that have greater activity and are more successful at blunt-end ligation (Figure 1-6; Stevenson 2014; Dr Robins, unpublished results). The competitive ligase selection protocol enables researchers to screen larger libraries of DNA ligases in a high-throughput approach without requiring purification of individual ligase variants. This selection protocol was applied by Dr Robins to enrich for lead variants from two linker libraries, derived from ligase-cTF and ligase-p50, over five rounds. From the ligase-cTF library, a single variant denoted ligase-CA3 predominated, comprising 42% of the library by the end of the fifth round (Figure 1-6). In comparison, the ligase-p50 linker library did not produce a “clear winner” with five variants showing equivalent enrichment across two biological replicates.

This chapter aimed to apply one further round of selection to each of the two biological replicate plates from the ligase-p50 library, to identify whether any single ligase variant present from Round 5 showed further signs of enrichment.

3.2 RESULTS

The competitive ligase selection is a complex protocol that, if all steps work as intended, takes one week to complete from the growth of the initial culture plate through to the storage of successful ligase variants. The two biological replicate plates from the end of Round 5 denoted Replicate A and B (Figure 1-6) were individually completed from the overnight culture growth through to the selection of ligase variants following transformation. Originally it was hoped that two more rounds of selection were to be performed to gain better insight into the enrichment of the ligase-p50 linker library. Due to issues with the whole plasmid PCR, as well as DNA sequencing, only one round of selection was conducted.

3.2.1 DNA Substrates for Ligation Reaction

Throughout the development of the competitive ligase selection protocol, the least reproducible aspect of the protocol was the successful amplification of the entire 7 kb plasmid to form a phosphorylated blunt-end linear amplicon that could be used as the template for the ligation reaction. A key requirement in the development of the competitive ligase selection was that the same bacterial culture acted as a source of the DNA and the ligase enzyme that could be combined in an individual ligation reaction (Figure 1-3, steps 1-2). To ensure that there was sufficient representation of all clones from the previous selection round, the benchmark number of successful amplifications per 96 well plate was set to be greater than 90 linear plasmid amplicons.

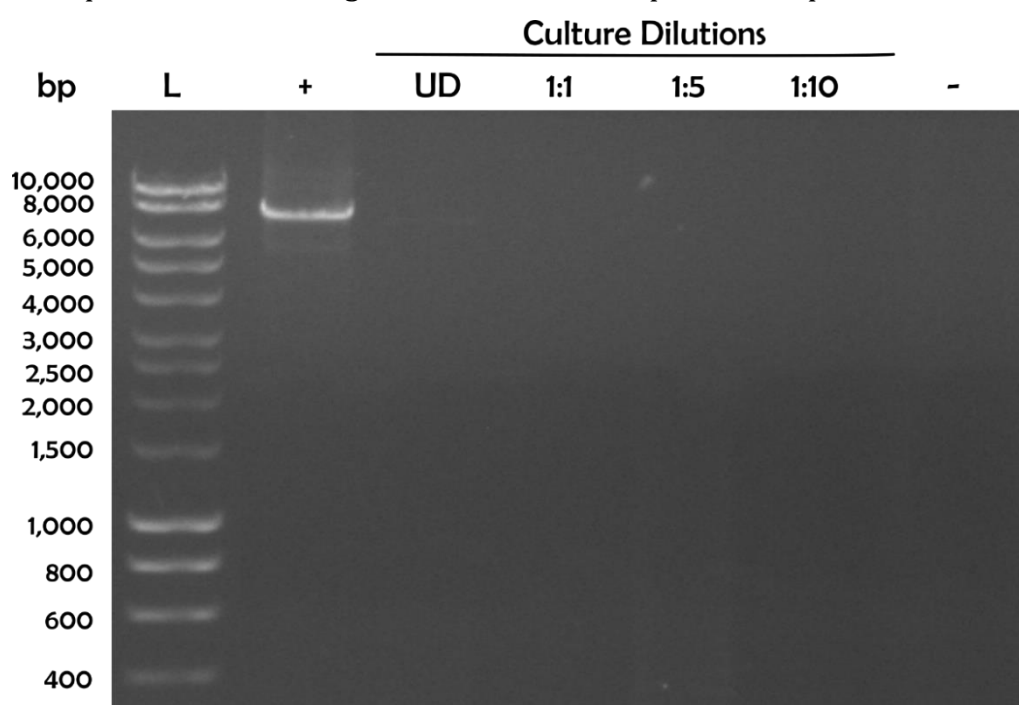


Figure 3-1 Initial efforts to achieve whole plasmid amplicons were only successful from purified DNA, not diluted culture templates.

PCR reactions were set up according to Section 2.9.1. The DNA template for each reaction contains 1 μ L of 10 $\text{ng}\cdot\mu\text{L}^{-1}$ of purified DNA (+), undiluted (UD) cell culture, diluted cell cultures with sterile molecular biology grade water to the indicated concentrations, or sterile molecular biology grade water (-). L denotes the Hyperladder 1 kb marker that was used as a DNA size standard. Purified DNA was mini-prepped according to Section 2.6.4

A major variable in the optimisation of the competitive ligase selection was to identify a dilution of cell culture that would permit whole plasmid DNA amplification. This is an aspect that the previous researcher Dr Robins had spent a substantial amount of time optimising, discovering that batch-to-batch variations in PCR master-mixes or even changing PCR machines could make the difference between successful and unsuccessful amplification (Dr Robins, unpublished results). Substantial variation was also

encountered in this project. To first determine the dilution factor of the starting DNA template for whole plasmid amplification, a serial dilution from the overnight culture was performed (Figure 3-1). Purified DNA for ligase-p50 was used as a positive control to confirm whole plasmid amplification (Figure 3-1, lane +). None of the PCRs from cell culture resulted in DNA amplification of the 7 kb plasmid, irrespective of the dilution factor. Since there was no amplification from any of the initial culture dilutions, the overnight cell cultures were left to grow to a higher cell density. Empirical testing identified that cell cultures grown for 18 or more hr and then diluted 1:5 with sterile molecular biology grade water were often able to successfully act as templates for PCR, i.e. resulted in a higher number of successful amplifications of the 7 kb linear amplicon, however, this was still unreproducible on a larger scale.

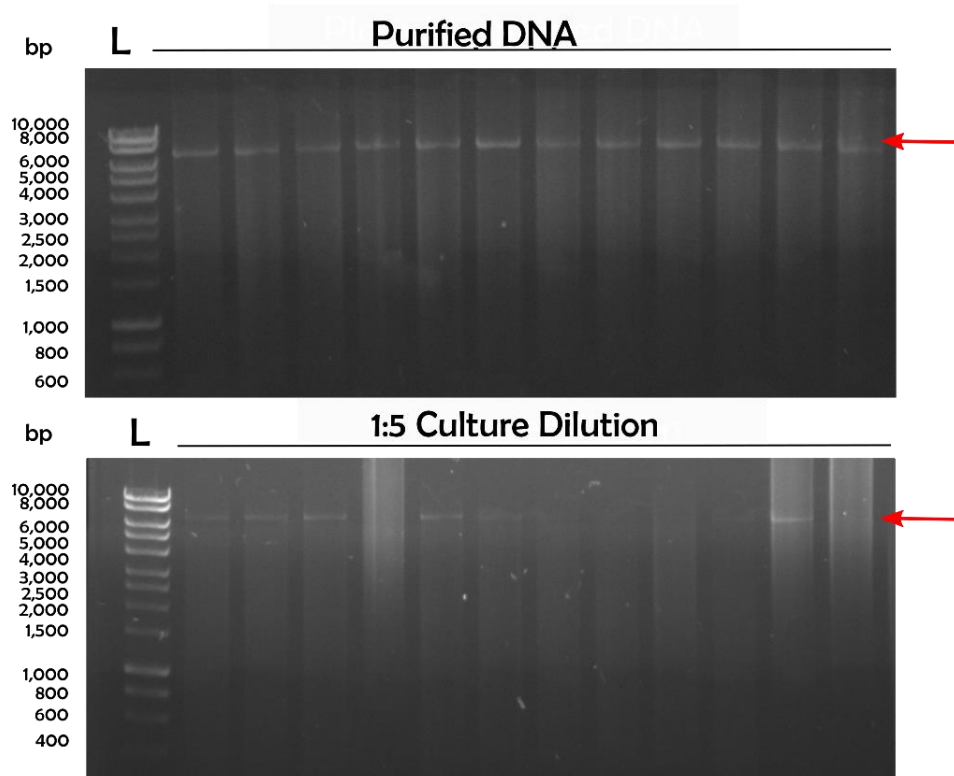


Figure 3-2 Small-scale amplification of whole plasmid amplicons.

1 μL of 2 $\text{ng}\cdot\mu\text{L}^{-1}$ of purified DNA or 1:5 dilution of 18 hr cell culture with sterile molecular biology grade water was set up and run according to Section Figure 3-22.9.1. L denotes the Hyperladder 1 kb that was used as a DNA size standard. Purified DNA was mini-prepped according to Section 2.6.4. The red arrow indicates the expected size of the linearized whole plasmid amplicon at ~ 7 kb.

To compare the reproducibility of whole plasmid amplification on a larger scale, 12 replicates of either 2 $\text{ng}\cdot\mu\text{L}^{-1}$ purified DNA or 1:5 dilution of 18 hr cell culture were set up. Both DNA templates successfully enabled the whole plasmid amplification however, the purified DNA template yielded a greater number of successful reactions, and more

prominent bands (Figure 3-2). As evident, all reactions from a purified DNA template were successful, whereas, only 50% from the culture dilution templates were successful. Due to time-constraints and a high number of reactions that continued to fail from diluted cell cultures, the two biological replicate plates from the end of Round 5 were purified (by mini-prep), diluted to $2 \text{ ng}\cdot\mu\text{L}^{-1}$ and used as the template in Round 6 for the competitive ligase selection. Following this, each biological replicate plate yielded the minimal requirements for the number of successful amplicons for the sixth round of competitive ligase selection to be conducted (Figure 3-3). The whole plasmid amplicons were successfully amplified and produced clean specific bands that were $\sim 7 \text{ kb}$ (Figure 3-3).

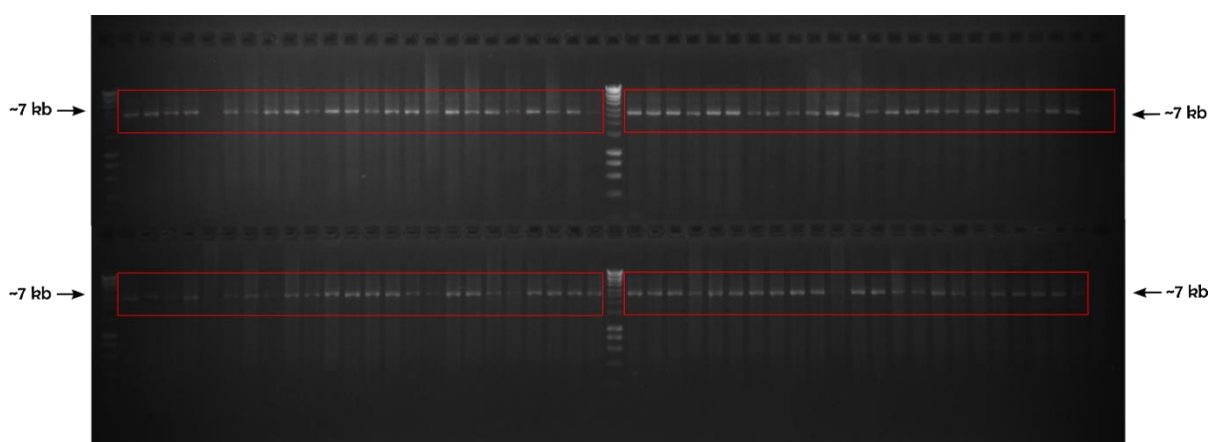


Figure 3-3 Example of the amplification of whole plasmid amplicons for one biological plate used as the DNA substrates for the ligation reaction in Round 6 of the competitive ligase selection.

The 96 well PCR from the first biological replicate plate was set up according to 2.9.1. 90 (or more) whole plasmid amplicons, excluding the negative control (last lane on the bottom gel), were successfully amplified for each replicate plate, as indicated within the red boxes.

3.2.2 Diagnostic Linker PCR Confirmation

Following ligation of the amplicons by the expressed ligase from each culture lysate (Figure 1-3, steps 2-3), the remainder of the competitive ligase selection protocol was completed. This included the removal of bound ligase by proteinase K treatment, the removal of the methylated DNA template by *DpnI*, and the dephosphorylation of unligated linear templates. Finally, the ligated, circularised DNA was used to transform *E. coli* DH5 α cells and plated on chloramphenicol supplemented agar plates (Figure 1-3, steps 4-7).

After a successful transformation, colonies were picked randomly from agar plates to confirm the presence of a linker variant using simple diagnostic PCR reaction by amplifying across the linker region (Figure 3-4, Panel A). Verified ligase-expressing cells were identified by the presence of a band at ~500 bp using primers that bound to the region at the end of T4 DNA ligase and the start of p50. Colony screening was used to identify a sufficient number of ligase linker variants to establish the start (or end) of a selection round. Each replicate plate required a total of 95 colonies or individual linker variants. In this study, approximately 150 colonies were screened from each transformation for each replicate plate, of which, approximately 1/3 were likely cause of failed amplifications (Appendix S4). Additional colonies were screened until a full plate of variants was confirmed. The colony screening is therefore an essential step in the competitive ligase selection to ensure that a full plate is established prior to the completion of additional selection rounds or DNA sequencing to reduce bias or influence enrichment results.

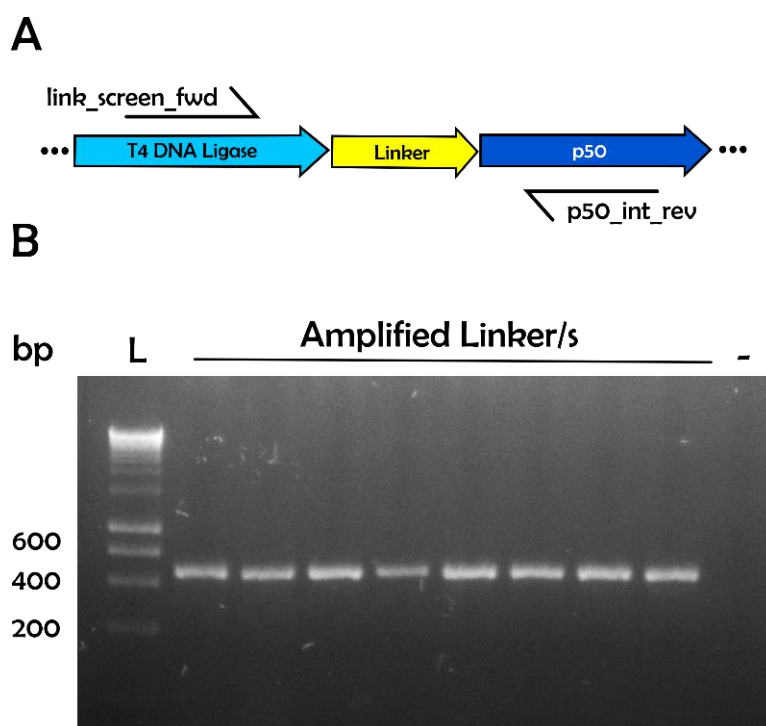


Figure 3-4 PCR amplification of ligase linker variants at the end of the competitive selection round.

A) Schematic representation of the PCR reaction performed using the primers `link_screen_fwd` and `p50_int_rev` to amplify across the linker region generating a generic PCR product to verify the presence of a ligase-p50 variant in transformed colonies.

B) An example of successfully amplified ~500 bp ligase linkers from a colony PCR after the resolution of bands on a 2% agarose gel. L denotes Hyperladder 1 kb marker that was used as a DNA size standard, and - denotes negative control with sterile MilliQ water used to create the PCR master-mix.

3.2.3 Determining the Linker Region Sequence

To determine the linker sequence in each selected ligase, variants were plasmid-prepped and were sent to Macrogen Inc. for sequencing using the primer that binds upstream of the linker region (link_screen_fwd). To view the enrichment of the linker variants after five rounds of competitive ligase selection, Dr Robins sent 48 variants from each biological replicate plate, thereby confirming 96 linker regions. To allow for equal comparison of enrichment after the sixth round of selection, the same number of cultures from each plate were plasmid-prepped and sent for sequencing. The linker regions from 48 ligase variants from the first biological replicate plate were successfully identified (explained in detail in Section 3.2.4). Unfortunately, the sequences of the linker regions from the second biological replicate plate could not be determined, despite the sequencing company Macrogen Inc. making multiple attempts to do so. The progression into COVID-19 category 4 lockdown in NZ precluded any repeat purification of the plasmid templates for additional repeats.

Immediately before the lockdown, a final attempt to determine the sequence of the linker region was made, independent of sequencing. For this, a two-step PCR method was devised, using the forward and reverse primers used to create each linker variant. The principle behind the reaction was that whole plasmid PCR primers used to create the designated linker could be combined with a diagnostic PCR primer and used as a presence/absence verification to confirm a specific linker (Figure 3-5). Firstly, the

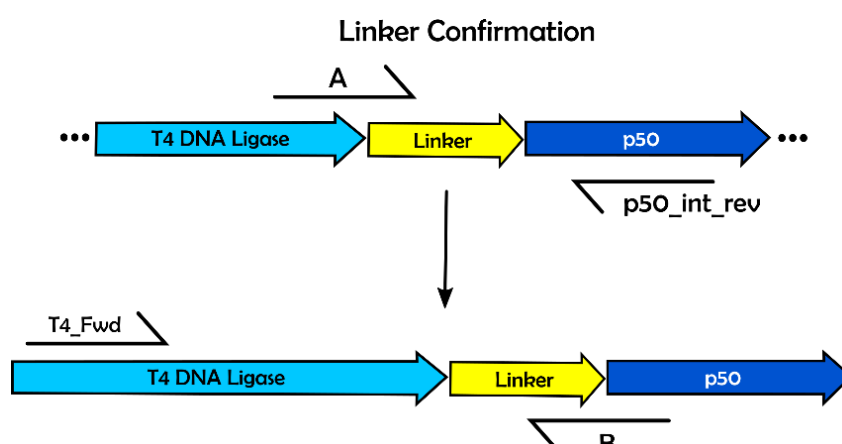


Figure 3-5 Schematic representation of the work-flow to confirm the sequence of ligase variants using PCR reactions.

The ligase linker variants were created using whole plasmid PCR by varying combinations of forward and reverse primers to create the linker of interest. Using the same primers, a two-step method was applied to confirm the specific linker variant. First, the reverse primer from generic linker screening, p50_int_rev, and forward primer (A) of the respective ligase variant/s was used to confirm half of the linker. Next, the second half of the linker was determined by using T4_Fwd and the reverse primer of the respective ligase variant. The information from the two reactions was combined to confirm the linker.

reverse primer from generic linker screening, p50_int_rev and the forward primer used to create the linker (i.e. Fwd_P50_PA, for ligase-PPA17; Figure 3-5, Primer A) were combined in a PCR reaction to confirm half of the linker. In a second PCR reaction the primers, T4_Fwd and reverse primer used to create the linker of the respective ligase variant (rev_PA12; Figure 3-5, Primer B) were combined in a reaction to confirm the second half of the linker.

The forward and reverse primers used to create each ligase linker variants had different lengths compared to the generic diagnostic primers used to confirm for colonies at the end of the competitive ligase selection, (up to 50 nucleotides for the primers used to create the linker variant versus 20 nucleotides for the diagnostic primers). To determine whether the PCR was possible using these primer sets, a gradient PCR was performed ranging from 52-70°C. The gradient PCR was performed using one of the enriched variants from the fifth round of selection, ligase-PPA17 with ligase-p50 as a reference for non-specific amplification to determine the sensitivity of the PCR.

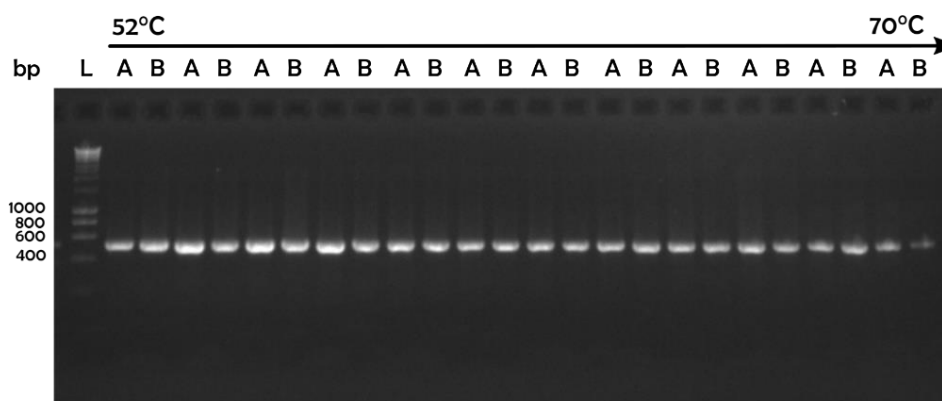


Figure 3-6 Validating the annealing temperature for the ligase-PPA17 linker with a gradient PCR amplification.

A 72 μ L PCR master-mix containing 1 \times Biomix™, 0.5 μ M Fwd_p50_PA, 0.5 μ M p50_int_rev, and 144 ng of DNA for either ligase-PPA17 or ligase-p50 was combined and aliquoted across 12 tubes. A gradient PCR was performed with annealing temperatures ranging from 52-70°C for 20 s for 30 cycles. The resulting amplicons for ligase-PPA17 (A) and ligase-p50 (B) were resolved on a 2% agarose gel.

The gradient PCR resulted in the successful amplification of every reaction, producing a band around the expected size of 500 bp (Figure 3-6). Unfortunately, ligase-p50, a non-specific target was amplified at all temperatures, including 70°C. The same experiment was repeated using T4_Fwd and rev_PA12 however, did not result in any DNA amplification (data not shown). Owing to the large difference in the primer sizes, as well as having a large common binding site (i.e. on T4 DNA ligase and p50), it is unlikely that the linker region could be determined without DNA sequencing. Thus, at the time of writing the second biological replicate of linker sequences remains undetermined.

3.2.4 Analysis of Variants Enriched in Round 6 of the Ligase-P50 Linker Selection

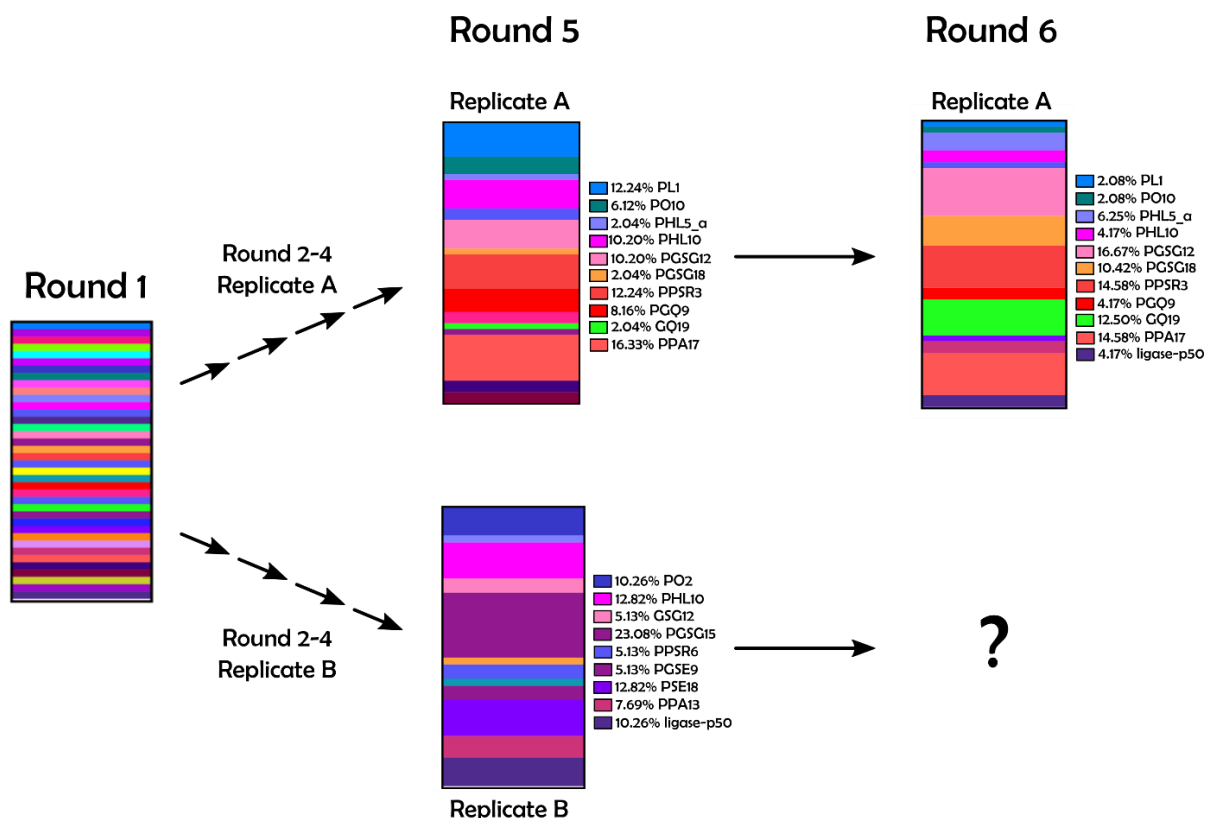


Figure 3-7 Enrichment of engineered ligase variants from the ligase-p50 linker library from Round 5 to Round 6.

One additional round of competitive ligase selection was applied to each of the two biological replicate plates from Round 5. The colonies after transformation of *E. coli* were sequenced, analysed, and tallied according to their encoded linker variant. The final data was inputted as a parts of a whole analysis on GraphPad Prism 8 to visually show the enrichment of ligase.

Following the completion of Round 6, there was no compelling evidence for the latent emergence of any individual ligase variant (Figure 3-7, distribution for Replicate A, Round 5 *versus* 6). Of the variants that were confirmed from the first biological replicate, there were minor changes to the enrichment of some of the variants. For example, variants ligase-PHL10 and ligase-PL1 showed a decrease from 10.2% and 12.2% in Round 5, to 4% and 2% in Round 6, respectively (Figure 3-7, Round 5 and Round 6). Conversely, ligase-PGSG18 and ligase-PGQ19 showed increases from 2% in Round 5 to 10.4% and 12.5% in Round 6, respectively. Linker variant PGSG12 showed the greatest improvement from 10.2% in Round 5 to 16.7% in Round 6. Finally, ligase-PPA17 and ligase-PPSR3 did not show a markable change. Ultimately, none of the ligase variants constituted >20% of the total variants. The results from the ligase-p50 linker library were in clear contrast to those for the ligase-cTF linker library which yielded a clear 'lead candidate', ligase-CA3.

3.3 CHAPTER 3 DISCUSSION

The goal of the research described in this chapter was to complete an additional round of competitive ligase selection on the ligase-p50 linker library to investigate the possibility of the emergence of a superior ligase-p50 variant. The competitive ligase selection is a high-throughput method to enrich for ligase variants that are capable of improved blunt-end activity. It was envisaged that the outcome from an additional round of selection may guide the choice of ligase variant/s to be critically evaluated using the fluorescence-based ligation assay in the subsequent chapter.

3.3.1 Whole Plasmid PCR Amplification

Throughout the development and optimisation of the competitive ligase selection, the whole plasmid PCR amplification was the most difficult aspect of the protocol, requiring substantial optimisation to successfully amplify the 7 kb plasmid from an overnight cell culture (Stevenson 2014; Dr Robins, unpublished results). An advantage of the competitive ligase selection is that it does not require the purification of each ligase variant, instead utilising the cell lysate as both a DNA template and expression culture. However, during the completion of the sixth round of selection, the cell culture lysate was unable to reproducibly produce enough successful amplicons to progress to the ligation reactions, therefore purified (mini-prepped) DNA from each well of the corresponding Round 5 plate/s was ultimately used as the template (Figure 3-1, Figure 3-2 & Figure 3-3). Ideally, the whole plasmid amplification would have performed identically to those performed by Dr Robins for Rounds 2-5, in terms of using diluted cell lysate as a template, however due to time-constraints, this was not achievable. Whilst using purified DNA for a single round of selection in this study was more reliable and time-efficient, over multiple replicate plates or subsequent rounds of selection, the additional cost of extracting DNA at the end of each round would significantly increase and may not be a feasible screening approach for larger libraries.

A future measure that could reduce the failures in the whole plasmid PCR amplification would be to change the vector that the ligase variants are cloned into. The current plasmid, pCA24N contains ~4,500 bp prior to the insertion of a ligase gene. Use of a smaller vector could reduce some of the issues. One such example would be to build from a pUC19 vector which has a high copy number however, is nearly half the size at ~2,700 bp. The initial development of the competitive ligase selection was performed in a pUC19 vector and aimed to use blue-white screening as a method to compare the relative

enrichment of two ligase variants (Stevenson 2014). This worked, showing that pUC19 was capable of whole plasmid amplification from within the ampicillin resistance gene and could be re-used as an alternative to reduce the total plasmid size.

In a similar vein, it should be noted that DNA ligases greatly range in size with the smallest identified being only 298 amino acids from PBCV-1 through to mammalian DNA ligases which can contain >900 amino acids (Ellenberger and Tomkinson 2008). The extensive size range for DNA ligases would become a difficult feature to address in the competitive ligase selection if this was applied to screening ligases identified within metagenomic libraries, or through genome mining. As these scenarios could lead to the preferential enrichment of smaller DNA ligases rather than the best-performing ones, it may be necessary to stratify candidates into different starting libraries based on their relative size. As a general rule, it is important to ensure the size of the whole plasmid remains relatively constant within a library.

3.3.2 Round 6 Ligase-p50 Linker Library Enrichment

Following the completion of the sixth round of selection on the ligase-p50 linker library, there was little evidence for the emergence of a superior ligase variant from the single biological replicate plate sequenced (Figure 3-7). After the sixth round of selection, the sequence verified ligase variants from Replicate A showed varying changes from Round 5 (Figure 3-7). Ligase-PPSR3 and ligase-PPA17, the two highest recovered variants from Round 5 did not show notable changes to enrichment and remained in the top three of the recovered variants in Round 6. The most dramatic changes from Round 5 to Round 6 were the decrease from 10.2% to 4.4% for ligase-PHL10 and the increase from 10.2% to 17.4% for ligase-PGSG12. After one additional round of selection, it was evident that there was no 'lead candidate' established, however this is not absolute without sequence confirmation for Replicate B which contained a very different selection of ligase variants at the end of Round 5. These results vastly contradict those produced in the ligase-cTF linker library which displayed a preference for the sole enrichment of ligase-CA3 after five rounds (Figure 1-6). Over multiple rounds of selection, ligase-CA3 would have been able to establish a greater number of colonies at the end of each round, therefore over multiple rounds the variant would be distinguished as more active than the rest of the library.

At the end of the fifth round, >50% of the initial linkers were still present within the two ligase-p50 replicate plates. A possible explanation for the trend displayed in the ligase-

p50 linker library is that several ligase linker variants were all capable of improved blunt-end ligation. The competitive ligase selection enriches for ligase variants that are more efficient at ligating a double-stranded break in the *cam^R* gene. The better a ligase variant is at circularising DNA, the greater the number of resistant colonies there are within the library following transformation into *E. coli*. If the ligase-p50 variants were equally as efficient as one another, then additional rounds would not improve the odds of it being selected from the final transformation. It must be considered that the outcome of the ligase-p50 evolution was a purely stochastic one driven by chance outcomes from small starting populations exhibiting comparable levels of activity.

An interesting contrast can be seen when comparing the linker regions enriched in the ligase-cTF and ligase-p50 libraries, which exhibited differences in the lengths and flexibilities of the linkers. These were the two criteria that Dr Robins varied upon designing and creating the libraries. The results from the ligase-cTF variants showed a preference for shorter linker regions that are less than 10 amino acids long and classified as flexible linkers (Figure 1-6; Appendix S1 & S2). The short linker, linker-NLS-cTF was also shown to produce the greatest amount of blunt-end ligation in the small-scale linker experiment (Wilson et al. 2013; Table 1-2). In comparison, the enriched ligase-p50 linker variants were predominantly longer linkers that were greater than 10 amino acids ranging to 19 amino acids and had no real preference for flexibility or rigidity (Figure 3-7). It may be that the ligase-cTF combinations function better with small flexible linkers between them, whereas the ligase-p50 prefers a longer linker. For example, it is possible that the DNA-binding domain for ligase-cTF functions most efficiently when proximal to T4 DNA ligase to increase the affinity and efficiency of the ligase; but that the ligase-p50 combination requires greater separation to prevent steric clashes or undesirable conformational changes during ligation. However, in the absence of any structural evidence to support this is purely speculative. We reasoned that if this was the case, none of the selected variants would exhibit substantial improvements relative to the starting ligase-p50 form. Whilst the second biological plate of linkers remains unconfirmed, the variant ligase-PPA17 was selected for protein purification and activity analysis in Chapter 4; because of its consistent recovery in both Rounds 5 and 6 (Figure 3-7).

Chapter 4: Quantification of Ligation

4.1 INTRODUCTION

In previous research led by Assoc. Prof. Patrick, the method for quantification of DNA ligation was performed using an agarose gel-based assay (Wilson et al. 2013). The agarose gel-based assay was a laborious and time-consuming process that produced inconsistent results (Assoc. Prof. Patrick, *personal communication*). Moreover, this assay required frequent preparation of DNA substrates, which could take up to a week to produce an adequate amount of DNA to perform a single assay (for this, the DNA substrates were grown from a culture, the DNA extracted, amplified, digested, and gel purified). As an alternative, our collaborator, Dr Adele Williamson of the University of Waikato, described a method to quantify DNA ligation that utilises a synthesised oligonucleotide with a 5' fluorescent dye tag. This enables visual comparison of the conversion of a 20 nt DNA substrate into a 40 nt product that is resolved by gel electrophoresis. The aims of the research described in this chapter were to establish and optimise the fluorescence-based ligation assay in our lab, so that the lead ligase variant from Round 5 (ligase-CA3) of the ligase-cTF library previously prepared by Dr Robins and the ligase-p50 library described in the previous chapter could be evaluated, including comparison with their predecessor enzymes. The engineered DNA ligases to be evaluated in this chapter were therefore T4 DNA ligase, ligase-cTF, ligase-p50, ligase-CA3, ligase-PPA17 as well as p50-ligase, the better p50 ligase variant from Wilson et al. 2013.

4.2 PURIFICATION AND CONFIRMATION OF LIGATION ACTIVITY FOR ENGINEERED LIGASES

4.2.1 Confirmation of Ligation Activity *in vivo*

Following the selection of the six variants that we wanted to evaluate ligation activity, it was important to ensure that they were still active *in vivo* and could retain their ability to seal single-stranded DNA nicks. As an initial test to confirm that the engineered ligases were functional *in vivo*, the genes encoding the variants to be assayed in this study were transformed into the temperature-sensitive ligase strain, *E. coli* GR501. At the permissive temperature, 28°C, growth of this strain can occur owing to the presence of an intact endogenous ligase, LigA (Dermody, Robinson, and Sternglanz 1979; Lavesa-Curto et al.

2004). However, the endogenous LigA present in GR501 contains an amino acid substitution that causes the protein to unfold upon raising the temperature to the non-permissive temperature, 42°C. For bacterial growth and survival, an alternative ligase must be present for DNA replication to occur. We reasoned that if the selected ligase variants were non-functional *in vivo*, this would result in no bacterial growth at the non-permissive temperature, 42°C, as was the case with expression of a control (i.e., non-ligase) protein, TrpCF (Figure 4-1; right). Conversely, all engineered ligase variants were able to support colony formation at 42°C, indicating that they could complement the non-functional endogenous LigA and rescue bacterial growth (Figure 4-1), verifying that the ligases were active and functional.

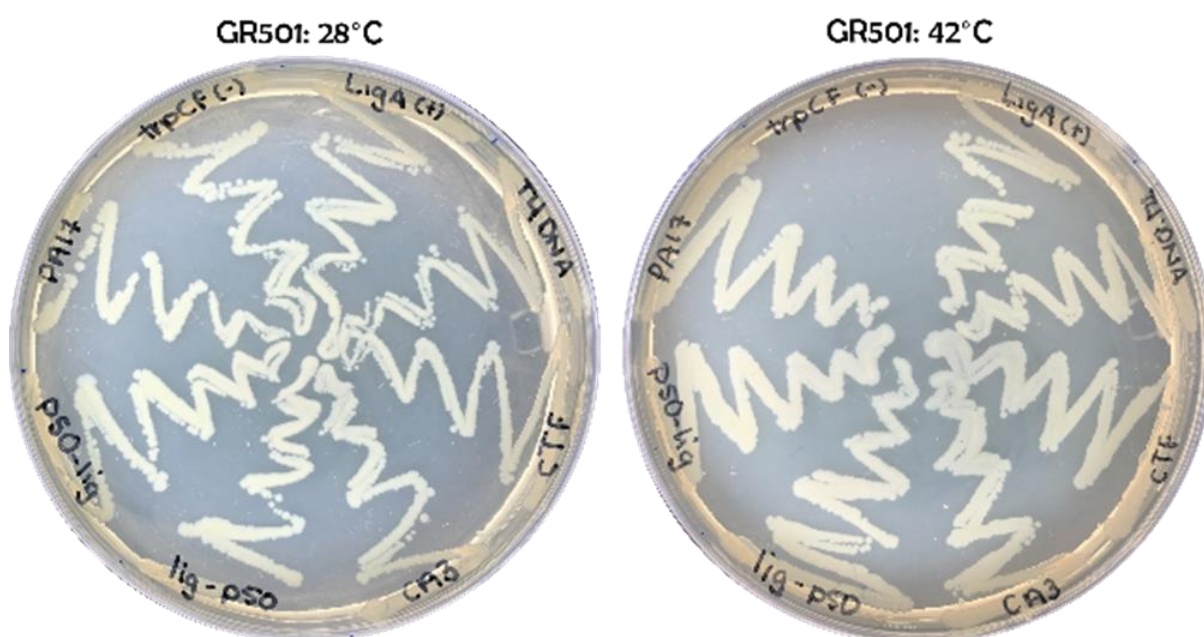


Figure 4-1 Testing whether engineered DNA ligases relevant to this study can complement the temperature-sensitive ligase LigA in *E. coli* GR501.

E. coli GR501 was transformed with vector, pCA24N, containing an unmutated *ligA* (+), *trpCF* (-), or the 6 engineered DNA ligase variants (T4 DNA ligase, ligase-cTF, ligase-CA3, ligase-p50, p50-ligase, and ligase-PPA17). The recovered cell cultures were incubated overnight on LB agar amended with chloramphenicol at 28°C. In each case, an individual colony was streaked from the 28°C overnight plate and regrown at 28°C (left) and 42°C (right).

4.2.2 Expression and Purification of Ligase Variants

Introducing changes to the linker region has the potential to alter the level of expression and purification of proteins (Chen, Zaro, and Shen 2013; George and Heringa 2003). To identify suitable isocratic elution concentrations for each ligase variant, a 0-500 mM imidazole gradient elution was performed using an Äkta purifier system. In this system, an excess imidazole concentration is passed through an IMAC column to displace and elute His-tag proteins (as per all ligases in this study). T4 DNA ligase, ligase-CA3, and ligase-p50 were evaluated in this manner to estimate the imidazole concentration

causing elution from the chromatogram gradient (Figure 4-2; pink line). T4 DNA ligase (Figure 4-2; B) and ligase-CA3 (Figure 4-2; E) both eluted at ~30% (150 mM) imidazole (as did the ligase-CA3 predecessor enzyme, ligase-cTF). 150 mM imidazole was therefore used as the imidazole concentration for all future isocratic protein purifications of T4 DNA ligase and the ligase-cTF variants.

Ligase-p50 was the sole enzyme tested to identify a preferred imidazole concentration for all ligase-p50 variants, because a 'top variant' had not yet been selected at this stage. Ligase-p50 (Figure 4-2; H) produced a more complex elution profile, exhibiting a shoulder peak off the end of non-specific protein elution. Ligase-p50 eluted at a 40% (200 mM) imidazole concentration, 10% less than the non-specific protein peak (Figure 4-2; G). To minimise non-specific protein contamination, a second wash step at 125 mM imidazole was introduced prior to elution of ligase-p50-variants at 200 mM imidazole. Finally, prior to storage at -80°C, all ligase variants were concentrated in a 30 kDa molecular cut-off column to remove the lower molecular weight proteins shown in the SDS-PAGE gels (Figure 4-2, E and H). The ligase variants were desalted into KPhos storage buffer and stored at -80°C. Unfortunately due to time constraints to complete the activity assays, ligase-PPA17 was not evaluated with the AKTA purifier system and eluted with the same imidazole concentration/s as ligase-p50.

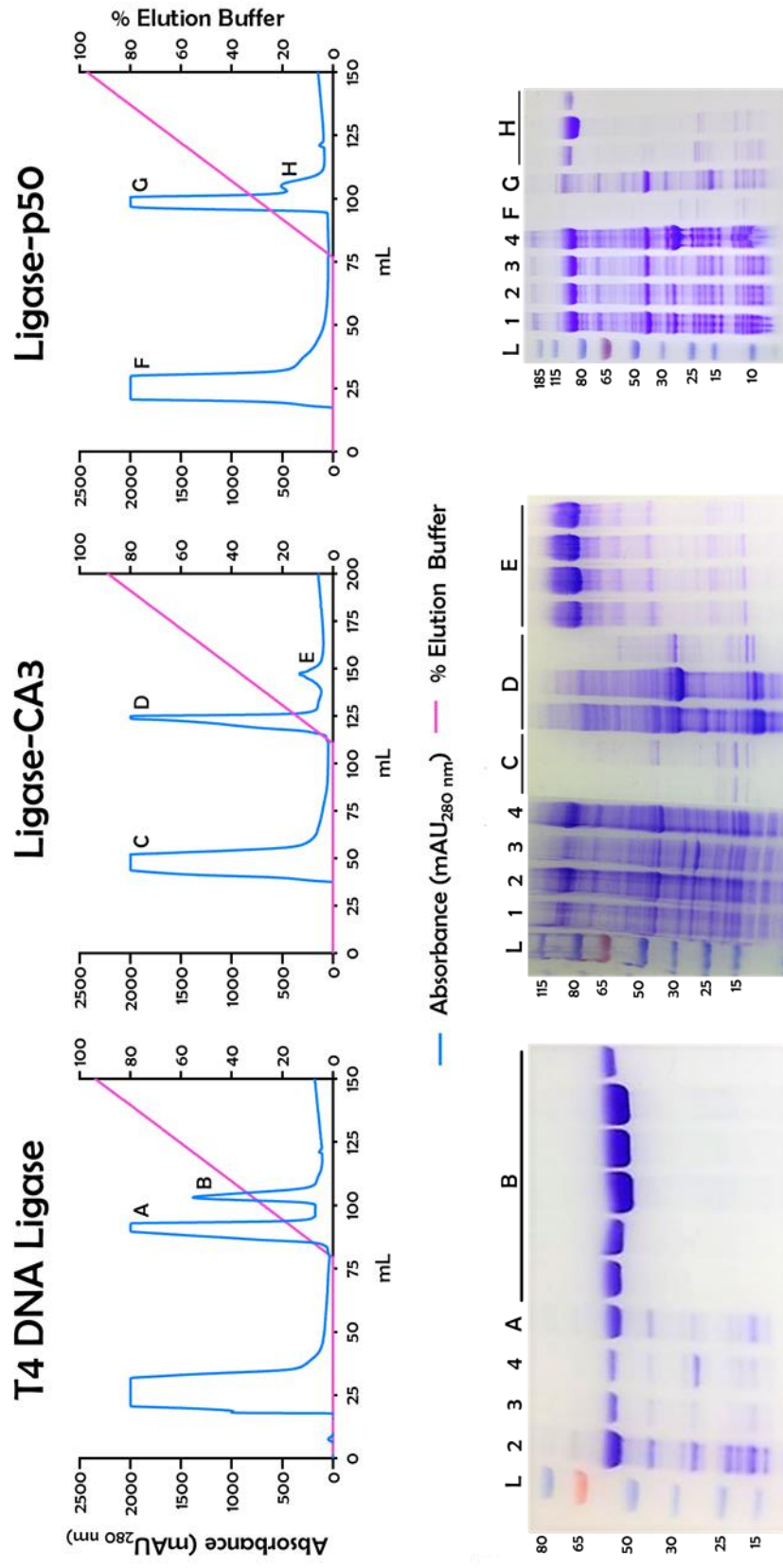


Figure 4-2 Chromatograms (top) and SDS-PAGE gels assessing protein purification of T4 DNA ligase, ligase-CA3 and ligase-p50.

A) The chromatograms depict the UV absorbance (blue line) and elution profile of the three ligase variants on a 0-100% gradient with a 500 mM imidazole elution buffer (pink line). Lanes in the SDS-PAGE gel denote L) PageRuler Protein Ladder 1) cell lysate pre-induction, 2) cell lysate post-IPTG induction, 3) sonicated cell lysate, 4) soluble cell lysate. Letters denote the corresponding fractions from the chromatogram. The expected size of each ligase variant was 57.1, 91.5, and 94.6 kDa, for T4 DNA ligase, ligase-CA3, and ligase-p50, respectively.

4.3 DESIGN AND OPTIMISATION OF FLUORESCENCE-BASED LIGATION ASSAYS

Our collaborator, Dr Adele Williamson, described a fluorescence-based ligation assay that visualises a set concentration (80 nM) of a fluorophore-modified oligonucleotide, denoted L1, in ligation assays (Williamson, Rothweiler, and Leiros 2014). The fluorescence-based ligation assay permits for the direct comparison of blunt-, cohesive-end, and single-stranded DNA nick-sealings by alternating the complementary oligonucleotides in each ligation reaction (Figure 4-3, blue shaded box). That is, in each case the assay measures the proportion of the 20 nt fluorescent L1 substrate ligated to a non-fluorescent L2 substrate, forming a 40 nt fluorescent product. The complementary oligonucleotides have been established for either blunt-end ligation (Figure 4-3, Panel B), cohesive-end ligation (Figure 4-3, Panel C), or nick-sealing ligation (Figure 4-3, Panel N)

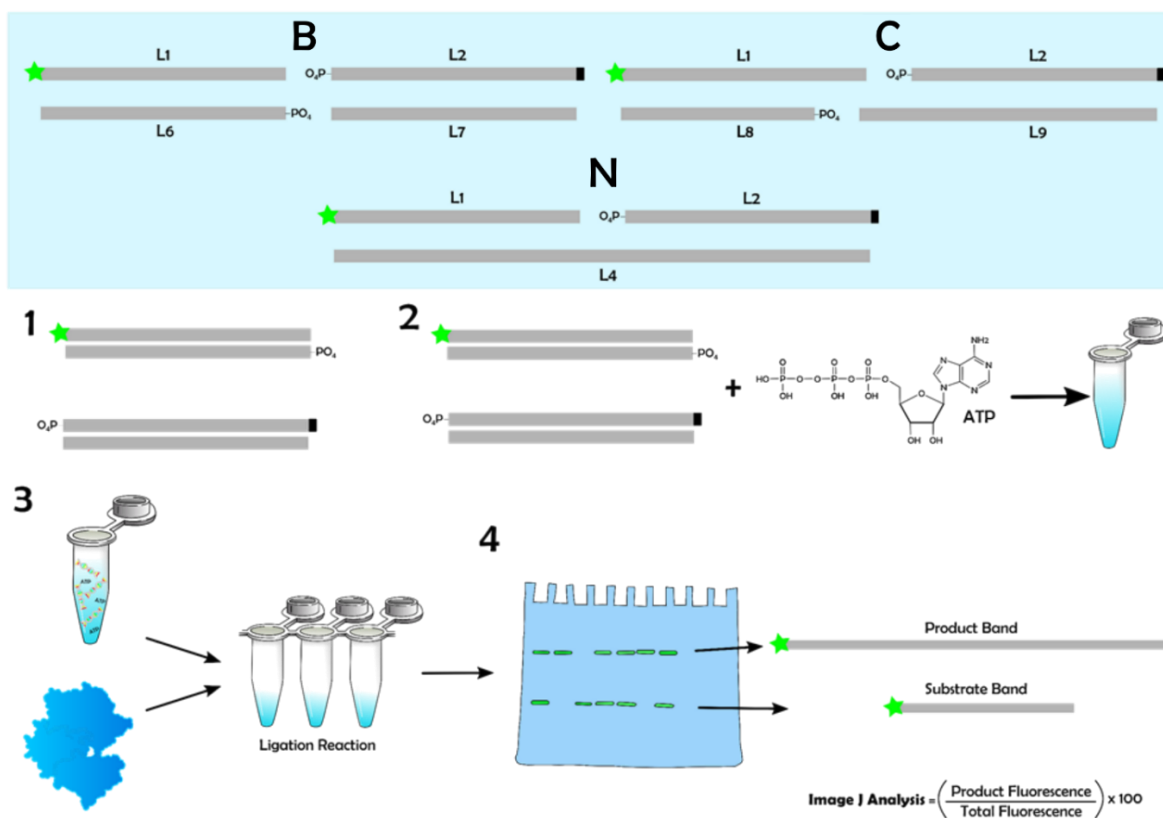


Figure 4-3 Schematic representation of the different fluorescence-based ligation assays.

The oligonucleotides in the blue shaded box are the different combinations used for a given ligation assay. Specified are the oligonucleotides combinations for blunt-end ligation (B), cohesive-end ligation (C), and single-stranded DNA nick ligation (N). Oligonucleotides contain either a 5' fluorescent dye label signified by the green star, a 5' phosphate as marked, or 3' amino blocker modified signified by the black box. Oligonucleotides are first annealed together to form double-stranded DNA substrates (1). These are then combined with ATP cofactor, to form the ligation substrate master-mix (2). The master-mix is divided into PCR tubes and the ligase is added to each tube to initiate the ligation reaction (3), after which the resultant reaction is resolved by denaturing electrophoresis on a TBE-PAGE gel (4). The outcome of the reaction is quantified using Image J by calculating the percentage of product band fluorescence intensity relative to the total reaction fluorescence.

N). This allows for many possible reactions to be evaluated including the relative activities of a single ligase variant for different ligation substrates, the relative activities of multiple ligase variants for the same ligation substrates, and/or the effects of varying different reaction parameters including time and temperature (Williamson, Rothweiler, and Leiros 2014). We made only one minor change to the base protocol described by Williamson, Rothweiler, and Leiros (2014): A 3' amino acid blocker (3' AmMO) modification was added to the end of the 5' phosphorylated oligonucleotide, L2, preventing the possibility of concatamerisation of dsDNA in the blunt-end ligation reaction. A schematic of the fluorescence-based ligation reaction is shown in Figure 4-3 (steps 1-4). (Not shown in this figure is the additional fluorescently-labelled control oligonucleotide control L3, the purpose of which is described in the following paragraph).

To predict the separation and locations of the 20 nt substrate and 40 nt product bands on a TBE-PAGE gel, various combinations of complementary oligonucleotides were annealed to the fluorophore-labelled oligonucleotides and resolved. This experiment was performed in the absence of a DNA ligase to serve as a template for the remainder of the gels in this study, which were assumed to contain varying intensities of substrate and product depending on the efficiency of each ligase variant. The TBE-PAGE gel was imaged and analysed using ImageJ to generate an AUC plot to display the corresponding substrate and product peaks that would be used to quantify ligation in later experiments (Figure 4-4, Panel C).

To visualise the location of all possible substrate bands post-electrophoresis, the 20 nt substrate, L1, was annealed to the complementary oligonucleotides for a blunt-end reaction (L6; Figure 4-4 Panel B, lane 2), cohesive-end reaction (L8, Figure 4-4 Panel B, lane 3), or nick-sealing assay (L4, Figure 4-4 Panel B, lane 4). Because this was a denaturing gel, all combinations yielded the same lower oligonucleotide band irrespective of the complementary oligonucleotide that L1 had been incubated with (Figure 4-4 Panel B). To predict the gel locations of all possible product bands, the oligonucleotide 'L3' was synthesized, which is the 40 nt oligonucleotide representing a single-stranded DNA product that would form from the ligation of L1+L2 (Figure 4-4 Panel B, lanes 5-8). L3 was annealed to the complementary oligonucleotides for a blunt-end reaction (L6; Figure 4-4 Panel B, lane 6), cohesive-end reaction (L8, Figure 4-4 Panel

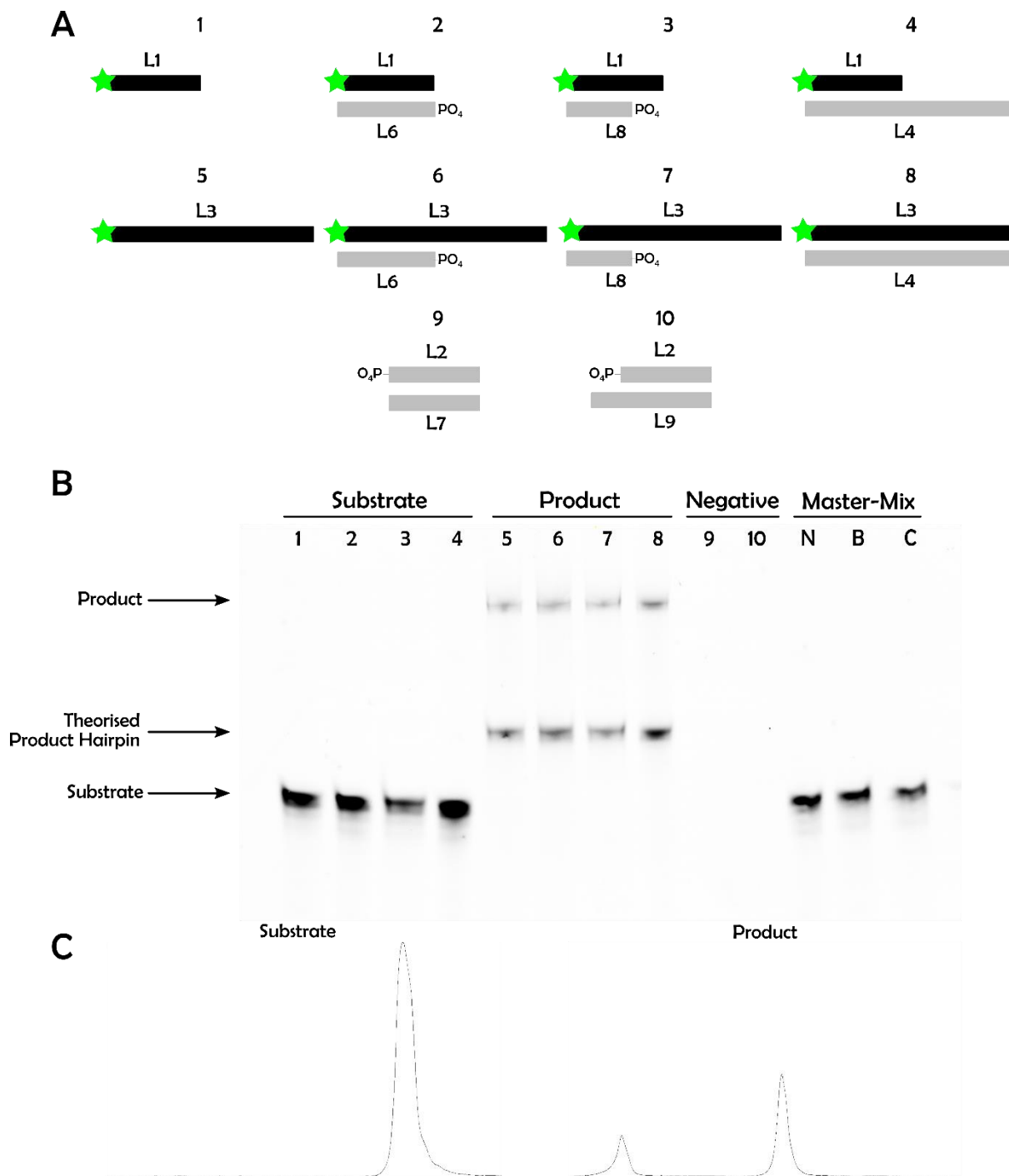


Figure 4-4 Oligonucleotide separation of substrate and product on a TBE-PAGE gel.

A) Schematic representation of the oligonucleotide combinations corresponding to the lanes 1 to 10 in Panel B. The black oligonucleotide is indicative of the oligonucleotide visualised on the TBE-PAGE gel. The schematic for the oligonucleotide combinations for the ligation substrate master-mixes can be found in the blue shaded box in Figure 4-3.

B) All reactions were established in a total volume of 22.5 μ L with their respective oligonucleotide combinations. Master-mixes were set up as specified in Section 2.11.3: N denotes nicked master-mix, B denotes blunt master-mix, C denotes cohesive master-mix.

C) The gel image from Panel B was analysed in ImageJ to generate an area under the curve (AUC) plot. The left-hand panel shows a single AUC profile for the substrate bands taken from lane 1, representing lanes 1-4; and the right-hand panel shows product bands taken from lane 5, representing in lanes 5-8.

A, lane 7), or nick-sealing assay (L4, Figure 4-4 Panel B, lane 8). The product fluorophore bands in lanes 5-8 were located higher on the gel than those observed for the substrates, with two bands present per lane (Figure 4-4 Panel B). The presence of a second product band was initially unexpected, however sequence analysis using the IDT oligo analyser tool indicated that the 40 nt oligonucleotide product can form self-hairpins between 40 and 60°C, which likely explains the additional band (Figure 4-4 Panel B, lanes 5-8; Appendix S5). To confirm that no fluorescence would be exhibited without the presence of the fluorophore, L2 was annealed to either L7 to form a blunt-ended substrate, or L9 to form a four base cohesive-ended substrate. As expected, no fluorescent signal was observed in either case (Figure 4-4 Panel B, lanes 9 and 10, respectively). Finally, to create the final ligation substrate master-mix: for blunt-end ligation L1+L6 were combined with L2+L7 (Figure 4-4 Panel B, lane B); for cohesive-end ligation L1+L8 were combined with L2+L9 (Figure 4-4 Panel B, lane C); and for nick-sealing L1 was combined with L2+L4 (Figure 4-4 Panel B, lane N). No product bands from the final ligation substrate master-mix were displayed in lanes N, B and C, supporting that all product formation was caused by the addition of ligase and not spontaneous ligation or association (Figure 4-4, Panel B).

Many early TBE-PAGE gels did not resolve the three bands as effectively as the gel displayed in Figure 4-4. Numerous efforts yielded curved band patterns caused by the outer wells running faster than the inner wells (Figure 4-5A), whereas others generated more than three bands (Figure 4-5B), meaning that the substrates and products were unable to be conclusively identified. Previous research indicated that curved gels result from uneven heat distribution across the gel during electrophoresis (Summer, Grämer, and Dröge 2009). One of the methods suggested to mitigate this was adhering a metal plate to the gel plate, so that this can more equally distribute the heat. Ultimately a simpler solution was found to be effective: TBE running buffer was pre-heated to 55°C to maintain a constant temperature prior to electrophoresis. In addition, to protect against possible fluorophore quenching or degradation, the gel tank was run in a polystyrene box (providing a dark environment, as well as insulating the heat). The resulting gels displayed bands running at the three predicted sizes, enabling for ligation quantification (Figure 4-5C). Increasing the temperature of the running buffer to 55°C was additionally unable to remove the theorised product hairpins displayed in Figure 4-4. The imaged TBE-PAGE gel was uploaded to ImageJ and an AUC analysis was conducted to determine the fluorescence intensity of bands present in each lane of the

gel (Figure 4-5D). DNA ligation was calculated as the fluorescence intensity of the sum of the two product bands (represented by the two left-most peaks in Figure 4-5D) divided by the total fluorescence intensity for all three peaks.

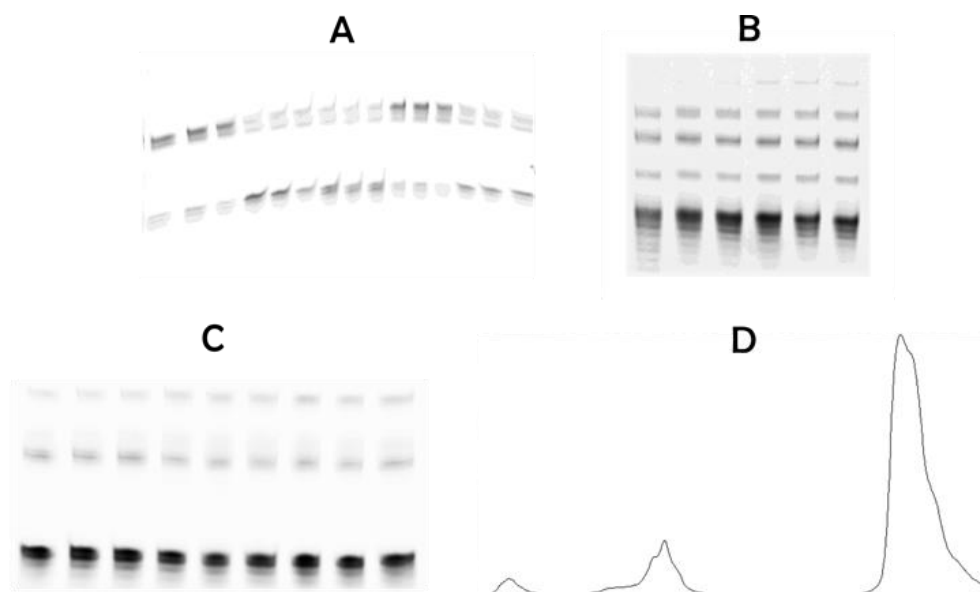


Figure 4-5 Example of pre-optimised and optimised TBE-PAGE gels.

Many of the pre-optimised gels either unevenly resolved causing a curvature of bands across the gel (A) or did not resolve into the 3 bands (B) seen in Figure 4-4. Optimisation of the gel running conditions introduced the addition of a pre-heating step of the 1 x TBE running buffer to 55°C as well as running the gel tanks ran in the dark. An example of an optimised TBE-PAGE gel is displayed in C. Quantification of DNA ligation was calculated from the AUC plot generated on Image J (D).

One additional test was performed to optimise the gel assay conditions. A critical step in the competitive ligase selection protocol was previously found to be disruption of the protein-DNA complex formed during the ligation reaction by the DNA-binding domain (Wilson et al. 2013). The inclusion of a DNA-binding domain or protein increases the affinity and interaction of engineered DNA ligases toward the DNA, which can result in the enzyme remaining bound to the DNA following ligation. To confirm that the ligation product band would run at the same location on a gel regardless of whether the ligase remained bound, a proteinase K dose-response was performed to remove any bound ligase following the ligation reaction. A 4 hr blunt-end ligation reaction was performed in duplicate with 5 μM ligase-CA3. All reactions were quenched with formamide loading dye and 0.5 μL of 0-20 $\text{mg}\cdot\text{mL}^{-1}$ working stocks of proteinase K was added, followed by a heat-inactivation incubation at 65°C for 20 min. Increasing the proteinase K concentration resulted in a decrease in total ligation product (Figure 4-6) but did not alter the location of the product or substrate bands on the gel. The reduction in overall ligation in samples that had been treated with proteinase K was unexpected, however for this reason, proteinase K was excluded from future ligation assays.

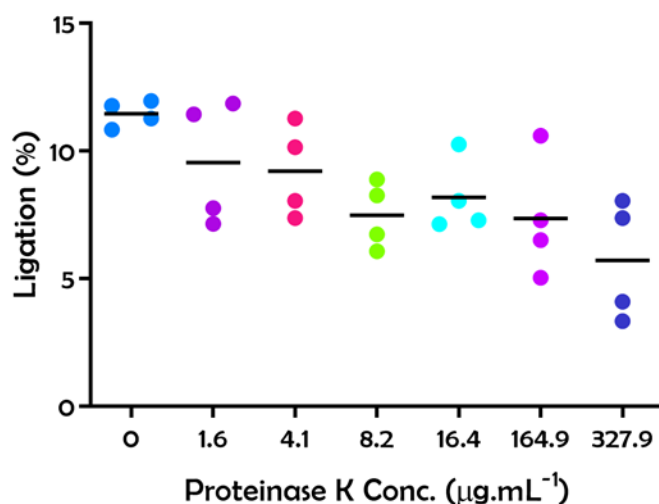


Figure 4-6 Evaluation of proteinase K concentration following fluorescence-based ligation reaction.

Reactions were set up as specified in Section 2.11.3 and incubated for 4 hr at 25°C. All reactions were heat-inactivated with the addition of 5 μL loading dye at 95°C and 0.5 μL of 0.1, 0.25, 0.5, 1, 10, or 20 mg.mL^{-1} proteinase K was added to each tube to give the final dilution series as per the X-axis. Reactions were incubated at 65°C for 20 min, and heat-inactivated for 10 min at 95°C. Black lines represent the mean ligation of substrate to product, expressed as a percentage. Reactions were performed in technical duplicates with two biological replicates.

4.4 CONFIRMING LIGATION ACTIVITY *IN VITRO*

DNA ligases are responsible for catalysing the repair of single-stranded nicks to restore the DNA backbone, in particular during DNA replication and recombination (Shuman 2009). In addition to confirming their abilities to promote survival *in vivo* in *E. coli* GR501, a single-strand DNA nick-sealing assay was performed to validate that the engineered ligase variants would retain activity *in vitro* using the fluorescence-based ligation assay. The ligase variants were purified by IMAC with a peristaltic pump and eluted based on the preferred imidazole concentrations identified above.

The optimal ligation temperature for T4 DNA ligase under physiological conditions is 37°C (Ferretti and Sgaramella 1981). The ligation reaction was initially performed at 37°C for 30 s, however, all ligation reactions went to 95% completion therefore any differences that may have been present could not be distinguished (data not shown). The temperature was therefore lowered to 25°C and the ligation repeated for 30 s. In this case, regardless of the ligase variants, all ligations yielded a similar extent of DNA ligation with ~25% of nicked DNA substrate being converted to product (Figure 4-3). This experiment showed that endogenous DNA ligation activity of single-stranded nick-sealing was retained *in vitro* following purification.

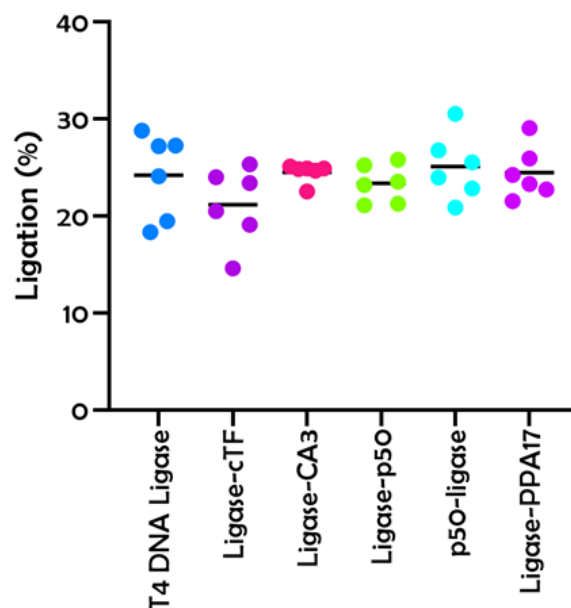


Figure 4-7 Engineered ligase variants maintain DNA nick-sealing ligation activity in the fluorescence-based ligation assay.

A single-strand nicked DNA ligation reaction was set up as per Section 2.11.3. Assays were incubated with 5 μ M of each candidate ligase variant for 30 s at 25°C. Black lines represent mean ligation of substrate to product, expressed as a percentage. Ligations were performed in technical duplicates with three biological replicates.

4.5 PRELIMINARY TIME-COURSE LIGATION ASSAYS REVEALED EXTENSIVE BATCH-TO-BATCH VARIATION

Following the visualisation of ligation bands and optimisation of the gel electrophoresis conditions, a ligation assay was performed at various time points to evaluate blunt- and cohesive-end ligation. Three variants including the lead variant ligase-CA3 from the ligase-cTF library, and its predecessor enzymes T4 DNA ligase and ligase-cTF were evaluated. This assay gave extremely variable results, indicating that further optimisation of reaction conditions would be required to enable reliable comparisons of variant activities. All proteins were purified by IMAC with a peristaltic pump, concentration quantified by nanodrop spectrophotometry and visually compared by SDS-PAGE to view purity and estimate concentration. Ligation reactions had been performed within 2-3 days of purification at 16°C and 25°C. The 16°C temperature was selected because it is the standard cohesive-end ligation temperature to maintain ligation activity but avoid melting the base overhangs (Lund, Duch, and Pedersen 1996), whereas, 25°C was chosen because this was the selection temperature for the ligation reaction during the competitive ligase selection protocol. When the activities of T4 DNA ligase, ligase-cTF, and ligase-CA3 were compared, it was evident that there was vast batch-to-batch variation between replicates which became more prominent

cumulatively from 20 min to 4 hr (Figure 4-8, Panel A). Cohesive-end ligation did not produce any consistent pattern across the three time-points for all three variants with ligation activity ranging from 0% up to 90% for some variants. In comparison, blunt-end ligation increased with duration (Figure 4-8, Panel A), however, produced clusters that were clearly distinguishable by their replicate batch, as illustrated for ligase-CA3 (Figure 4-8, Panel B).

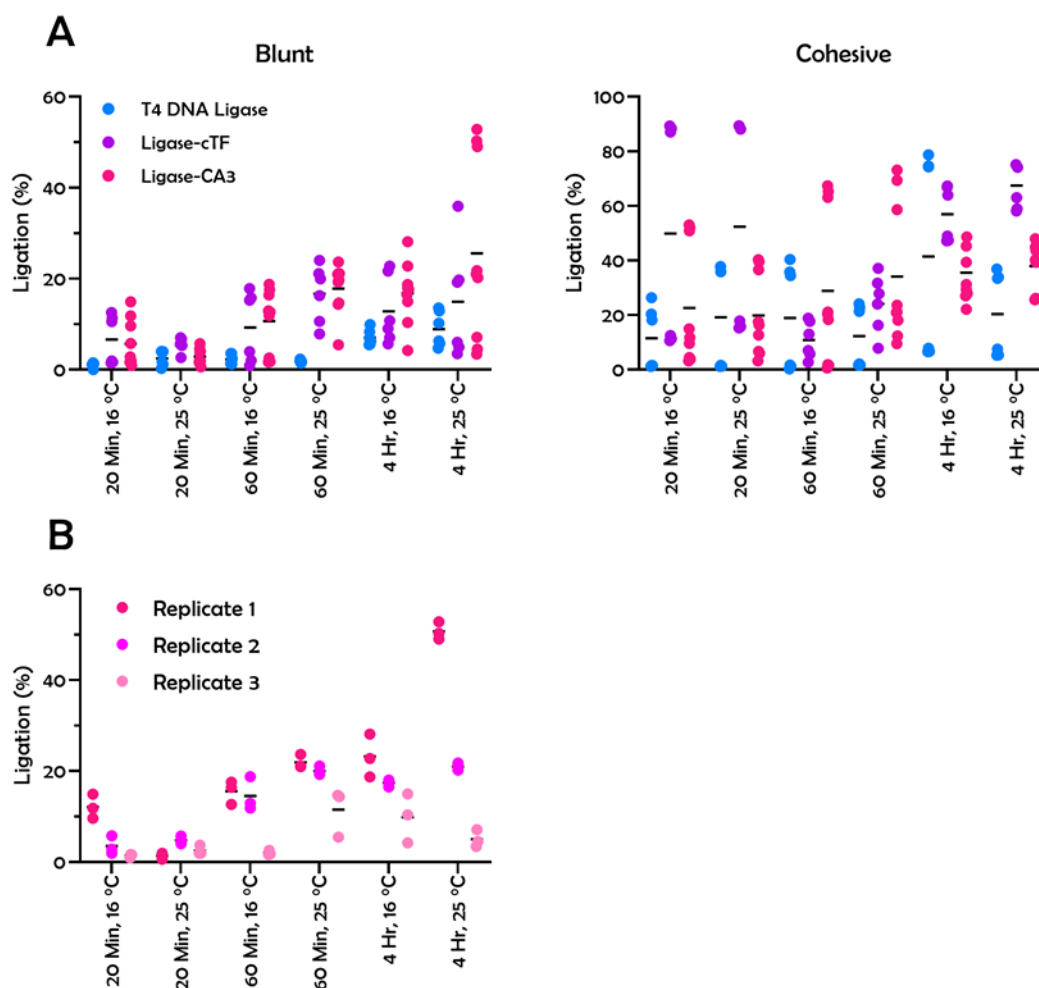


Figure 4-8 Initial efforts to plot a time-course of ligation activity for T4 DNA ligase, ligase-cTF, and ligase CA3.

A) Reactions were set up according to Section 2.11.3 and incubated for indicated times, and temperatures. Reactions were performed with 5 μ M enzyme in technical triplicates with two biological replicates for T4 DNA ligase and ligase-cTF, and three biological replicates for ligase-CA3. Black lines represent mean ligation of substrate to product, expressed as a percentage.

B) Data from the blunt-end ligation results presented in Panel A for ligase-CA3 was separated and colour-coded to distinguish each of the three biological replicates, revealing substantial batch-to-batch variation.

All ligation reactions were performed from the same ligation substrate master-mix stock, with each replicate batch coming from the same buffer solutions. This suggested that the

issue was more likely to be a consequence of purified ligase buffer or storage conditions. Owing to time-constraints, a detailed analysis of varying incubation duration and temperature parameters was no longer feasible. The remainder of this study therefore used incubation temperature of 25°C and durations of 4 hr for blunt-end ligation and 60 min for cohesive-end ligation, as these conditions were empirically determined to provide an acceptable period to observe differences to ligation activity between the assorted ligase variants.

4.6 PROTEIN BUFFER AND STABILITY OPTIMISATION

It is known that many proteins are not optimally suited for industrial conditions which often leads to decreased stability and activity (Eijsink et al. 2005). Protein instability is typically caused by isolation of the protein from its native environment which is often affected by diverse factors including temperature, metal ions or cofactors, and pH. Since there are large number of conditions that could be assessed to improve ligation, this section of work investigated a limited range of storage conditions and buffer compatibilities for the engineered DNA ligases, guided by recent literature.

4.6.1 Protein Storage

An essential requirement following protein purification is to identify storage conditions that maintain activity (Simpson 2010). Typically, it is recommended that purified proteins are stored in the short-term at 4°C, or -20°C for labile proteins. Furthermore, to prevent protein denaturing or degradation, long term storage of protein might require temperatures of -20°C or -80°C. Typically, commercial enzymes are stored at -20°C with 50% glycerol to prevent from freezing as this reduces the number of freeze-thaw cycles, decreasing opportunities for activity loss.

An experiment was set up to evaluate the effects of short-term storage temperature of ligase-CA3 at 4°C, -20°C, or -80°C (Figure 4-9). On the day of purification, blunt-end ligation was consistent for each biological replicate with ~10% conversion of substrate to product (Figure 4-9A). The two biological replicates for cohesive-end ligation separated into two distinct clusters. One replicate exhibited activity on day 0 and after storage at -20°C and -80°C on day 1, whilst the other did not produce any ligation on subsequent days. After one week storage at any temperature, DNA ligation was reduced in all cases with both blunt- and cohesive-end ligation exhibiting negligible activity, less

than 5% ligation. It was therefore considered that the current KPhos storage buffer might be affecting protein stability, causing the decreased ligation activity.

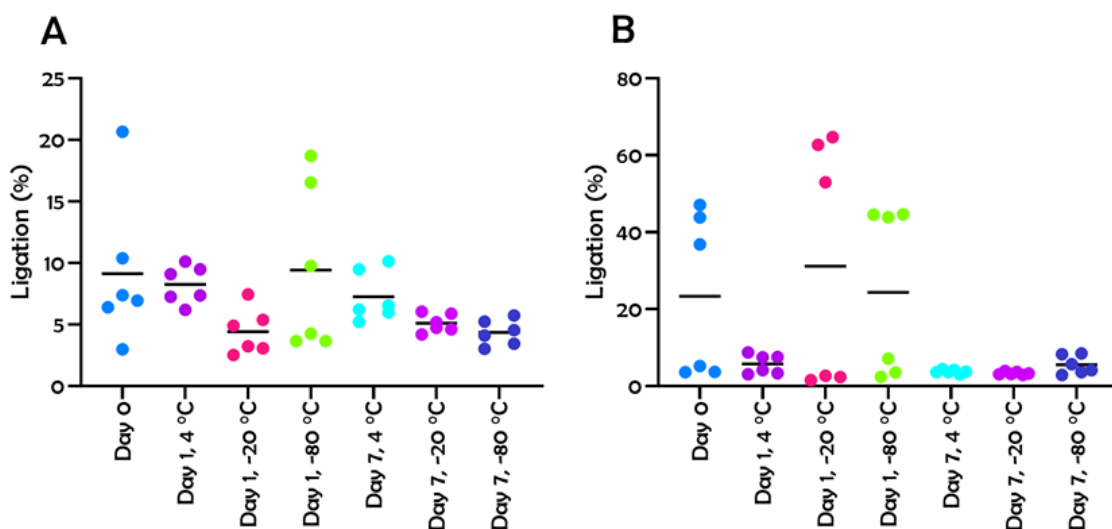


Figure 4-9 Impact of storage on the activity of ligase-CA3 as measured by the fluorescence-based ligation assay.

Ligation reactions were performed as specified in Section 2.11.3. Blunt-end ligation (Panel A) was incubated for 4 hr and cohesive-end ligation (Panel B) for 60 min. Assays were performed in technical triplicates with two biological replicates. Black lines represent mean ligation of substrate to product, expressed as a percentage.

Since T4 DNA ligase is already a commercially available enzyme, a comprehensive search of the current storage conditions for this enzyme was conducted, with the results summarised in Table 4-1. All companies use a Tris-HCl buffer at pH 7.4-7.5 with 50-60 mM KCl in 50% glycerol. DTT and EDTA were typically present at 1 mM and 0.1 mM, respectively. Three out of four of the commercial storage buffers contained entirely the same reagents, with NEB and Promega comprising the same reagents and concentrations. The previous storage buffer, KPhos (containing 50 mM potassium phosphate, pH 7.0, 200 mM NaCl, and 10% glycerol), was therefore compared to NEB storage buffer (containing 10 mM Tris-HCl, pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, and 50% glycerol).

Table 4-1 Commercial T4 DNA ligase storage buffers

Reagent	Jena Bioscience	NEB or Promega	Roche	Thermo-Fisher Scientific
Tris-HCl	10 mM, pH 7.4	10 mM, pH 7.4	20 mM, pH 7.5	20 mM, pH 7.5
KCl	50 mM	50 mM	60 mM	50 mM
DTT	1 mM	1 mM	5 mM	1 mM
EDTA	0.1 mM	0.1 mM		0.1 mM
BSA	200 $\mu\text{g}\cdot\text{mL}^{-1}$			
Glycerol	50%	50%	50%	50%

A large expression culture of ligase-CA3 was grown to increase the protein yield over previous efforts. This provided enough enzyme to split into two equal volumes, each having sufficient protein concentration to directly compare the ligation activities following storage in two different buffers (KPhos *versus* NEB buffer).

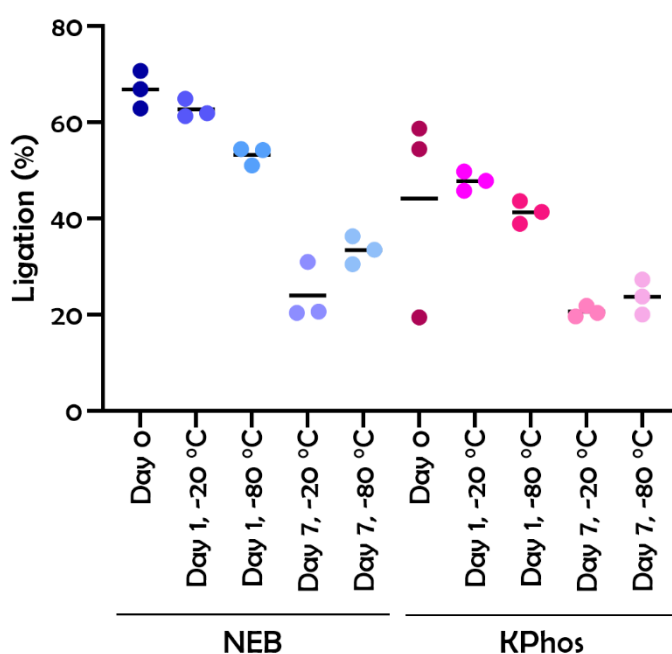


Figure 4-10 Comparison of ligase-CA3 blunt-end ligation activity in two storage buffers.

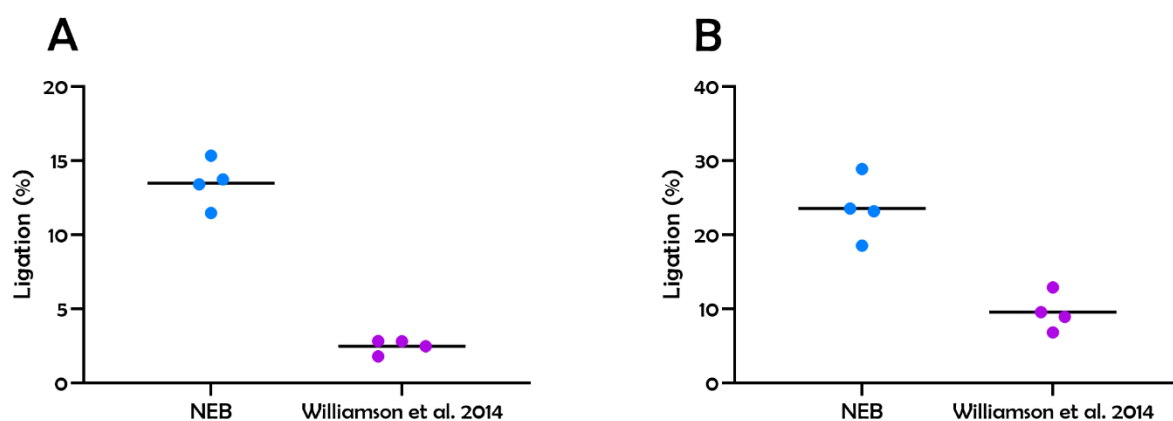
A large expression culture of ligase-CA3 was grown and separated into two equal soluble fractions. The protein was purified, and 1.5 mL of eluted protein was loaded on to a pre-equilibrated desalt columns containing KPhos storage buffer or 2 \times NEB storage buffer. Ligase-CA3 in NEB storage buffer was concentrated to 10 μM and diluted 1:1 in 100% glycerol. Blunt-end ligation reactions were set up according to Section 382.11.3 and incubated for 4 hr at 25°C. Black bars represent mean ligation, expressed as a percentage. The assay was performed as a single triplicate experiment.

On all occasions, ligase-CA3 resulted in greater ligation activity when the enzyme was stored in the NEB storage buffer (Figure 4-10). Over the duration of a week, the enzyme activity declined as a similar rate from both storage buffers at either -20°C or -80°C. The enzyme exhibited more ligation post-storage at -20°C after one day than at -80°C, but more at -80°C after one week. From this experiment we can conclude that the NEB

storage buffer conditions are more suited to maximise ligation efficiency compared to KPhos storage buffer. All future purifications were stored in NEB storage buffer and ligation reactions were performed the next day to minimise loss of activity.

4.6.2 Ligase Buffer Comparison

Following amendment to storage limitations for ligase-CA3 to the NEB T4 DNA ligase storage buffer, a final experiment sought to optimise the pH and sodium chloride concentration within the ligase reaction buffer. The initial ligase reaction buffer used in this work came from the protocol for fluorescence-based ligation assay described by Williamson et al. (2014). When this was compared to the NEB ligase reaction buffer it was observed that the reaction buffers were very similar with the key difference being that the Williamson et al. (2014) buffer contained 50 mM Tris-base, pH 8.0 with 50 mM NaCl, *versus* 50 mM Tris-HCl, pH 7.5, with no NaCl for the NEB buffer. Both reaction buffers additionally contained 50 mM MgCl₂, 10 mM DTT, and 1 mM ATP which are



essential cofactors and stabilising reagents (Section 1.2.1).

Figure 4-11 Comparison of ligase-CA3 ligation activity in either the NEB or Williamson et al. 2014 reaction buffers.

Ligation reactions were set up according to Section 2.11.3 and incubated at 25°C. Blunt-end ligation (Panel A) was incubated for 4 hr, whereas, cohesive-end ligation (Panel B) was incubated for 60 min. Black bars represent mean ligation, expressed as a percentage. The assay was performed in as two biological replicates with technical duplicates.

In the presence of NaCl, ligation was reduced by approximately 3-fold for both blunt- and cohesive-ends (Figure 4-11). Previous research has shown that salt concentrations greater than 200 mM (in particular NaCl or KCl) can inhibit DNA ligation (Hayashi et al. 1985; Raae, Kleppe, and Kleppe 1975). Whilst the final salt concentration is relatively low, if ligation master-mix is insufficiently mixed, this may increase the presence of salt in some samples which could possibility have contributed to the decreases in ligation

observed in Figure 4-8 and Figure 4-9. This is unlikely because the ligation substrate master-mix was thoroughly mixed (by pipetting or vortexing) prior to distribution, however, cannot be ruled out.

4.7 PROTEIN QUANTIFICATION

Section 4.6 provided evidence that the current commercial T4 DNA ligase reaction and storage buffers are better suited for DNA ligation than the original Wilson et al. (2013) and Williamson et al. (2014) buffers. However, the extent of ligation within these experiments was inconsistent between assays, particularly for blunt-end ligation. For example, the data from blunt-end ligation from Figure 4-8 to Figure 4-10, performed for 4 hr at 25°C with 5 µM ligase-CA3 was collated and graphed to show the distribution of the percentage ligation of substrates to product (Figure 4-12). Since some experiments were repeated over several days, only the day 0 results were included. Although there is variability in the buffers used which may have led to variability in the total ligation, there are clearly two distinct clusters that are not buffer specific.

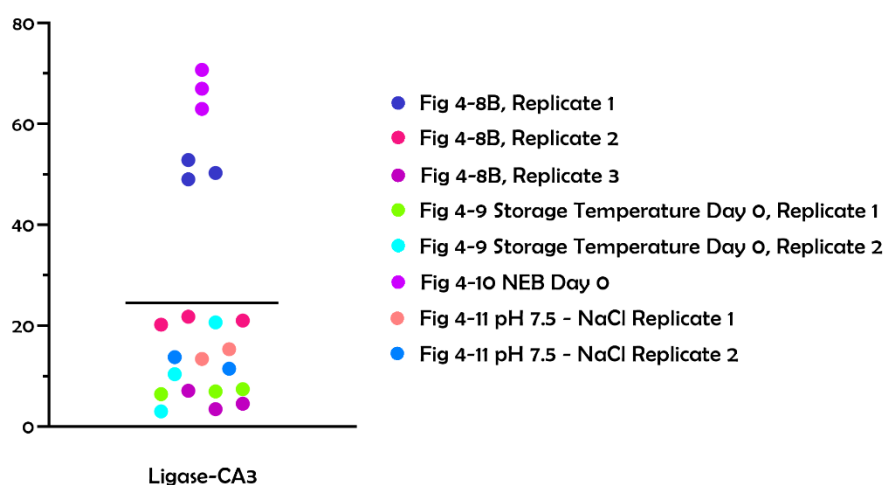


Figure 4-12 Collation of blunt-end ligation data from Figures 4-8 to 4-11 in this study performed at 4 hr, 25°C.

For experiments where multiple days were evaluated, the day 0 ligations were selected for incorporation. Each colour denotes a different replicate performed in triplicate.

Previously, all ligases were purified by IMAC using a peristaltic pump and fractions at relevant steps were collected and run on an SDS-PAGE. Protein concentrations were calculated using a nanodrop spectrophotometer and visualised for purity off the SDS-PAGE. The amount of protein required to generate a 5 µM enzyme stock solution was calculated based off the protein concentration ascertained by spectrophotometry and diluted accordingly in storage buffer. The final proteins concentrations were re-recorded

by the nanodrop spectrophotometer and then assayed or surplus protein stored at -80°C . The protein concentration equivalent to $5\ \mu\text{M}$ is $0.29\ \text{mg}\cdot\text{mL}^{-1}$ for T4 DNA ligase, and $0.46\ \text{mg}\cdot\text{mL}^{-1}$ for either ligase-cTF or ligase-CA3.

To assess the possibility that protein concentrations had been inaccurately measured, stored ' $5\ \mu\text{M}$ ' stock solutions from these previous experiments were assessed on an SDS-PAGE gel alongside a $1\ \text{mg}\cdot\text{mL}^{-1}$ BSA protein standard. This SDS-PAGE analysis indicated unequal protein concentrations for each sample, and that many samples were not at the predicted protein concentrations (Figure 4-13), as specified in the previous paragraph. For example, the samples from Replicate 1 of T4 DNA ligase and ligase-CA3 from Figure 4-8, contained more protein than equivalent for a $5\ \mu\text{M}$ enzyme stock. T4 DNA ligase contained $\sim 1\ \text{mg}\cdot\text{mL}^{-1}$ protein, $3\times$ that of the protein concentration required for $5\ \mu\text{M}$, whereas ligase-CA3 contained greater than $1\ \text{mg}\cdot\text{mL}^{-1}$ in reference to the size and intensity of $1\ \text{mg}\cdot\text{mL}^{-1}$ BSA. For Replicate 2, T4 DNA ligase contained almost $10\times$ the required protein concentration, whereas, ligase-cTF and ligase-CA3 contain approximately the correct amount of protein for a $5\ \mu\text{M}$ enzyme stock. The third ligase-CA3 replicate contains even less $0.5\ \text{mg}\cdot\text{mL}^{-1}$ protein concentration with the band being almost imperceptible, therefore was not an adequate protein concentration for a $5\ \mu\text{M}$ enzyme stock. The relative protein concentrations inferred qualitatively from Figure 4-13 were consistent with the variations in data evident in Figure 4-8, where the first replicate produced the greatest conversion of substrate to product, whilst the third

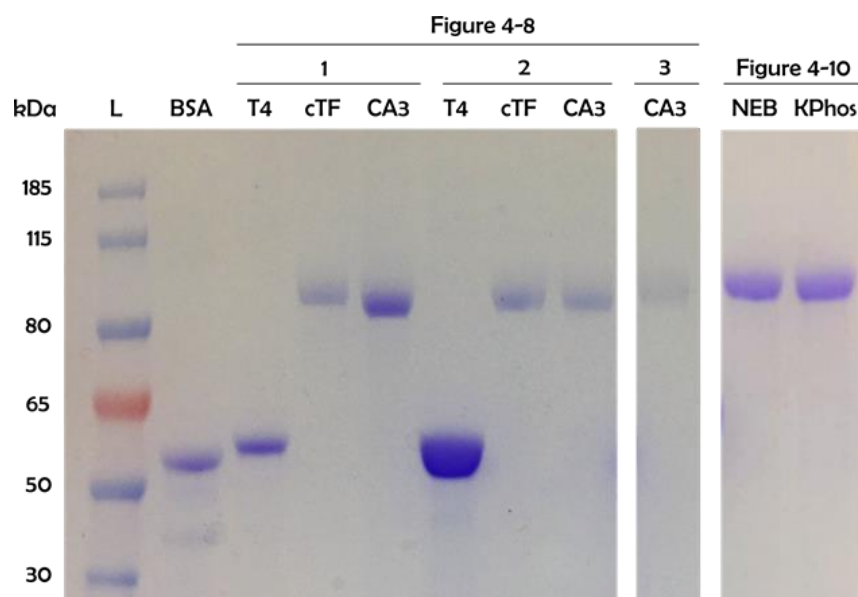


Figure 4-13 SDS-PAGE to assess relative protein concentrations from a selection of the ligase variants assayed in Figure 4-8 and 4-10.

$5\ \mu\text{L}$ of the $5\ \mu\text{M}$ stock concentration of each ligase variant was combined with $10\ \mu\text{L}$ $3\times$ SDS-PAGE loading dye and reactions were performed according to Section 2.8. L is PageRuler Protein Ladder. A $1\ \text{mg}\cdot\text{mL}^{-1}$ BSA was included as a protein concentration standard control.

replicate exhibited almost no ligation activity. Similarly, the results in the buffer comparison in Figure 4-10 can be explained by much higher protein concentration, which produced much higher levels (60-80%) of blunt-end ligation.

To more accurately and reproducibly determine the protein concentration present in each ligase preparation, additional protein quantification methods were employed. For the final ligation assay, all purified ligase variants were concentrated to an equivalent protein concentration for a 10 μ M stock, and measured by nanodrop and UV-Vis spectrophotometry, and a Bradford assay before an average concentration was calculated. The average protein concentration was further analysed by SDS-PAGE to visually compare the intensity next to the protein standard, BSA, then diluted to 5 μ M with an equal volume of 100% glycerol.

4.8 EVALUATING THE LIGATION ACTIVITIES OF ENGINEERED DNA LIGASES

The final aim of this project was to evaluate the ligation activity of the lead engineered DNA ligases from the competitive ligase selection. Following five rounds of competitive ligase selection, ligase-CA3 from ligase-cTF linker library was the clear lead candidate from the library (Figure 1-6). Comparatively, the ligase-p50 linker library did not produce a prolific enrichment of a single variant (Figure 1-6 & Figure 3-7). The linker variant, ligase-PPA17 was therefore chosen for evaluation because it was both highly and consistently recovered at the end of Rounds 5 and 6. Additionally, the original linker variants developed by Wilson et al. (2013), ligase-cTF and ligase-p50 were evaluated to substantiate the changes to ligation activity following competitive selection protocol. Furthermore, the reciprocal fusion construct, p50-ligase, had previously displayed greater ligation activity compared to ligase-p50 (Wilson et al. 2013). p50-ligase was additionally evaluated to reveal whether the competitive ligase selection could enrich for a ligase variant with greater ligation activity from a lower start point. Finally, T4 DNA ligase was the enzyme responsible for ligation in all engineered ligase variants in this study. The unmodified ligase was also assayed and acts as the reference baseline to show overall improvements for DNA ligation.

Following the optimisation conducted throughout this chapter, all six ligase variants were purified by IMAC using a peristaltic pump. The protein concentration was then calculated by averaging the three quantitative methods followed by visual confirmation next to a BSA standard by SDS-PAGE. All proteins were then diluted accordingly and stored in NEB storage buffer. The day after purification, ligation assays were performed

in triplicate in NEB T4 DNA ligase reaction buffer and the results summarised in Figure 4-14.

The overall results from this assay displayed increasing blunt-end ligation at each stage of the development process (Figure 4-14, Panel A). This suggested that better DNA ligases for this application had been successfully developed from both the ligase-cTF and ligase-p50 progenitors, however the improvement in activity of ligase-CA3 over its ligase-cTF progenitor was small and not significant ($p = 0.996$). It should be noted that the original dataset for ligase-CA3 contained an outlier value indicating much higher activity confirmed by a box and whisker plot analysed with a Tukey's test (Appendix S6). This was calculated as an outlier because the value was more than 1.5 times the interquartile range from either the first or third quartile (Tukey 1977), and hence the individual datapoint was excluded. Had the outlier data been retained, the mean activity would have been 17.5%, resulting in a 1.5-fold increase over T4 DNA ligase ($p = 0.001$; Appendix S6). Nevertheless ligase-CA3 displayed a 1.4-fold mean increase in blunt-end ligation compared to T4 DNA ligase ($p = 0.007$). However, this was less than the activity exhibited by the original ligase-p50 developed by Wilson et al. (2013) which also exhibited 1.5-fold improvement in blunt-end ligation activity over T4 DNA ligase ($p = 0.001$), whilst its reciprocal, p50-ligase displayed a 1.7-fold increase over T4 DNA ligase ($p < 0.001$). Finally, the enriched ligase-p50 linker variant, ligase-PPA17 exhibited the greatest improvement in blunt-end ligation, with a 1.8-fold increase over T4 DNA ligase ($p < 0.001$). Overall, the four variants evaluated T4 DNA ligase, ligase-cTF, ligase-CA3, and ligase-p50, yielded consistent data and low standard deviations (SD), however, p50-ligase and ligase-PPA17 exhibited greater spread of data and higher SD (Figure 4-14; Table 4-2). In contrast, cohesive-end ligation activity displayed greater variation in ligation activity (Figure 4-14, Panel B) with five out of six exhibiting relatively high SD (Table 4-3). The large SD makes it difficult to draw firm conclusions for cohesive-end ligation activity, however overall the variant selected by Dr Robins (ligase-CA3) exhibited the highest level of activity and showed a significant improvement over T4 DNA ligase ($p = 0.003$).

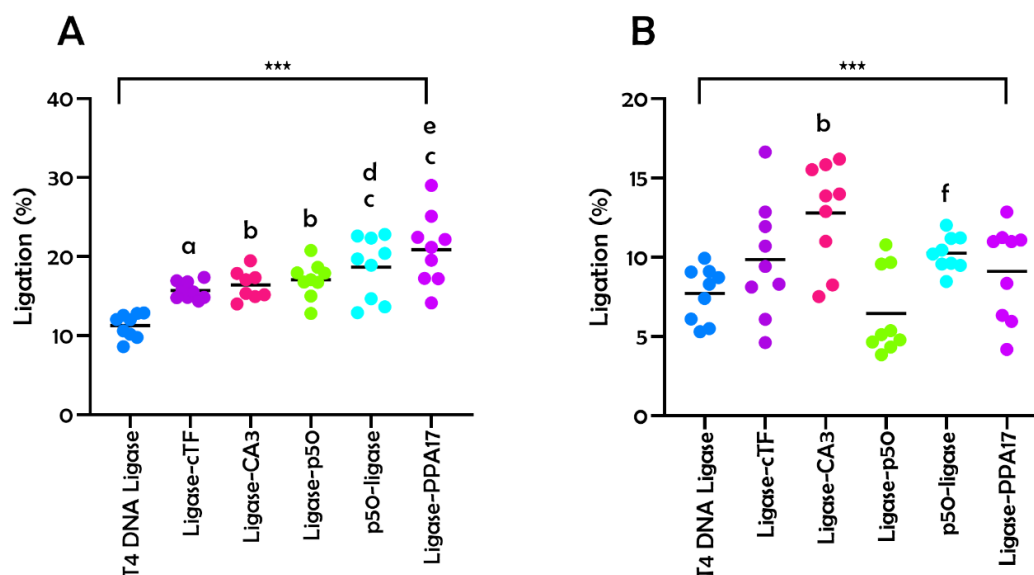


Figure 4-14 Ligation activity of engineered DNA ligase variants.

Ligation reactions were performed as specified in Section 2.11.3. Blunt-end ligation (Panel A) was incubated for 4 hr and cohesive-end (Panel B) for 60 min. Assays were performed in technical triplicates with three biological replicates. Black lines represent mean ligation of substrate to product, expressed as a percentage. The data was analysed using a one-way ANOVA with multiple comparisons in GraphPad Prism 8 to determine statistical significance. Statistical significance for the one-way ANOVA is signified by *** representing $p < 0.001$. The multiple comparisons test was performed with a Tukey's test. a: $p < 0.05$ compared to T4 DNA ligase, b: $p < 0.01$ compared to T4 DNA ligase, c: $p < 0.001$ compared to T4 DNA ligase, d: $p < 0.01$ compared to ligase-cTF, e: $p < 0.05$ compared to ligase-CA3, f: $p < 0.001$ compared to ligase-CA3.

Table 4-2 Blunt-end ligation activity over 4 hr at 25°C (from Figure 4-14, Panel A)

Ligase Variant	Average Ligation (%)	Standard Deviation	nM of Enzyme/nM L1 Ligated ^a
T4 DNA Ligase	11.3	1.5	55.4
Ligase-cTF	15.7	1.1	39.8
Ligase-CA3	16.4	1.8	38.1
Ligase-p50	17.1	4.0	36.6
p50-ligase	18.7	2.2	33.5
Ligase-PPA17	20.9	4.5	29.9

^a All ligation reactions contained a final concentration 80 nM fluorophore-labelled oligonucleotide, L1, and 500 nM ligase enzyme. The ratio of 'nM of Enzyme' to 'nM L1 Ligated' or number of ligations/enzymatic events was calculated by dividing the enzyme concentration in the assay by the average nM of L1 ligated.

Table 4-3 Cohesive-end ligation activity over 60 min at 25°C (from Figure 4-14, Panel B)

Ligase Variant	Average Ligation (%)	Standard Deviation	nM of Enzyme/nM L1 Ligated ^a
T4 DNA Ligase	7.7	1.7	81.0
Ligase-cTF	9.9	3.7	63.4
Ligase-CA3	12.8	3.2	48.9
Ligase-p50	6.5	2.7	96.7
p50-ligase	10.3	1.1	61.0
Ligase-PPA17	9.1	3.0	68.6

^a All ligation reactions contained a final concentration 80 nM fluorophore-labelled oligonucleotide, L1, and 500 nM ligase enzyme. The ratio of 'nM of Enzyme' to 'nM L1 Ligated' or number of ligations/enzymatic events was calculated by dividing the enzyme concentration in the assay by the average nM of L1 ligated.

Table 4-4 Nick-sealing ligation activity over 30 s at 25°C (from Figure 4-7)

Ligase Variant	Average Ligation (%)	Standard Deviation	nM of Enzyme/nM L1 Ligated^a
T4 DNA Ligase	24.2	4.4	25.8
Ligase-cTF	21.2	4.0	29.5
Ligase-CA3	24.5	1.0	25.5
Ligase-p50	23.4	1.9	26.7
p50-ligase	25.1	3.4	24.9
Ligase-PPA17	24.5	2.7	25.5

^a All ligation reactions contained a final concentration 80 nM fluorophore-labelled oligonucleotide, L1, and 500 nM ligase enzyme. The ratio of 'nM of Enzyme' to 'nM L1 Ligated' or number of ligations/enzymatic events was calculated by dividing the enzyme concentration in the assay by the average nM of L1 ligated.

All ligation reactions in this study contained the same concentration of ligase (500 nM) and fluorescent oligonucleotide (80 nM), therefore comprised a total of 6.25 molecules of ligase for every molecule of L1. This permitted the generation of additional data - the number of ligations performed within each ligation. Ideally, the total number of ligation events per reaction would be closer to, or 1 (i.e. each molecule of ligase was catalysing one ligation event; note it is also possible that an enzyme may be able to perform multiple ligation events while others may not perform a single ligation). The data for blunt-, cohesive-end, and nick-sealing activity was calculated based off the average ligation events depicted in Figure 4-7 and Figure 4-14, and summarised in, Table 4-2, Table 4-3 & Table 4-4, respectively. The blunt-end ligation activity indicated that it took 55 molecules of T4 DNA ligase to catalyse one ligation event, whereas our best ligase variant ligase-PPA17 only took 30 molecules of enzyme to catalyse one ligation event within 4 hr (Table 4-2). Likewise, for cohesive-end ligation, all six variants required between 49 and 97 molecules of enzyme to catalyse one ligation event within 60 min (Table 4-3). Ligation activity is relatively low for these two reactions, however T4 DNA ligase more efficiently catalyses single-stranded nicks compared to blunt- and cohesive-end substrates (Lohman, Chen, and Evans 2011). This was verified in the nick-sealing assay with 25-30 molecules of each engineered ligase enzyme performing one ligation event within 30 s, irrespective of whether they contained an additional DNA-binding domain (Table 4-4). This provides evidence that a higher number of T4 DNA ligase (molecules of enzyme) could retain efficient ligation activity for nicked DNA substrates over a shorter duration. Whilst we were unable to establish the same level of activity for blunt-end ligation as single-stranded nick sealing, we have displayed significant improvements to the ligation efficiency following multiple ligase rounds of selection.

4.9 DISCUSSION

In this chapter, we sought to establish and optimise the fluorescence-based ligation assay to evaluate the capabilities of T4 DNA ligase, ligase-cTF, ligase-CA3, ligase-p50, p50-ligase, and ligase-PPA17 to catalyse blunt-, cohesive-end, and single-stranded nicks. The main objective of this research was to investigate whether the competitive selection protocol could successfully enrich for ligases with improved blunt-end ligation activity without addition of macromolecular exclusion molecules (such as PEG), with the overall aim of generating a ligase variant that could be used in commercial applications to improve molecular biology experiments such as NGS.

4.9.1 Purification of Ligase Variants

The three ligase variants, T4 DNA ligase, ligase-CA3, and ligase-p50 were evaluated to identify the optimal imidazole concentration for protein elution and purity (Figure 4-2). The chromatograms showed that T4 DNA ligase, and the two ligase-cTF variants evaluated in this study could be eluted with a 150 mM imidazole concentration, whereas, ligase-p50 required was eluted at 200 mM imidazole, however required an additional 125 mM imidazole wash step to remove non-specific protein contaminants. Ligase-CA3 would therefore be a better commercial enzyme for large-scale purification because it requires fewer purification steps and buffers, thereby decreasing manufacturing costs.

The in-depth analysis of the imidazole elution concentrations for T4 DNA ligase, ligase-CA3 and ligase-p50 was shown to be beneficial because these variants produced more consistent blunt-end ligation in Figure 4-14 (Panel A). The three variants (and ligase-cTF) resulted in more consistent reliable data as shown by smaller SD. In comparison, p50-ligase and ligase-PPA17 used the elution concentrations for ligase-p50. The results from these variants exhibited much greater spread of individual data points and higher SD. Not individually determining imidazole elution concentrations may have contributed to this, therefore future work would be to evaluate p50-ligase and ligase-PPA17 on a 0-100% elution gradient to more accurately determine the elution concentration for these variants.

4.9.2 Optimisation of Ligation Reactions

Generally, to determine an optimal storage buffer, a thermal shift assay is performed to screen a range of buffers conditions and assess protein denaturing or unfolding. However, because T4 DNA ligase is already a commercially available enzyme, we chose

to use a pre-optimised storage buffer that is already known to be compatible and retain ligation activity. The buffer and pH of a solution are important parameters in biochemical and biological applications as these can influence the recovery and purity of proteins, as well as the rate of chemical reactions (Ferreira et al. 2015). Initially, ligase variants were purified into KPhos storage buffer, as previously employed by Wilson et al. (2013), however, when the engineered ligase variants were stored in KPhos storage buffer, the ligation activity was shown to be varied between batches (Figure 4-8 to Figure 4-10). Whilst this was later determined to have been caused by inaccurately measured protein concentrations, there were several advantages to changing away from a phosphate buffer to a Tris buffer.

Previous research revealed that phosphate buffers can be active participants in many biochemical processes and reactions. Phosphate buffers sequester metal-ions such as calcium and magnesium, therefore inhibiting metal-ion dependent enzymes (Green and Sambrook 2018). In addition, phosphate buffers can inhibit enzymes that use phosphate esters and anhydrides as substrates (Ferreira et al. 2015), therefore can interfere in molecular biology reactions including restriction enzyme DNA cleavage and DNA ligation (Green and Sambrook 2018). DNA ligases are Mg^{2+} -dependent and enzymatically activated by a nucleophilic reaction with a phosphate anhydride from a high energy molecule (Figure 1-1; Shuman 2009). The adenylated (or active) ligase next needs to recognise an exposed 5' phosphate prior to the formation of the phosphodiester bond. It is likely that storage of DNA ligases in a phosphate buffer might lead to reduced ligation activity.

Furthermore, the presence of high (200 mM) salt concentrations can also inhibit DNA ligation (Hayashi et al. 1985; Raae, Kleppe, and Kleppe 1975). The KPhos storage buffer that was initially employed contained 200 mM NaCl. This buffer therefore contained two known components that can inhibit ligation activity. These factors may account for the 20% difference in ligation observed between KPhos and the NEB buffer (Figure 4-10). The last buffer optimisation step was investigating the differences in the ligation reaction buffer (Figure 4-11). The main difference was that initial ligation buffer described in Williamson et al. (2014) contained a final concentration of 50 mM NaCl, whereas the NEB buffer did not. The study of DNA Lig E family of enzymes and their mutants were shown to tolerate high concentrations up to 500 mM of NaCl (Williamson, Rothweiler, and Leiros 2014). However, T4 DNA ligase has previously shown not to tolerate >200 mM salts including NaCl, KCl, CsCl, or LiCl, resulting in low or reduced ligation activity (Raae,

Kleppe, and Kleppe 1975), and our observation of superior ligation activity in NEB buffer was consistent with this (Figure 4-11). Moreover, the optimisation of the ligase reaction buffer matched the same conditions that were competitive ligase selection protocol to enrich each variant (this enrichment was performed at 25°C in 1× NEB T4 ligation reaction buffer; Section 2.9.3). Vitality stands to reason that the ligation reaction buffer used in the competitive ligase selection should be the same as that used in the fluorescence-based ligation assays, to reproduce the conditions which activity was selected for.

4.9.3 Quantification of Ligation Activity

In the field of directed evolution, a common truism is “You get what you select (or screen) for” (Zhao and Arnold 1997). Our results show that after five rounds of selection, we enriched a superior ligase variant, ligase-PPA17 that was capable of improved blunt-end ligation compared to T4 DNA ligase and the other five variants (Figure 4-14, Panel A). Ligase-PPA17 produced the greatest ligation activity with a 1.8 fold-increase in mean blunt-end ligation activity over T4 DNA ligase ($p < 0.001$). In addition, ligase-PPA17 produced a 1.2-fold increase over its predecessor, ligase-p50 which provided strong evidence that the variant exhibited improved blunt-end ligation activity ($p = 0.067$). Finally, ligase-PPA17 displayed a 1.1-fold increase compared to the superior p50 variant (p50-ligase) from Wilson et al. (2013). Our results suggest that the competitive ligase selection protocol can enrich for ligase variants with improved blunt-end ligation activity. We have additionally shown that we can take a ligase-variant from a lower starting activity, ligase-p50, and with several iterations of the competitive ligase selection enrich for a variant with greater activity than p50-ligase, the superior ligase variant identified in Wilson et al. (2013).

A somewhat surprising outcome was that the ligase variant that exhibited the greatest enrichment over five rounds of selection, ligase-CA3, was not the best-performing variant overall for blunt-end ligation (Figure 4-14, Panel A). Ligase-CA3 did not provide evidence for improved blunt-end activity over its predecessor, ligase-cTF ($p = 0.996$), displaying a 1% mean increase in total ligation. Analysis of the raw data using a box and whisker plot with Tukey’s test (Appendix S6) showed a clear outlier had been produced with 10% more ligation greater than the mean total blunt-end ligation recorded in Table 4-2, or a value that was 1.5 times the interquartile range from the first or third quartile (Tukey 1977). Considering that ligase-CA3 was the variant used for buffer optimisation,

and the other eight remaining datapoints were reliably within 5% of one another, this gave increased certainty that this specific reaction was an outlier that should be excluded. Ideally, further repeats would be performed to gain confidence around the final datasets and replace the excluded dataset, however exclusion from the laboratory due to the Covid-19 outbreak precluded this.

Both blunt- and cohesive-end ligation activity displayed statistically significant changes across the six variants ($p < 0.001$). Whilst there is considerable variance in cohesive-end ligation activity data, it is possible that ligase-CA3 was substantially improved in cohesive-end ligation over T4 DNA ligase ($p = 0.003$), despite this activity not having been directly selected for (Figure 4-14, Panel B). However, judgments on this should be reserved for now.

Finally, modifications to T4 DNA ligase with either DNA-binding domain and their respective linkers were not shown to change the enzyme's ability to perform its primary reaction of single-strand nick sealing (Figure 4-3; Table 4-4). All six variants produced similar nick-sealing activities ranging from 21-25% over 30 s. Within 30 s, 1 in every 25 molecules of ligase within the reaction could ligate a single-stranded nick. While no qualitative enzyme kinetic experiments were performed in this study, it is clear that we have not been able to evolve a "perfect enzyme". Albeit, we have evolved for improved blunt-end ligation without the addition of macromolecular crowding agents, producing a better DNA ligase variant capable of this activity, for which would be advantageous in NGS techniques and molecular cloning where blunt-end substrates are integral components within the workflow.

Chapter 5: Key Findings, Future Directions and Conclusions

5.1 RESEARCH MOTIVATIONS

The overarching aim of the DNA ligase project in the Patrick and subsequently Ackerley labs has been to deliver more effective and efficient DNA ligases to market. The first breakthrough was showing that we can improve T4 DNA ligase activity by fusing a DNA-binding domain to the enzyme, increasing its affinity for DNA substrates (Wilson 2009; Wilson et al. 2013). From this study, a small-scale linker experiment demonstrated the potential for this variable to modulate activity, leading to the concept and design of a competitive ligase selection protocol to enable a high-throughput approach to screen for more efficient DNA ligase variants (Stevenson 2014). This selection was optimised, and two linker libraries (derived from ligase-cTF and ligase-p50) were designed and constructed by Dr Robins, after which five rounds of the optimised competitive ligase selection protocol were applied to each library. The ligase-cTF library yielded clear enrichment of a single variant, ligase-CA3, however the ligase-p50 library did not produce a lead linker variant (Figure 1-6).

The initial focus of the research in this thesis was to apply an additional round of selection to the ligase-p50 linker library to establish whether a lead linker variant would emerge. Subsequently, the enriched DNA ligases from five or six rounds of selection were to be evaluated using a fluorescence-based ligation assay. This assay was established in our lab and optimised during this study.

5.2 KEY FINDINGS

5.2.1 Identifying Improved Ligase-p50 Linker Variants

Following one additional round of competitive ligase selection, there was no evidence for emergence of a lead linker variant in the ligase-p50 linker library (Figure 3-7), based off the results produced in Replicate 1. All increases and decreases in the relative levels of linker variants from Round 5 appeared consistent with stochastic variation, and no linker variant exceeded 20% of the recovered variants. As there were complications in sequencing Round 6, Replicate 2, the final data may not reflect the final changes to

enrichment at the end of Round 6. Given the uncertainties, one of the most highly represented linker variants, ligase-PPA17, was selected for further investigation in the fluorescence-based ligation assay.

5.2.2 Elucidating Protein Buffer Conditions for Improved Ligation Activity

Our initial ligation activity assays produced inconsistent and unreproducible ligation activity data (Figure 4-8). Having identified variabilities in protein concentrations as being a major cause of the disparities between variants, this problem was resolved by more meticulously measuring of protein concentration by averaging various quantification techniques to obtain a more accurate final protein concentration. In addition, we demonstrated that the buffer conditions were unfavourable for ligation activity. The initial ligation storage buffer was a potassium phosphate buffer containing NaCl. Phosphate buffers have been previously identified as buffers that interfere with many biochemical processes including DNA ligation (Green and Sambrook 2018). Additionally, NaCl and other monovalent ions have been found to inhibit T4 DNA ligase (Hayashi et al. 1985; Raae, Kleppe, and Kleppe 1975). With the guidance of current commercial T4 DNA ligase buffers, our results showed improved ligation activity in NEB storage and reaction buffers (Figure 4-10 & Figure 4-11).

5.2.3 Evaluating Ligation Activity of Engineered DNA Ligase Variants

All six variants retained high affinity and efficiency for single-stranded nick-sealing, reporting equivalent levels of ligation activity. However, the evolved ligase linker variants from the ligase-cTF and ligase-p50 linker libraries displayed contrasting results following five rounds of selection. While all five of the engineered DNA ligase variants exhibited improved blunt-end ligation activity over T4 DNA ligase (Figure 4-14), ligase-CA3 did not provide clear evidence for improved blunt-end ligation over its predecessor, ligase-cTF. In contrast, ligase-PPA17 exhibited a significant improvement for blunt-end ligation over T4 DNA ligase and provided strong evidence that the variant was significantly improved from its predecessor, ligase-p50. Cohesive-end ligation was not the ligation activity selected for by the competitive ligase selection and the six variants reported higher variance for this ligation activity after 60 min. The inconsistencies in the data made it difficult to make firm conclusions on changes to cohesive-end ligation, however it was clear that ligase-CA3 exhibited greater activity compared to the other five variants.

5.3 CRITICAL EVALUATION OF METHODOLOGY

5.3.1 Competitive Ligase Selection Protocol

The competitive ligase selection is a directed evolution strategy designed to provide a compartmentalised self-replication reaction to select for enzyme variants by interacting with their own gene, thereby acting as their own selection pressure. Like many directed evolution screening strategies, it may take several weeks to months to reap the benefits of improved enzymatic activity however, this selection can increase the throughput analysis of ligation activity. Ligase variants do not need to be individually purified and assayed, instead they can be expressed directly from raw cell lysate and our protocol has successfully proven that it can eliminate undesirable variants and enrich for improved blunt-end ligation activity.

However, these advantages only accrue if the protocol can be run consistently. The whole plasmid PCR step has historically been a troublesome aspect of the competitive ligase selection protocol, displaying reproducibility complications, and requiring considerable optimisation (Stevenson 2014; Dr Robins, unpublished results). While this resulted in a protocol that was successful for Dr Robins, the whole plasmid PCR was unable to be reproducibly replicated in this study for a sixth selection round from a cell culture, and ultimately led to the whole plasmid amplicons being amplified from a purified DNA template (Figure 3-1 to Figure 3-3). There are several variables that might change between researchers which could influence the ability to successfully amplify the whole plasmid amplicons (e.g. interpretation of high cell density, accuracy and differences in pipettes or lab equipment, or day-to-day environmental conditions/temperatures). It is therefore recommended that each researcher should perform a small number of optimisation experiments to maximise the reproducibility and success of the reaction from cell cultures.

In regard to the overall aim of the DNA ligases project i.e. delivery of more effective and efficient DNA ligases to market, the adaptable nature of the ligation reaction in this protocol provides an opportunity to compete several different DNA ligases to identify the best for a particular condition. For example, T4 DNA ligase comes from a mesothermic organism and has optimal ligation activity at 37 °C (Ferretti and Sgaramella 1981). However, DNA ligation is routinely performed at much lower temperatures between 16-25 °C. Consequently, there is a sigmoidal relationship between DNA ligation and temperature, for which increasing temperature results in decreased ligation (Ferretti

and Sgaramella 1981). Ferretti and Sgaramella (1981) showed that temperatures greater than 16 °C or 25 °C for blunt- and cohesive-end ligation respectively, resulted in greater than 50% loss of total ligation. A cold active DNA ligase would likely offer advantages activity over T4 DNA ligase, as psychrophilic organisms have an optimal temperature ranging between 0 - 20 °C (Morita 1975). Different candidates could be competed against one another at a lower ligation temperature than used in this study to enrich for the best. Alternatively, the competitive ligase selection could even be combined with a metagenomic screening approach (e.g., screening cloned soil DNA in the temperature-sensitive ligase strain GR501 used in Section 4.2.1 of this study) to first discover entirely new DNA ligases and then select for those that behave most optimally under preferred conditions.

In a similar fashion to strategies used to evolve DNA polymerases to preferentially incorporate specific nucleotides in PCR reactions (Tubelvičiūtė and Skirgaila 2010), the competitive ligase selection could be used to preferentially select for ligase variants with specific ligase fidelity (i.e. better at ligating A/T or G/C). Previous research investigating the base pairing specificity of various DNA ligases showed that T4 DNA ligase exhibited increased ligation for DNA substrates with a A/T end rather than G/C; T3 DNA ligase preferred substrates with a G/C end, whereas Human DNA ligase III had no nucleotide preference (Bauer et al. 2017). The whole plasmid PCR may be modified to change the end base pair specificity and select for a specific ligation reaction. This would be beneficial if selecting for alternative DNA ligases for library preparation of NGS techniques which require the ligation of adaptor molecules to an adenine nucleotide at the end of fragmented DNA. In conclusion, the competitive ligase selection has numerous potential uses and applications to enrich for improved DNA ligases.

5.3.2 Evaluating Ligation Activity using the Fluorescence-Based Ligation Assay

The evaluation of DNA ligation using the fluorescence-based ligation assay was major focus of this study and offered several advantages over the previous agarose-based ligation assays used in Wilson et al. (2013). Firstly, the use of synthesized oligonucleotides meant that the DNA substrates for ligation assays did not have to be prepared by time-consuming and low yield plasmid digestion and gel extraction (as was the case in Wilson et al. 2013). This approach also allowed for the same template oligonucleotides (L1 and L2) to be used at the equivalent concentrations and combined

with different complementary oligonucleotides to form specific substrates or reactions (i.e. blunt- or cohesive-end, or single-stranded nick sealing).

Electrophoretic mobility shift assays are the most popular technique for studying protein-nucleic acid interactions, including DNA ligation (Bowater et al. 2015). This is because it is a simple method that can detect size differences with relatively high sensitivity. However, one limitation of this assay is that the size of the oligonucleotides used is much smaller than that of the DNA ligase fragments used in the molecular cloning experiments or adaptor attachment in the library preparation step for NGS techniques. For example, the current size of Illumina sequencing libraries ranges from 200-500 bp (Bronner et al. 2009), or 10 times the size of that used in our assay. While polyacrylamide (or TBE-PAGE) gels can resolve nucleic acids between 1-1000 bp, it is more cost-effective and time-efficient to work with smaller oligonucleotides (Hellman and Fried 2007). It will be necessary to confirm that lead ligases from the competition assay be validated for DNA ligation with larger fragments; for this, the blue-white colony screening and the multiplex NGS adaptor assay performed in Wilson et al. (2013) could be used to model alternative molecular biology applications (explained in detail in Section 5.4).

Gel electrophoretic mobility shift assays are laborious and time-consuming. This was a key driver behind the design of the competitive ligase selection protocol to allow for higher throughput and analysis of ligase variants and activity. However, it is essential to quantitatively compare the activity of these variants after several iterations of selection. While it is possible to increase the throughput for the number of ligation reactions performed (e.g. in a 96 well plate), the purification of ligase variants, as well as number of wells on a gel places limits on their evaluation. These factors constrained the number of ligase variants that were analysed in this study.

5.4 FUTURE DIRECTIONS

A limitation that may have affected the enrichment of ligase-CA3 was that neither the original ligase-cTF variant, nor the improved blunt-end variant, ligase-NLS-cTF from the Wilson et al. (2013) were not included in the ligase-cTF library. Thus, it may have been that ligase-CA3 was indeed the best variant in the starting library, but that there were no candidates in the pool offering significant improvements on ligase-cTF. Alternatively, it is possible that additional repeats (not possible with the Covid-19 lockdown) would have

indicated that the 'outlier' data discarded on the basis of a Tukey test (Appendix S7) was in fact a truer representation of ligase-CA3 activity.

As mentioned in the Introduction (Section 5.3.2), one limitation for the fluorescence-based ligation assay was number of variants that could be realistically assessed. Nevertheless, the results from Round 5 and 6 of the ligase-p50 linker library were not able to enrich of lead variant, therefore it may be beneficial to evaluate additional linker variants. Ligase-PPSR3 and ligase-PGSG12 were two variants that were consistently selected in Round 5 and 6 from Replicate Plate A (Figure 3-7). In addition, ligase-PGSG15 from Round 5 Replicate Plate B showed substantial enrichment with 23% and was the best ligase variant in Round 5 for that plate. Lastly, ligase-PHL10 was overall the most enriched linker from Round 5 (Figure 1-6). Evaluating additional variants would provide a more comprehensive assessment of whether ligase-PPA17 was truly a better ligase variant from the ligase-p50 library.

Lastly, it would be advantageous to evaluate our improved engineered ligase variants in molecular biology applications that they would be used in. In addition to an agarose-based ligation assay, our previous study analysed DNA ligation using a blue-white colony screening assay and an NGS assay (Wilson et al. 2013).

Molecular cloning is a fundamental which utilises a DNA ligase to ligate fragments into a cut vector backbone (Carroll 1993; Jackson, Symons, and Berg 1972). This is a key process in molecular biology and has been the driver of many biotechnological advances. The blue-white colony screening assay evaluates the number of white colonies produced following transformation of *E. coli* with the resultant ligation reactions performed by a candidate DNA ligase. A white colony is formed when lacZ α , encoding β -galactosidase is disrupted by insertion of a DNA fragment indicating a successful ligation. The more effective a ligase is at ligating the DNA fragment into a cut linear vector backbone, the greater number of white colonies present following transformation. Ligase-cTF was previously the best variant in this assay (Wilson et al. 2013), therefore it would be interesting to determine the differences following the enrichment of ligase-CA3 in this particularly experiment.

The NGS assay is more complicated assay and additionally cannot be performed in our lab at Victoria University of Wellington because we do not have the facility or equipment to do so. In 2019, we enquired with the Auckland Genomics Facility about repeating the

NGS assay performed in Wilson et al. (2013) with our improved variants. The layout for this assay would follow the standard library preparation workflow for NGS libraries however, ligation of adaptors would be replaced by our engineered ligase variants rather than the commercially supplied T4 DNA ligase and kits. A large sample of genomic DNA would be purified and fragmented before DNA end-repair (Head et al. 2014; Quail et al. 2008). Next, the 5' prime end of DNA would be phosphorylated, and the 3' prime end of DNA adenylated resulting in a poly-A tail for adaptor attachment. The DNA would then be equally split into equal fractions for the number of ligase variants and the respective biological replicates. Each individual ligation assay would have a different adaptor to ligate to fragmented DNA. Finally, all ligation reactions would be purified and sequenced. However, instead of analysing the fragmented DNA, we can use bioinformatic tools deconvolute and count the number of adaptors based on the unique adaptor assigned to each variant. As one of the major goals of this project was to improve blunt-end DNA ligation activity to overcome the kinetic limitations of T4 DNA ligase, particularly for NGS which often results in sub-optimal library formation and therefore poor sequencing reads, it would be appropriate to evaluate activities in this context.

5.5 CONCLUDING REMARKS

This thesis demonstrated that the competitive ligase selection can select for improved DNA ligase variants. While we were unable to select a clear lead candidate from the ligase-p50 library, our results in the fluorescence-based ligation assay showed that this exhibited superior blunt-end ligation activity over the five other variants, including T4 DNA ligase and its predecessor ligase-p50. Although we did not produce significant improvements for the lead candidate from the ligase-cTF library, ligase-CA3, the competitive ligase selection has provided evidence to prove that it enriches for the condition it is selecting for. However, further work is needed, as proposed above to determine whether the improved DNA ligase variants are sufficiently good to substitute the commercial T4 DNA ligase for applications in molecular biology and biotechnology where superior blunt-end ligation enzymes are required.

References

- Abil, Zhanar, and Andrew D. Ellington. 2018. "Compartmentalized Self-Replication for Evolution of a DNA Polymerase." *Current Protocols in Chemical Biology* 10(1): 1–17.
- Argos, Patrick. 1990. "An Investigation of Oligopeptides Linking Domains in Protein Tertiary Structures and Possible Candidates for General Gene Fusion." *Journal of Molecular Biology* 211(4): 943–58.
- Armstrong, J, R. S Brown, and A Tsugita. 1983. "Primary Structure and Genetic Organization of Phage T4 DNA Ligase." *Nucleic Acids Research* 11(20): 7145–56.
- Arnold, Frances H. 2018. "Directed Evolution: Bringing New Chemistry to Life." *Angewandte Chemie - International Edition* 57(16): 4143–48.
- Baeuerle, Patrick A, and Thomas Henkel. 1994. "Function and Activation of NF-KB in the Immune System." *Annual Review of Immunology* 12: 141–79.
- Bauer, Robert J et al. 2017. "Comparative Analysis of the End-Joining Activity of Several DNA Ligases." *PLOS One* 12(12): e0190062.
- Becker, Andrew, Glenda Lyn, Aialcolm Geffer, and Jerard Hurwitz. 1967. "The Enzymatic Repair of DNA, II. Characterization of Phage-Induced Sealase." *Proceedings of the National Academy of Sciences of the United States of America* 58: 1996–2003.
- Berg, Kristel, Ingar Leiros, and Adele Williamson. 2019. "Temperature Adaptation of DNA Ligases from Psychrophilic Organisms." *Extremophiles*: 1–13.
- Bowater, Richard P. et al. 2015. "Biophysical and Electrochemical Studies of Protein-Nucleic Acid Interactions." *Monatshefte fur Chemie* 146(5): 723–39.
- Bronner, Iraad F, Michael A Quail, Daniel J Turner, and Harold Swerdlow. 2009. "Improved Protocols for Illumina Sequencing." In *Current Protocols in Human Genetics*, , 10.1002/0471142905.hg1802s62. www.sagescience.com] (April 6, 2020).
- Carroll, William. L. 1993. "Introduction to Recombinant-DNA Technology." *American Society for Clinical Nutrition* 58: 249–58.
- Chen, Xiaoying, Jennica L. Zaro, and Wei Chiang Shen. 2013. "Fusion Protein Linkers: Property, Design and Functionality." *Advanced Drug Delivery Reviews* 65(10): 1357–69.
- Cheng, Chonghui, and Stewart Shuman. 1997. "Characterization of an ATP-Dependent DNA Ligase Encoded by Haemophilus Influenzae." *Nucleic Acids Research* 25(7): 1369–74.
- Cherepanov, Alexey V., and Simon De Vries. 2003. "Kinetics and Thermodynamics of Nick Sealing by T4 DNA Ligase." *European Journal of Biochemistry* 270(21): 4315–25.
- Chytil, Milan, and Gregory L Verdinef. 1996. "The Rel Family of Eukaryotic Transcription Factors." *Current Opinion in Structural Biology* 6: 91–100.
- Cozzarelli, N. R., Norman E. Melechen, Thomas M. Jovin, and Arthur Kornberg. 1967. "Polynucleotide Cellulose as a Substrate for a Polynucleotide Ligase Induced by Phage T4." *Biochemical and Biophysical Research Communications* 28(4): 578–86.
- Dermody, James J, Gary T Robinson, and Rolf Sternglanz. 1979. "Conditional-Lethal Deoxyribonucleic Acid Ligase Mutant of Escherichia Coli." *Journal of Bacteriology* 139(2): 701–4.
- Eijsink, Vincent G.H., Sigrid GÅseidnes, Torben V. Borchert, and Bertus Van Den Burg. 2005. "Directed Evolution of Enzyme Stability." *Biomolecular Engineering* 22(1–3): 21–30.
- Ellenberger, Tom, and Alan E Tomkinson. 2008. "Eukaryotic DNA Ligases: Structural and Functional Insights." *Annual Review Biochemistry* 77: 313–38.
- Ferreira, Carlos M.H., Isabel S.S. Pinto, Eduardo V. Soares, and Helena M.V.M. Soares. 2015. "(Un)Suitability

- of the Use of PH Buffers in Biological, Biochemical and Environmental Studies and Their Interaction with Metal Ions-a Review." *Royal Society of Chemistry Advances* 5(39): 30989-3.
- Ferretti, Luca, and Vittono Sgaramella. 1981. "Temperature Dependence of the Joining by T4 DNA Ligase of Termini Produced by Type II Restriction Endonucleases." *Nucleic Acids Research* 9.
- Fujita, T, G P Nolan, S Ghosh, and D Baltimore. 1992. "Independent Modes of Transcriptional Activation by the P50 and P65 Subunits of NF-Kappa B." *Genes & Development* 6: 775-87.
- Gefter, Malcolm L, Andrew Becker, and Jerard Hurwitz. 1967. "The Enzymatic Repair of DNA, I. Formation of Circular Lambda DNA." *Proceedings of the National Academy of Sciences of the United States of America* 58: 240-47.
- Gellert, Martin. 1967. "Formation of Covalent Circles of Lambda DNA by E. Coli Extracts." *Proceedings of the National Academy of Sciences of the United States of America* 57: 148-55.
- George, Richard A, and Jaap Heringa. 2003. "An Analysis of Protein Domain Linkers: Their Classification and Role in Protein Folding." *Protein Engineering* 15(11): 871-79.
- Ghadessy, Farid J, Jennifer L Ong, and Philipp Holliger. 2001. "Directed Evolution of Polymerase Function by Compartmentalized Self-Replication." *Proceedings of the National Academy of Sciences of the United States of America* 98(8): 4552-57.
- Gibson, Daniel G. et al. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nature Methods* 6(5): 343-45.
- Green, Michael R., and Joseph Sambrook. 2018. "Buffers." *Cold Spring Harbor Protocols* 2018(2): 82-85.
- Green, Michael R, and Joseph Sambrook. 2019. "Ligation and Ligases." In *Cold Spring Harbor Protocols*, Cold Spring Harbor Laboratory Press, 583-88.
- Hayashi, Ken'ichiro, Masako Nakazawa, Nobutsugu Hiraoka, and Akira Obayashi. 1986. "Regulation of Inter- and Intramolecular Ligation with T4 DNA Ligase in the Presence of Polyethylene Glycol." *Nucleic Acids Research* 14(19): 7617-31.
- Hayashi, Ken'ichiro, Masako Nakazawa, Yukuo Ishizaki, and Akira Obayashi. 1985. "Influence of Monovalent Cations on the Activity of T4 DNA Ligase in the Presence of Polyethylene Glycol." *Nucleic Acids Research* 13(9): 3261-71.
- Head, Steven R. et al. 2014. "Library Construction for Next-Generation Sequencing: Overviews and Challenges." *BioTechniques* 56(2): 61-77.
- Hellman, Lance M., and Michael G. Fried. 2007. "Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions." *Nature Protocols* 2(8): 1849-61.
- Ho, C Kiong, James L Van Etten, and Stewart Shuman. 1997. "Characterization of an ATP-Dependent DNA Ligase Encoded by Chlorella Virus PBCV-1." *Journal of Virology* 71(3): 1931-37.
- Howes, Timothy R L, and Alan E Tomkinson. 2012. "DNA Ligase I, the Replicative DNA Ligase." *Subcell Biochemistry* 62: 10.1007/978-94-007-4572-8_17.
- Jackson, David A., Robert H. Symons, and Paul Berg. 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 of the United States of America* 69(10): 2904-9.
- Jain, Jugnu et al. 1995. "A Similar DNA-Binding Motif in NFAT Family Proteins and the Rel Homology Region." *Journal of Biological Chemistry* 270(8): 4138-45.
- Kawashima, Eric H, Laurent Farinelli, and Pascal Mayer. 2005. "Method of Nucleic Acid Sequencing."
- Kim, Hyun Kyu, and Suk Tae Kwon. 1998. "Cloning, Nucleotide Sequence, and Expression of the DNA Ligase-Encoding Gene from Thermus Filiformis." *Molecules and Cells* 8(4): 438-43.
- Kitagawa, Masanari et al. 2005. "Complete Set of ORF Clones of Escherichia Coli ASKA Library (a Complete Set of E. Coli K-12 ORF Archive): Unique Resources for Biological Research." *DNA Research* 12(5):

291–99.

- Kuchner, Olga, and Frances H. Arnold. 1997. "Directed Evolution of Enzyme Catalysts." *Trends in Biotechnology* 15(12): 523–30.
- Lavesa-Curto, Manuel et al. 2004. "Characterization of a Temperature-Sensitive DNA Ligase from *Escherichia Coli*." *Microbiology* 150: 4171–80.
- Leemhuis, Hans, Ronan M. Kelly, and Lubbert Dijkhuizen. 2009. "Directed Evolution of Enzymes: Library Screening Strategies." *International Union of Biochemistry and Molecular Biology Life* 61(3): 222–28.
- Leemhuis, Hans, Viktor Stein, Andrew D. Griffiths, and Florian Hollfelder. 2005. "New Genotype-Phenotype Linkages for Directed Evolution of Functional Proteins." *Current Opinion in Structural Biology* 15(4): 472–78.
- Lehman, I R. 1974. "DNA Ligase: Structure, Mechanism, and Function." *Science* 186(4166): 790–97.
- Linnarsson, Sten. 2010. "Recent Advances in DNA Sequencing Methods - General Principles of Sample Preparation." *Experimental Cell Research* 316(8): 1339–43.
- Loakes, David et al. 2009. "Evolving a Polymerase for Hydrophobic Base Analogues." *Journal of the American Chemical Society* 131(41): 14827–37.
- Lohman, Gregory J S, Lixin Chen, and Thomas C Evans. 2011. "Kinetic Characterization of Single Strand Break Ligation in Duplex DNA by T4 DNA Ligase." *Journal of Biological Chemistry* 286(51): 44187–96.
- Lohman, Gregory J S, Stanley Tabor, and Nicole. M Nichols. 2011. "DNA Ligases." In *Current Protocols in Molecular Biology*, , 3.14.1-3.14.7.
- de Lumley, Marie et al. 2004. "A Biophysical Characterisation of Factors Controlling Dimerisation and Selectivity in the NF-KB and NFAT Families." *Journal of Molecular Biology* 339(5): 1059–75.
- Lund, Anders H., Mogen Duch, and Finn Sjou Pedersen. 1996. "Increased Cloning Efficiency by Temperature-Cycle Ligation." *Nucleic Acid Research* 24(4): 800–801.
- Luo, J. 1996. "Improving the Fidelity of *Thermus Thermophilus* DNA Ligase." *Nucleic Acids Research* 24(15): 3071–78.
- Macián, Fernando, Cristina López-Rodríguez, and A Rao. 2001. "Partners in Transcription: NFAT and AP-1." *Oncogene* 20: 2476–89.
- Margulies, Marcel et al. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437(7057): 376–80.
- Martin, Ina V, and Stuart A Macneill. 2002. "ATP-Dependent DNA Ligases." *Genome Biology* 3(4): reviews3005.1–3005.7.
- May, Michael J, and Sankar Ghosh. 1997. "Rel/NF-KB and IκB Proteins: An Overview." *Cancer Biology* 8: 63–67.
- Mierzejewska, Karolina et al. 2014. "Structural Basis of the Methylation Specificity of R.DpnI." *Nucleic Acids Research* 42(13): 8745–54.
- Morita, Richard Y. 1975. "Psychrophilic Bacteria." *Bacteriological Reviews* 39(2): 144–67.
- Nair, Pravin A. et al. 2007. "Structural Basis for Nick Recognition by a Minimal Pluripotent DNA Ligase." *Nature Structural and Molecular Biology* 14(8): 770–78.
- Narumi, Issay et al. 2004. "PprA: A Novel Protein from *Deinococcus Radiodurans* That Stimulates DNA Ligation." *Molecular Microbiology* 54(1): 278–85.
- Ogawa, Tohru, and Tuneko Okazaki. 1980. "Discontinuous DNA Replication." *Annual Review Biochemistry* 49: 421–57.
- Olsen, Mark, Brent Iverson, and George Georgiou. 2000. "High-Throughput Screening of Enzyme Libraries."

Current Opinion in Biotechnology 11(4): 331–37.

- Pascal, John M., Patrick J. O'Brien, Alan E. Tomkinson, and Tom Ellenberger. 2004. "Human DNA Ligase I Completely Encircles and Partially Unwinds Nicked DNA." *Nature* 432(7016): 473–78.
- Patrick, Wayne M., and Jonathan M. Blackburn. 2005. "In Vitro Selection and Characterization of a Stable Subdomain of Phosphoribosylanthranilate Isomerase." *Federation of European Biochemical Societies Journal* 272(14): 3684–97.
- Pheiffer, Barbara H., and Steven B. Zimmerman. 1983. "Polymer-Stimulated Ligation: Enhanced Blunt- or Cohesive-End Ligation of DNA or Deoxyribooligonucleotides by T4 DNA Ligase in Polymer Solutions." *Nucleic Acids Research* 11(22): 7853–71.
- Quail, Michael A et al. 2008. "A Large Genome Centre's Improvements to the Illumina Sequencing System." *Nature Methods* 5(12): 1005–10. <http://www.illumina.com/pages.ilmn?ID=204> (February 11, 2019).
- Raae, Arnt J., Ruth K. Kleppe, and Kjell Kleppe. 1975. "Kinetics and Effect of Salts and Polyamines on T4 Polynucleotide Ligase." *European Journal of Biochemistry* 60(2): 437–43.
- Reddy Chichili, Vishnu Priyanka, Veerendra Kumar, and J. Sivaraman. 2013. "Linkers in the Structural Biology of Protein-Protein Interactions." *Protein Science* 22(2): 153–67.
- Rossi, Rossella, Alessandra Montecucco, Giovanni Ciarrocchi, and Giuseppe Biamonti. 1997. "Functional Characterization of the T4 DNA Ligase: A New Insight into the Mechanism of Action." *Nucleic Acid Research* 25(11): 2106–13.
- Schneider, Caroline A, Wayne S Rasband, and Kevin W Eliceiri. 2012. "NIH Image to ImageJ : 25 Years of Image Analysis." *Nature Methods* 9(7): 671–75.
- Sevier, Carolyn S., and Chris A. Kaiser. 2002. "Formation and Transfer of Disulphide Bonds in Living Cells." *Nature Reviews Molecular Cell Biology* 3(11): 836–47.
- Sgaramella, V, J H Van De Sande, and H G Khorana. 1970. "Studies on Polynucleotides, C. A Novel Joining Reaction Catalyzed by the T4-Polynucleotide Ligase." *Proceedings of the National Academy of Sciences* 67(3): 1468–75.
- Shuman, Stewart. 2009. "DNA Ligases: Progress and Prospects." *Journal of Biological Chemistry* 284(26): 17365–13769.
- Shuman, Stewart, and Christopher D. Lima. 2004. "The Polynucleotide Ligase and RNA Capping Enzyme Superfamily of Covalent Nucleotidyltransferases." *Current Opinion in Structural Biology* 14(6): 757–64.
- Shuman, Stewart, and Beate Schwer. 1995. "RNA Capping Enzyme and DNA Ligase: A Superfamily of Covalent Nucleotidyl Transferases." *Molecular Microbiology* 17(3): 405–20.
- Siebenlist, Ulrich, Guido Franzoso, and Keith Brown. 1994. "Structure, Regulation and Function of NF- κ B." *Annual Review of Cell Biology* 10: 405–55.
- Simpson, Richard J. 2010. "Stabilization of Proteins for Storage." *Cold Spring Harbor Protocols* 5(5).
- Sriskanda, V, and S Shuman. 1998. "Chlorella Virus DNA Ligase: Nick Recognition and Mutational Analysis." *Nucleic Acids Research* 26(2): 525–31.
- Sriskanda, Verl, Richard W. Moyer, and Stewart Shuman. 2001. "NAD⁺-Dependent DNA Ligase Encoded by a Eukaryotic Virus." *Journal of Biological Chemistry* 276(39): 36100–109.
- Sriskanda, Verl, and Stewart Shuman. 2002a. "Role of Nucleotidyl Transferase Motif V in Strand Joining by Chlorella Virus DNA Ligase." *Journal of Biological Chemistry* 277(12): 9661–67.
- Sriskanda, V, and S Shuman. 2002b. "Role of Nucleotidyltransferase Motifs I, III and IV in the Catalysis of Phosphodiester Bond Formation by Chlorella Virus DNA Ligase." *Nucleic Acids Research* 30(4): 903–11.
- Stevenson, Luke Jordan. 2014. "Development of a Novel Selection Strategy to Enable Directed Evolution of

DNA Ligase Enzymes." Victoria University of Wellington.

- Subramanya, Hosahalli S., Aidan J. Doherty, Stephen R. Ashford, and Dale B. Wigley. 1996. "Crystal Structure of an ATP-Dependent DNA Ligase from Bacteriophage T7." *Cell* 85(4): 607–15.
- Sugino, Akio et al. 1977. "Interaction of Bacteriophage T4 RNA and DNA Ligases in Joining of Duplex DNA at Base-Paired Ends." *Journal of Biological Chemistry* 252(11): 3987–94.
- Summer, Heike, René Grämer, and Peter Dröge. 2009. "Denaturing Urea Polyacrylamide Gel Electrophoresis (Urea PAGE)." *Journal of Visualized Experiments* (32).
- Tabor, Stanley. 2001. "DNA Ligases." In *Current Protocols in Molecular Biology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., Unit 3.14.1-3.14.4.
- Tanabe, Maiko et al. 2012. "Structure-Based Mutational Study of an Archaeal DNA Ligase towards Improvement of Ligation Activity." *ChemBioChem* 13(17): 2575–82.
- Tanabe, Maiko, Sonoko Ishino, Yoshizumi Ishino, and Hirokazu Nishida. 2014. "Mutations of Asp540 and the Domain-Connecting Residues Synergistically Enhance *Pyrococcus Furiosus* DNA Ligase Activity." *Federation of European Biochemical Societies Letters* 588(2): 230–35.
- Tanabe, Maiko, Yoshizumi Ishino, and Hirokazu Nishida. 2015. "From Structure-Function Analyses to Protein Engineering for Practical Applications of DNA Ligase." *Archaea* 2015: 10.1155/2015/267570.
- Terpe, K. 2003. "Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems." *Applied Biochemistry and Microbiology* 60: 523–33.
- Timson, David J, Martin R Singleton, and Dale B Wigley. 2000. "DNA Ligases in the Repair and Replication of DNA." *Mutation Research* 460: 301–18.
- Tubeleviciute, Agne, and Remigijus Skirgaila. 2010. "Compartmentalized Self-Replication (CSR) Selection of *Thermococcus Litoralis* Sh1B DNA Polymerase for Diminished Uracil Binding." *Protein Engineering, Design and Selection* 23(8): 589–97.
- Tukey, John. 1977. Pearson *Exploratory Data Analysis*.
- Urban, Manuela B, R Schreck, and Patrick A Baeuerle'. 1991. "NF-KB Contacts DNA by a Heterodimer of the P50 and P65 Subunit." *European Molecular Biology Organization Journal* 10(7): 1–817.
- de Vega, M. et al. 2010. "Improvement of Φ 29 DNA Polymerase Amplification Performance by Fusion of DNA Binding Motifs." *Proceedings of the National Academy of Sciences of the United States of America* 107(38): 16506–11.
- Wang, Yan et al. 2004. "A Novel Strategy to Engineer DNA Polymerases for Enhanced Processivity and Improved Performance in Vitro." *Nucleic Acids Research* 32(8): 1197–1207.
- Weiss, Bernard, and Charles C Richardson. 1967. "Enzymatic Breakage and Joining of Deoxyribonucleic Acid, I. Repair of Single-Stranded Breaks in DNA by an Enzyme System from *Escherichia Coli* Infected with T4 Bacteriophage." *Proceedings of the National Academy of Sciences of the United States of America* 57(4): 1021–1208.
- Weiss, Bernard, Ann Thompson, and Charles C Richardson. 1968. "Enzymatic Breakage and Joining of Deoxyribonucleic Acid." *Journal of Biological Chemistry* 243(17): 4556–63.
- Weller, Geoffrey R et al. 2002. "Identification of a DNA Nonhomologous End-Joining Complex in Bacteria." *Science* 297(5587): 1686–89.
- Wilkinson, A., J. Day, and R. Bowater. 2001. "Bacterial DNA Ligases." *Molecular Microbiology* 40(6): 1241–48.
- Williamson, Adele, Miriam Grgic, and Hanna-Kirsti S Leiros. 2018. "DNA Binding with a Minimal Scaffold: Structure-Function Analysis of Lig E DNA Ligases." *Nucleic Acids Research* 46(16): 8616–29.
- Williamson, Adele, Erik Hjerde, and Tim Kahlke. 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): 274–90.

Williamson, Adele, Ulli Rothweiler, and Hanna-Kirsti Schrøder Leiros. 2014. “Biological Crystallography Enzyme-Adenylate Structure of a Bacterial ATP-Dependent DNA Ligase with a Minimized DNA-Binding Surface.” *Acta Crystallographica* 70: 3043–56.

Wilson, Robert H. 2009. “Engineering a Chimeric T4 DNA Ligase for Enhanced Performance in Vitro.” Massey University.

Wilson, Robert. H. et al. 2013. “Engineered DNA Ligases with Improved Activities in Vitro.” *Protein Engineering, Design & Selection* 26(7): 471-478.

Xiao, Han, Zehua Bao, and Huimin Zhao. 2015. “High Throughput Screening and Selection Methods for Directed Enzyme Evolution.” *Industrial and Engineering Chemistry Research* 54(16): 4011–20.

Zhao, Huimin, and Frances H. Arnold. 1997. “Combinatorial Protein Design: Strategies for Screening Protein Libraries.” *Current Opinion in Structural Biology* 7(4): 480–85.

Zimmerman, S B, J W Little, C K Oshinsky, and M Gellert. 1967. “Enzymatic Joining of DNA Strands: A Novel Reaction of Diphosphopyridine Nucleotide.” *Proceedings of the National Academy of Sciences of the United States of America* 57: 1841–48.

Appendix

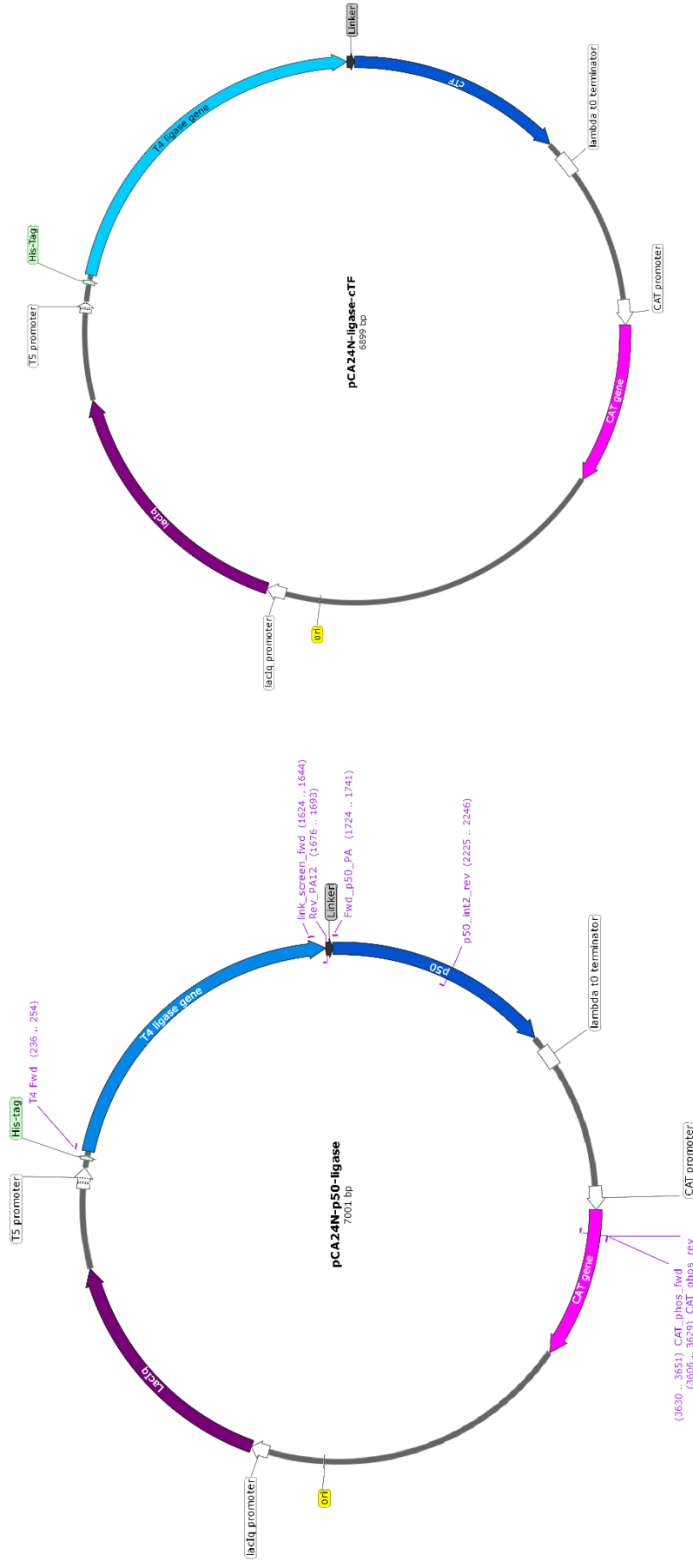
S1 Ligase-cTF Linker Library

Variant	Linker Sequence	Linker Length	Linker Type
CL1	GTPTGEF	7	Natural Linker
CL3	GTPTPTGEF	9	
CL5	GTPTPTPTGEF	11	
CL9	GTPTPTPTPTPTGEF	15	
CL11	GTPTPTPTPTPTPTPTGEF	17	
CL13	GTPTPTPTPTPTPTPTPTGEF	19	
CO2	GTSGGG	6	Original Linker
CO10	GTSGGGSGGGSGGGSGGG	18	
CHL5	EAAAKEAAAK	10	Rigid - Helical
CHL10	EAAAKEAAAKEAAAK	15	
CHY1	EAAAKAGSGGSGAEAAAK	18	
CPA5	PAPAP	5	Rigid – Non-Helical
CPA9	PAPAPAPAP	9	
CPA13	PAPAPAPAPAPAP	13	
CPA17	PAPAPAPAPAPAPAPAP	17	
CGSG3	GSG	3	Flexible
CGSG6	GSGGSG	6	
CGSG9	GSGGSGGSG	9	
CGSG12	GSGGSGGSGGSG	12	
CGSG15	GSGGSGGSGGSGGSG	15	
CGSE9	GSEGSEGEG	9	
CGSE15	GSEGEGGSEGSEGEG	15	
CGQ5	QGGQ	5	
CGQ9	QGGQGGQGGQ	9	
CGQ13	QGGQGGQGGQGGQ	13	
CGQ15	QGGQGGQGGQGGQGGQ	15	
CA2	AA	2	
CA3	AAA	3	
CA4	AAAA	4	
CA5	AAAAA	5	
CPSR3	PSR	3	Unknown
CPSR6	PSRGGG	6	
CPSR8	PSRGGGGG	8	

S2 Ligase-p50 Linker Library

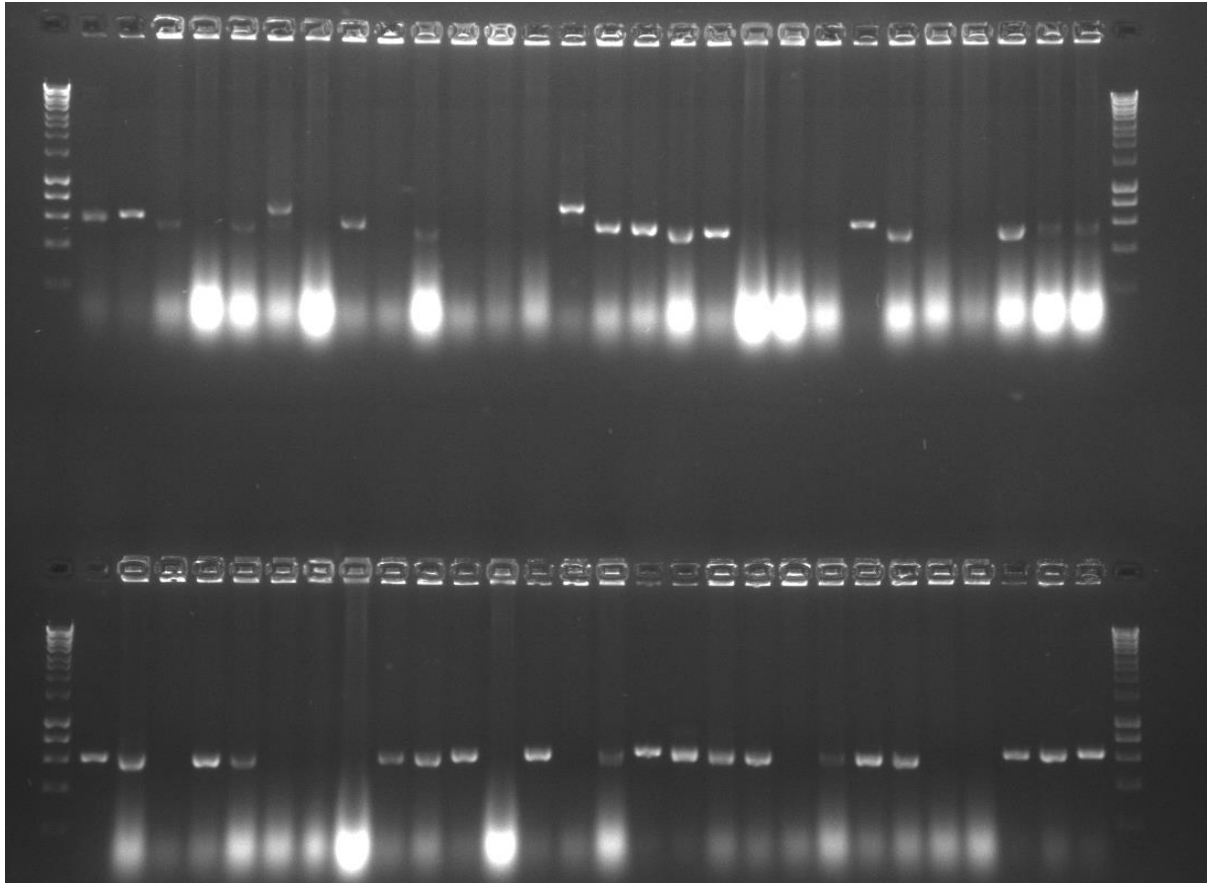
Variant	Linker Sequence	Linker Length	Linker Type	
PL1	GTPTGEF	7		
PL3	GTPTPTGEF	9		
PL5	GTPTPTPTGEF	11	Natural Linkers	
PL9	GTPTPTPTPTPTGEF	15		
PL11	GTPTPTPTPTPTPTGEF	17		
PL13	GTPTPTPTPTPTPTPTGEF	19		
Ligase-p50	GTSGGGSGGG	10		
PO2	GTSGGG	6	Original Linker	
PO10	GTSGGGSGGGSGGGSGGG	18		
PO30	GTSGGGSGGGSGGGSGGGSGGGSGGGSGGG	30		
PHL5	EAAAKEAAAK	10		Rigid – Helical
PHL5_a	EAAAEAAAK	9		
PHL10	EAAAKEAAAKEAAAK	15		
PHY1	EAAAKAGSGGSGAEAAAK	18		
PPA5	PAPAP	5	Rigid – Non-Helical	
PPA9	PAPAPAPAP	9		
PPA13	PAPAPAPAPAPAP	13		
PPA17	PAPAPAPAPAPAPAPAP	17		
PGSG3	GSG	3		Flexible
PGSG6	GSGGSG	6		
PGSG9	GSGGSGGSG	9		
PGSG12	GSGGSGGSGGSG	12		
PGSG15	GSGGSGGSGGSGGSG	15		
PGSG18	GSGGSGGSGGSGGSGGSG	18		
PGSE9	GSEGSEGEG	9		
PGSE15	GSEGGSEGSEGEG	15		
PGSE18	GSEGSEGGSEGSEGEG	18		
PGQ5	QQGQ	5		
PGQ9	QQGQGGGQ	9		
PGQ13	QQGQGGGQGGGQ	13		
PGQ15	QQGQGGGQGGGQGGGQ	15		
PGQ19	QQGQGGGQGGGQGGGQGGGQ	19		
PA2	AA	2		
PA3	AAA	3		
PA4	AAAA	4		
PPSR3	PSR	3	Unknown	
PPSR6	PSRGGG	6		
PPSR8	PSRGGGGG	8		

S3 Plasmids maps for ligase-CTF and ligase-p50.
















The ligase-p50 plasmid/s were used in the competitive ligase selection for Round 6. This also contains the locations of the primers used in this study. Both plasmids were used for the expression and purification of T4 DNA ligase without a DNA-binding domain and the engineered DNA ligase variants

S4 Example of a Round 6 Diagnostic Linker PCR Gel



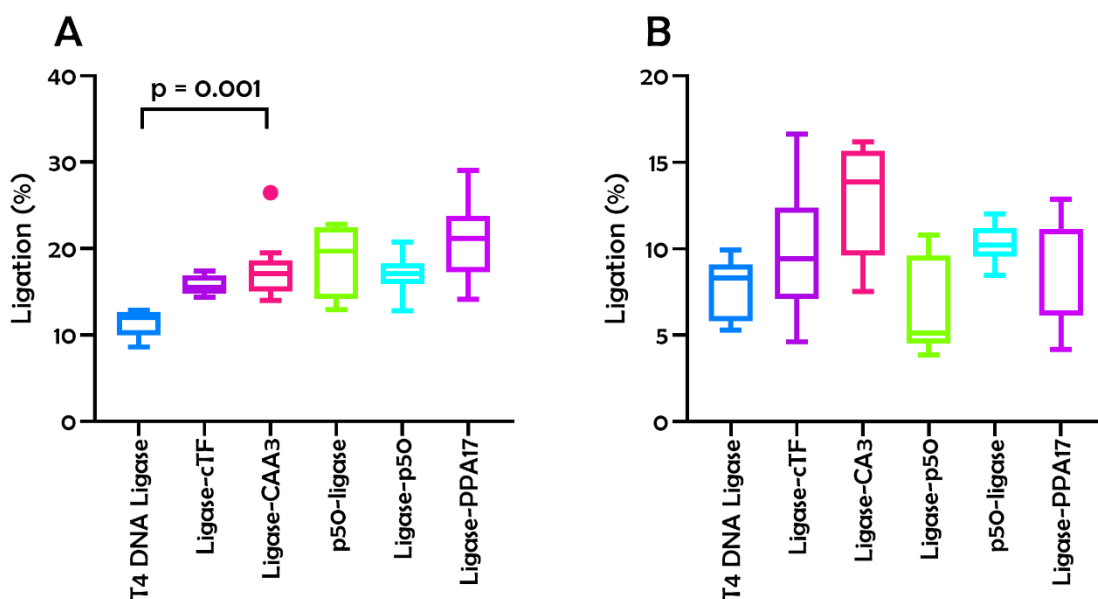
S5 Oligonucleotide analysis results from IDT for the 40 nt oligonucleotide product

Structures

structure	Image	ΔG (kcal.mole ⁻¹)	T _m (°C)	ΔH (kcal.mole ⁻¹)	ΔS (cal.K ⁻¹ .mole ⁻¹)	Output
1		-7.61	58.2	-76	-229.37	Ct Det
2		-7.59	57.6	-77.1	-233.13	Ct Det
3		-6.03	50.6	-76.3	-235.68	Ct Det
4		-5.7	49.8	-74.2	-229.75	Ct Det
5		-5.43	45.6	-84.2	-264.18	Ct Det
6		-5.09	43.8	-86	-271.38	Ct Det
7		-5.06	43.5	-86.7	-273.81	Ct Det
8		-4.87	47.7	-68.8	-214.44	Ct Det
9		-4.69	47.8	-66	-205.65	Ct Det
10		-4.66	49.8	-60.6	-187.62	Ct Det
11		-4.35	42.3	-79.2	-251.04	Ct Det
12		-4.34	52	-52.3	-160.84	Ct Det
13		-4.33	54.3	-48.3	-147.49	Ct Det

Following the oligonucleotide separation on a 20% denaturing TBE-PAGE in Section 4.3, Figure 4-4, the product oligonucleotide sequence was inputted to the IDT oligo analyser tool to determine whether the 40 nt product could form hairpins above 40°C. Thirteen possible hairpin structures were predicted.

S6 Raw data of blunt-end (A) and cohesive-end (B) ligation data from Figure 4-14



Data analysis of ligation activity of engineered DNA ligases displayed a possible outlier for ligase-CA3 for blunt-end ligation. To determine whether the potential outlier would result in a statistically significant change, a Tukey's test was performed in GraphPad Prism 8 on the raw dataset and the results were displayed in a box and whisker plot. For cohesive-end ligation (Panel B), there were no outliers produced. In comparison, for blunt-end ligation (Panel A), one of the nine replicates were identified as an outlier, therefore removed from the final analysis in Figure 4-14. A one-way ANOVA with Tukey's multiple comparisons to determine statistical significance for ligase-CA3 over T4 DNA ligase ($p = 0.001$).