ABSTRACT

Title of Dissertation: USING DNA LOOPING PROTEINS TO ENHANCE HOMOLOGY DIRECTED REPAIR IN VIVO FOLLOWING A CAS9 INDUCED DOUBLE STRAND BREAK

Ian Theodore Ferencz, Doctor of Philosophy, 2024

Dissertation directed by: Associate Professor Jason D. Kahn, Department of Chemistry and Biochemistry

Genome engineering methods that start with a CRISPR/Cas9 targeted genomic DNA double strand break proceed through cellular DNA repair mechanisms after the induction of the break. Imprecise nonhomologous end-joining (NHEJ) is useful for knockouts, but precise homology-directed repair (HDR) is necessary for gain of function changes. NHEJ tends to be more efficient, so directing the cell to knock in a precise sequence via HDR is an active area of research. The system we have designed uses a bivalent protein to recruit HDR donor DNA to the site of a specific DNA double strand break induced by a Cas9/sgRNA nuclease. Previously described leucine zipper dual-binding (LZD) proteins are

used because they are small and stable. The system was designed to reduce the effort needed for screening, shorten the time required for the repair process, and/or decrease the amount of donor DNA needed, reducing potential off-target effects. We developed a model system in *Saccharomyces cerevisiae* to measure gene disruption and HDR frequencies in yeast that contain combinations of non-replicating donor DNA plasmid or linear DNA, expression plasmids of four LZD variants, and a plasmid expressing Cas9 and an sgRNA targeting either the AGC1 or ADE2 genes. The donor DNA includes a gene coding for G418 resistance in yeast. It also includes an INV-2 site recognized by the C-terminal DNA binding domain of LZDs adjacent to suboptimal regions of homology to the target gene. The N-terminal DNA binding domain of the LZDs recognizes an endogenous CREB site near the target gene. The desired recombinants are scored by their inability to grow on acetate as a sole carbon source (for AGC1) or their red color (ADE2), accompanied by resistance to G418. We believe that LZD enhancement can become a simple and valuable adjunct to any other method of improving the efficiency of HDR, in any system. We were able to show that the inclusion of LZD73 in recombination experiments increased the number of colonies presenting with the desired phenotype and genotype nearly eight-fold in the absence of a designed DNA break. We also provide evidence suggesting that the presence of LZD73 has a slight positive effect on the efficiency of Cas9 targeting. Desired recombinants were recovered after Cas9/sgRNA cleavage in an experiment where there was no apparent recombination in the absence of LZD73. Future work on this project includes optimization of the homologous sequences to improve background recombination so a more quantitative measure of the improvement observed in the presence of LZD proteins. This work can be transferred laterally to enhance other recombination-based methods in other organisms: the LZD proteins could be analogous to an adjuvant that increases overall efficiency.

USING DNA LOOPING PROTEINS TO ENHANCE HOMOLOGY DIRECTED REPAIR IN VIVO FOLLOWING A CAS9 INDUCED DOUBLE STRAND BREAK

by

Ian Theodore Ferencz

Dissertation submitted to the Faculty of the Graduate School of the

University of Maryland, College Park in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

2024

Advisory Committee:

Associate Professor Jason D. Kahn, Chair/Advisor

Associate Professor Emeritus Douglas Julin

Associate Professor Paul Paukstelis

Associate Professor Myles Poulin

Professor Jonathan Dinman, Dean's Representative

Dedication

To my parents,

my siblings,

and Pongo,

I could not have made it here without all of your love, support, and inspiration.

To Samantha,

You provided the daily motivation I needed to get through the hardest of times. Without your constant encouragement I may not have made it here today.

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Jason Kahn for allowing me to pursue my degree in his lab and providing all the help I needed along the way.

I am grateful to my committee, not only the members who are written on this document but also those who have helped in the past and moved on from the university to pursue their own path.

I am grateful to Dr. Jonathan Dinman and Dr. Alexandra Olson who allowed me use of their space and taught me the ways of yeast genetics so I could continue this project.

I would not be here without the help of Emma Neiswender, who put in more hours than was expected of her to make sure the yeast were always being looked after.

To Jackie and Jessica who worked on this project before I joined UMD I say thanks for laying the groundwork and developing an idea that could become reality.

I would like to thank the entire Poulin Lab who I have shared a floor with all these years, not only for the advice and scientific support, but for casual conversation and friendship.

To Dr. Jason Hustedt, I will always be thankful for you being brutally honest and helping me make a smooth transition back into academia, and for being there any time of day or night even after you graduated.

I would like to send special thanks to every administrative and staff member in the department that I have dealt with. This list is extensive but above all I'd like to thank Dolores, Carl, Renee, Cathy, Jerry, Vera, and Tyrik. You have always been more than helpful, and I will remember it forever.

I am also grateful to have been able to teach for some of the faculty members throughout my time here. Specifically, I would like to thank Dr. Lee Friedman for always having my back and putting so much trust in me with shaping the minds of tomorrow.

I would like to thank Dr. Paul Nix of Princeton Separations Inc., who provided an ElectroSEP kit to help with troubling DNA purifications.

I am also very thankful to the Graduate School at UMD for providing funding opportunities throughout my time here.

Finally, I would like to thank all my family and friends who have had to put up with me talking about science and school for as long as I can remember.

Table of Contents

List of Tables

List of Figures

Commonly Used Terms and Abbreviations

2μ – Yeast 2-micron Plasmid

ADE2 – Gene coding Phosphoribosylaminoimidazole Carboxylase, Involved in Adenine Biosynthesis

ade2-10 – Example of a knockout in the ADE2 gene

AFM – Atomic Force Microscopy

AGC1 – Gene coding Aspartate-Glutamate Carrier 1 Protein

Agc1p – Aspartate-Glutamate Carrier 1 Protein

AIR – 5'-Phosphoribosyl-5-Aminoimidazole

ampR – Gene coding Ampicillin Resistance

AP-1 – Activator Protein 1

ARP8 – Actin Related Protein 8

ATP – Adenosine Triphosphate

AUC – Analytical Ultra Centrifugation

BamHI – Restriction Enzyme from NEB

BCL11A – Gene coding B-Cell Lymphoma/Leukemia 11A Protein

BclI – Restriction Enzyme from NEB

bZip – Basic Leucine Zipper Domain

CAG repeat – Repeat of the trinucleotide sequence Cytosine, Adenine, Guanine

CAIR – 5'-Phosphoriboyl-4-Carboxy-5-Aminoimidazole

Cas – CRISPR Associated

CD – Circular Dichroism

CEN/ARS – Centromere/ Autonomously Replicating Sequence

CEN4 – Centromere Sequence from Chromosome IV

CKY8 – commonly used laboratory yeast strain

CMV - Cytomegalovirus

CREB – cAMP Response Element Binding Protein

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

crRNA – CRISPR RNA

dam – DNA Adenine Methylase

DBD – DNA Binding Domain

DMD – Duchenne's Muscular Dystrophy

DNA – Deoxyribonucleic Acid

dNTPs – Deoxynucleotide Triphosphates

DSB – Double Strand Break

EDTA – Ethylenediaminetetraacetic Acid

FPLC – Fast Performance/ Protein Liquid Chromatography

G418 – Geneticin

GAP promoter – General Amino Acid Permease Promoter sequence

GCN4 – Gene coding General Control Nonrepressible 4 Protein

GFP – Green Fluorescent Protein

gRNA/ sgRNA – Guide/ Short Guide RNA

HD – Huntington's Disease

HDR – Homology Directed Repair

HDV – Hepatitis D Virus

his3-Δ200 – Example of a knockout in the HIS3 gene

HJ – Holliday Junction

HR – Homologous Recombination

IDT – Integrated DNA Technologies

IMAC – Immobilized Metal Affinity Chromatography

InDels – Insertions and/or Deletions

INV-2 – Inverted CREB DNA binding sequence

KanMX Cassette – Sequence including promoter and terminator for the *kanR* gene

kanR – Gene coding Kanamycin or Geneticin Resistance

LEU2 – Gene coding ß-Isopropylmalate Dehydrogenase, Involved in Leucine Biosynthesis

leu2-3,112 – Example of a knockout in the LEU2 gene

LiOAc – Lithium Acetate

LZD – Leucine Zipper Dual DNA Binding Protein

MAGE – Multiplex Automated Genome Engineering

MATa/MATα – *S. cerevisiae* Mating Type

MBP – Maltose Binding Protein

MET25 Promoter – O-Acetyl Homoserine-O-Acetyl Serine Sulfhydrylase Promoter Sequence

NEB – New England Biolabs

NHEJ – Non-Homologous End Joining

PAA – bZip Fragment of the GCN4 protein

PAM – Protospacer Adjacent Motif

pCM173 or pCM173::eV – shuttle vector used for regulated *in vivo* protein expression in yeast

PCR – Polymerase Chain Reaction

PEG – Polyethylene Glycol

PEI - Polyethyleneimine

pEmpty-Control – shuttle vector used for *in vivo* Cas9 expression in yeast

pG2 – shuttle vector used for *in vivo* Cas9 and HDV ribozyme-sgRNA expression in yeast

pMal-c2T – bacterial vector used for protein purification of LZD proteins

pML107 – shuttle vector used for *in vivo* Cas9 and sgRNA expression in yeast

PMSF – Phenylmethylsulfonyl Fluoride

PNK – Polynucleotide Kinase

pRK793 – bacterial vector used for protein purification of TEV protease

PSI – Princeton Separation Inc.

pUC57 – bacterial vector used for common cloning experiments

RNA – Ribonucleic Acid

RNP – Ribonuclear Protein

rSAP – Recombinant Shrimp Alkaline Phosphatase

RVD – Repeat Variable Diresidue

SbfI – Restriction Enzyme from NEB

SC – Synthetic Complete media

SCD – Sickle Cell Disease

SD-Amino Acid – Synthetic Dropout Media, the amino acid(s) listed after are not included in the formulation of that media for auxotrophic control

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SNP – Single Nucleotide Polymorphism

SNR52 – Small Nucleolar RNA 52

SOC – Super Optimal Broth + Glucose

SV40 NLS – Simian Virus 40 Nuclear Localization Sequence

SwaI – Restriction Enzyme from NEB

T4 – Escherichia Virus T4 bacteriophage

TAL – Transcription Activator Like

TALEN – Transcription Activator Like Effector Nuclease

TBE – Tris Borate EDTA buffer

TEF – Translation Elongation Factor

tetO – Tetracycline Operon

TetOff – Tetracycline Off Expression System

tetR – Gene coding Tetracycline Resistance

TEV – Tobacco Etch Virus

tracrRNA – Transactivating CRISPR RNA

$tRNA^{tyr}$ – Tyrosine specific Transfer RNA

TRP1 – Gene coding Phosphoribosylanthranilate Isomerase, Involved in Tryptophan Biosynthesis

trp1-Δ901 – Example of a knockout in the TRP1 gene

tTA – Tetracycline Transactivating Protein

URA3 – Gene coding Orotidine-5'-Phosphate Decarboxylase, Involved in Pyrimidine Biosynthesis

ura3-52 – Example of a knockout in the URA3 gene

UTR – Untranslated Region

UV – Ultraviolet

VP16 – α Trans-Inducing Protein

yEGFP – Yeast Enhanced Green Fluorescent Protein

yJD100 – yeast strain obtained from Dr. Dinman

yJD1112 – yeast strain obtained from Dr. Dinman

yJD52 – yeast strain obtained from Dr. Dinman

YMC2 – Gene coding Yeast Mitochondrial Carrier 2 Protein

YP(A)D – media containing Yeast extract, Peptone, (Adenine sulfate), and Dextrose

YPAA – media containing Yeast extract, Peptone, Adenine sulfate, and Acetate

YPAO – media containing Yeast extract, Peptone, Adenine sulfate, and Oleate

ZFN – Zinc Finger Nuclease

Chapter 1: Introduction

Section 1.1: DNA as the Source for Genetic Information

When DNA was discovered to be the basis for genetic information, it opened the door for scientists to begin understanding how changes in DNA can lead to disease and genetic conditions. The structure of DNA lends itself to containing large amounts of information in a relatively small package, which by means of transcription and translation becomes RNA and finally protein[1]. Transcription of DNA to RNA takes place in a one-to-one manner where RNA polymerases use the antisense strand of DNA as a template to create a copy of the sense strand. The only sequence difference between the two strands is that DNA uses thymine while RNA uses uracil. This RNA molecule then undergoes translation in the ribosome in a three-to-one manner, where every three bases of RNA, also called a codon, codes for a single amino acid. There are a total of 64 possible codons coming from the different combinations of bases and these code for 20 different amino acids. This discrepancy in the codons and amino acids is the reason for degeneracy in the code, whereby multiple codons code for the same amino acids[2]. Additionally, stop codons that signal the end of translation do not code for an amino acid. From all this it is clear that when changes happen to the base DNA molecule it can have wide ranging downstream effects.

Not all changes are negative, because not all bases in DNA are coding bases or otherwise important to genome function. Noncoding DNA refers to everything else in the genome that does not eventually code for protein, whether that is regulatory sequence DNA, DNA that is transcribed to make non-coding RNA, or nonfunctional DNA. The types of change that DNA can undergo also can affect how much of an effect it will have downstream. Due to the redundant nature of codons, small changes or single nucleotide

polymorphisms (SNP) may not change a protein's amino acid composition, and if it does the effect could be localized to a single amino acid. In nonfunctional non-coding regions of DNA, the impact of an SNP is fairly low. The removal or addition of nucleotides – not in multiples of three – in a sequence can lead to frame shifting mutations which in most cases leads to a complete change of the codons downstream of the affected change, and therefore much greater change to the coded protein. These changes are also impactful when they appear in noncoding areas of DNA, as they could affect promoters or splice sites which will affect nearby functional or coding sequences.

Three prevalent genetic diseases caused by mutations, or alterations to the genome, are Duchenne Muscular Dystrophy (DMD), Huntington's Disease (HD), and Sickle Cell Disease (SCD), each of which is caused by a slightly different type of DNA change. DMD is caused primarily by large insertions or deletions in the dystrophin gene leading to downstream frameshifting and loss of protein production[3]. Decreased levels of the dystrophin protein eventually lead to muscle weakness and respiratory failure. HD is a neurodegenerative disorder caused by the insertion of many copies of the trinucleotide sequence cytosine, adenine, guanine (CAG) in the huntingtin gene[4]. This triplet repeat expansion is a common feature seen in many other genetic diseases. The mutant huntingtin protein has many negative effects in the brain, eventually leading to decreased motor function and death. SCD is one of the most abundant genetic disorders in the world and affects over 5 million people worldwide. The cause of SCD is a SNP that changes a single adenine to a thymine, leading to a glutamate to valine mutation in position 6 of the ß-globin protein[5,6]. This mutant hemoglobin polymerizes and causes the red blood cell to assume a sickling shape, which leads to multifarious health problems for the carrier. These

conditions as well as many other genetic diseases do not have cures; the primary course of action for affected individuals is management of symptoms.

Section 1.2: Genome Engineering

By the 1970s scientists had already begun work on modifying proteins in living organisms from bacteria to higher level eukaryotes by way of genetic modification[7,8]. This change relates back to the central dogma in that changing the DNA of an organism leads to changes in the proteins expressed in that system. The idea of genetic engineering originally came to be as a way to express non-native proteins in prokaryotes for *in vitro* assays[9]. In 1973 Stanley Norman Cohen and Herbert Boyer introduced an antibiotic resistance gene to a DNA plasmid cut with restriction enzymes, then transformed it into *E. coli*. The resulting bacteria were able to live in the presence of that antibiotic, becoming one of the first instances of genetic engineering, and one that is still heavily used in the laboratory today. Moving into genome engineering this work was taken further with the introduction of recombineering using the Lambda Red system[10] and multiplexed automated genome engineering (MAGE)[11], for the introduction of exogenous genes through recombination of homologous sequences in bacterial systems.

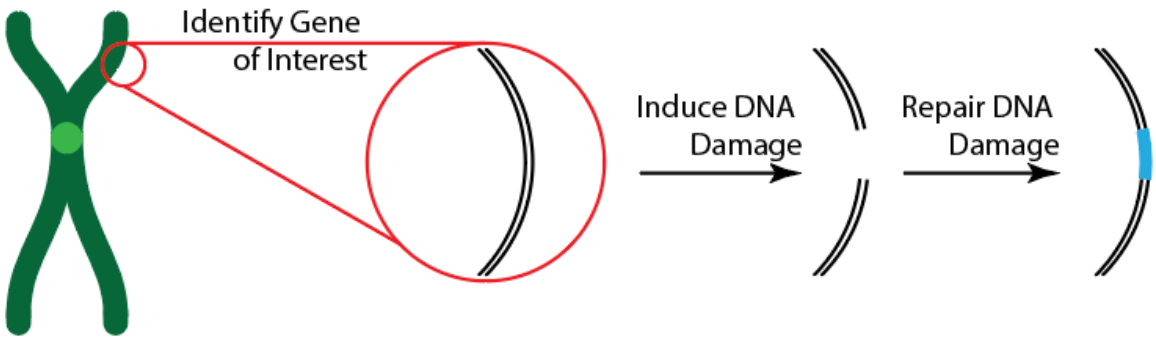

Figure 1.1 Genome Engineering in Practice: Genes in the chromosome are first identified for possible targeting. Then they are specifically cleaved to induce DNA double strand breaks that are later repaired to modify function.

In order to break down what genome engineering is more specifically, it needs to be understood what the steps of the process are and what the end goal is. As in Figure 1.1, initially a gene is targeted—this could be a mutant gene locus to be reverted to wild type, for example—and then the best way to alter that DNA has to be determined [7]. The most straightforward way is to induce a DNA double strand break (DSB). After the DNA has been broken, DNA repair pathways can be exploited to fix the DNA either though precise methods, like homology directed repair (HDR), or imprecise methods like nonhomologous end joining (NHEJ). The repaired gene now has a new function or has lost its function, based on the method chosen.

In the Cohen and Boyer experiment, they used restriction enzymes that are specific to 4-6 bp regions of DNA and cut every copy of that target sequence with high efficiency. In organisms with larger genomes that have many copies of that small target sequence, different enzyme systems are required to induce DSBs. These enzymes are separated into four main groups: meganucleases, zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), and CRISPR (clustered regularly interspaced short palindromic repeats) associated nucleases.

Meganucleases are similar to restriction enzymes in that they are mostly naturally found, but instead of short target sequences, these recognize and cut anywhere from 18-40 base sequences[12]. The chance of a random 18 base sequence showing up multiple times is very low and thus these enzymes are very specific. These enzymes are commonly used for recombination experiments when the sequence being targeted is near to or overlapping a naturally occurring meganuclease restriction site. ZFNs and TALENs are both engineered nucleases that combine a sequence specific DNA recognition motif fused to a non-specific catalytic domain to cleave DNA. The most commonly used catalytic domain of these

nucleases is the nuclease domain of the FokI restriction enzyme, which is active only as a dimer, not as a monomer. In ZFNs the DNA recognition domain is comprised of three to six zinc finger repeats, small protein motifs that each recognize three-nucleotide sequences[13,14]. Work done by Nikola Pavletich and Carl Pabo in 1991, and later by Carlos Barbas in the early 2000s, introduced the basis for generating all 64 iterations of trinucleotide sequences which could then be used for generation of ZFNs to target any DNA sequence. TALENs have a DNA binding domain that is based on the *Xanthomonas* bacteria TAL effectors. The proteins contain a series of repeat variable diresidue (RVD) regions that each recognize a specific nucleic acid base[15–17]. Combinations of these RVD sequences provides to a programmable TALEN that has high specificity for a specific DNA sequence. Both ZFNs and TALENs require two engineered DNA binding domains (DBD) so as to recognize both strands of the target DSB cut site. The two bound DBDs allow assembly of a complete, dimeric (and therefore active) FokI nuclease targeting both DNA strands. Figure 1.2 depicts the differences between these two nucleases in addition to the CRISPR system.

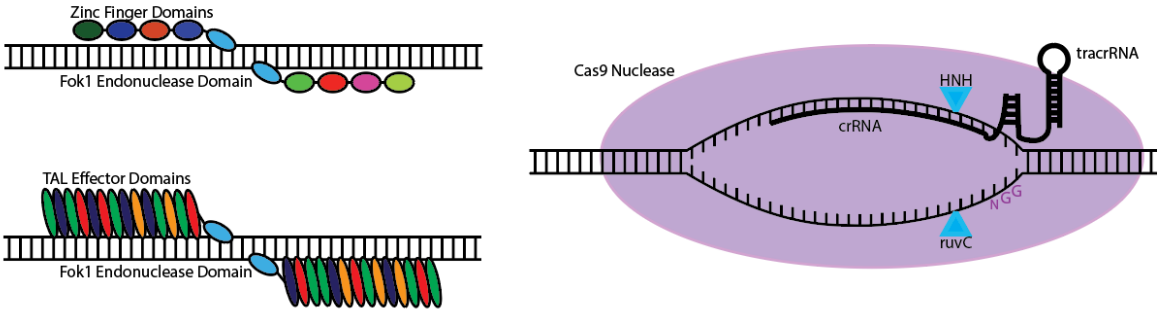

Figure 1.2 ZFNs, TALENs, and CRISPR: The major operational difference among the nucleases is the recognition of nucleic acids either through zinc-finger domains, TAL effector domains, or crRNA.

Section 1.2.1: CRISPR

The latest method being used to initiate targeted DSBs is the use of a CRISPR/ CRISPR associated (Cas) system. Discovered in the 1980s, CRISPR sequences in bacteria

were found to be associated with a form of immune response against phages[18,19]. Upon initial infection with phage, Cas proteins – which are polymerases, helicases, and nucleases – process the phage genome and integrate short sequences from it into the bacterial genome. These segments are used as recognition elements to fight the phage if it invades again. Along with the Cas proteins, two different RNA sequences are required for the system, the CRISPR RNA (crRNA) and the transactivating CRISPR RNA (tracrRNA). The crRNA, generated from transcription and processing of the sequence that contains bits of integrated phage genome, contains the genetic information necessary for targeting the nuclease to the phage genome through Watson-Crick base pairing. The tracrRNA hybridizes to the crRNA and also acts as a scaffold to complex with Cas9, as in Figure 1.3.

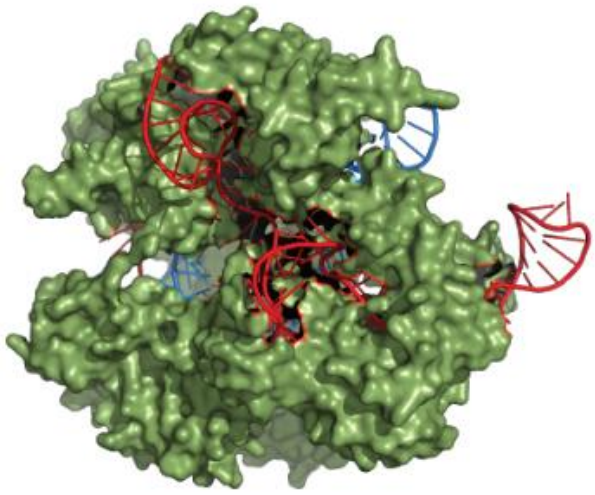

Figure 1.3: The Cas9 + sgRNA Complex: The Cas9 protein is depicted in green, sgRNA is depicted in red, and the target sequence is depicted in blue. Molecules rendered in PyMol from the crystal structure 7OX8[20].

Another important aspect of the CRISPR process is the Protospacer Adjacent Motif (PAM) sequence, a three to five base pair region in the phage (or other target) genome that acts as a necessary part of the signal for binding of Cas9. The PAM sequence is not present

in the CRISPR repeat region, so the Cas nucleases do not target the host genome. PAM sequences are also specific to the nucleases being used. For example, the most commonly used Cas9 nuclease, from *Streptococcus pyogenes,* uses the PAM sequence NGG. (Here N stands for any of the 4 nucleotides.)

The CRISPR system has been streamlined to fit the needs of genome engineering – for which Jennifer Doudna and Emmanuel Charpentier won the Nobel Prize in Chemistry in 2020 – requiring only the Cas9 nuclease and a synthetic guide RNA (gRNA or sgRNA) as shown in Figure 1.4. The sgRNA is a fusion between the tracrRNA and a short 20 base sequence that functions like the crRNA, meaning the sgRNA can both complex directly with the Cas9 and also target the nuclease to a genomic site. This makes experimental design based on Cas9 much more adaptable than designs using ZFNs or TALENs, since the sgRNA is independent of the nuclease protein and can be swapped out for another sequence without the complication of making a new fusion protein. Aside from RNA stability and secondary structure, the only limitation to sequence design is the proximity to a PAM sequence in the genomic DNA.

Two other methods that have been developed for making small changes in the genome using CRISPR tools are base editing and prime editing. Base editing uses nucleoside deaminases fused to nCas9 (a modified Cas9 that can only induce a single strand break) that can make single base changes in the genome. This is a very useful tool for scientists to be able to make changes without the need for inducing full double strand breaks[21]. There are limits to the technology though, as not all bases can be modified into all other bases. Prime editing uses the same nCas9 fused to a reverse transcriptase domain and a prime editing guide RNA. This tool expands on the usage of the CRISPR toolbox facilitates more substantial modifications than base editors[22].

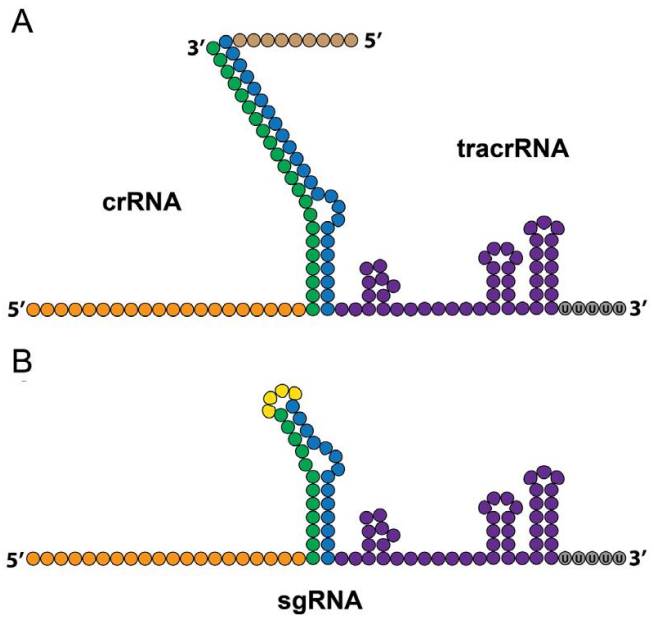

Figure 1.4 crRNA + tracrRNA (top) vs. sgRNA (bottom): (A) In the natural system, the two RNA molecules are held together with base pairing interactions only. (B) The engineered RNA loop (in yellow) allows the entire RNA to be expressed as a single molecule. (Figure from Nowak 2016)[23]

Section 1.2.2: Repairing DNA Double Strand Breaks

An overabundance of DNA damage will lead to cell death, and thus DNA repair is essential not only to successful genome editing, but also to cell survival. The two primary pathways that are used to repair DSBs in DNA are NHEJ and HDR[24], shown in Figure 1.5. These repair pathways are innate in most forms of life and compete with each other whenever DNA damage occurs. NHEJ is thought of as the quicker and more readily available method of repair, as it occurs without a homologous template to copy, and it can occur anytime throughout the cell cycle (though primarily in G1 when there is no homologous template present)[25]. In the NHEJ mechanism, the cut ends of DNA are first bound by the Ku70/80 heterodimer and are stabilized by chaperone proteins before processing by a number of enzymes before the ends are ligated by DNA Ligase IV, which can join mismatched DNA or DNA that includes gaps[26,27]. This leads to a heterogeneous

mixture of products, which is why NHEJ is referred to as error prone and is the desired pathway when looking to knockout or silence a gene, as the mechanism tends to lead to indels (insertion-deletion events), causing gene disruption. NHEJ is not useful for genome engineering experiments where a unique product is desired or where a new sequence is to be introduced.

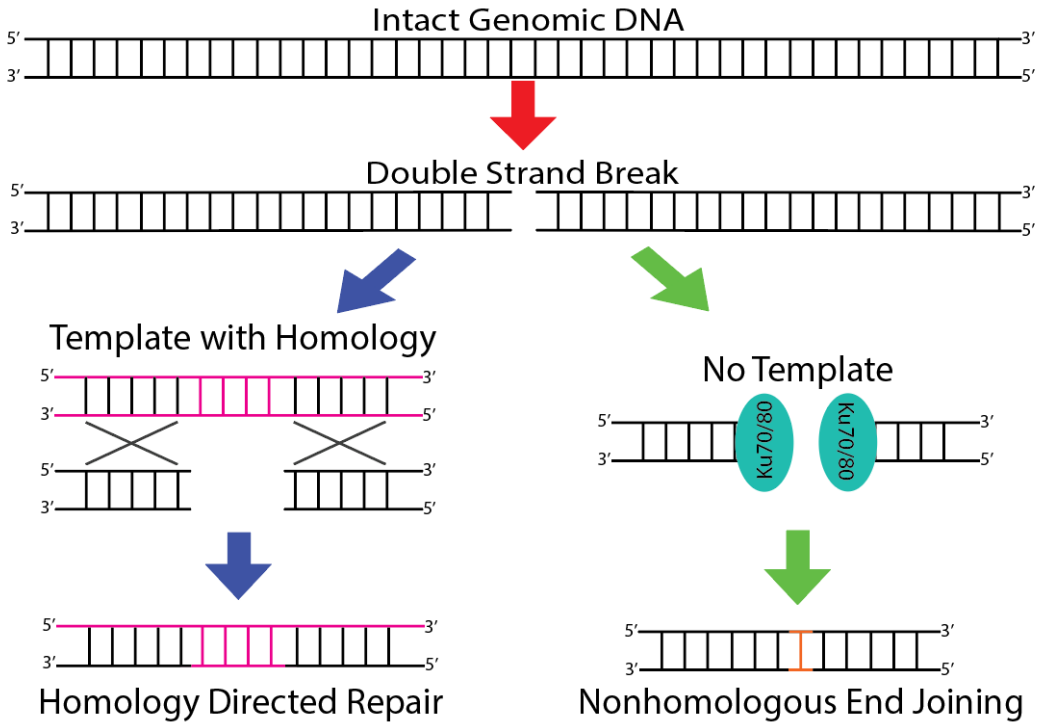

Figure 1.5 Repair Pathways HDR vs. NHEJ: After a DSB the DNA can be repaired by HDR with a homologous template leading to transfer of the genetic material in the donor template or repaired without a template leading to a loss of gene function through NHEJ.

HDR or homologous repair (HR) uses a segment of homologous DNA as a template to copy cross the region of the double strand break. This homology between the intact DNA donor and the end of the damaged DNA innately allows the cell to repair the broken strand with an identical sequence to the previous content, but it can also replace the central segment with a different allele or an entirely different sequence. Sister chromatids are the typical endogenous homologous template. They are most available during the S and G2 cell

cycles[28,29]. I describe the process with the yeast protein names, but all systems have corresponding enzymatic activities.

The HDR process relies on resection of the 5' DNA ends by the MRX complex as a checkpoint to signal that NHEJ can no longer occur[30] (Figure 1.6B). The free 3' DNA ends are then bound by RAD51 and used to invade the intact template, creating a Holliday junction (HJ)[31,32]. Polymerase and other repair enzymes extend the 3' DNA end until a resolution of the Holliday junction occurs. The strands can then complete the repair when the gaps are filled by more repair enzymes. This is the simplest form of homologous recombination and does not include a crossover event. If the first Holliday junction is not resolved rapidly and the second end 3' end of the resected DNA is captured, this will lead to the formation of a second Holliday junction. Repair enzymes will fill in the missing DNA and upon completion the two junctions will be resolved leading to a crossover event. The crossover just indicates that the source of the newly repaired DNA comes from both molecules instead of just one.

In genome engineering the HDR mechanism can be used to introduce long sequences of DNA to allow for gain of function mutations, and small sequences can be used as template for replacement of DNA containing a mutation, which could then lead to a return to wild type function. The addition of a homologous donor DNA is required for HDR methods, but due to partial homology, there is the potential that donor DNA can integrate at genome locations that are not the desired target. Off-target editing events are a major concern for all genome editing platforms; we assume that enhancing the efficiency of on-target integration could allow for decreasing the amount of donor DNA needed and thereby decrease off-target integration.

The choice of repair pathway in organisms is not one size fits all. Yeast like *Saccharomyces cerevisiae* tend to prefer HDR over NHEJ, though current research shows that yeasts are quite adept at performing NHEJ[33], whereas it was previously thought that it was a rare process. On the other hand, humans and most other organisms tend to perform NHEJ much more frequently than HDR. This is due to the previously mentioned factors such as the relative simplicity of NHEJ and the ability for it to occur during all parts of the cell cycle.

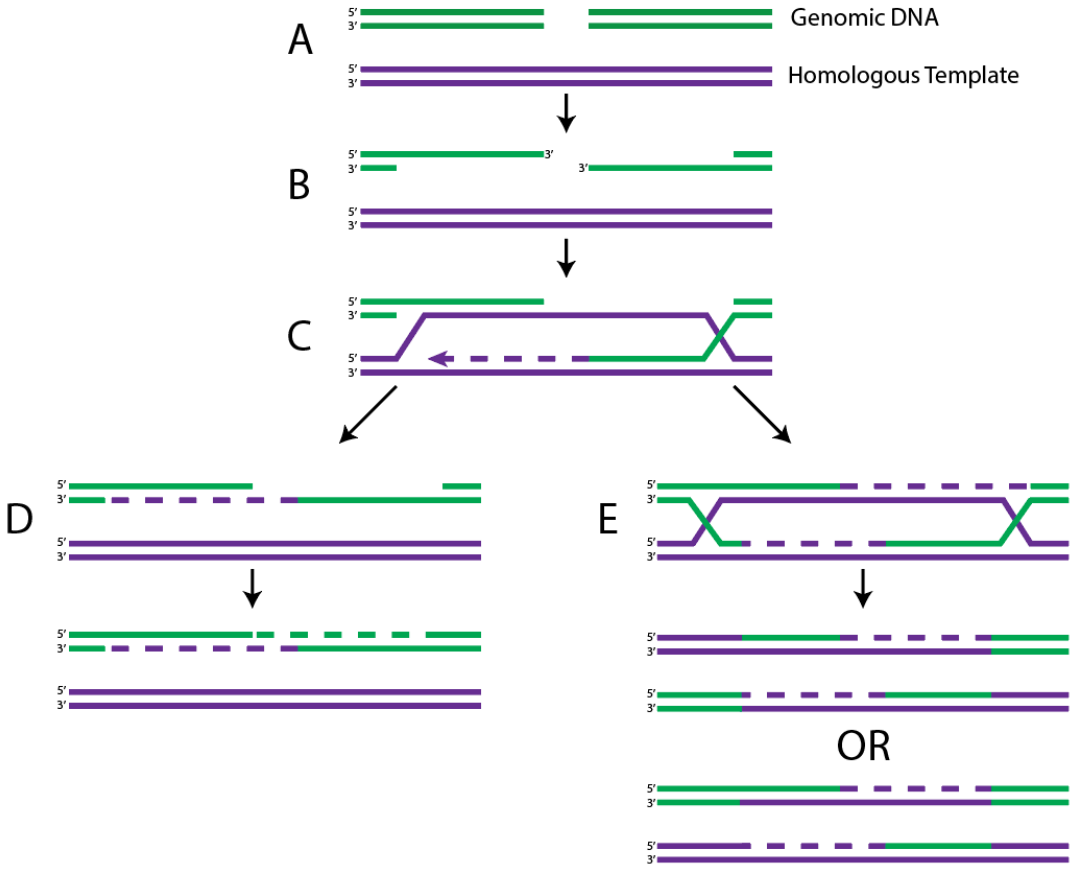

Figure 1.6 Simplified HDR Pathway: Genomic DNA is shown in green with the homologous template in purple and newly synthesized DNA with dashed lines. (A) A DNA double strand break in the genome. (B) End resection leading to exposed 3' DNA ends. (C) Strand invasion into the homologous template and formation of a Holliday junction. (D) Resolution of the single HJ and completed repair of the genomic DNA. (E) Second end capture leading to formation of a second HJ. Resolution of both HJs leads to either a crossover or non-crossover recombination event.

Section 1.2.3: Gene Therapy for Treatment of Genetic Disease

When genome engineering crosses into the realm of treatment for genetic conditions it is commonly referred to as gene therapy, though it utilizes many of the same steps and functions as genome engineering. Gain, change, and loss of function are all ways in which gene therapy aims to treat genetic conditions, with no "one size fits all" solution being appropriate in all cases. The genetic diseases mentioned above are all candidates for this kind of treatment in one way or another. In the case of HD, if it were possible to edit out the extraneous copies of the CAG repeat at an early enough time in the individual's life it might be possible to remove the risk of disease entirely. For DMD, scientists are currently looking at CRISPR as a way to repair the mutant copy of the dystrophin gene by correcting mutations specific to the patient using wild type homologous sequences as a template[3,34].

The most recent news on this topic comes from treatments being successfully used to treat SCD. Scientists have developed a way to use CRISPR to inactivate a repressor, BCL11A, that blocks the production of fetal hemoglobin. BCL11A is inactivated via CRISPR-mediated NHEJ within a splicing enhancer sequence. Blocking synthesis of the repressor protein allows the patient to produce fetal hemoglobin, which competes with ß-globin for assembly with the alpha chain. The patients effectively present as carriers of the sickle cell trait, instead of as homozygotes, thus removing the highly negative effects of having sickled blood cells[35–38]. This gene therapy protocol is by no means a small feat, but it is a complex workaround for a problem caused by a single base change in the genome.

Section 1.3: Current Methods for Improving HDR Efficiency

This all leads to the fact that the efficiency of HDR for genetic modification is fairly low, otherwise it would be the preferred method for treatment of some genetic conditions. As mentioned before, homologous templates may be incorporated into the genome in

locations other than where the desired target is. The introduction of a precise DSB improves efficiency and specificity due to the initiation of some kind of repair in the presence of a donor, but the efficiency of that combination still is not enough to ensure 100% desired recombination. The nucleases being used for genome engineering are also not infinitely accurate and can also recognize sequences with a number of bases that are changed[39]. This problem has become easier to address with the development of improved CRISPR systems and modified variants of commonly used nucleases. There is a nickase variant of Cas9 that has one of the two catalytic domains inactivated so it can only cleave or nick a single strand of the DNA molecule[40]. By combining this with a second nickase Cas9 and guide RNA the user has more control of the target site, requiring two enzymes to cut in close proximity, while retaining the simplicity of designing two gRNA sequences instead of engineering two fusion proteins like the ZFNs and TALENs.

Many groups have further attempted to address this problem by introducing minor modifications to the general system, with the end goal being to make recombination using a homologous template more efficient so that it can be a viable therapeutic mechanism. Initial attempts focused on making NHEJ less facile by either modulating the cell cycle chemically or introducing small molecule inhibitors for NHEJ related enzymes, thus turning off the NHEJ pathway completely[41]. This is an effective approach for *in vitro* cell experiments, but if implemented *in vivo* could have a wide-ranging negative effect on the organism, as DSBs occur frequently even when there is no attempt to induce them.

Other groups have attempted to address the efficiency issue by trying to package all of the pertinent components together. Expression and purification of Cas9 and sgRNA and packaging them as a ribonuclear protein complex (RNP) aims deliver the nuclease and targeting sequence simultaneously in the host cells. Similarly, donor DNA sequences have

been chemically linked to Cas9 or the sgRNA in an attempt to force the donor DNA to go wherever the nuclease does, ensuring that there will be no off-target integration of the donor DNA[42,43]. These methods both require proteins and RNAs or chimerae produced *in vitro* to be introduced into cells via electroporation or some other method of direct injection.

Section 1.4: Use of DNA Looping Proteins to Increase the Local Concentration of Donor DNA

To address the off-target incorporation of donor DNA and more importantly increase the efficiency of HDR, our current research efforts involve a DNA looping protein in targeting the donor DNA. This protein has two unique DNA binding domains that recognize different DNA sequences. By designing a donor DNA that contains one of the binding sites while targeting a chromosomal gene bearing the other DNA binding site, the donor may localize preferentially to the site of the DSB before the DNA is cleaved. Since the donor DNA is already present when the DNA break occurs, the repair pathway could partition towards HDR instead of NHEJ. Figure 1.7 depicts a model for how this looping protein acts in the proposed system. This solution also has the advantage of being modular: as the looping protein is not doing anything to alter the host cell machinery, it could work synergistically with other methods that work to either repress NHEJ or improve HDR. A similar method of enhancement has been looked at before using the LexA-Fkh1p fusion protein, with an estimated fivefold improved recruitment of donor[44].

To evaluate the proposed targeting method, experiments were performed in the yeast *S. cerevisiae*. Yeast cells carry out HDR efficiently, so a key component of this study was to enhance HDR relative to an expected baseline level by using CRISPR and/or incorporating our looping protein. To assay the model system, both the loss of function of a semi-essential gene and the gain of a new function were employed. Both the target and the

insert are scorable markers that either change yeast cell color or the ability to live in the presence of antibiotics. The model system was tested entirely *in vivo,* using plasmids that express the looping protein, Cas9, and an sgRNA along with either nonreplicating circular donor DNA or linear donor DNA.

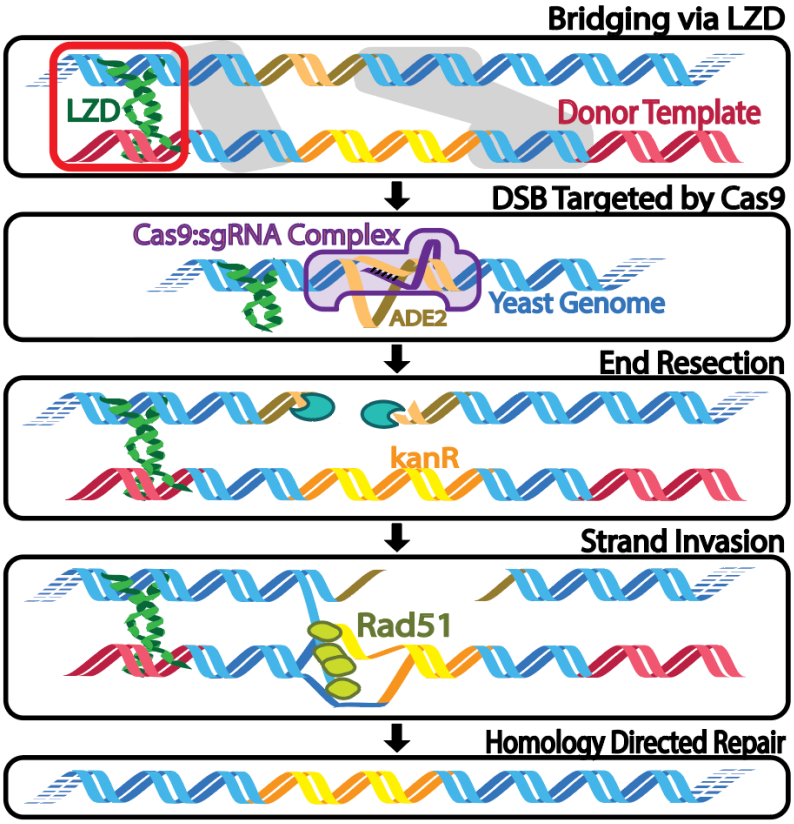

Figure 1.7 Bridging DNA with LZD: The looping protein LZD (in green) binds to both the genomic DNA and donor DNA, acting as a recruiter to bring the molecules together. Cas9 (in purple) cleaves the genome and since the donor is already nearby the HDR mechanism can predominate over NHEJ. Regions of homology are noted in the first pane in gray.

Section 1.4.1: Leucine Zipper Dual DNA Looping Proteins (LZD)

Members of the Leucine Zipper Dual DNA binding (LZD) family of proteins, developed by Dr. Dan Gowetski in the Kahn lab, were used in this experiment[45]. LZDs are proposed to be the most useful DNA looping proteins for this experiment by virtue of their small size and rigidity, their high stability and refoldability, their lack of Cys residues, and

the capability to purify them from bacteria or potentially synthesize them chemically. As seen in Figure 1.8, the protein is proposed to be a homodimer of alpha helices that form a coiled-coil domain in the leucine zipper portion of the sequence, with DNA binding sites at each end of the protein recognizing two different DNA sequences. The N-terminal DBD binds the CREB DNA sequence, which is the binding site recognized by the yeast transcription factor GCN4, and the C-terminal DBD of LZD recognizes and binds to INV-2, which is the binding sequence for the reverseGCN4 peptide developed by Martha Oakley's lab[46]. Since each end recognizes a different DNA sequence there is potential for using this as a targeting mechanism in the scope of an HDR experiment; one end of the protein can bind to the template DNA, and the other end of the protein can bind to the genome near a DSB site.

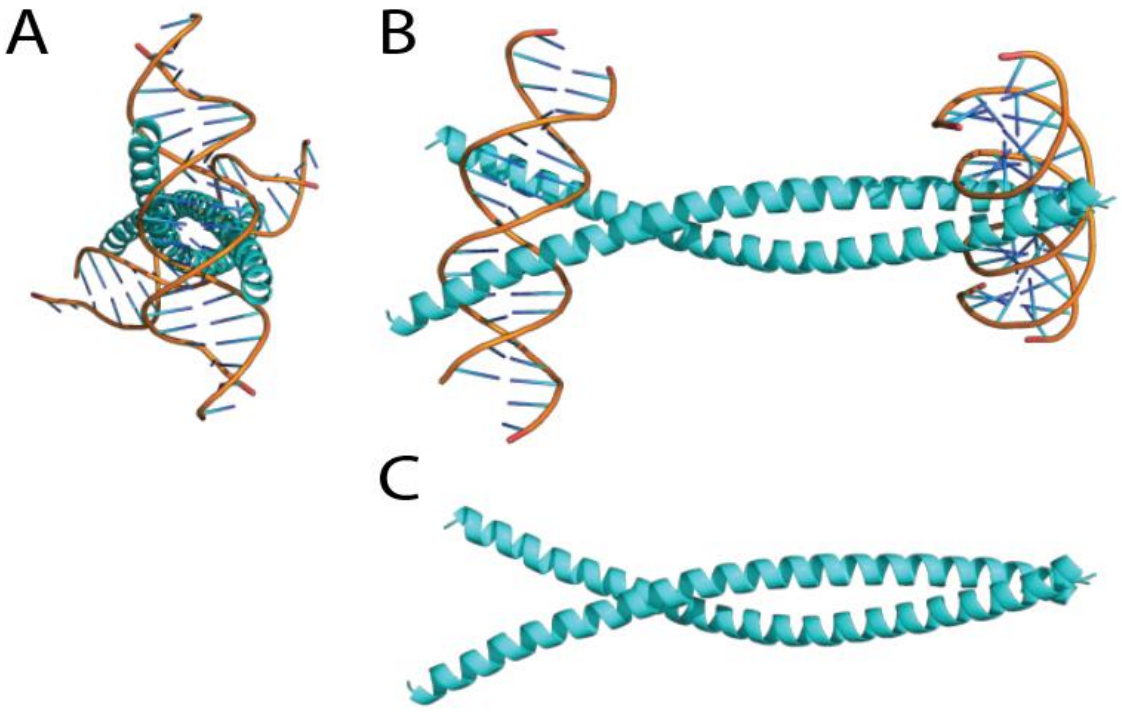

Figure 1.8 Models of the LZD73-DNA complex: (A) Head on view of the protein with DNA to see the angular orientation between the two DNA molecules. (B) Side view of the protein with DNA showing the coiled-coil structure. (C) Side view of the protein without DNA. Molecules rendered in PyMol.

The LZD family of proteins were designed with four different linking regions between the DNA binding domains. The coding sequences and protein sequences for all four variants are found in appendix 1. LZD proteins are named by the number of amino acids between the binding domains, with LZD66 being the smallest and LZD87 being the largest. The varied length of the proteins impacts the relative angle between the two bound pieces of DNA as can be seen in Figure 1.9.

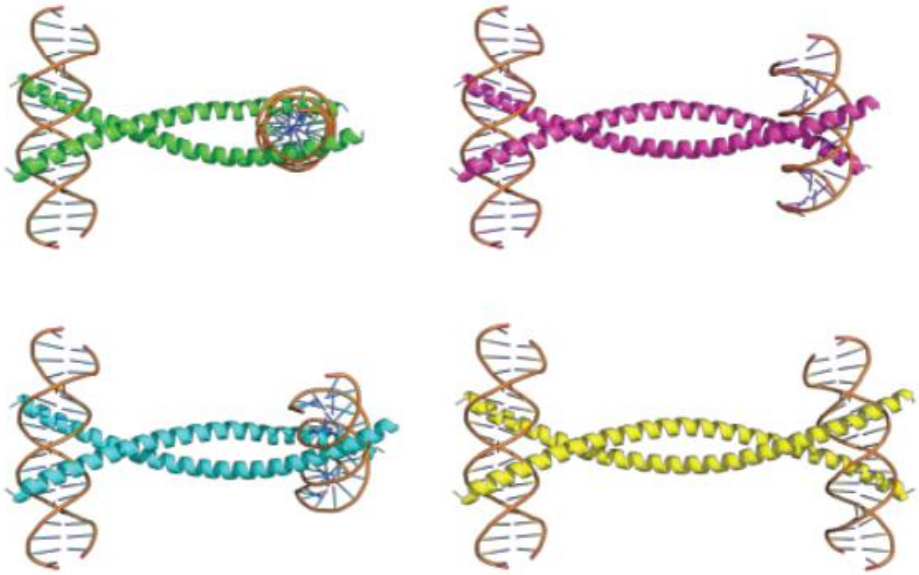

Figure 1.9 Models for LZD Length Variants: As the length of the coiled-coil region of the proteins increases, the relative end-on angle between the two DNA molecules decreases. In LZD66 (green) the DNA molecules are perpendicular. In LZD87 (yellow) the DNA molecules are parallel. LZD73 (blue) and LZD80 (magenta) have intermediate orientations. All protein models are rendered in PyMol.

LZD proteins in contrast to the LexA-Fkh1p fusion protein target two specific DNA sequences through the two different DNA binding domains. The fusion protein uses LexA as a DNA binding domain, but the Fkh1p domain binds to phosphorylated threonine residues[44]. Many of the proteins implicated in homologous recombination contain phosphorylated threonines, but there are likely to be other phosphorylated threonines present in the cell which could detract from the maximum recruitment potential.

Section 1.5: Designing and Testing the Recombination System

Developing a system that allows for the control and measurement of recombination while incorporating the LZD proteins was not trivial. The remainder of this document describes the efforts that went into the development of each individual component and then the final combination of all the components and the tests for enhanced HDR. A detailed but brief description of the system is laid out in chapter 2, with each of the plasmid systems described separately in subsequent chapters 3-5. chapter 6 describes the results of testing the system in yeast and all recombination outcomes that were seen. Finally, the document ends discussing the implications of this work and the future of the system, with an emphasis on improvement and optimization of different components.

Chapter 2: Experimental Design

Section 2.1: Overview

In order to determine if the LZD proteins are beneficial in the realm of homologous recombination the idea was tested in a system capable of handling the various components. Yeast systems have been a reliable eukaryotic model for recombination since the 1980s, with the first successful transformations being performed back in 1978 if not earlier[47]. The budding yeast *Saccharomyces cerevisiae*, besides being the prototypical eukaryotic genetic system, has commonly been used in labs for protein expression and purification in addition to more modern uses as a model for CRISPR systems. Yeast have a specific advantage over other eukaryotic systems in that they replicate much more quickly and show the effects of gene modifications much sooner than plants, insects, or invertebrates.

Section 2.2: Designing an *in vivo* Expression System

In order to properly evaluate the effect LZD has on homologous recombination, a system was first designed that allowed for the testing of parameters with enough controls that any differences could be interpreted. The key to having a successful system is a model organism that can handle the biological changes that are going to be induced, while also being relevant to an eventual target organism in which the method could be used. The long-term end goal of this work is to show that with the inclusion of small DNA looping proteins, homology directed repair can be enhanced to the point it will be a more efficient and effective tool, suitable for human therapeutic use.

The best model organism to start with then is yeast, since they are eukaryotic and also have been well studied as an organism that can handle CRISPR gene modification. The specific yeast being used in this project is *Saccharomyces cerevisiae* or budding yeast, which has been used for plasmid transformation experiments since the late 1970s[28]. Methods for

retaining plasmids in yeast often use auxotrophy as a form of selective control, whereby the yeast strain being used is unable to produce specific organic compounds needed for growth. These auxotrophic genes are often non-essential amino acids but can also be genes involved in biosynthesis of other biological components like uracil and adenine. Plasmids are retained by the yeast when grown on media lacking the organic compound, as the yeast are able to produce the compound while retaining a knocked-out chromosomal copy of the auxotrophic marker. For example, as shown in Figure 2.1, the *S. cerevisiae* strain CKY8 has the genotype *MATα ura3-52 leu2-3,112*. This means that if it is grown on selective media not containing the amino acid leucine (SD-Leu) the yeast does not grow. However, if a plasmid such as pML107 that contains an intact LEU2 coding sequence is transformed into the CKY8 yeast, it is able to grow on the media lacking leucine.

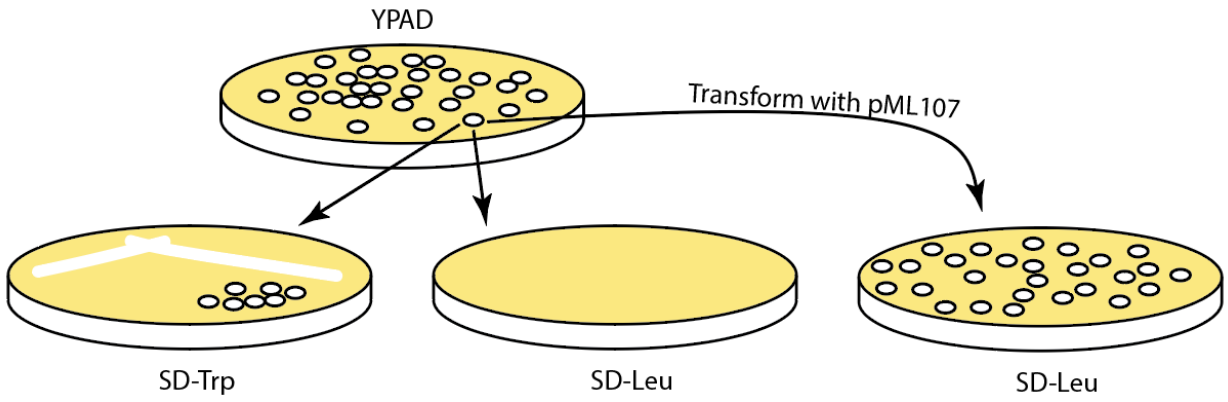

Figure 2.1 Auxotrophy: CKY8 (MATα ura3-52 leu2-3,112) yeast is grown on rich media (YPAD) and then a single colony is streaked onto media lacking either tryptophan or leucine only. The yeast grow in the absence of tryptophan but not in the absence of leucine. Growth in the absence of leucine is recovered when the pML107 plasmid containing the LEU2 gene is transformed into the yeast.

Yeast is also a great model for this experiment due to the large number of genes that have been identified and have a known function. In order to help in the screening efforts it is beneficial to induce a gene knockout (by homologous recombination) that is apparent

without the need for sequencing each and every organism that grows. In this way it benefits from identifying a gene that has a visible phenotype and exploit that for screening. One of the many ways to control the yeast growth is through the media they are plated on, so when looking at genes to target there was an emphasis placed on phenotypes that could be easily seen with a change in media conditions. The gene that was initially chosen is the AGC1 gene, which codes for the Aspartate Glutamate Carrier 1 protein (Agc1p). In addition to being an amino acid transporter, Agc1p was claimed to be essential for the cell's use of acetate as a sole carbon source. Successfully knocking out this gene was expected to be scoreable by the inability of the yeast to grow on media that contain acetate as a sole carbon source in place of more traditional dextrose.

Section 2.2.1: Cas9 and sgRNA Expression Plasmid

With a model organism selected, the other aspect of the system to consider is how the different necessary components are going to be introduced. Cas9 and an sgRNA need to be present in order to induce DSBs in a targeted manner, and the two common ways of incorporating this are through RNP injection or *in vivo* expression[48]. The process of injecting RNPs involves first expressing the Cas9 nuclease and sgRNA *in vitro* and then forming a lipid nanoparticle around the Cas9/sgRNA complex. *In vivo* expression of the components just involves the design and transformation of a plasmid that is retained by an auxotrophic marker and has a yeast origin of replication. Each method adds a layer of complexity to the system. RNP formation is usually done with a kit that complexes in-house produced sgRNA with a provided Cas9. In this regard, a system needs to be in place for generating sgRNA. With plasmid generation for the *in vivo* transformation, the lab was already set up and equipped to handle such experiments, so that is the method used herein.

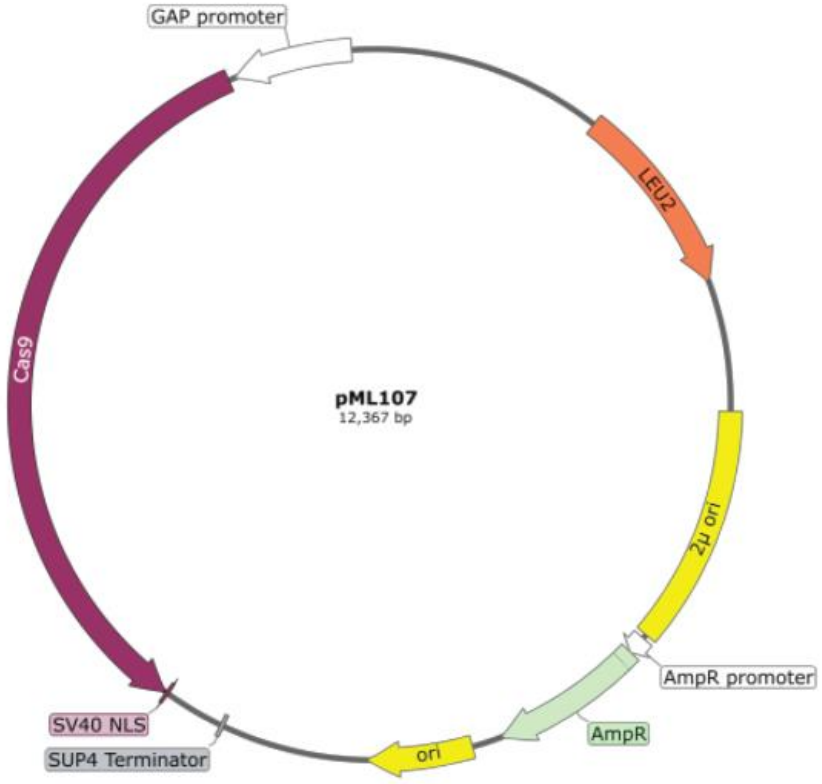

Figure 2.2 pML107 Plasmid Map

The initial plasmid vector used for the Cas9/sgRNA component was pML107 (Figure 2.2), designed in the Wyrick lab[49]. The plasmid is retained by yeast through the LEU2 auxotrophic marker and contains an origin of replication from the yeast 2-micron plasmid (2μ) making it a yeast episomal plasmid with a high copy number. The constitutively active GAP promoter controls the Cas9 nuclease, the mRNA for which is present as a pre-spliced sequence, preventing activity in organisms that don't have RNA splicing mechanisms. Finally, the Cas9 is fused at the protein's C-terminus to the simian virus 40 nuclear localization signal (SV40 NLS), which tags the protein to be transported into the nucleus of the cell. The sgRNA is expressed from a cassette separate from the Cas9 nuclease, controlled by an RNA specific promoter and a nonspecific terminator. The scaffolding tracrRNA fragment is coded into the plasmid prior to the terminator while the crRNA

portion is flanked by restriction enzymes for easy exchange if new gene targets are identified. The nomenclature we use for pML107 (and pG2, see below) derivatives is e.g. pML107::AGC1 for a plasmid containing an sgRNA targeting the AGC1 gene.

The plasmid also contains useful features for cloning in bacteria, notably an ampicillin resistance gene (*ampR*) and bacterial origin of replication. This simply makes the cloning and exchanging of sequences in the vector easier to perform by doing that in bacteria compared to yeast.

Section 2.2.2: Homologous Donor Plasmid

The second component needed for the system is the donor DNA. The construct (Figure 2.3) in question was required to have two homologous regions to the yeast genome, on the 5' side and the 3' side of the proposed DSB. For scoring, the donor DNA contained a gene with a new function placed between the two homologous regions. Typical homologous recombination experiments recommend a minimum of 40 bases of homology in order to facilitate efficient recombination. Since yeast are proficient at HDR the initial template that was designed utilized homologous regions that are below this, at 25 bases. The reduction of homology was intended to hamper the yeast's natural proficiency at HDR and lead to greater accessible dynamic range for any effect of LZD. The final element needed in the donor DNA is a (replaceable) LZD binding site so that the protein can be introduced to recruit the donor to the site of the DSB via specific protein-DNA binding interactions.

The new gene function added through this plasmid is the Kanamycin resistance (*kanR*) gene, as part of the KanMX cassette that contains a TEF promoter and TEF terminator needed for gene expression in yeast[50]. The resistance gene allows for the identification of yeast that can grow in the presence of antibiotics, independently of their ability to use acetate as a carbon source. In prokaryotes the *kanR* gene imparts resistance

to kanamycin, but in eukaryotes it provides resistance to another antibiotic called geneticin, or G418. G418 is an aminoglycoside inhibitor that prevents translation elongation, thus impeding protein synthesis.

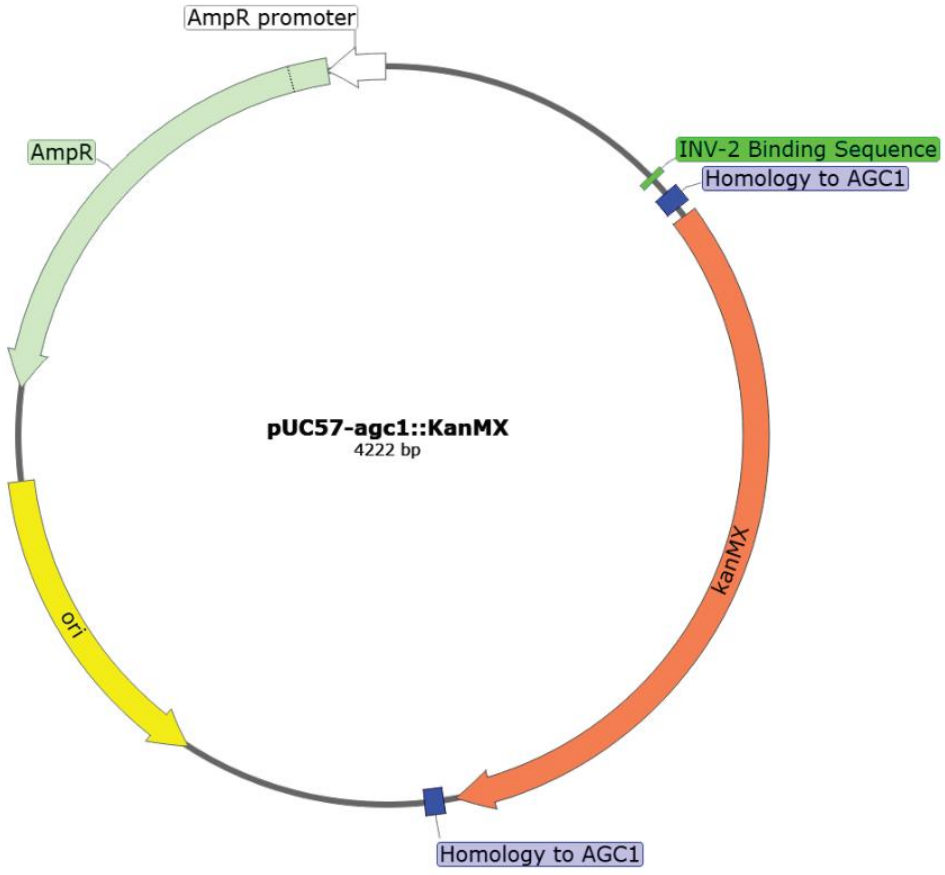

Figure 2.3 pUC57-agc1::KanMX Plasmid Map

Important to note about this plasmid is that the design does not include a yeast origin of replication. This is intentional, as the inclusion of an origin would allow the plasmid to replicate and give the yeast resistance to G418 without undergoing homologous recombination. Similar to the pML107 plasmid, there are features key for letting this plasmid be retained and manipulated in bacteria for cloning and modifying as needed in the future.

Section 2.2.3: LZD Expression Plasmid

The final plasmid component in this system is one that expresses the LZD proteins. This plasmid contained the coding sequence for the LZD proteins under the control of an inducible promoter in order to prevent excessive expression of the protein. It has been previously shown in the Kahn lab that overexpression of LZD in bacterial systems leads to decreased fitness for the host organism[45]. The rationale is that since LZD is a DNA looping protein and it has the capacity to bind two DNA sequences it interferes with the replication process and leads to inhibition of growth.

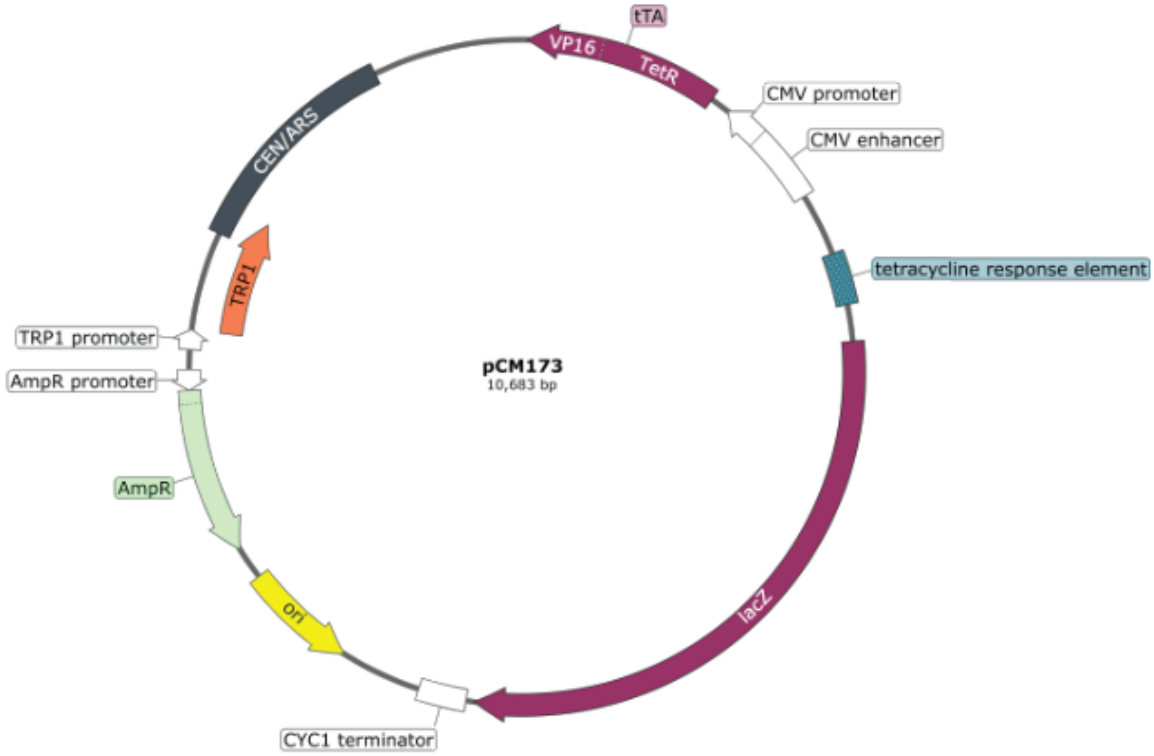

Figure 2.4 pCM173 Plasmid Map

The plasmid used is pCM173 (referred to as pCM173::eV [for empty vector] when unmodified from its original form) from the Herrero group (Figure 2.4)[51]. This plasmid uses a tetracycline off (TetOff) system for inducible expression of the desired gene. In this system there is constitutive activation by cytomegalovirus (CMV) promoter for the

tetracycline transactivator (tTA) which is the fusion of a tetracycline resistance gene (*tetR*) and the activating domain of the herpes virus VP60. This tTA is able to interact with the *tetO* operator and induce expression of downstream genes. In the presence of tetracycline and its derivatives – like doxycycline (structures shown in Figure 2.6) – the tTA is bound by the antibiotic and no longer interacts with the operator, thus turning off expression of downstream genes[52]. As shown in Figure 2.5, the absence of doxycycline induces protein expression. The gene of interest can be cloned into the vector through restriction digestion in the multiple cloning site. This system is essential to the survival of the yeast, as overproduction of our protein could lead to decreased fitness similar to what was described in bacteria.

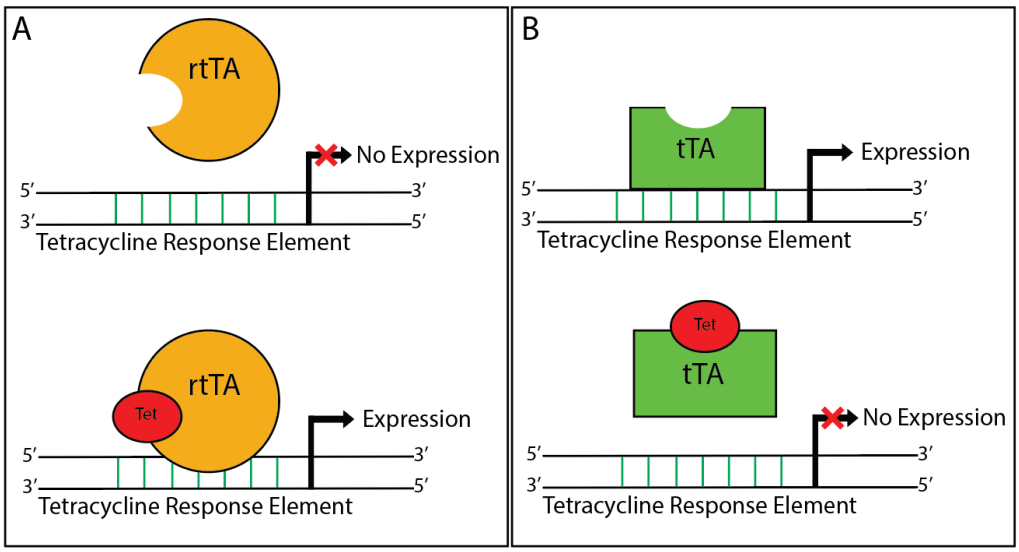

Figure 2.5 TetOn vs. TetOff System: (A) TetOn system, the presence of tetracycline turns on expression of downstream genes. (B) TetOff system (present in pCM173::eV), the presence of tetracycline turns off expression of downstream genes.

This plasmid contains a TRP1 auxotrophic marker for retention in yeast as well as the yeast centromere sequence CEN4. The centromere sequence makes this a low copy number plasmid but also ensures stability in the vector so that it won't be lost during replication. As with all other plasmids used in this experiment, there is a bacterial origin of

replication and the *ampR* gene to allow for working up and modifying the plasmid in *E. coli* as needed.

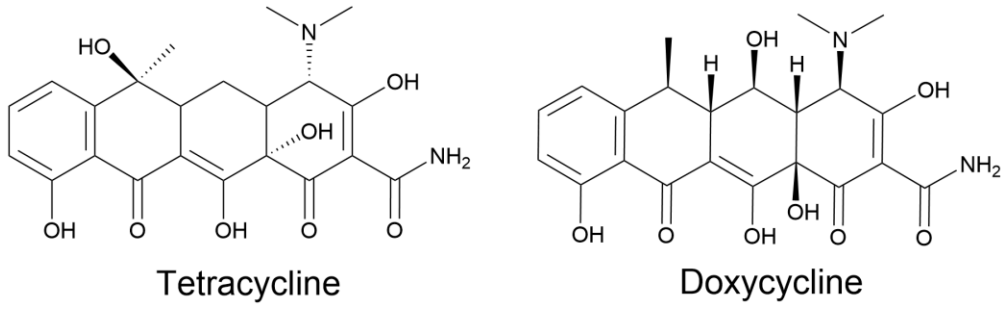

Figure 2.6 Tetracycline and the Derivative, Doxycycline

The pCM173::eV vector was used as the basis for new plasmids that express several varieties of LZD proteins or controls. The four different length variants of LZD described in chapter 1 were all cloned individually into the pCM173 vector backbone, affording e.g. pCM173::LZD73. We also cloned a single-ended DNA binding variant, known in our lab as PAAW, which is the bZip domain of GCN4. The final vector used in the experiments is the unmodified pCM173::eV as a control.

Section 2.3: *Saccharomyces cerevisiae* Selection and Confirmation

Based on the plasmids used in this experiment a yeast strain was needed with the appropriate auxotrophic markers. Dr. Jonathan Dinman generously provided access to his many yeast strains. The strain selected is yJD52 (*MATa his3-Δ200 leu2-3, 112 trp1- Δ901 ura3-52 ade2-10*). The *leu2* and *trp1* knockouts are most important since the pML107 plasmid has the LEU2 gene and pCM173::eV has the TRP1 gene.

The yeast strain genotype was confirmed by growing onto rich media (containing all amino acids) from a glycerol stock of the yJD52 yeast and then streaking single colonies onto media lacking leucine, tryptophan, or both, as well as media containing those amino

acids. The results indicated that the yeast were in fact knockouts for the genes listed, confirming that they were suitable for downstream experiments.

Section 2.4: Transformations

To best assess the effect that LZD has on this system a stepwise transformation scheme was implemented, visually represented in Figure 2.7. The process started with the unmodified yJD52 yeast which are transformed with each of the six different pCM173 variants and plated on synthetic dropout tryptophan (SD-Trp) media, with 100 µg/mL Doxycycline (Dox) to prevent expression of LZD. These transformants were then grown in SD-Trp with 100 µg/mL Doxycycline and then prepared for glycerol stocks. This glycerol stock of yeast became our new starting point for further transformations.

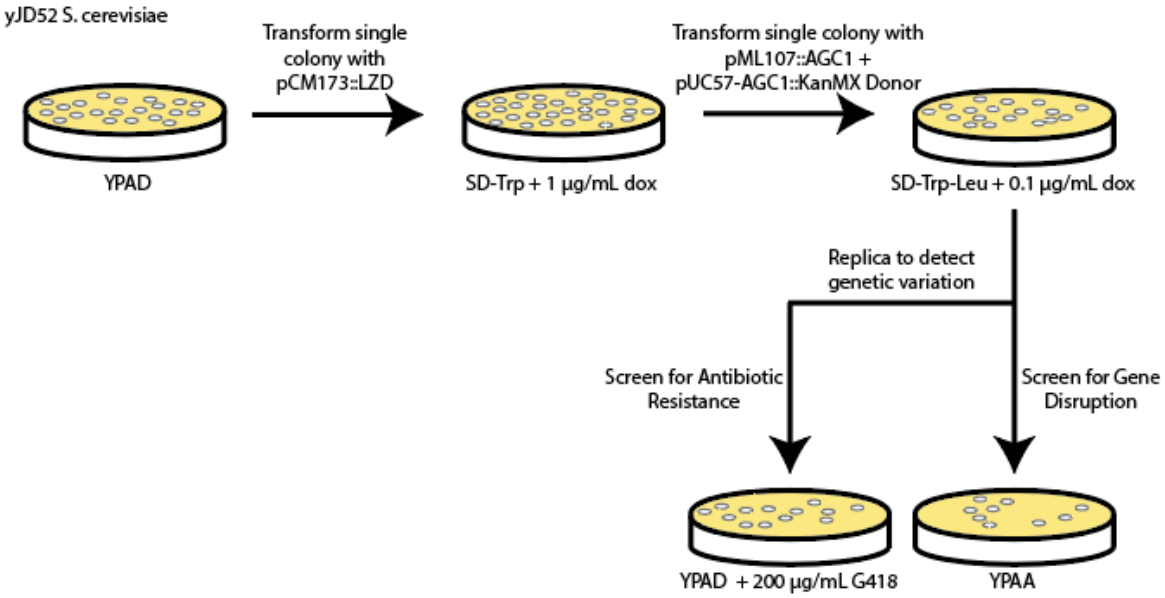

Figure 2.7 Visual Depiction of the Transformation Workflow: Transformed yeast are given 3-5 days to grow on new media before moving to next steps. The final replica plates are compared to the parent plate for determination of phenotypes.

Secondary transformations are grown in SD-Trp overnight with 1 µg/mL Doxycycline to allow for expression of LZD, before being mixed with the pML107::AGC1 plasmid and the KanMX donor DNA. In doing so, the yeast are preloaded with our protein

so that when donor is introduced into the cells it should be bound and localized prior to the Cas9 being expressed and inducing DSBs. This second transformation was then plated onto SD-Trp-Leu with 1 µg/mL Doxycycline so that both pCM173 and pML107 plasmids are retained.

Once these plates grow for 3-5 days, they are processed by replica plating for the desired phenotypes. The parent plate is stamped onto velveteen squares and then transferred to two different plates. The first plate contained acetate instead of dextrose, the second plate had G418 antibiotic. Desired recombinants should survive on G418 containing media, but not on acetate substituted media. By comparing growth patterns on the three plates, the phenotypic changes can be determined.

There are a number of possible recombination outcomes in this system. They are summarized in Table 2.1 below. The desired product is outcome #5 noted in green, where the ability to grow on acetate has been lost and the ability to grow on G418 is gained. It should be noted that outcome #4 also shows this same phenotype. In order to differentiate the two, genotypes were confirmed via colony PCR or sequencing to determine the actual status of the AGC1 gene.

Table 2.1 Possible Recombination Outcomes

Outcome #	Genomic Change	Grows on Acetate?	Grows on G418?
1	No Change	Yes	No
2	Off Target Insert	Yes*	Yes
3	NHEJ	No	No
4	NHEJ + Off Target Insert	No	Yes
5	HDR	No	Yes

Confirmation of the recombinants genotypes is done through colony PCR (polymerase chain reaction). Primers were designed up and downstream of the desired DSB site to amplify the region of interest with a known length. Additional primers were designed to specifically target the KanMX insert for additional coverage. Amplification

reactions were performed on yeast colonies that were resuspended in water. Once PCR reaction completed the samples were then analyzed via agarose gel electrophoresis, given the knowledge of the possible band lengths.

Section 2.4.1: Controls

The scheme in Figure 2.7 is the ideal test scenario including all three major components of our system – LZD, Cas9/sgRNA, and the donor DNA. But other transformations needed to be completed in order to determine changes relative to a baseline level of repair/recombination when modifications are done. Each of the components impacts the system as a whole and it is imperative to know what those changes are so that proper data analysis can be performed.

The easiest to start with is the exclusion of pCM173 variants or LZD. This was done in multiple ways. The first was taking our base yeast strain and transforming in pML107 and donor DNA without any pCM173 plasmid. This transformation only introduces the Cas9/sgRNA and donor to observe what the background level of recombination is. On a very real level this is the experiment that is being performed around the world in research labs to generate CRISPR knockouts of many organisms. Another way that this was assessed in our system used the pCM173::eV vector with no LZD coding sequence. This transformation worked best due to the ability to keep all of the media consistent throughout the different transformations and still gives the same information about background or baseline levels of recombination.

Another control is leaving out the pML107 plasmid and seeing if the inclusion of LZD alone is enough to enhance homologous recombination in the yeast. In doing so the transformation started from our pCM173 variant glycerol stocks and just added the donor DNA to the transformation mixture. This transformation posed problems since there is no

selection when the donor is added to the transformation. It is not a yeast replicating plasmid, and has no auxotrophic marker, so the yeast is not motivated to retain the plasmid. Changes can only be seen if the donor alone was sufficient for performing recombination.

Lastly, controls of just transforming in single plasmids into the original yeast strain and observing what happens. This was essential in the initial transformations of pCM173::LZD as the doxycycline concentration had to be modulated in order to determine what the optimal conditions were for growth but also LZD expression. This control also served as confirmation that the pML107 plasmid functions as expected and what the background levels of DNA repair are in the yeast. We streaked successful transformants with the pML107::AGC1 plasmid onto media containing acetate instead of dextrose to determine if Cas9 has properly targeted the AGC1 gene and the yeast have lost the ability to grow on acetate.

Table 2.2 shows the possible outcomes based on different input components in the system. The table is color coded to show outcomes with high likelihood in green and those with a low likelihood in red. For example, the ideal HDR outcome #5, is expected to be seen at the highest rate when all the system components are present. If any of the components are removed, the amount of HDR should decrease compared to the situation with all components being present. Similarly, taking a look at outcome #1, no change, it would be expected to be seen most when there are no components present.

Table 2.2 Expected Frequency of Recombination Outcomes								
Plasmids				Recombination Outcomes				
Cas9	sgRNA	LZD	Donor	1	2	3	4	5
+	+	+	+	-	+	+	+	+++++
+	+	+	-	-	-	++	+	-
+	+	-	+	-	++	+	+	+++
+	+	-	-	-	-	++	+	-
-	-	+	+	+	+	-	-	++
-	-	+	-	+++++	-	-	-	-
-	-	-	+	+	++	-	-	+
-	-	-	-	+++++	-	-	-	-

Chapter 3: Confirming *in vivo* Inducible Expression of LZD

Section 3.1: Overview

The primary difference between the proposed system and what researchers already do is the inclusion of LZD peptides. As previously described, the LZD proteins are small rigid dual DNA binding proteins that have specific binding domains on the N- and C-termini. Introducing LZD into the cell could be done through one of two main methods: electroporation of pre-purified proteins accompanying a Cas9 RNP or endogenous expression *in vivo* from a plasmid vector. Since the effect of LZD proteins is the main focus of this innovation, it is important to ensure they are present in all reactions. For this reason, a plasmid-based approach was used, with a vector that enables auxotrophic selection and controlled expression of our protein as shown in Figure 3.1. This chapter introduces the in-depth design and construction of expression vectors based on the pCM173::eV plasmid. This chapter introduces of the original protocol used as well as the development of an alternative protocol(s) for the construction of the LZD expressing plasmid.

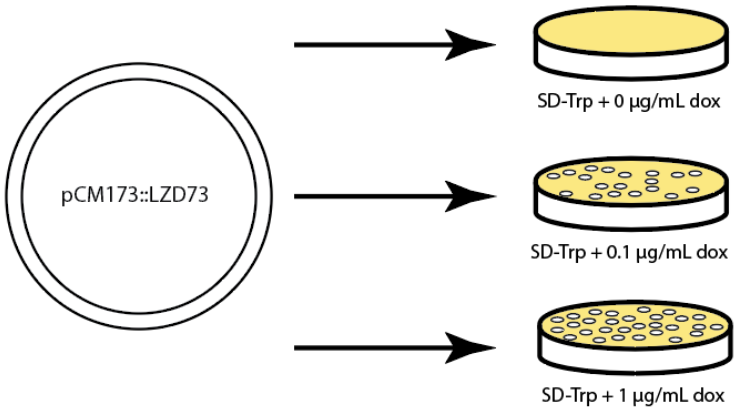

Figure 3.1 Controlled Expression of LZD: This artistic rendition of plates shows that there is an inverse relationship between doxycycline concentration and LZD concentration. With higher antibiotic concentration there is a lower LZD concentration and thus more growth. With lower antibiotic concentration there is a higher LZD concentration and thus less growth.

Section 3.2: Intended Construction of LZD Expressing Vectors

In order to have controlled expression of the LZD proteins *in vivo,* a vector containing an inducible expression system was required. A paper from the Herrero group described a set of vectors with a system to control protein expression using varying tetracycline concentrations[51]. The vector in question – pCM173::eV – contains the machinery for the TetOff system, which had previously been shown to work for controlling gene expression in yeast. The TetOff system functions by using the tetracycline operator DNA (*tetO*) next to a promoter upstream of the gene of interest and a tTA as the effector that promotes induction. The tTA protein is a fusion between the VP16 effector domain of herpes simplex virus and a bacterial *tetR* gene. This fusion protein binds to the *tetO* element in the absence of tetracycline or its derivatives, such as doxycycline, which was used due to its relative lack of toxicity in *Saccharomyces*. In the presence of doxycycline, the tTA protein binds the antibiotic and therefore is released from the *tetO* promoter region, turning off expression of the protein of interest.

The vector also includes a centromere / autonomously replicating sequence (CEN/ARS) so that it can replicate on its own in yeast. It is retained in *trp1* knockout yeast strains through relief of tryptophan auxotrophy via the TRP1 gene on the plasmid. For bacterial cloning and insertion of desired gene sequences the plasmid also contains an *ampR* gene and a bacterial origin of replication. Restriction enzyme sites for SbfI and BamHI, located on either side of a LacZ gene driven by tTA, are used for excising a portion of the plasmid and allowing the vector backbone to be ligated to another insert – in this case coding for the LZD proteins – to create a new plasmid (Figure 3.2).

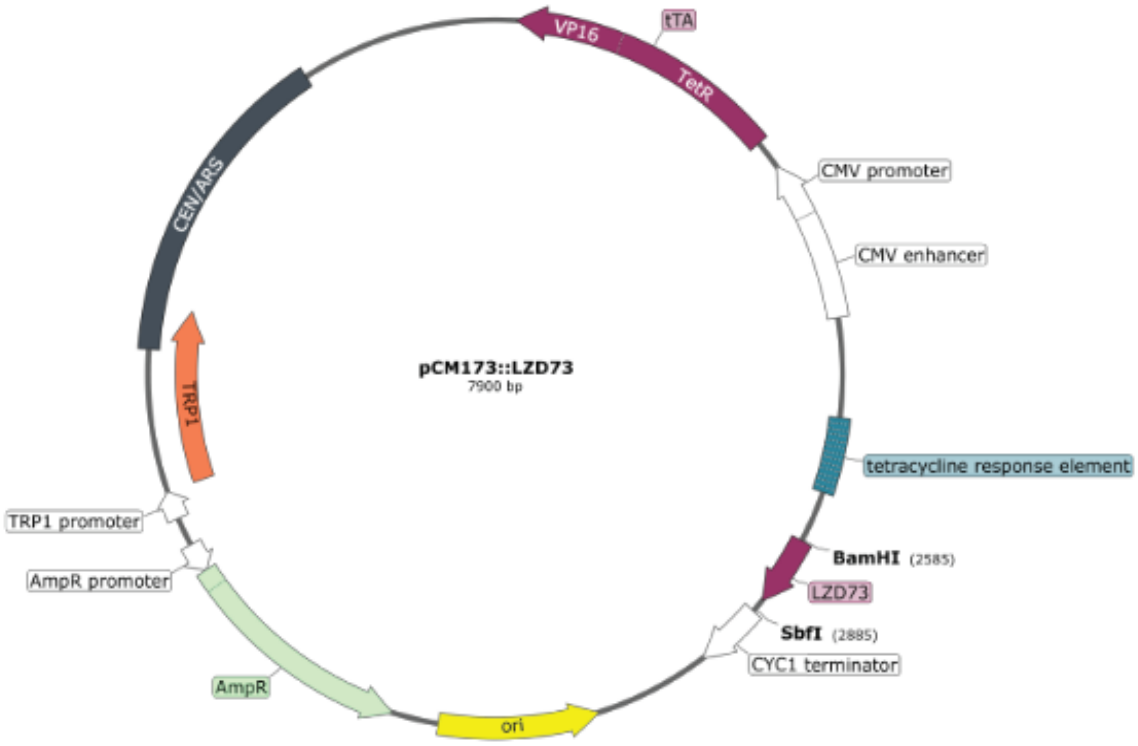

Figure 3.2 pCM173::LZD73 Plasmid Map

To generate the insert for each of the LZD coding sequences, PCR was used to make usable quantities of all five variants of the gene sequence (four LZD length variants and one with only the N-terminal DNA binding domain). The template for the PCR is the pMal-c2T::LZD series of plasmids (sequences found in appendix 1 section A1.2), which are used for bacterial expression of LZD as described in appendix 3. The 5' ends of the PCR primers contained sequences that include the SbfI and BamHI restriction enzyme sites so that PCR products could be digested and used for downstream ligation experiments. After restricting the vector and PCR products, T4 DNA ligase was used to attempt to generate the new plasmids, as shown in Figure 3.3, followed by transformation into *E. coli* and isolation of plasmids.

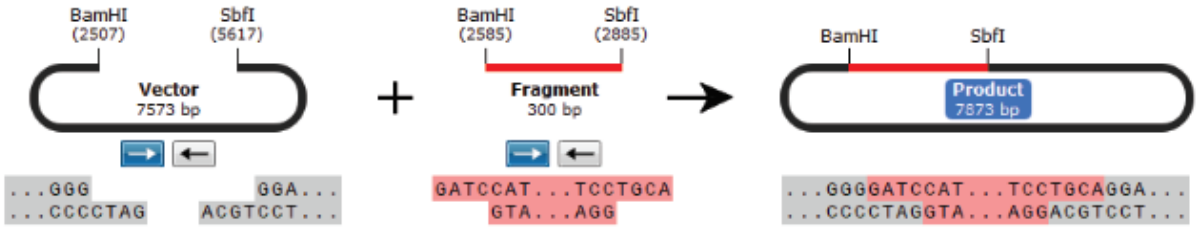

Figure 3.3 pCM173::LZD73 Cloning Visual Workflow: Shown are the compatible DNA ends from the doubly digested pCM173::eV vector, and the Fragment – pMal-c2T::LZD73 PCR product, which should ligate together forming the complete plasmid.

The plasmid had been designed with BamHI and SbfI restriction sites that were intended to be used to excise and replace the LacZ gene in pCM173::eV. The generation of doubly cut vector appeared to not be problematic, as it could be seen from analytical gels that there were two distinct bands compared to the intact plasmid.

Despite the basic biochemistry principles being solid the results of the transformations were always the unmodified pCM173::eV vector. PCR reaction products and restriction digestion were confirmed by gel electrophoresis prior to ligation reactions being initiated. The successful restriction of the PCR products, however, was hard to detect even with the use of polyacrylamide gels due to the difference in length of products being less than five base pairs. The first thought as to the cause of the failed ligation was that the PCR products were not being effectively cut with the enzymes, so new PCR primers were designed to extend the length of the PCR products further past the border of the restriction enzyme sites. Based on the New England Biolabs (NEB) website there should have been no problems with two bases present past the restriction sites, changes were made to extend the area to ten bases to ensure the difference could be seen on a gel as well. New PCR products were obtained but the results remained the same – only unmodified pCM173::eV was cloned.

In addition to the new PCR primers an attempt was made to dephosphorylate the vector with recombinant shrimp alkaline phosphatase (rSAP) to prevent ligation of the two original vector components. The samples were then heat treated to inactivate the rSAP and a phenol extraction was performed to ensure the enzyme was inactivated and purified away from the vector before introducing the insert and ligase. The subsequent reaction did not lead to successful transformants.

The next attempt to generate the proper plasmid was to extract the backbone of the vector from an agarose gel to purify away the old insert and mix the backbone with new insert to generate only one possible vector product. The Qiagen gel extraction kit was then used, with very little success recovering the DNA. The vector backbone is 7600 base pairs long, well within the stated acceptable range of the Qiagen kit (they claim 80% recovery of DNA up to 10 kbp), but any DNA eluted off the column was not detectable through ultraviolet (UV) measurements or gel electrophoresis. Since only a small amount of DNA was needed it was again mixed with the insert, but again there were no successful transformants.

Another attempt was made to purify the backbone away from the restricted fragment using an ElectroSEP kit under development by the company Princeton Separation Incorporated (PSI). A visual workflow of the product is shown in Figure 3.4. It works in a similar way to standard but cumbersome electroelution protocols, except in this method there is a proprietary GeneStix® membrane that is inserted into the agarose gel and the DNA from the band of interest is run into the membrane (Figure 3.4B). The membrane is further processed by eluting with buffer for collection of the DNA. The yield for this protocol was much greater than using a traditional Qiagen gel extraction kit: the band was visible both on a confirmation gel and UV. However, the ligation reaction using this method was

successful only once. One of the constructs was confirmed to be the correct sequence, but the other four constructs failed to successfully join.

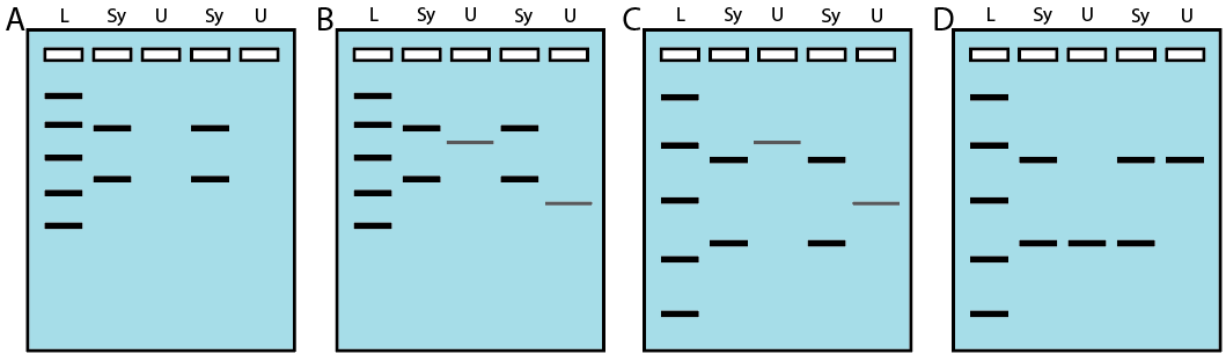

Figure 3.4 DNA Extraction with ElectroSEP and GeneStix®: A) Agarose gel with ladder (L), samples stained with SYBR Gold (Sy) and unstained samples (U). The samples shown in lanes 2 and 4 are the same, samples shown in lanes 3 and 5 are the same. B) After visualizing with blue light, the GeneStix® are inserted into the gel below the band of interest. C) Gel is run until band of interest is past the GeneStix® as seen by blue light illumination. D) Gel is stained in EtBr to confirm removal of desired band. (Figure adapted from ElectroSEP protocol material)

The final method for generating the LZD expression plasmids was to use HiFi assembly, a product from NEB reported to be an improvement over the Gibson assembly reaction. The process uses overlapping DNA ends that are resected by a 5' to 3' exonuclease to allow annealing of the 3'-tailed single strands in the overlaps, then filled by a DNA polymerase to fill the gaps, then ligation to re-seal the DNA and create a circular plasmid. The high efficiency and fidelity of the HiFi mix makes for an easy conversion from vector and insert to constructed plasmid. For this method the vector was treated as it was previously, requiring only restriction digestion rather than purification, as the excised insert cannot regenerate a circular product with the vector backbone. The inserts for the HiFi reaction were generated using a different set of PCR primers that had longer 5' tails that are homologous to the region of the vector outside of the restriction sites (including 20 bp overlaps with the vector, as described by the NEB web resource, as seen in Figure 3.5).

This is essential for the function of the exonuclease to chew back the insert allowing for base pairing interactions in the HiFi reaction. Pairing the restricted vector and the new PCR product in the HiFi reaction, all five variants of the pCM173::LZD vector expressing different LZD proteins were generated. The sequences of all the constructs are reported in appendix 1 section A1.3. All constructs are readily retransformed and maintained in *E. coli*, so the difficulty in preparing them appears to be due to the assembly process, not toxicity or stability once in the cell.

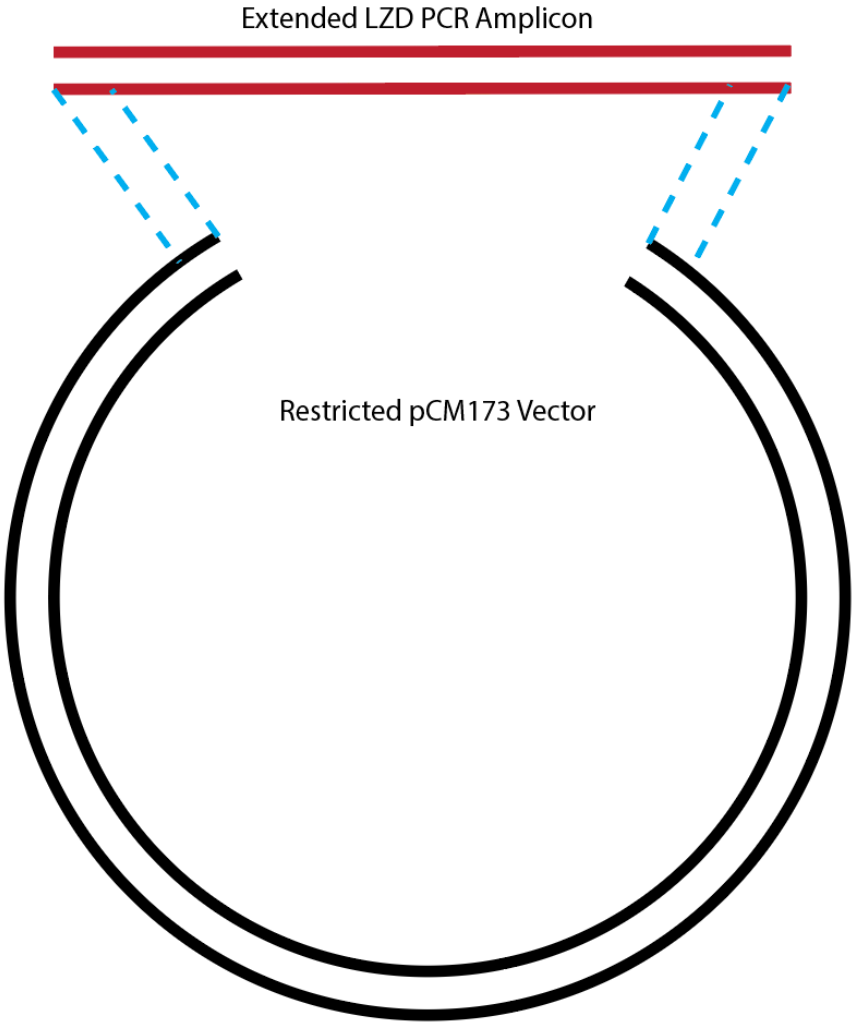

Figure 3.5 NEBuilder® HiFi Plasmid Construction: Blue dotted lines indicate required homologous regions (20 bp) for the activity of the HiFi enzyme mix to join the two DNA fragments. (Figure adapted from NEBuilder® HiFi DNA Assembly Protocol)

Section 3.3: Determination of Optimal Expression Conditions

From previous work done in the Kahn lab it is known that LZD can be toxic to the host organism in high enough concentrations[45]. When LZD proteins are expressed as His-tagged fusions in *E. coli*, the bacteria must be grown at lower temperatures (30 °C compared to 37 °C) and for a much longer period of time than bacteria expressing the single-ended GCN4 variant (called PAA). Currently in the lab, LZD is expressed in *E. coli* as a maltose binding protein (MBP) fusion, but there is a large difference in the cell mass obtained after inducing the double-end binding LZD proteins compared to PAA.

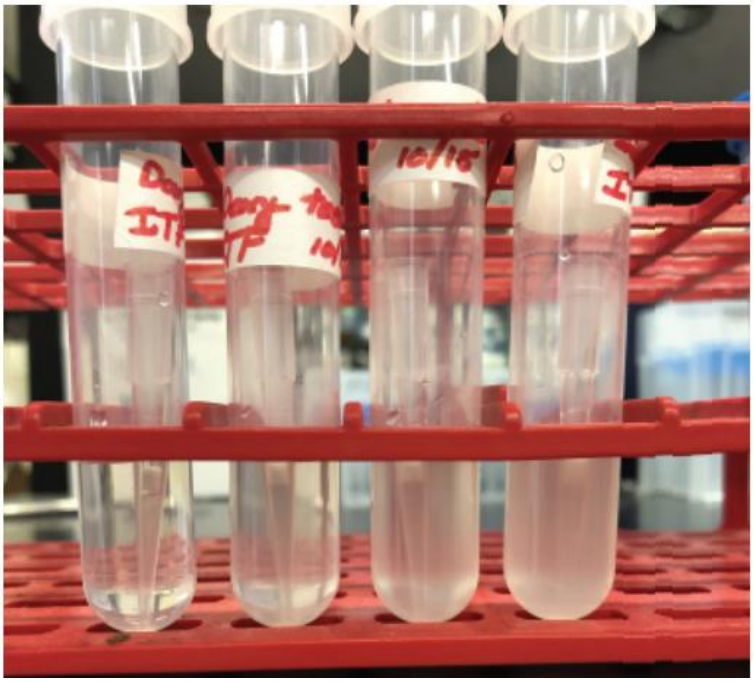

Figure 3.6 Doxycycline Test with pCM173::LZD73: Single colonies of yJD52 transformed with pCM173::LZD73 were inoculated and grown in 3 mL SD-Trp media with varying levels of doxycycline (0, 0.01, 0.1, 1 µg/mL). Turbidity of samples increases as the concentration of antibiotic increases, indicating that LZD at high concentrations negatively affects the cells' ability to survive.

In order to test if LZD proteins could be expressed in yeast without being lethal to the host the pCM173::LZD plasmids were transformed into yJD52 yeast (*MATa his3-Δ200 leu2-3, 112 trp1- Δ901 ura3-52 ade2-10*), generously provided to us by Dr. Dinman.

Transformation used a standard lithium acetate protocol, with plating onto Synthetic Defined media without tryptophan (SD-Trp) with 10 µg/mL of doxycycline (dox). At this concentration, doxycycline is not harmful to the yeast[53], but the authors of the pCM173::eV paper showed it to be a concentration that completely inhibits activation, presumably via releasing the tTA protein from *tetO*. Once successful transformants were dispersed and grown in liquid culture also containing 10 µg/mL doxycycline, the cells were collected and resuspended in glycerol with fresh media for long term storage at –80 °C. Additional colonies were grown in SD-Trp media with varying concentrations of doxycycline to see if there is an effect on growth and survival – the tubes from this experiment are in Figure 3.6. Without any antibiotics (i.e. high expression of LZD), there was no growth visible in the liquid culture. As the level of the antibiotic is increased, the OD660 of the samples is also increased. This indicated that LZD is also toxic in yeast, but with the appropriate amount of doxycycline, cell growth and can be maintained. The observed OD readings for each doxycycline concentration are listed in Table 3.1 below. We infer that at 0.1 µg/mL doxycycline there is enough expression of the LZD proteins to affect growth, but to date we have not been able to directly demonstrate expression via Western blot or proteomics.

Table 3.1 Doxycycline Test Growth

[Doxycycline] µg/mL	OD 660
0	0.01
0.01	0.11
0.1	0.53
1	0.84

Section 3.4: Materials and Methods

All enzymes used in reactions including restriction digestions and PCR were obtained from New England Biolabs (NEB). Purification kits were obtained from Qiagen. Oligonucleotides used as primers were synthesized and obtained from Integrated DNA

Technologies (IDT). XL1-Blue *E. coli* were originally purchased from Agilent. All other reagents used were purchased from Fisher Scientific. The polymerase chain reaction was run in an MJ Research PTC-200 Gradient Thermal Cycler.

The pCM173::eV vector was purchased from Euroscarf (ACCNO P30314) and shipped as a bacterial stab in agar. *E. coli* from the stab were grown on standard LB agar plates supplemented with 100 µg/mL ampicillin (LB-amp) to retain the plasmid. Individual colonies were then inoculated and grown overnight in 5 mL of LB-amp liquid media. Bacterial suspensions were then processed through Qiagen Miniprep or Midiprep protocols to purify pCM173 plasmid stocks for future use. Plasmids were stored at -20 °C. Plasmids containing the LZD coding sequences were previously prepared this way by myself and Dr. Jason Hustedt, and had been stored at -20 °C. Prior to storage, the plasmids were quantitated on a Cary UV/vis spectrometer to measure the purity of nucleic acids, and then run on 0.7 % agarose gels with a TBE (Tris, boric acid, and EDTA) running buffer to confirm size and a single band. The gels were stained after the electrophoresis was run using Ethidium Bromide (EtBr) and visualized on a UV transilluminator. Plasmids were also sequenced at Genewiz (now Azenta) using M13 forward and reverse primers to confirm the correctness of the plasmid sequences.

To generate doubly cut pCM173::eV, a reaction using the restriction enzymes BamHI and SbfI was performed. Both enzymes have 100 % activity in CutSmart buffer provided by NEB, so the reaction contained 1 × CutSmart buffer, 20 units of each restriction enzyme, and 2 µg of DNA in a total 50 µL reaction. Reactions were incubated at 37 °C for 1 hour before being cooled to room temperature. Digestion of the plasmids were confirmed by electrophoresis on 0.7 % agarose gels as previously described.

To generate the LZD coding sequence as an insert, PCR was used to amplify from the pMal-c2T plasmid templates containing each of the different coding sequences of interest. The PCR primers were designed with 5' tails that contained the requisite sequences to be cut with the same enzymes used to cut the vector, thus generating complementary overlaps for ligation. Reaction volumes of 50 µL were used and included 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each primer (forward and reverse), 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (30 repeats):

98 °C – 10 seconds

68 °C – 20 seconds

72 °C – 30 seconds

Final Extension: 72 °C – 120 seconds

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. PCR products were subsequently digested with BamHI and SbfI in the same manner as pCM173::eV in order to generate complementary sticky ends for ligation into the vector backbone.

Doubly restricted pCM173::eV and the individual LZD coding sequences were then combined in 3:1 molar ratio (insert to vector) along with 1 × T4 Ligase buffer (containing 1 mM ATP) and 400 units of T4 DNA Ligase. Reactions were incubated overnight at either 16 °C, room temperature, or 37 °C. DNA ligase was heat inactivated at 65 °C and then reactions cooled to room temperature before being used in bacterial transformations.

Transformations were performed using standard electrocompetent cell methods or chemically competent cell methods. For electrocompetent cell transformations, 50 µL aliquots of electrocompetent *E. coli* cells stored at –80 °C were thawed on ice before having 2 µL of ligation reaction pipetted in and mixed by resuspension. Cell suspensions were left to incubate on ice for 10 minutes before being transferred to electroporation cuvettes with a 0.1 cm path length and loaded into the electroporation device (Eppendorf). A pulse of 1700 V was applied to the cells and the time constant was recorded. Suspensions with a time constant above 4.0 were deemed likely to survive and thus were resuspended in 1 mL of super optimal broth with glucose (SOC) and left to incubate at 37 °C for recovery. Suspensions with a time constant below 4.0 were discarded, as those samples had been found not to yield colonies. For chemically competent cell transformations, *E. coli* competent cell stocks were thawed on ice before having 5 µL of ligation reaction pipetted in and mixed by resuspension. Cell suspensions were left to incubate on ice for 30 minutes before being transferred to a 42 °C water bath for 30 seconds to heat shock the cells. Cell suspensions then sat on ice for minimum 2 minutes before being resuspended in 1 mL of SOC and left to recover at 37 °C while shaking. After an hour of recovery, 250 µL aliquots of cell suspensions were spread onto LB-amp agar plates and left in a 37 °C oven to incubate overnight.

Successful plates yielded bacterial colonies that were inoculated into 5 mL of LB-amp liquid media and grown overnight for Minipreps. 3 mL of cell suspension was used for the plasmid prep and the remaining 2 mL left for inoculation of larger cultures for Midipreps. Isolated plasmids from the preps were subject to restriction digestion with BamHI and SbfI and then analyzed on a 0.7% agarose gel to determine if the DNA fragments were of the correct size. Minipreps with properly sized fragments were then sent

to Genewiz (now Azenta) for Sanger sequencing using M13 forward and reverse primers. Analyzed sequence data from Genewiz was run through Chromas software to confirm provided sequences matched the provided sequencing traces. The sequence data obtained were then aligned to the plasmid maps using SerialCloner software. For successful sequences, the leftover 2 mL of cell suspension was used to inoculate 50 mL of LB-amp liquid media which would then be processed with a Qiagen Midiprep kit to obtain larger quantities of pure plasmid for future experiments.

To perform the dephosphorylation reaction with rSAP, half of the restriction reaction for pCM173::eV was moved to a new tube after incubation and allowed to cool to room temperature. To it, 2 units of rSAP were added and the mixture resuspended by pipetting. This mixture was then incubated at 37 °C for 1 hour before heat inactivation of the enzyme at 65 °C for 45 minutes. This reaction was then ready to resume the previously mentioned workflow and move onto ligation.

The two methods for DNA extraction from agarose gels both used commercially available kits: Qiagen's QIAquick Gel Extraction Kit (ID:28704) and Princeton Separations' ElectroSEP and Gene Stix (ES-100) which was provided by Dr. Paul Nix of Princeton Separations. Upon extraction and purification of DNA from agarose gels the samples were run on a Cary UV/vis spectrometer to confirm the presence of DNA, and then run on a second 0.7 % agarose gel to confirm that DNA of the correct size was extracted and present. The DNA was then ready to resume the previously mentioned workflow and move onto ligation.

To perform the NEBuilder® HiFi DNA Assembly reaction, a kit was obtained from NEB with the HiFi master mix. Vector DNA for pCM173::eV was doubly digested with BamHI and SbfI as previously described, with no further workup necessary. The insert

DNA for LZD coding sequences was generated using primers with longer 5' ends that are homologous to the vector ends (sequences are in appendix 2 section A2.1.1). Since the portion of the primers that base paired to template remained unchanged, the PCR cycling conditions did not have to be modified. Reactions were run with the same mixture composition as previously stated. PCR products were confirmed to be the proper length by electrophoresis on 0.7 % agarose gels as previously described. DpnI digestion was still performed to remove template DNA so as to not interfere with the final reaction. Insert and vector DNA were mixed in a 4:1 molar ratio, respectively and then incubated with 1 × HiFi master mix in a 20 µL total reaction at 50 °C for 1 hour. Reactions were then allowed to cool to room temperature before being used in bacterial transformations as previously described.

Yeast transformations were performed using a protocol modified from the standard lithium acetate / polyethylene glycol method described by Geitz and Woods[54,55]. For these initial experiments the yeast used was yJD52, obtained from Dr. Jonathan Dinman. Yeast were grown on rich media containing yeast extract, peptone, adenine sulfate, dextrose (YPAD), and agar for 2-3 days at 30 °C until isolated colonies began to form. Individual colonies were then grown in 3 mL of YPAD liquid media overnight at 30°C while shaking. A 500 µL aliquot of the yeast suspension was collected via centrifugation at 4000 × g for 3 minutes. The supernatant was decanted, and the pellet was washed with 200 µL of 0.1 M lithium acetate in 1 × Tris EDTA buffer (LiOAc/TE) by resuspension. The mixture was then centrifuged again at 4000 × g for 3 minutes. The supernatant was again decanted, and the pellet resuspended with 100 µL of the LiOAc/TE buffer. Then 20 µL of salmon sperm ssDNA (stock concentration 10 mg/mL) that had been boiled was added, as well as 500 µL of polyethylene glycol (PEG)/LiOAc/TE buffer (10 % PEG, 0.1 M LiOAc, 1 × TE), and the mixture was mixed well to dissolve the pellet. Once the pellet is dissolved, 500 ng of

plasmid DNA was added and mixed well by pipetting. The mixture was incubated at 30 °C for 1 hour, then at 42 °C for 15 minutes. The transformations are then spun down at 4000 × g for 5 minutes. The supernatant was removed, and cells were resuspended in 1 mL of water before being spun down again. Then finally 60 µL of water was added, the pellet was resuspended, and the mixture was spread onto SD-Trp agar plates with 10 µg/mL doxycycline. The plates were incubated at 30 °C for 3-5 days until single colonies began to form.

To make the glycerol stocks, single colonies are incubated in 3 mL of SD-Trp liquid media with 10 µg/mL doxycycline overnight. A 1 mL aliquot of the suspension is spun down, and the supernatant is discarded. The cells are resuspended in 500 µL of fresh SD-Trp liquid media with 10 µg/mL doxycycline. An additional 500 µL of 50 % glycerol is added to the tube and mixed well by pipetting. The cells are safe to store at -80 °C in a final concentration of 25 % glycerol.

To determine optimal doxycycline concentrations for expressing LZD proteins, single colonies of yJD53::LZD73 were inoculated into four separate culture tubes containing 3 mL of SD-Trp liquid media. The different tubes had different concentrations of doxycycline (0, 0.01, 0.1, 1 µg/mL). The tubes were shaken in an incubator at 30 °C overnight before being observed the following day for turbidity. A growth curve was run in a similar manner with much smaller volumes and a wider range of antibiotic concentrations. The sample volumes for the growth curve were 300 µL, and all samples were made from a single culture that was diluted to an initial OD660 of 0.01. The concentrations used for this experiment were 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0 µg/mL. The plate reader was kept at a constant 30 °C and was set up to shake the cells every 10 minutes right before measurements were taken. Measurements were taken every 10 minutes for 48 hours.

Section 3.5: Discussion

In the construction of the LZD expression plasmids various issues were encountered. First, the restriction enzyme based cloning protocol that was expected to be routine proved unsuccessful. Generation of the LZD coding sequence as a PCR amplicon did not seem to be the problem, but ligation reactions yielded only intact pCM173::eV as the product after miniprepping plasmids. Dephosphorylating the vector ends with rSAP to prevent self-ligation gave no colonies, suggesting that the PCR product might be the issue as opposed to the vector: if the PCR product had been properly restricted with BamHI and SbfI, it should have provided the 5' phosphates required for ligation.

Purifying the vector backbone to make self-ligation impossible was also difficult. Electrophoretic elution of the 7600 bp backbone worked to an extent: DNA was obtained, but the elution buffer used appears to contain a ligase inhibitor as the subsequent ligation reactions were only successful a small number of times.

At this same time, preparation of the LZD insert was investigated to ensure that it was getting cut by the restriction enzymes in the proper way. The main problem here was the size of the amplicon and the small size of the ends being removed were difficult to detect even on an acrylamide gel. The LZD73 amplicon was originally 305 bp in length, and after removal of the ends with BamHI and SbfI the new length of the product was supposed to be 300 bp. Many attempts were made to run high percentage, high cross linked acrylamide gels to view the small bp changes with no success. The NEB website has a page dedicated to restriction reactions near the end of DNA fragments, which notes that both BamHI and SbfI should have at least 4 bp to either side for 50-100 % cleavage. This is the highest reported efficiency given on their table with the next rating being 20-50 % cleavage. To solve this potential problem, new PCR primers were designed that had additional bases at

the 5' ends so that there would be more DNA for the restriction enzymes to bind and less chance of fraying or foldback. It was these new PCR products and the restriction products that were successful, though only once, when combined with the vector fragments isolated from the ElectroSEP kit.

The final method used to eventually construct all of the pCM173::LZD variants was the use of the NEBuilder® HiFi Assembly kit. The ease and simplicity of the HiFi reaction is something that was used in generating new plasmids throughout the life of this project.

With the constructed plasmids in hand, it was time to assess if they would express in yeast and if so, do they have any impact on fitness. It is known from work done by Dr. Daniel Gowetski that the LZD family of peptides can have a toxic effect on bacteria[45]. The belief is that because the protein can bind DNA twice – either looping the same DNA molecule or bringing two DNA molecules together – it inhibits replication via figuratively or literally tying the chromosome up in knots, leading to cell death. This is the major reason behind the need to engineer controlled expression of LZD in the yeast system. To assess the effect on yeast the pCM173::LZD plasmids needed to be transformed into the yeast. Every precaution was made to ensure the level of doxycycline was high enough to inhibit expression of the proteins while glycerol stocks were generated, and initial transformations were performed. Once stocks of the yeast were in hand, yeast containing pCM173::LZD73 were grown in liquid culture with different concentrations of doxycycline. This was a quick way to determine if the antibiotic concentration had any effect. The yeast grown with no doxycycline showed no growth when the tubes were checked the next day and yeast grown with 1 µg/mL of doxycycline showed the most growth. The two middle concentrations (0.01 and 0.1 µg/mL) showed an amount of growth between that of the two extremes. The density of cells observed from 0.1 µg/mL was greater than that of 0.01 µg/mL, so that was the

selected concentration moving forward. The same experiment was also run on yeast that contained the unmodified pCM173::eV and there was no noticeable effect when the doxycycline concentration was varied, inferring that the observed decrease in growth has something to do with LZD expression.

Other experiments were planned to go along with this and show definitively that LZD was being expressed, and further optimizing the concentration of doxycycline to maximize LZD in the cells. A western blot was going to be run comparing extracts from yeast with high and low concentration doxycycline. Unfortunately, there was a lack of success when lysing yeast cells with several different methods. Attempts were made using Zymolyase to break down cell walls and generate spheroplasts, using kits specifically designed for western blot preparation, and using bead beating for mechanical lysis of the yeast. Samples that were processed never showed reasonable A280 values indicating proteins present, and even when run on SDS-PAGE there was no indication that any proteins were present in the prepared solutions.

The other experiment that would have given more information on the effect of LZD was a growth curve analysis. With the 96-well plate format there could have been more in-depth analysis done into what the true ideal concentration of doxycycline required for optimal LZD expression. A growth curve was attempted but the results were inconclusive, as many of the wells became occluded with too many cells and the readings just dropped to zero. If this were to be run again, a smaller concentration of starting cells would be used and there would be more consistent shaking of the plate to ensure cells do not fall out of solution and stick to the bottom of the wells.

Though it took some time to construct them, all five of the pCM173::LZD plasmids have been prepared in quantity in the event that a large number of transformations or

further cloning need to be done in the future. It is seen that the absence of doxycycline leads to the inability to grow cells, which means there is some confidence in saying LZD has biological activity in yeast. The plasmid is also stable as is evident by the ability to generate glycerol stocks that retain the ability to grow on SD-Trp media – plates confirming growth on media are shown in Figure 3.7. This is an important aspect of the project as it limits the number of components needed to be introduced in any single transformation.

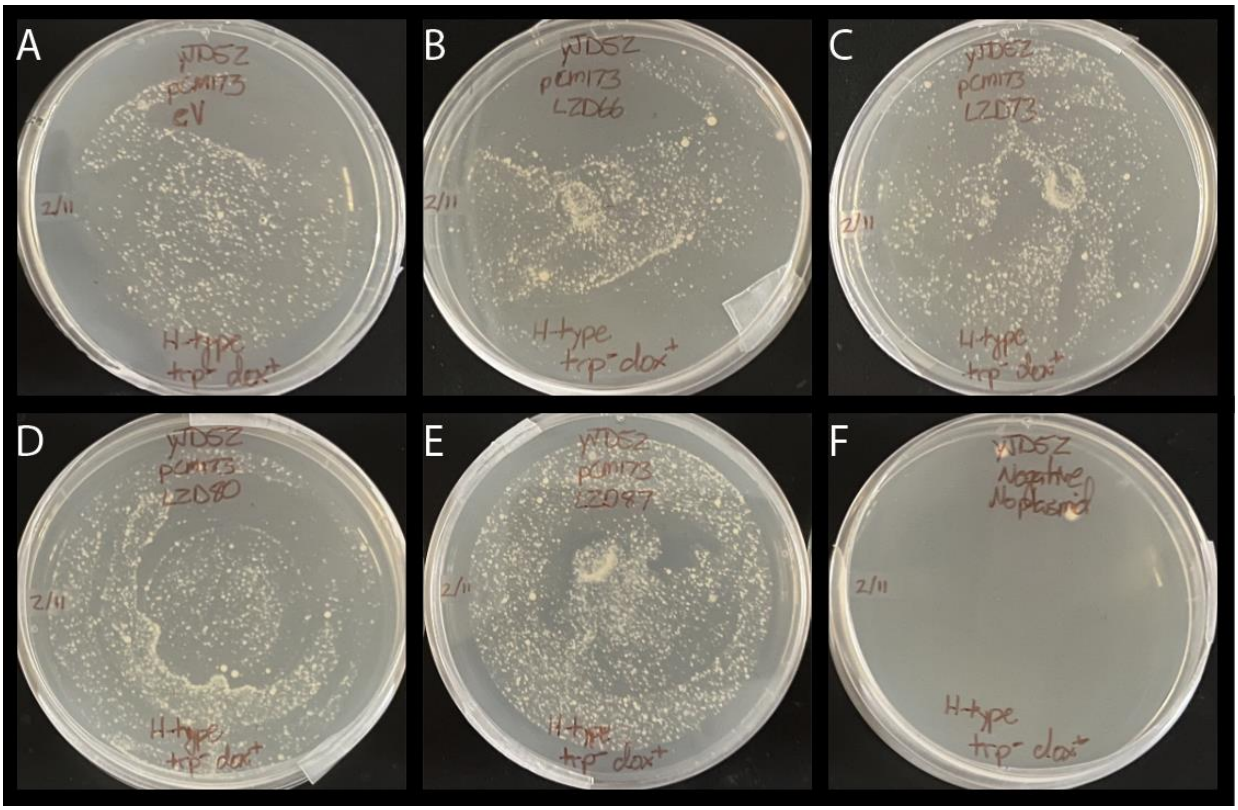

Figure 3.7 Confirmation of Yeast Growth in the Presence of Doxycycline and pCM173::LZD Variants: Plates are labeled with specific pCM173::LZD variants, and the labels are reproduced in this legend. All variants show growth on SD-Trp media with 1 µg/mL doxycycline. (A) yJD52 transformed with pCM173::eV. (B) yJD52 transformed with pCM173::LZD66. (C) yJD52 transformed with pCM173::LZD73. (D) yJD52 transformed with pCM173::LZD80. (E) yJD52 transformed with pCM173::LZD87. (F) Negative control plate – no input DNA was used in the transformation, and the untransformed yJD52 cannot grow in the absence of tryptophan.

Chapter 4: Confirming *in vivo* Expression of Cas9 Nuclease and sgRNA

Section 4.1: Overview

Another essential component in this system is an effective and targeted way to create DNA double strand breaks. While this project could have been completed similarly with TALENs or ZFNs, the decision to use a CRISPR/Cas based system came down to its relative ease of use, the ability to modify the target gene as needed more readily, and the broader utility to the community. In modifying a target sequence for TALENs or ZFNs there is a non-trivial amount of work necessary to generate new fusion proteins. With a CRISPR system the Cas9 protein and the sgRNA can be expressed separately *in vivo,* and the only modification needed to change the gene target is a sequence change at the plasmid level.

Section 4.2: Construction of pML107::AGC1

As previously discussed, the methods for introducing Cas9 nuclease and sgRNA are varied in the genome editing community. There are direct injection or electroporation methods that use pre-formed RNP complexes of enzyme and RNA, while other groups tend to use *in vivo* expression of both components through plasmids and viral vectors. Modifications have been made to more traditional two plasmid systems – one plasmid for expressing Cas9 while another plasmid expresses the sgRNA – combining them to make a single plasmid capable of expressing both necessary components. A paper from the Wyrick lab describes the pML107 plasmid[49], a shuttle vector capable of constitutive expression of both Cas9 and sgRNA in yeast. The vector backbone contains an *ampR* gene and a bacterial origin of replication for manipulation in *E. coli*. Retention and replication in yeast is possible due to leucine auxotrophy and the LEU2 gene present in the plasmid along with a segment of the yeast 2μ plasmid that is responsible for replication at a high copy number.

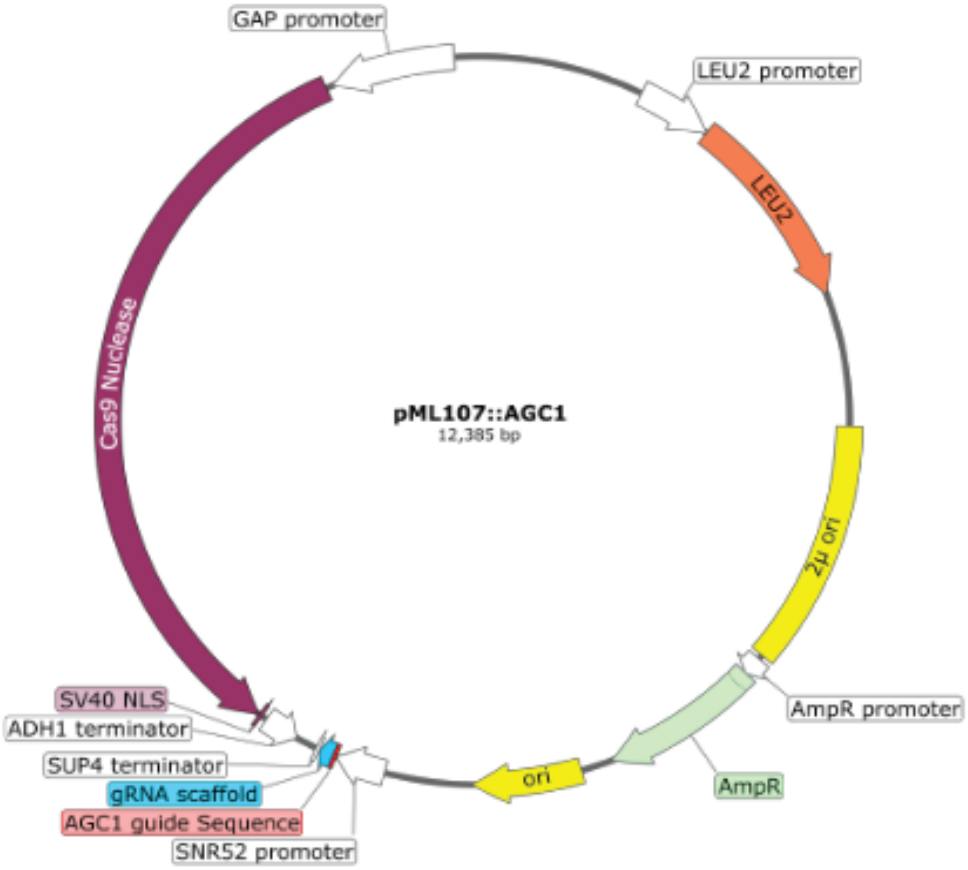

Figure 4.1 pML107::AGC1 Plasmid Map

Expression of the Cas9 nuclease is controlled by the general amino acid permease (GAP) promoter, which is constitutively active in yeast. To prevent Cas9 from having activity in bacterial cloning vectors, the RNA sequence expressed by the plasmid has an intron that is spliced out when in yeast but not bacterial systems. Additionally, the C-terminus of the nuclease has the SV40 NLS, which should bring the protein into the nucleus of the cell where it is required to act on the genome.

Expression of the sgRNA is controlled by the small nucleolar RNA 52 (SNR52) promoter, which is also constitutively active in yeast. For user modification there are convenient restriction enzyme sites immediately following the promoter sequence and

before the built in tracrRNA part of the sequence. The enzymes used are notably SwaI – a blunt end cutter – and BclI – an enzyme sensitive to DNA adenine methylase (*dam*) methylation. To select suitable crRNA sequences the Wyrick lab has also provided a website designed to search your gene of interest and select sequences that are adjacent to a recognized PAM sequence. The paper instructs users to synthesize a duplex pair of oligonucleotides as the insert, with a 4 base overhang to match the BclI site restriction fragment, and then perform a ligation reaction to insert new crRNA sequence into the restricted backbone as shown in Figure 4.2.

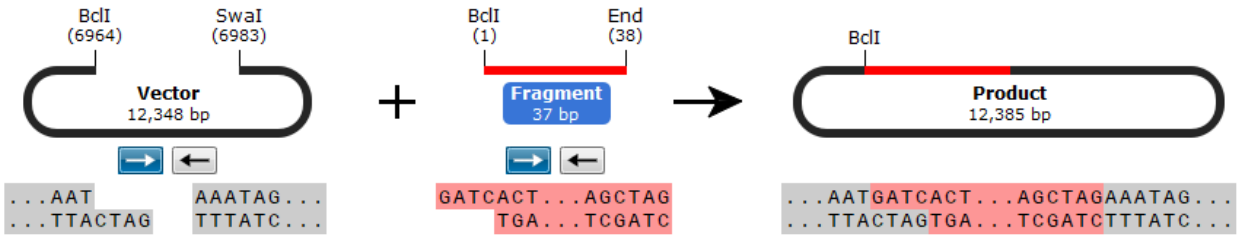

Figure 4.2 pML107::AGC1 Cloning Visual Workflow: Shown are the compatible DNA ends from the doubly digested vector pML107, and the fragment, annealed ssDNA oligonucleotides. These should ligate together forming the complete plasmid.

This restriction cloning method never produced viable bacterial colonies despite many repeated efforts. An alternative protocol was found from the Xin Gan lab at Nanchang University in China, who also reported having difficulties performing the original cloning protocol, specifically in regard to working with *dam*⁻ bacteria and blunt ended cloning, so they devised a method of doing whole plasmid amplification to generate the new vector by PCR – see Figure 4.3[56]. In this method, outward oriented PCR primers have 5' tails that comprise the crRNA fragment. After cycling the template is removed by DpnI, and linear PCR product is phosphorylated with T4 polynucleotide kinase (PNK) and then ligated. This method was successful, though the size of the plasmid – 12,300 base pairs – is not the ideal PCR template, and there was a lot of sequencing that needed to be

done to ensure the integrity of key sequences throughout the plasmid. After thorough confirmation of the plasmid, testing began on the Cas9/sgRNA activity in knocking out the target gene.

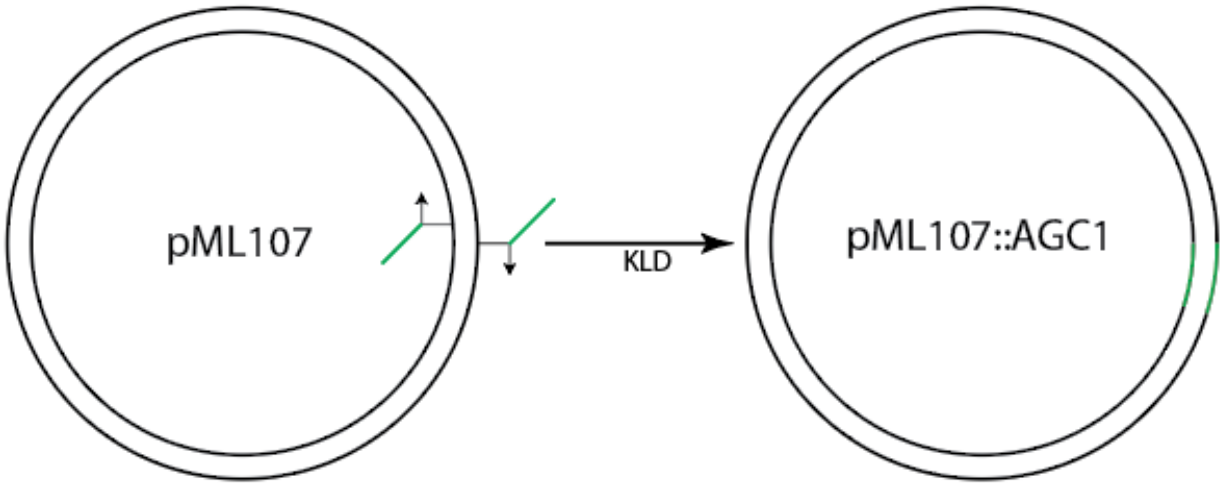

Figure 4.3 Whole Plasmid Amplification Followed by KLD: Homologous regions of primers are offset to remove a portion of the template from amplification. The 5' tails of the primers introduce new sequences. The linear product is then phosphorylated ("Kinased") and the ends are joined through blunt-end Ligation, forming the desired product. DpnI is used to destroy any remaining starting material.

Section 4.3: Plasmid Modifications

Throughout testing of the Cas9 activity and targeting of AGC1 a few issues that were impeding progress arose. First was the greatly reduced number of yeast colonies that grew upon transformation with the pML107::AGC1 plasmid, relative to the other plasmids being used. The root cause of this low efficiency could have been any number of factors including large plasmid size, overactive nuclease activity, or poor sgRNA targeting. Changes were made to the pML107 plasmid to alleviate the issues being seen based on literature on the subject. The next few sections describe these changes, included attempting

to express Cas9 through an inducible promoter, changing the gene target from AGC1 to ADE2, and finally protecting the sgRNA.

Section 4.3.1: Inducible Control of Cas9 Nuclease (MET25 Promoter)

The first change that was made to the pML107::AGC1 plasmid was to exchange the constitutive GAP promoter for the inducible MET25 promoter[57] as shown in Figure 4.4. As the promoter's name suggests, it is activated by the absence of the amino acid methionine and over ninety percent of downstream activity is shut off in the presence of only 0.5 g/L methionine. It was thought that this ability to control Cas9 by modifying media conditions would be enough to prevent overactivity of Cas9 if this were in fact the problem.

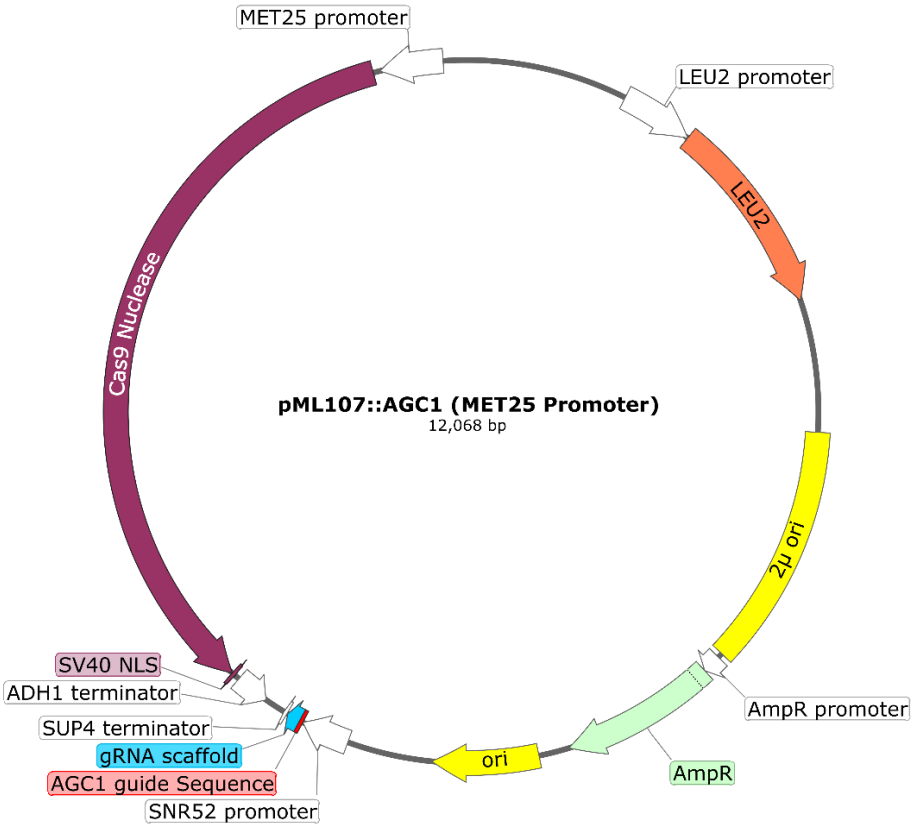

Figure 4.4 pML107::AGC1 (MET25 Promoter) Plasmid Map

The MET25 promoter sequence was obtained from Genewiz fragment gene synthesis and cloned into pML107::AGC1 that had been restricted to first remove the GAP promoter, using the HiFi assembly kit from NEB. Sequencing confirmed that the reaction was successful, and the DNA was then used to transform yeast. Regardless of methionine concentration the number of yeast colonies obtained was still insufficient for experiments. Even at high concentrations of methionine there is still leaky expression of downstream genes, but nonetheless this experiment suggested that nonspecific Cas9 activity is not the origin of pML107::AGC1's apparent toxicity.

Section 4.3.2: Change of Target Gene in the Yeast Genome

The next attempt to generate a plasmid that could be used to provide Cas9 activity was to change the target gene; this was pursued in parallel with modifying regulation of Cas9. The AGC1 gene was originally selected as our chromosomal target, as the knockout phenotype should have been easy to detect. Based on the literature, in the presence of acetate as a sole carbon source – replacing dextrose entirely – the only colonies that should grow are those that have an intact AGC1 gene[58]. To detect an AGC1 knockout would require replica plating to screen for colonies that grew on rich media but not on the selective media. The knockout should depend on a DSB being introduced by the Cas9/sgRNA, but not on the choice of HDR or NHEJ as a repair mechanism.

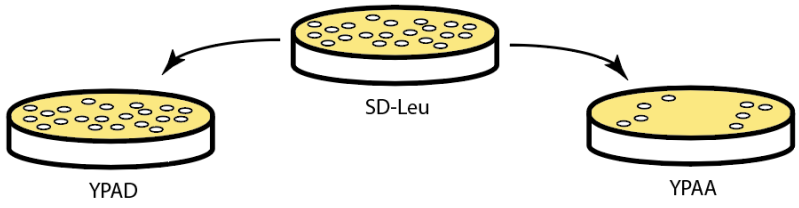

Figure 4.5 Screening for AGC1 Gene Function: Yeast transformed with pML107::AGC1 on SD-Leu are replica plated onto YPAD and YPAA. All colonies should grow on YPAD but only those that have an intact AGC1 gene should grow on YPAA.

While troubleshooting the pML107::AGC1 plasmid, a control yeast train was obtained from Euroscarf (Y07400 Genotype BY4741; *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YPR021c::kanMX4*) in which the AGC1 gene had been knocked out via replacement with a Kanamycin resistance cassette. This strain phenocopies our target HDR product, so it could be used to confirm the acetate$^-$ phenotype of AGC1 knockout yeast, an experiment depicted in Figure 4.5. The yeast successfully grew on rich media containing G418 antibiotic, but it also grew on acetate$^-$ media as shown in Figure 4.6 below. At the same time additional literature was found that indicated the AGC1 gene may not be the sole cause of a yeast's ability to grow on acetate substituted media; in retrospect the molecular explanations offered for the supposed AGC1 phenotype had never been convincing. The Palmieri group showed that the yeast mitochondrial carrier 2 (YMC2) gene is actually more responsible for this phenotype than AGC1[59]. This information, along with the lack of success with the concurrent construction of the pML107::AGC1 plasmid, led to a change in the target gene for our experiments.

Figure 4.6: Control Yeast on YPAA: Y07400 yeast from Euroscarf streaked onto YPAA media containing G418. As this yeast is an agc1 knockout it should not have been able to grow on media with only acetate as a carbon source. The ability to grow on G418 confirmed that the KanMX cassette is present in the yeast.

The newly chosen gene target was ADE2, which is a gene used in the purine biosynthetic pathway. This is a common laboratory control mutant, as a knockout in this gene is identifiable by a red pigment that mutant yeast produce in the absence of sufficient adenine in the media. The ADE2 gene codes for the enzyme phosphoribosylaminoimidazole carboxylase (AIR carboxylase)[60], which is responsible for catalyzing the sixth step in the *de novo* purine biosynthesis pathway. Without the enzyme, 5'-phosphoribosyl-5-aminoimidazole (AIR) is not converted to 5'-phosphoribosyl-4-carboxy-5-aminoimidazole (CAIR). This leads to build up of AIR in the vacuole of the yeast, where it is eventually oxidized into red pigment(s) that are visible on a plate as the reaction shows in Figure 4.7. Therefore, ADE2 knockouts are readily scored, again whether or not the change is caused by NHEJ or HDR.

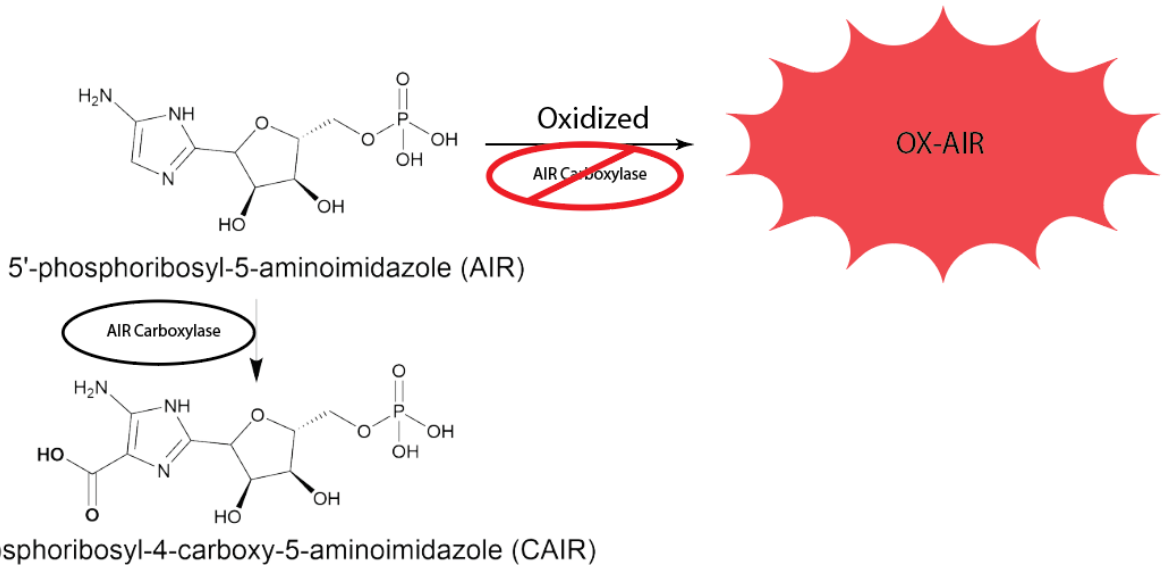

Figure 4.7 ADE2 Enzymatic Reaction: AIR Carboxylase (the product of the ADE2 gene) converts AIR into CAIR as part of the purine biosynthetic pathway in yeast. When ADE2 is knocked out, AIR accumulates in the cells and is oxidized in the vacuole. The oxidized product is red, and the accumulation leads to the cells visibly turning red.

To replace the sgRNA sequence complementary to AGC1 with one complementary to ADE2, new PCR primers were synthesized and then the previously described whole

plasmid PCR + ligation protocol was performed to generate the new pML107::ADE2 plasmid. In addition to constructing a new plasmid, the yeast host strain needed to be changed, as yJD52 is already an *ade2* mutant strain. Thanks to Dr. Dinman, a new yeast cell line yJD1112 (*MATa leu2 lys2 his4 trp1 ura3*) was obtained, which is a *trp1* and *leu2* mutant but has an unmodified copy of the ADE2 gene. As a reminder, the *trp1* phenotype is necessary for maintaining the LZD expression vectors based on pCM173, and yJD1112 was readily transformed with these plasmids in the absence of tryptophan. However, upon transformation of the pML107::ADE2 plasmid into yJD1112, both with and without pCM173::LZD, very few surviving colonies were seen, and none of them turned red as would have been expected from an ADE2 knockout.

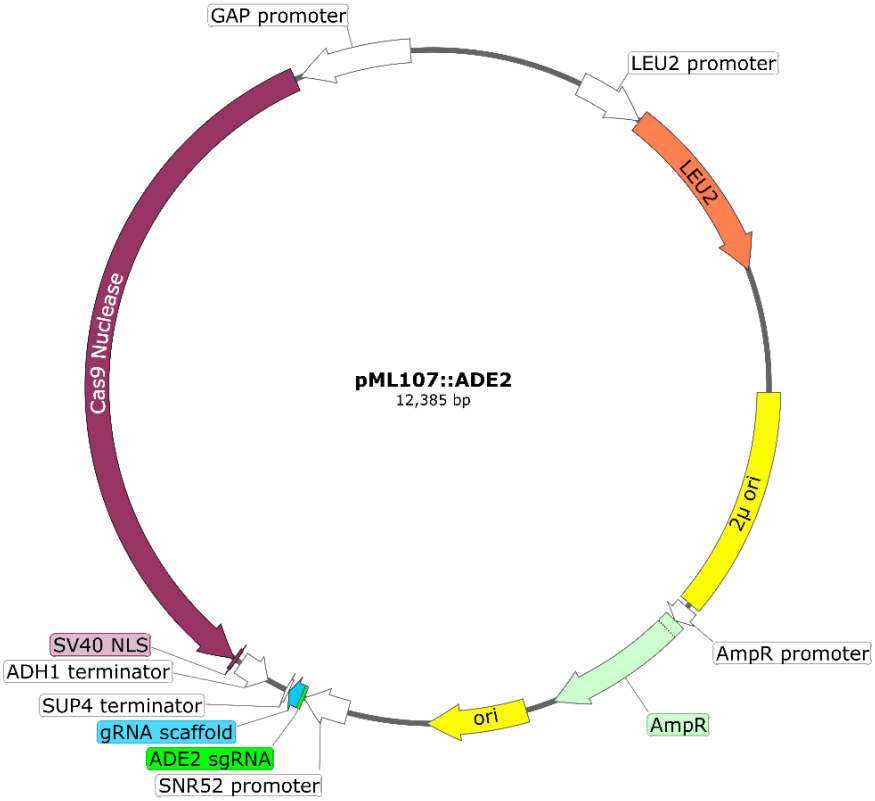

Figure 4.8 pML107::ADE2 Plasmid Map

Detection of the *ade2* knockout phenotype is not trivial. The yeast must be supplied with adenine in order to survive, but if they are given more than the minimum needed for growth the purine biosynthetic pathway is repressed and the AIR precursor for the red pigment is not synthesized. The *ade2* mutant yJD52 strain was grown with varying concentrations of adenine hemisulfate in the media and it was determined that 0.001% gave the most apparent red color in the shortest amount of time. At lower concentrations the yeast fail to grow, at concentrations above 0.001% the colonies are white as can be seen in Figure 4.9. Based on these findings it was concluded that if ADE2 had been knocked out upon transformation of pML107::ADE2 into yJD1112, it would have been detected.

Figure 4.9 Effects of ADE2 Knockout on Yeast: The two plates above show the same yeast (yJD52) but grown on media with different adenine concentrations. On the right, the media contains 0.003% adenine, on the left the media contains 0.001% adenine. The red coloring of the yeast is clear and easy to observe.

Section 4.3.3: Improved sgRNA Expression System (pG2)

A final modification to the Cas9/sgRNA expression system came from finding that unprotected sgRNAs tend to degrade and can lead to promiscuous nuclease activity. This explains the low number of colonies visualized, as unabated nuclease activity sooner leads

to cell death rather than infinite repair. One way that this has been averted is the use of 5' protecting groups that prevent degradation of the sgRNA targeting sequence. Jamie Cate's lab showed that by using a hepatitis D virus (HDV) ribozyme as a protecting group (shown in Figure 4.10) increased the intracellular levels of full sgRNA six-fold compared to no protecting group[44,61]. The group tested different promoters for the sgRNA and eventually settled on the $tRNA^{tyr}$ promoter, though the increase in activity between that and the SNR52 promoter was negligible. The HDV ribozyme has 5' cleavage activity that removes any promoter sequence making the use of any promoter in any system more seamless.

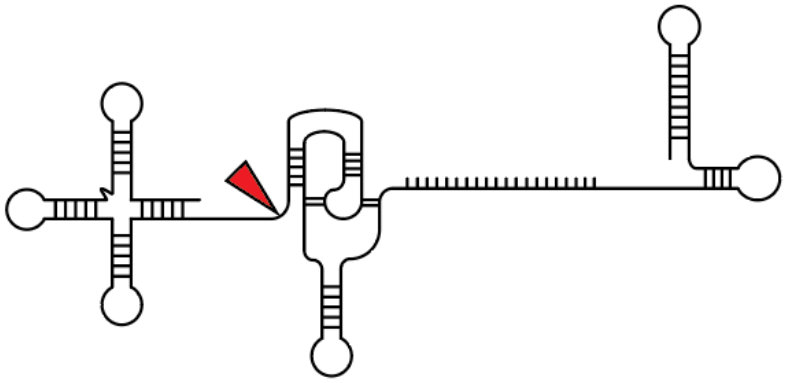

Figure 4.10 Protected sgRNA with 5' HDV Ribozyme: Cartoon representing the overall structure of the expressed sgRNA from pG2. The 5' tRNA is cleaved by the HDV Ribozyme at the red arrow. The 3' half of the molecule is the sgRNA targeting the ADE2 gene.

A plasmid was found called pG2 that implements this HDV-ribozyme protection strategy. The plasmid was acquired in addition to the control plasmid pEmpty-Control – which simply omits the gene targeting sequence – as a gift from Lars Steinmetz at Stanford, via Addgene (Addgene plasmid #162603; http://n2t.net/addgene:162603; RRID:Addgene_162603 and Addgene plasmid #162607; http://n2t.net/addgene:162607; RRID:Addgene_162607). The sgRNA sequence in pG2 is designed to target the ADE2 gene one base upstream of the previously designed sgRNA in pML107::ADE2.

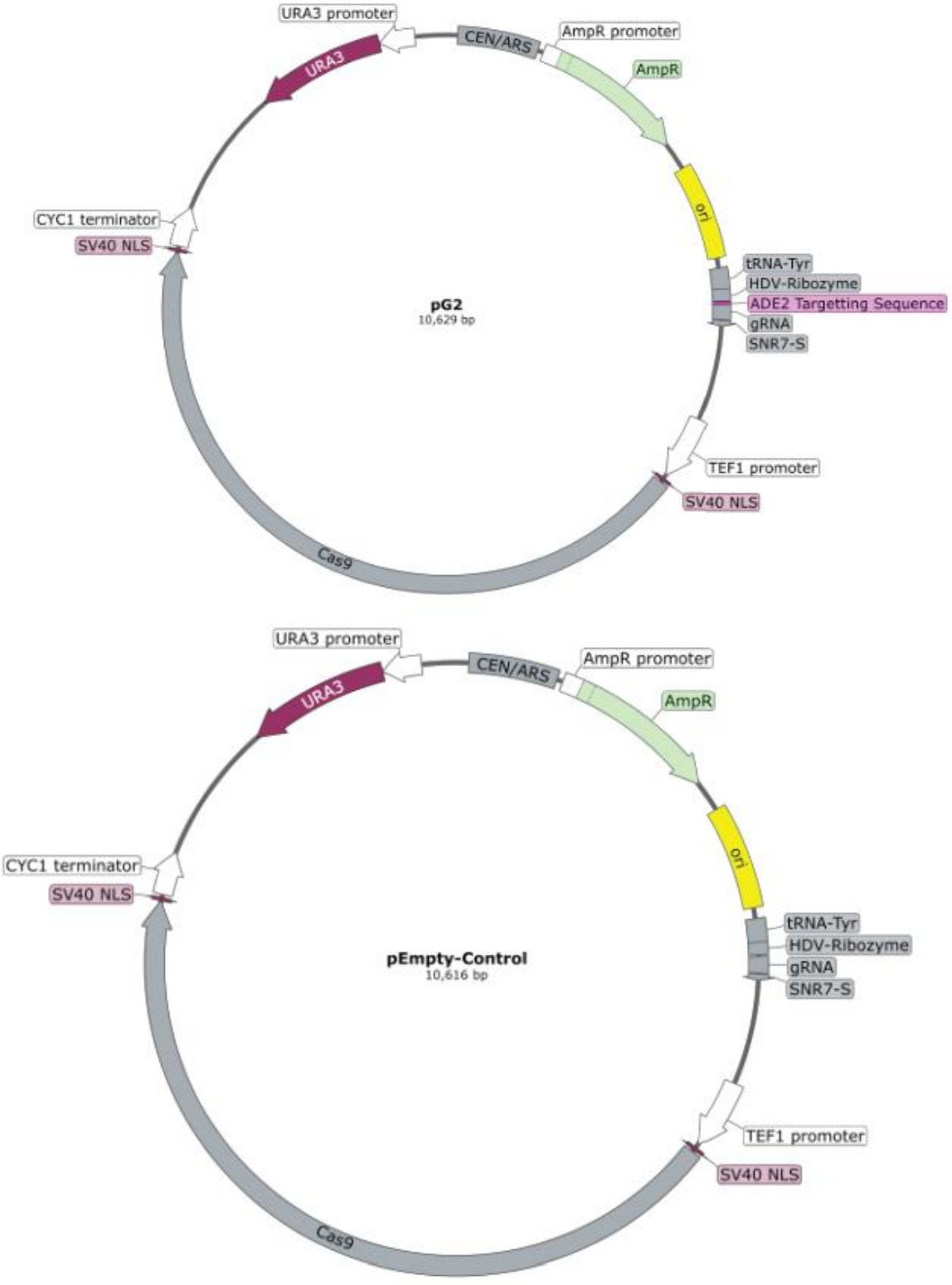

Figure 4.11 pG2 and pEmpty-Control Plasmid Maps

In addition to the 5'-HDV ribozyme and tRNA promoter there were a number of small differences between pG2 and the pML107 vectors, as they are not derived from the same parent vector. First instead of the 2μ origin, this plasmid again used the CEN/ARS origin of replication similar to the pCM173 vectors used for LZD expression, meaning that pG2 is a low copy number plasmid. The pG2 plasmid also uses URA3 instead of LEU2 as the auxotrophic marker. The plasmid pEmpty-Control is important as a control for the transformation efficiency of a large plasmid. It also offers the convenience of using the same uracil–media whether or not double strand break is being induced in the ADE2 gene.

In order to use the pG2 plasmid in experiments a new yeast strain needed to be obtained, with the help of Dr. Dinman. yJD100 (*MATa ura3-52 his3 trp1Δ K1*$^{+}$), which has both *ura3* and *trp1* knocked out and has an intact copy of the ADE2 gene fit all the genotype needs. Once the yeast was confirmed to have the correct genotype, the yeast was transformed with pG2. Over one hundred colonies grew on the SD-Ura plates including 0.001% adenine hemisulfate, with 50% - 75% of the colonies turning red, indicating the ADE2 gene had been knocked out through Cas9 nuclease activity. When the same experiment was repeated with pEmpty-Control no red colonies were seen.

Section 4.4: Materials and Methods

All enzymes used in reactions including restriction digestions and PCR were obtained from NEB. Plasmid and DNA purification kits were obtained from Qiagen. Oligonucleotides used as primers were synthesized by IDT. FragmentGenes were synthesized and obtained from Azenta. XL10-Gold Electrocompetent *E. coli* were purchased from Agilent. All other reagents used were purchased from Fisher Scientific. PCRs were run in an MJ research PTC-200 Gradient Thermal Cycler.

Plasmid pML107 was obtained as a gift from John Wyrick (Addgene plasmid # 67639; http://n2t.net/addgene:67639; RRID: Addgene_67639) in the form of a bacterial stab. *E. coli* from the stab were grown on standard LB-amp agar to retain the plasmid. Individual colonies were further inoculated overnight in 5 mL of LB-amp liquid media. Bacterial suspensions were then taken through Qiagen Miniprep or Midiprep protocols to purify plasmid stocks for future use and stored at -20 °C. Prior to storing the plasmids were quantitated on a Cary UV/vis spectrometer to note purity of nucleic acids, and then run on 0.7 % agarose gels with a TBE running buffer to confirm size and a single band. Once the run was completed, gels were stained with EtBr and visualized on a UV transilluminator. Plasmids were also sequenced at Azenta using M13 forward and reverse primers to confirm accuracy of the plasmid sequences. Later plasmids were sequenced by Plasmidsaurus, which provides affordable whole plasmid sequences using Oxford Nanopore Technologies' newest long-read sequencing technology.

The pML107 plasmid then was transformed into competent JM110 *E. coli*, which is a *dam*⁻ cell line. These transformations were performed chemically as previously described in chapter 3. The cells were obtained from a professor on campus who was unsure of the lineage (Dr. Scott Juntti, Dept. of Biology) so confirmation needed to be done to ensure that the purified plasmids were in fact unmethylated. To do so, 2 µL of Miniprepped pML107 plasmid from the JM110 cells was digested with DpnI, which specifically targets GATC sites that have been methylated by *dam*, and the plasmids remained undigested as desired. The plasmids were then digested for use in cloning in two steps, first with SwaI and then with BclI. For the SwaI reaction 1 µg of DNA was added to 1 × CutSmart buffer and 20 units of SwaI restriction enzyme. This reaction goes overnight at 25 °C. The following day the reaction is heat inactivated at 65 °C for 20 minutes before being cooled back down to

room temperature. Then BclI is added to the reaction mixture and incubated at 37 °C for an additional 2 hours. This reaction is then heat inactivated at 65 °C for 20 minutes. The doubly restricted vector was confirmed via a 0.7 % agarose gel as previously described in chapter 3. Since only 19 bp are excised, it was difficult to ensure that both restriction enzymes were able to cut the same DNA fragment, but it could be demonstrated that each one worked independently. Attempts were made to purify the linear fragment from the agarose gel without success. Processing of the vector for ligation reactions was done with the heat inactivated, doubly restricted product.

Oligonucleotides were selected to target the AGC1 (and later ADE2) gene based on the recommendations from the website provided in Dr. Wyrick's paper (updated link: https://wyrickbioinfo2.vetmed.wsu.edu/cgi-bin/crispr_search.pl). Received oligonucleotides were solubilized according to IDT's specifications and then purified via PAGE as follows: The DNA was run on a 15%, 29:1 polyacrylamide gel containing 8 M urea and 1 × TBE buffer. The running buffer was also 1 × TBE. The DNA was visualized by UV shadowing, and the band was excised was cut with a single use razor. The DNA slice was further cut with the razor and added to an Eppendorf tube. The mass of the gel slice was measured, three volumes of cut and soak buffer were added (300 mM Sodium Acetate, 1 mM EDTA, pH 8), and the mixture sat on a shaking table overnight. The supernatant was then collected and precipitated with ethanol. The purified samples were quantified on a Cary UV/vis spectrophotometer. The single stranded DNA was then annealed by mixing equimolar quantities together with an equal volume of 2 × T4 DNA ligase buffer, heating to 95 °C, and allowing a slow cool in a powered off heat block. The duplex was confirmed via polyacrylamide gel electrophoresis as previously described.

To ligate the double stranded insert and vector the two were mixed in a 40:1 molar ratio, with 1 × T4 DNA ligase buffer and 400 units of T4 DNA ligase. The reaction mixture was incubated at 16 °C overnight. The following day the reaction was heat inactivated at 65 °C and then reactions cooled to room temperature before being used in bacterial transformations. Bacterial transformations were carried out using the electrocompetent cell method described in chapter 3, except using XL10 Gold *E. coli*, which have a higher efficiency for larger plasmids. The transformation mixtures were plated onto LB-amp agar and incubated overnight at 37 °C. These transformations were unsuccessful.

Another protocol for generating the pML107::AGC1 (and later pML107::ADE2) plasmid was performed using whole plasmid amplification by PCR. Primers were designed with 5' tails containing the requisite recognition sequence for the AGC1 gene (and later the ADE2 gene) that would generate a linear amplicon identical to the desired plasmid. The reactions were performed in 50 µL reaction with 1 × Q5 reaction buffer, 200 µM dNTPs, 600 nM each primer (forward and reverse), 1 unit of Q5 DNA polymerase, and 8 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 60 seconds

Cycle (25 repeats):

98 °C – 10 seconds

55 °C – 30 seconds

72 °C – 8 minutes

Final Extension: 72 °C – 5 minutes

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed, reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. Samples were then treated

with T4 PNK to phosphorylate 5' ends for ligation. T4 PNK reaction mixtures used 1 × T4 DNA ligase buffer (which includes 1 mM ATP), 10 units of T4 PNK and 2 µL of PCR product. The reaction incubated at 37 °C for 30 minutes. The enzyme was then heat inactivated by incubating at 65 °C for 20 minutes. After cooling to room temperature 5 µL of the reaction mixture was mixed with an additional 1 µL of T4 DNA Ligase buffer (with ATP) and 400 units of T4 DNA ligase and left to incubate at 16 °C overnight. Bacterial transformations were carried out using the electrocompetent cell method described previously. The transformation mixtures were plated onto LB-amp agar and incubated overnight at 37 °C. As noted, this procedure was also used to generate the plasmid with the ADE2 targeting sgRNA.

The plasmid modification to introduce the MET25 promoter in place of the GAP promoter was performed in three steps. The GAP promoter was excised by restriction with the HindIII-HF and NcoI-HF enzymes. Both are 100% active in CutSmart, so the reaction was set up with 20 units of each enzyme, 1 × CutSmart buffer, and 1 µg of plasmid. The reaction incubated at 37 °C for an hour before being run on a 0.7 % agarose gel as previously described. With the size of the linear fragment confirmed, the reaction was purified of enzymes using a phenol/chloroform extraction, followed by ethanol precipitation. The MET25 promoter sequence was identified and a FragmentGene (Azenta) was designed that included homology to the pML107 plasmid so that a HiFi reaction could be performed. The insert and vector were mixed in a 2:1 ratio and added to a 1 × HiFi master mix solution. The 20 µL total reaction was incubated at 50 °C for 1 hour. Reactions were then allowed to cool to room temperature before being used in bacterial transformations as previously described.

Yeast transformations using the pML107 plasmids were performed using the protocol described in chapter 3, except the media used to plate out the pML107-containing yeast was SD-Leu instead of SD-Trp because of the LEU2 marker in pML107. To confirm the activity of Cas9 and the ability to detect the loss of function associated with AGC1 knockouts, yJD52 was transformed with 500 ng of pML107::AGC1 and plated onto SD-Leu agar. Colonies that grew on SD-Leu after the transformations were then replica plated onto rich media containing yeast extract, peptone, adenine sulfate, and acetate (YPAA).

Yeast transformations using the ADE2 targeting plasmid pML107::ADE2 required a new yeast strain, as yJD52 is already an ADE2 knockout strain. To this end yJD1112 was used which still contained a knocked out *trp1* and *leu2* gene so as to not disrupt the plasmid systems in place. The new yeast strain was transformed with pML107::ADE2 and instead of replica plating from SD-Leu to YPAA a change in color for the yeast that had become *ade2* knockouts should have been seen.

To test the ability of the MET25 promoter to regulate the activity of Cas9 the same series of transformations were used with a small modification to the plate conditions. In the absence of methionine, the promoter should induce expression of the Cas9 nuclease, whereas in the presence of 670 µM methionine, there should be no expression of Cas9. These two conditions, SD-Leu + Met and SD-Leu-Met, were used to plate out transformations with the pML107::ADE2+MET25p plasmid to compare the results.

To use pG2 and pEmpty-Control plasmids the yeast strain was changed again to match the new URA3 auxotrophic marker in the new plasmid. Dr. Dinman was gracious enough to let us use another strain yJD100, which contains an intact ADE2 gene and knockouts for the TRP1 gene and URA3 gene. Transformations were performed as

previously described to test the Cas9 activity of the new plasmid. Transformants were plated this time onto SD-Ura media to retain either the pG2 or pEmpty-Control plasmids.

Section 4.5: Discussion

As with the pCM173::LZD plasmids described in the previous chapter, the pML107 modified plasmid should have been easy to construct based on the published literature. The scheme involved annealing short oligonucleotides to create a duplex with one blunt end and one end with an overhang. This insert could then be ligated into the vector that had been cut with restriction enzymes BclI and SwaI. BclI leaves a 4 base overhang and SwaI is a blunt end cutter. Two major factors played into the problems with the original construction scheme: The necessity to have pML107 be replicated in a *dam*⁻ host and overcoming the low efficiency of blunt end ligation. The first problem was solved using JM110 competent *E. coli*, but the second proved insuperable.

The insert oligonucleotides were easy to anneal and should have posed no problems, but upon ligation with the double digested vector, there were no colonies. To address this, the ligation was performed at a 100:1 insert: vector molar ratio, but this still showed no success. Knowing that blunt ended ligation is highly inefficient, attempts were also made to phosphorylate the insert ends with T4 PNK as a way to aid in overall ligation success. The putatively phosphorylated insert was then used in the same ligation reaction, still with no success.

While in the literature a source was found that also struggled with the restriction cloning method of modifying pML107. This group used PCR to amplify the entire 12.3 kb plasmid, using 5' tailed primers that incorporate the sequence that normally would have been the oligo insert. The authors then took the long amplicon through a kinase, ligase and DpnI (KLD) protocol with a kit from NEB that combines all the enzymes into a single step.

This protocol was used through the PCR and after confirming an amplicon of the proper length, it was modified with an in-house protocol for KLD. In this method the amplicon is digested with DpnI first, followed by a kinase reaction with T4 PNK, finishing with an overnight ligation with T4 DNA Ligase. As was evident in the past, this homebrew developed KLD worked, and the desired plasmid was able to successfully be obtained.

Confirming the pML107::AGC1 plasmid had the right sequence was not as straightforward as confirming the pCM173::LZD variants. There was no unique restriction enzyme cleavage sequence in the 19-base excised portion of the original pML107, and similarly no unique cleavage sequence in the 31-base insert. The only universal sequencing primer (M13 Forward) is located over 5,000 bases away from the region of interest in the plasmid. To address this, two solutions were used that would indirectly determine if the sequence was at the very least not the original pML107, and at the very best indicate that the plasmid was properly constructed.

The first method involved retransforming and purifying the plasmid with JM110 *E. coli* and testing if the original BclI and SwaI digestion would work. Neither restriction enzyme sequence should be present in the construct after ligation with a new insert. This method worked, as the original pML107 stock and the pML107::AGC1 plasmid gave different restriction patterns on an agarose gel. The second method involved the use of different restriction enzymes around the modified location in the plasmid. Using the EagI and NheI enzymes, two differently sized products were expected when run on an agarose gel (304 bp and 322 bp). The difference between the two is not large, but on a 1% agarose gel there was enough to notice a difference between the two fragments.

At this point the plasmid could be transformed and the overall effect of Cas9 and knock out the AGC1 gene needed to be assessed. The transformations were plated onto SD-

Leu agar and grown for 3-4 days, at which point the plates were inspected only to find less than 15 colonies grew. This was in stark contrast to the previous transformations with pCM173::LZD where hundreds of colonies were present on selective media. The low number of colonies could be attributed to a relative lack of efficiency when it comes to NHEJ repair of the genome. If NHEJ is not effectively repairing DNA breaks, then it is reasonable to assume that the yeast could be experiencing a higher rate of death than if no DNA breaks were occurring. However, this trend of seeing low colony counts was still evident when using the unmodified pML107 plasmid, which shouldn't have been targeting any location in the genome. Aside from low colony counts, it was also troubling that the colonies that were collected and streaked onto acetate containing media were growing with no issues.

Addressing the issue of why all of the cells grew on acetate led to some interesting literature that complicated the *agc1* phenotype. There is apparently another gene (YMC2) that has a similar effect on yeast's ability to metabolize and use acetate as a replacement carbon source. This meant that to continue using AGC1 as the target gene would have entailed competition with a redundant pathway, such that *agc1* knockout yeast still grow on media containing only acetate. It is unclear to us how the *agc1* growth defect originally occurred and how we able to observe it occasionally: perhaps the regulation of YMC2 is complex or it is readily knocked out. With this information in hand, the target gene was changed to one that was better understood in the literature, as well as one that more readily will be seen if a gene knockout had occurred. ADE2 was selected because it is commonly used in laboratory experiments and is responsible for turning yeast red when disrupted. The red color comes from a buildup of pre-adenine compounds getting oxidized into a red chromophore in the vacuole of the yeast.

Exchanging the AGC1 for an ADE2 targeting sequences in pML107 was easy, as all that was required was a second PCR with new primers containing the ADE2 specific sequence. When plasmids were obtained, the full plasmid was sequenced using Plasmidsaurus. The sample submission drop box is available in Plant Sciences building at the top of the spiral staircase. The service offers just as rapid a turnaround time as Azenta or ACGT do, but the results are the entire plasmid, not just 1000 bases of Sanger sequencing from available primers. The service certainly made all subsequent sequencing efforts simpler, as there was no concerns as to whether there were anomalies in non-sequenced portions of the plasmid. Unfortunately, there was never evidence that the pML107::ADE2 sgRNA plasmid afforded cleavage of the ADE2 locus.

Another explanation for the low number of colonies seen was that an overactive nuclease could introduce repeated DSB events that overwhelm the yeast repair pathways, leading to replication through the DSB and cell death. This issue was addressed by swapping the constitutively active GAP promoter for the repressible MET25 promoter. Literature showed that using a concentration of ≥ 670 µM methionine should shut of the promoter and prevent expression of downstream proteins. In a perfect world this would have solved the problems and allowed the maintenance of glycerol stocked yeast containing a plasmid for controlled expression of LZD and also for the controlled expression of Cas9. Unfortunately, this was not the case: transformation with the pML107::ADE2 (MET25 Promoter) plasmid still gave very few colonies, and they were not red, regardless of the methionine concentration.

It was then considered that the sgRNA chosen is not specific enough to the ADE2 gene, and is causing multiple DSBs throughout the genome, which would also be too much damage for the yeast to overcome. This seems unlikely, because running the sgRNA

sequence through a sequence alignment tool like BLAST (Basic Local Alignment Search Tool) does not show additional matches to the 17-base sequence. However, it could not be ruled out entirely.

Another source targeting the ADE2 gene was found using a plasmid called pG2. The sgRNA targeted a site a single base away from the target in pML107::ADE2; the sgRNA was shifted forward one base to use a different PAM sequence. This plasmid also contained an HDV-ribozyme on the 5' end of the sgRNA, which was intentionally placed to be a protecting group and prevent degradation of the sgRNA sequence. With this knowledge and previous results, it is reasonable to think that the sgRNA sequences that were being expressed from pML107 variants were being degraded to some extent and making the Cas9 nuclease less selective in the sequences it was targeting. This leads to the same conclusion that an overactive nuclease would lead to more DSBs than the yeast could manage to repair, leading to cell death. In the end, pML107 and its derivatives were difficult to work with and never demonstrated specific Cas9 activity in our hands. As the focus of the work is the steps subsequent to inducing a double strand break, there was no downside to moving on from pML107 and using pG2.

When pG2 was transformed into the yJD100 yeast strain, a much larger number of yeast colonies (>100) grew when compared to that of pML107, and the majority of them were red. It should also be noted that when pEmpty-Control is transformed into the same yeast strain, a large increase in the number of colonies, (~ 500-1000) is seen, but none turn red. These results can be explained by the lack of Cas9/sgRNA activity coming from the control plasmid. Without nuclease activity there is less stress on the yeast, as there will be fewer induced DSBs, and naturally nothing should be targeting the ADE2 gene specifically,

so the yeast tend to grow better and none of them give the red color that is expected when pG2 is present.

In summary, the sgRNA protection provided from HDV sequence built into the pG2 plasmid, which is not present in the pML107 plasmid, is the reason that the pG2 plasmid is successful in knocking out the ADE2 gene. The plasmids are both over 10 kb, so the size of the plasmids does not play a role in the relative overall transformation efficiency. Additionally, the system has moved on from targeting the AGC1 gene due to irreproducible results with the mutants as well as the lack of concrete information in the literature to back up the claim that a single knockout would induce the correct phenotype. The switch to targeting ADE2 also affords the ability to do a single transformation and determine exactly how many colonies have the red phenotype, without needing to replica plate or restreak colonies. In short, a reliable Cas9 and sgRNA expression system that works in yeast was developed: this checks all the boxes needed to move forward with this project.

Chapter 5: Generation of Homologous Donor DNA

Section 5.1: Overview

The final component in this system is the DNA that will be used as a template for homologous recombination, also known as the donor DNA. Ther is the most freedom to design this component, as its only purpose is to provide a sequence that will be integrated into the genome, but not replicate or reside in the yeast on its own. There is no need for a eukaryotic origin of replication or auxotrophic markers in this plasmid, making it the only plasmid in our system that is not a shuttle vector. Linear DNA or single-stranded DNA can also be used as a donor template.

Section 5.2: Sequence Selection for Homologous Template

In designing the donor DNA, there were a few main points to keep in mind. First the plasmid, or the linear donor, needed to have a gene not present in our yeast strain to create a gain of function that could be screened through phenotypic changes. This was the KanMX cassette, a common yeast insert that contains the sequence for the TEF promoter and terminator flanking the coding sequence for the *kanR* gene. In eukaryotes the *kanR* or kanamycin resistance gene does not confer kanamycin resistance, but instead confers resistance to another antibiotic: geneticin, or G418, both molecules shown in Figure 5.1 below. G418 acts by binding to the ribosome of organisms and preventing protein synthesis, leading to death of the organism[62]. The inclusion of the accompanying promoter and terminator also means there is no need for in-frame insertions or other coordination with outside sequences: the KanMX cassette is self-contained.

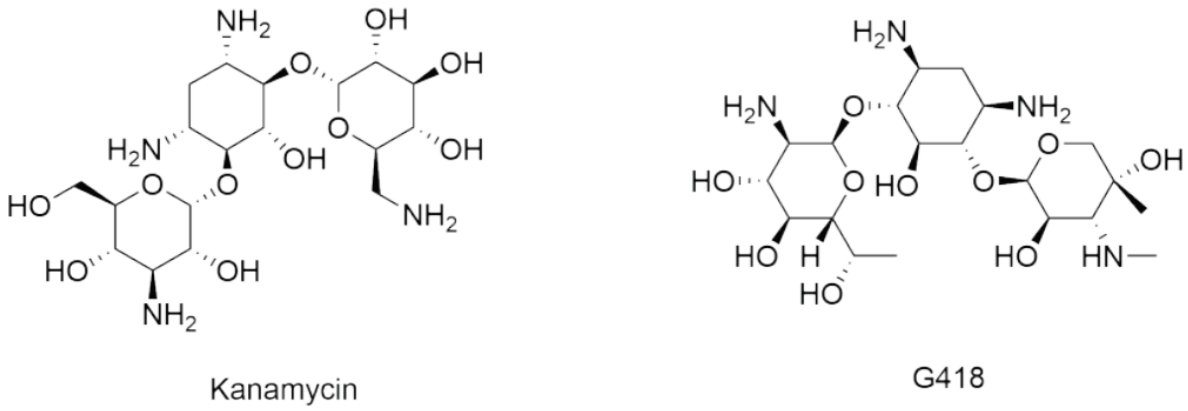

Figure 5.1 Aminoglycoside Inhibitors Kanamycin and G418

The next requirement for the donor is a pair of sequence regions flanking the integration cassette that are homologous to the target gene in the chromosome. Most recombination experiments in yeast are performed with homologous regions ranging from 40 base pairs up to 150 base pairs. Due to the reputation of yeast as being so proficient at HDR, it was important to make sure the background recombination was not so high that increased recombination would not be seen. For this reason, the initial design of the plasmid contained sequences with a sub-optimal length of 25 base pairs, with the hope that inclusion of the LZD protein would be able to overcome that deficiency and facilitate recombination relative to a low background.

These homologous regions were selected to be about 850 base pairs apart from each other, straddling the proposed cut site as shown in Figure 5.2. This distance is thought to be sufficient in allowing the loose ends of genomic DNA sufficient space to find their homologous counterparts in the donor DNA. Current literature has not reached a consensus for the ideal distance between homologous regions and DSB location, with some groups showing an increase in HR at 10 bp, while others indicate the distance can be over 100 bp[63,64]. An additional consideration for the positioning of homologous regions is that the

more of a gene that is removed in the HDR process, the less likely it is that the gene will revert and regain function. If HDR occurs, the Cas9 target sequence will be removed during the repair process, meaning that the nuclease activity would not act on the region and there should not be repeated rounds of repair or further genomic disruption.

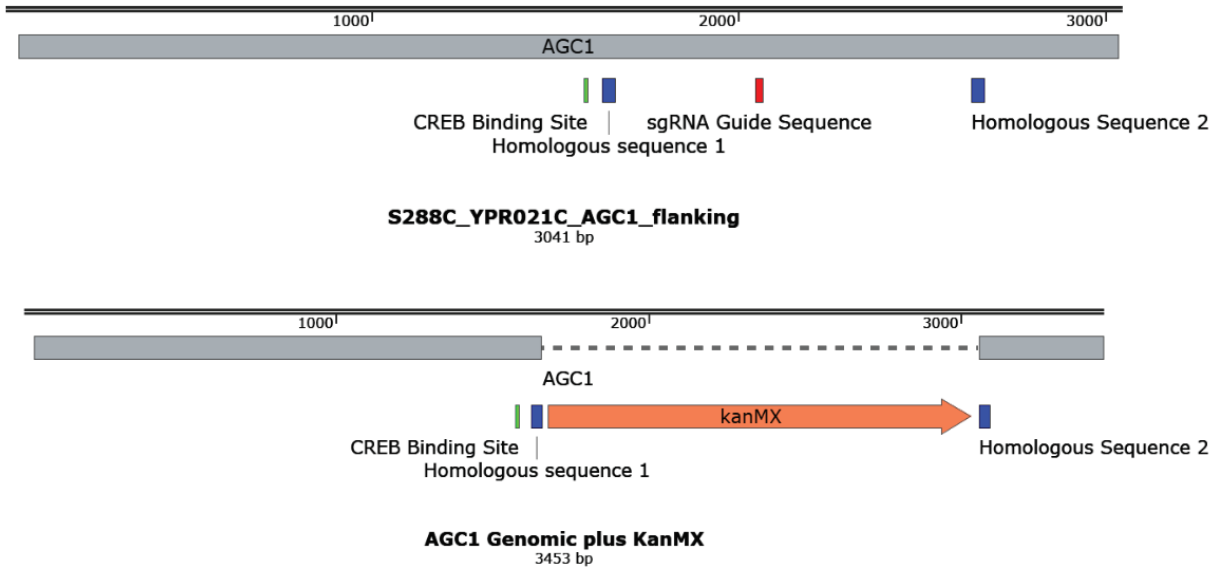

Figure 5.2 Annotated AGC1 Gene: The top sequence shows the intact AGC1 sequence with relevant features in our design. The blue sequences are the homologous sequences present in the pUC57-agc1::KanMX donor. The red sequence is the sgRNA guide sequence present in pML107::AGC1. The green sequence is the CREB binding site that would be bound by LZD73 upon expression from pCM173::LZD73. The bottom sequence shows where the KanMX cassette is designed to insert into the AGC1 gene.

Another element of the donor sequence is the INV-2 binding sequence. This is the DNA binding sequence for the C-terminal end of the LZD proteins and is less commonly found in *S. cerevisiae* than the CREB or AP-1 site that the N-terminal DNA binding domain recognizes. This short 10 base sequence was included in the donor to the 5' side of the KanMX cassette. This intentional decision was made because the closest CREB sequence in the genome is upstream of the proposed cleavage site. This design allows optimal positioning, as the homology between the donor and the 5' untranslated region (UTR) of

AGC1 is close to both the CREB site on the target and the INV-2 site on the donor. It is likely that the lengths and locations of the homologous regions and the locations of the CREB and INV-2 sites will need to be optimized as our method is developed further. Furthermore, different sequence combinations might be optimal for different LZD proteins.

The final considerations for the donor DNA are strictly for cloning and transformation purposes in both bacteria and yeast. Most importantly, there can be no auxotrophic markers or yeast origins of replication in this plasmid: it needs to be non-replicating in yeast so that antibiotic resistance (or relief of auxotrophy if a different cassette is chosen in the future) is not conferred just by transforming it into the organism. Similar to the other plasmids it is preferrable to include bacterial control elements like the *ampR* gene and origin of replication. Since this plasmid is going to be transformed into the system at a larger amount than the other components more of it will be needed, and the easiest way to do so is by cloning into bacteria. In some experiments the donor may also be amplified by PCR so that it does not contain any of the bacterial control sequences and use the amplicon to transform.

This original donor plasmid was generated by FragmentGene synthesis and subsequently cloned into a pUC57 vector by Genewiz/Azenta. The plasmid that was received was transformed into *E. coli* and then sequence confirmed before being used in experiments.

Section 5.3: Plasmid Modifications

Thinking forward to other experiments, there were a variety of sequence elements included in the donor design in order to accommodate future controls or changes that could potentially occur. One of those changes was building in the ability to swap out the INV-2 sequence for a CREB sequence or a random DNA sequence. The rationale behind this is to

make a donor DNA that is not be recruited to the DSB site in the genome. This functions as a control to show that the specific nature of the LZD binding domains specifically brings together donor and genome only when both unique sequences are present.

To make this change in the plasmid, PCR primers were designed with 5' tails containing either the 10 base CREB sequence, or a random 10 bases. The primers anneal to the template and generate a linear amplicon that is absent of the INV-2 binding site, and in its place contains either the CREB binding sequence or a random 10 base sequence, depending on the primers used. The PCR then generated linear blunt-ended products that would be phosphorylated with T4 PNK and then ligated. Transformations into *E. coli* were performed, and the plasmids isolated and confirmed by sequencing.

Section 5.3.1: Change Homologous Regions to Match New Gene Target

Another change to the donor DNA was modifying the regions of target homology to change the target gene from AGC1 to ADE2. The new homologous regions were selected, still at 25 base-pairs in length, and at a similar distance from the proposed cut site as before. The new 1500 bp gene sequence was generated by Genewiz/Azenta and then cloned into our pUC57 vector by a HiFi method. At this same time, another donor was generated containing a yeast enhanced green fluorescent protein (yEGFP)[65] sequence instead of the *kanR* gene – but still using the TEF promoter and terminator in the cassette. This was performed by HiFi cloning of a designed FragmentGene containing the GFP/TEF cassette with the same restricted pUC57 vector with ADE2 homology. All sequences were confirmed via gel electrophoresis restriction maps and sequencing.

Section 5.3.2: Extension of Homologous Regions

It was eventually confirmed that targeting ADE2 with Cas9/sgRNA, due to the appearance of red colonies, but there was no evidence of HDR, which would have been the

ability to grow in the presence of G418. It was then decided to extend the homologous regions to 45 base pairs each, which is more consistent with the literature in terms of minimum homology required for efficient HDR and make pUC57-ade2::KanMX v2 shown in Figure 5.3. This was done using HiFi assembly of the restricted donor vector and a PCR product generated by using primers to amplify out the KanMX gene with longer homologous ends.

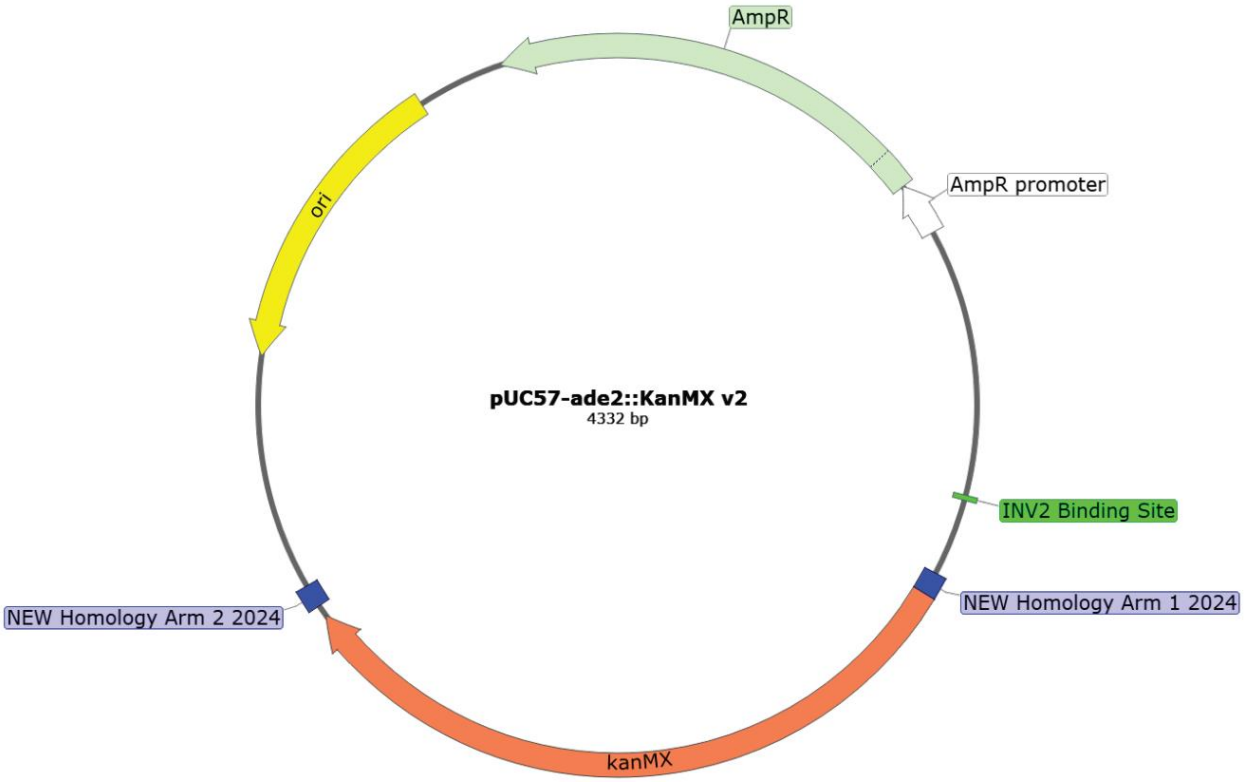

Figure 5.3 pUC57-ade2::KanMX v2 Plasmid Map

Section 5.3.3: Change the Gene Insert to a LEU2 Auxotrophic Marker

Due to the inhibition of translation by G418, it was thought that the antibiotic could prevent the expression of Cas9 nuclease, thus rendering the system useless if transformations were plated directly onto media containing the antibiotic. Replica plating or patching colonies after an initial transformation with the donor, which does add time and variability to the experiments, should be able to address this concern. Alternatively,

plating directly onto selective media should imparts more reproducibility and simplicity, as it should allow more time for HDR, improve the reliability and consistency of results, and therefore improve the ability to quantitate effects relative to background. To address this, an additional donor was made using the same homologous regions targeting ADE2 but with a LEU2 gene in place of the KanMX Cassette. This was done using the LEU2 sequence and its promoter from the previously used pML107 plasmid. The sequence was amplified by PCR using primers that had the same 5' tails as the oligos used to make the 40 bp homologous regions for the KanMX cassette in section 5.3.2 above. This pUC57-ade2::LEU2 plasmid functionally replaces the previous versions of the donor, but is not reliant on antibiotic selection, instead dropping out leucine from the media which allows for determination of the colonies that have undergone recombination.

Section 5.3.4: Planning for Future Changes to Donor DNA Plasmid

As many changes have already been made to the donor and more could be made in the future, it was important to design the donor with versatility in mind. Using the NEBuilder® HiFi reaction kit makes the process easy to do. To generate the pUC57 vector backbone, the restriction enzymes XbaI and BamHI are used to cleave the vector into a 2892 bp backbone and the 1440 bp insert (including the old homologous sequences). Both of these fragments can be visualized by agarose gel electrophoresis to confirm the reaction occurred.

The new insert should be generated through one of two methods. The first method is PCR amplification, but this is only useful if the donor gene is being kept the same. For example, in the case of modifying pUC57-ade2::KanMX v2 to pUC57-ade2::KanMX v3 this works because the donor gene in both instances is the *kanR* gene. The PCR will be performed in two separate reactions. The first will amplify the donor gene with new 40 bp

ends that are homologous to the new gene of interest as shown in Figure 5.4. The second PCR reaction will amplify that purified template and add an additional 40 bp to each end that will be homologous to the pUC57 vector generated earlier. This secondary PCR and the pUC57 vector can then be joined with the HiFi reaction to generate the new plasmid donor.

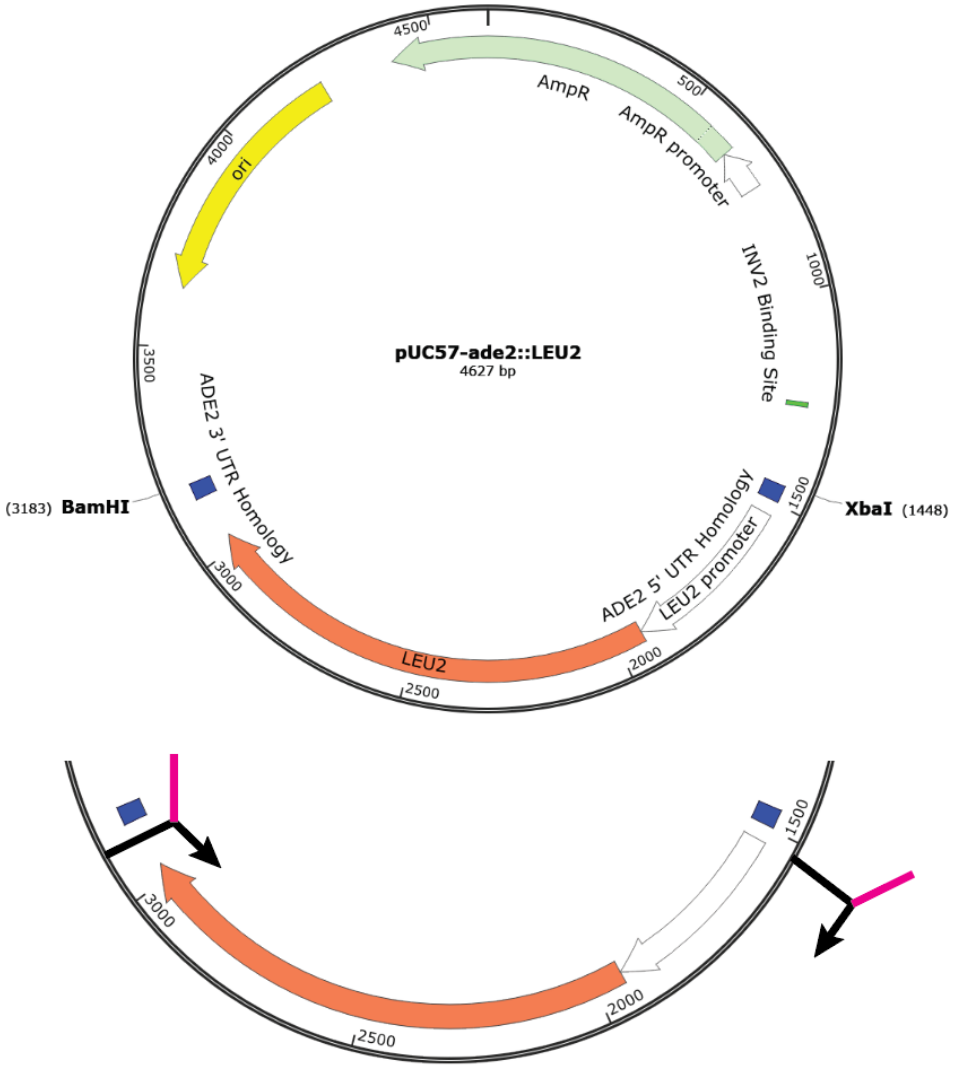

Figure 5.4 Generation of Future Donor Plasmids: Restriction enzymes shown in top map to generate pUC57 vector backbone. Bottom image depicts PCR primers (in black) with nonhomologous 5' tails (in magenta). The tails include new homology for a new genomic target. The amplicon from this PCR will be used in a subsequent PCR whereby the same process will add homology to the vector backbone for efficient HiFi reactions.

If the donor gene is going to be new, then a FragmentGene or similar product would be the best option. The likelihood of having the correct new gene sequence in house is unlikely, and as such having the insert synthesized saves the most time and effort. In designing the synthetic DNA sequence the homologous regions to the gene target should be included on both ends of the gene, followed by the homologous regions to the pUC57 vector. This will make the subsequent HiFi reaction easier to perform without additional workup.

Section 5.4: Use of Linear Donor

Another potential improvement to the donor is to introduce it as a linear DNA molecule instead of a circular plasmid. The linear product is in some instances more effective than a plasmid due to the increased availability of homologous sequences in the DNA ends, compared to a supercoiled plasmid[66]. The linear donor is generated as a PCR product using tailed primers that have the desired regions of homology. The benefit to this method is the ability to change the site of the homology at will simply by changing the sequences of the primer pairs. This method has the added bonus of requiring significantly less plasmid, as each PCR generates multiple transformations worth of DNA.

Section 5.4.1: Improved Linear Donor with Modular LZD Binding Sites

In conjunction with using the linear donor, it is also possible to vary the LZD binding site easily through careful choice of primers and an additional PCR step. In this method the plasmid donor is used as a template with six different possible PCR primers. There are three different forward primers and three different reverse primers. The different primers are either for the INV-2 sequence, CREB sequence, or no LZD binding site. In this way any forward primer can be used with any reverse primer to generate a total of nine different PCR products with any LZD binding site on either end, an example of this is shown in Figure 5.5. This modular donor is not essential to this project, but it could be

useful in future designs. This PCR could also be modified further to append other DNA binding protein binding sequences if other protein families were also implemented into a system similar to the one being used here.

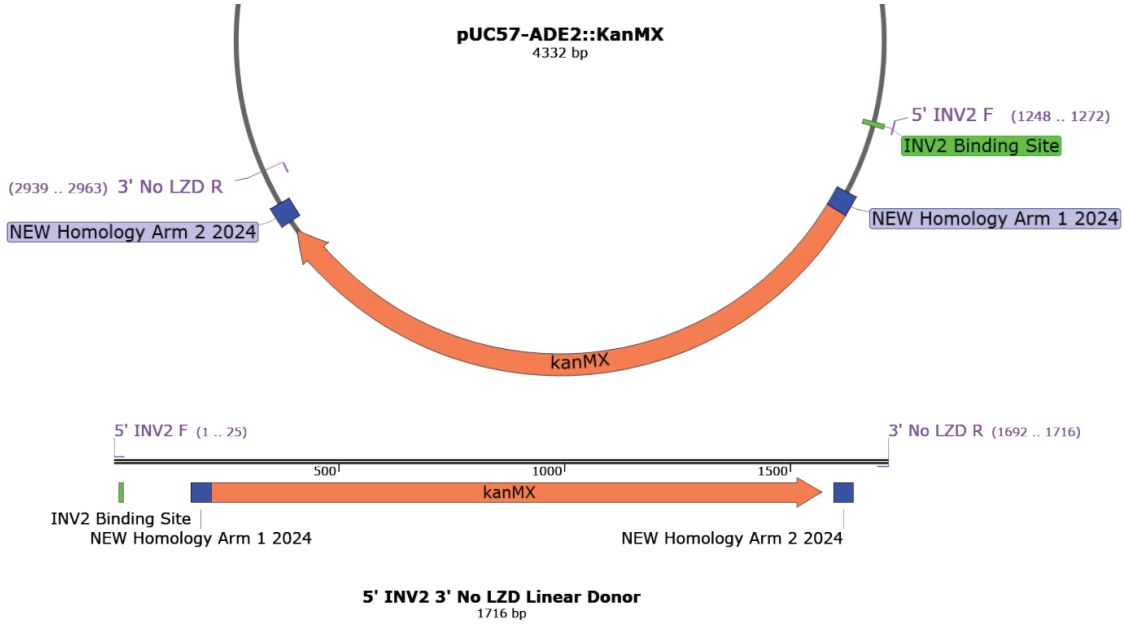

Figure 5.5 Generation of Linear Donor DNA: Template for generating the linear donor is the plasmid donor vector. The primers shown amplify through the donor gene and include the INV-2 binding site for LZD on the 5' side of the KanMX cassette. Alternative primers could be used that put either LZD binding site, or no LZD binding site on either end of the donor cassette.

Section 5.5: Materials and Methods

All enzymes used in reactions including restriction digestions and PCR were obtained from New England Biolabs (NEB). Purification kits were obtained from Qiagen. Oligonucleotides used as primers were synthesized by Integrated DNA Technologies (IDT). XL1-Blue Electrocompetent *E. coli* cells were purchased from Agilent. All other reagents used were purchased from Fisher Scientific. PCRs were run in an MJ Research PTC-200 Gradient Thermal Cycler.

Longer synthetic DNA sequences were designed by me and synthesized by Azenta (then Genewiz) as so-called FragmentGenes cloned into pUC57. This original donor,

pUC57-AGC1::KanMX, was the template used in all subsequent modifications including PCRs and HiFi reactions.

In order to exchange the INV-2 binding site in the plasmid for the CREB site or to remove it completely, oligonucleotides were designed to perform whole plasmid amplification PCR. The oligonucleotide used as the forward primer contained a 10-base 5' tail with either the CREB binding site (5'-ATCAGGTCAT) or a nonspecific 10-base sequence. The reverse primer was complementary to the template stand immediately beside the INV-2 binding site. The PCR was set up with 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each forward and reverse primer, 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (25 repeats):

98 °C – 10 seconds

66 °C – 25 seconds

72 °C – 4 minutes

Final Extension: 72 °C – 120 seconds

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed, reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. Samples were then treated with T4 PNK to phosphorylate 5' ends for ligation. T4 PNK reaction mixtures used 1 × T4 DNA ligase buffer (with ATP), 10 units of T4 PNK and 2 µL of PCR product. The reaction incubated at 37 °C for 30 minutes. The enzyme was then heat inactivated by incubating at 65 °C for 20 minutes. After cooling to room temperature 5 µL of the reaction mixture was mixed with an additional 1 µL of T4 DNA Ligase buffer (with ATP) and 400 units of T4

DNA ligase and left to incubate at 16 °C overnight. Bacterial transformations were carried out using the chemically competent cell method described previously in chapter 3. The transformation mixtures were plated onto LB-amp agar and incubated overnight at 37 °C. Single colonies were inoculated into 5 mL of LB-amp liquid media, grown overnight, and processed with Qiagen Miniprep and Midiprep kits as previously described. Sequencing of the whole plasmid was performed by Plasmidsaurus.

Replacing the regions of homology to change the target from AGC1 to ADE2 required another FragmentGene synthesis, this time with 25 bases of homology to ADE2 in the 5' and 3' positions relative to the KanMX cassette. The new FragmentGenes contained sequences that overlap with the pUC57 vector outside its BamHI and XbaI restriction sites. To prepare the vector for HiFi assembly, 2 µg of plasmid was restricted with 20 units each of BamHI and XbaI in the presence of 1 × CutSmart buffer. Reactions were incubated at 37 °C for 1 hour before being cooled to room temperature. Digested plasmids were confirmed by electrophoresis on 0.7 % agarose gels as previously described.

To perform the NEBuilder® HiFi DNA Assembly reaction a kit was obtained from NEB with the HiFi master mix. The FragmentGene insert was resuspended in 1 × TE buffer and an accurate concentration was determined by UV/vis spectroscopy. The insert and doubly digested vector DNA were mixed in a 4:1 molar ratio, and then incubated with 1 × HiFi master mix in a 20 µL total reaction at 50 °C for 1 hour. Reactions were then allowed to cool to room temperature before being used in bacterial transformations as previously described. This process was also repeated using a different FragmentGene product that included a donor cassette with green fluorescent protein (GFP) instead of *kanR* as the selectable/screenable gene. All steps to construct the plasmid remained the same, just with the use of a different insert.

In order to modify the regions of homology to ADE2, oligonucleotides were designed to PCR out the KanMX cassette. The oligonucleotides contain either 45 bases in the 5' tails homologous to regions 10 bases to the right and left of the DSB site (for the final product pUC57-ade2::KanMX v2) or 40 bases in the 5' tails that are homologous to the 5' and 3' UTR of the ADE2 gene (for the final product pUC57-ade2::KanMX v3). The PCR was set up with 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each forward and reverse primer, 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (35 repeats):

98 °C – 10 seconds

68 °C – 20 seconds

72 °C – 30 seconds

Final Extension: 72 °C – 120 seconds

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. This template was used to perform a HiFi reaction with the same restricted vector of pUC57 used to make the previous donor plasmid. All the same conditions were kept for the reaction. Successful transformations were further processed by Miniprep and then a Midiprep once sequences were confirmed.

Obtaining the LEU2 gene and its promoter was done using PCR primers with 5' tails including 40 bp of homology to the 5' and 3' UTR of ADE2 as in the previous experiment. The template for the reaction was the pML107 plasmid that had previously been used for

expressing Cas9; here it is simply a convenient source for the LEU2 marker. PCR reaction mix and cycling conditions were the same as listed above. Reactions were analyzed on 0.7 % agarose gels as previously described and used in transformations into *E. coli.*

To generate linear donor products, oligonucleotides were designed to be complementary to the furthest ends of the two homologous regions in the plasmid. The PCR was set up with 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each forward and reverse primer, 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The plasmid template for this reaction could be either of the donors with 40 bp of homology to ADE2 (KanMX or LEU2). The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (35 repeats):

98 °C – 10 seconds

68 °C – 20 seconds

72 °C – 30 seconds

Final Extension: 72 °C – 120 seconds

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. This new linear PCR product served as the template for a secondary PCR reaction to add LZD binding sites to either the 5' or 3' end of the amplicon.

The PCR was set up with 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each forward and reverse primer, 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (35 repeats):

98 °C – 10 seconds

68 °C – 20 seconds

72 °C – 30 seconds

Final Extension: 72 °C – 120 seconds

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Since the template is not methylated it could not be digested with DpnI and it was just left in the reaction mixture. Once the size of the amplicon was confirmed the linear template was ready to use in yeast transformations.

Section 5.6 Discussion

Construction of the donor plasmid began in two software programs Serial Cloner and SnapGene. The freedom to design the entire plasmid was beneficial in ensuring that the features required for experiments were all included and appropriately located relative to each other. Homologous regions were selected in the AGC1 gene approximately 850 bases from each other, with the DSB site biasing slightly towards the 5' homologous region. The homologous regions were designed to be 25 bp in length, which is smaller than the 40 bp length that is normally recommended for recombination experiments. The rationale here is that the system should not be predisposed to a high background rate of homologous recombination. Handicapping the system would show a greater increase in the rate of recombination, so a change could be seen regardless of how small, and it is easily attributed to the inclusion of LZD proteins.

The next feature that needed to be in the donor was the INV-2 DNA binding sequence. LZD binds to the naturally occurring CREB and AP-1 sites in yeast, so the design needed to provide an INV-2 binding sequence for the C-terminus. The CREB site in the

genome relative to the DSB site in 40 bases outside of the first homologous region. In the donor sequence the INV-2 site was placed outside of the first homologous region 31 bases away.

The final feature that was needed in the donor was the gene that would impart function if recombination occurred. For this the KanMX cassette was selected as it is commonly used in yeast genetic experiments when knocking out genes and confirming disruption has occurred with the gain of antibiotic resistance. In eukaryotes the *kanR* gene does not provide resistance against kanamycin, instead it provides resistance against geneticin, or G418. Geneticin is part of a family of aminoglycoside inhibitors and functions by preventing elongation in protein synthesis. This mechanism did engender some potential problems for the system, as the donor DNA and the Cas9 expression plasmid were transformed into the yeast at the same time. If the transformants were plated directly onto G418 containing media there would be little time and ability for the yeast to express the Cas9 nuclease necessary to test the system. For this reason, all transformations were first plated onto media without G418, and then subsequently replica plated or patched (meaning transferred one by one and streaked) onto media with G418 to determine resistance.

The initial construct of the donor template cloned in a pUC57 vector. Many variations of the plasmid were made as described above, notably with different LZD binding sites, and with different homologous regions. The 25 base homologous regions showed little to no background recombination in the absence of LZD, and still showed little recombination with LZD present. This was the main driving force behind extending the regions to a more accepted 40 bases when gene targets were switched.

When the target gene was changed to ADE2 the donor had to be adjusted so that the homologous regions were near the new DSB site. The first construct of the donor used

homologous regions that 5 bp away from the proposed cut site, significantly closer than in the AGC1 system. This was due to literature pointing out that proximity of the homologous regions has an effect on the success of recombination.

Different groups performing HDR experiments have reported varying levels of success using a linear donor compared to a plasmid donor. Primers were designed that would allow for the amplification of a region of the donor encompassing the homologous regions and the INV-2 binding site. Additional primers with 5' tails that could be used interchangeably to swap out the LZD binding site on the amplicon were also designed. In doing so it was possible to create nine different amplicons based on whether an LZD binding site is programmed on the 5' end or 3' end, different binding sites on either end, or no binding sites at all. These different donor amplicons were not used, but they are designed and available to be used in the future of this project.

Literature that was found after most of the experiments were complete showed that using an auxotrophic marker gene as the donor in recombination experiments would obviate the need for replica plating[67]. The experiment described used a PCR amplicon containing the TRP1 sequence and promoter. The donor had 40 bases of homology though they were located in the 5' and 3' UTR of the ADE2 gene. The experiment did not use any external source for a DSB and just relied on the homology between the genome and donor to facilitate recombination. The transformants in that experiment were plated directly onto SD-Trp media with an undisclosed amount of adenine. The results showed that a number of colonies that grew were red and were confirmed by colony PCR to be the expected recombination products.

Using this reference, a donor was designed and constructed with the LEU2 gene and its promoter in place of the KanMX cassette to be used in experiments. The same

homologous regions depicted in the previous source were also used to hopefully replicate the results they saw. Because the PCR primers designed for the linear donor annealed outside of the homologous regions in the plasmid there was no need to design new primers, and all of the combinations were still able to be used with this new donor model.

To summarize, a series of fully synthetic donor DNA were constructed for use in experiments including homologous regions, LZD binding site, and gain of function gene to insert into the yeast genome, which are all noted in Table 5.1. Through trial and error, the donor was adjusted as needed, as in extending the homologous regions from sub-optimal to a more reasonable length consistent with the literature. The homologous regions were also swapped out as needed to adapt to the changes experienced while testing other aspects of the system. And finally, the effector gene that would impart function to the yeast if recombination proceeded was changed.

Table 5.1 Donor DNA Comparison

Donor Name	Gene Target	Homology Length	Distance From Target Cut Site (5')	Distance From Target Cut Site (3')
pUC57-agc1::KanMX	AGC1	25 bp	386 bp	587 bp
pUC57-ade2::KanMX	ADE2	25 bp	386 bp	587 bp
pUC57-ade2::GFP	ADE2	25 bp	386 bp	587 bp
pUC57-ade2::KanMX v2	ADE2	45 bp	10 bp	10 bp
pUC57-ade2::KanMX v3	ADE2	40 bp	32 bp	1688 bp
pUC57-ade2::LEU2	ADE2	40 bp	32 bp	1688 bp

Chapter 6: Experimental Results Exploring Facilitated Recombination in Yeast Cells

Section 6.1: Overview

The previous chapters have described the construction of a set of plasmids or DNA fragments to be used in experiments testing their effect on the efficiency of HDR by potentially increasing the local concentration of a DNA donor in the neighborhood of a Cas9-induced DSB. This chapter will describe the experiments that used said plasmids and discuss the results.

Section 6.2: Results

Previous experiments described transformation of single plasmid components into yeast with the intention of confirming the activity of a Cas9/sgRNA and optimizing the expression of LZD proteins. The experiments described here combine multiple plasmids and donor DNA in order to determine the effect that LZD has on the overall progress of Cas9/sgRNA cleavage and recombination-mediated repair. The transformations include controls that should account for sample-to-sample variation as well as omitting each individual component to confirm any specific enhancement activity from LZD proteins.

Section 6.2.1: LZD Plus Donor Transformation (Presented at the Biophysical Society Meeting, February 2023)

The first set of data collected used only pCM173::LZD variants and pUC57-agc1::KanMX plasmid. At this time, the system was still targeting the AGC1 gene, with the first-generation donor plasmid. This experiment was designed to determine if there was any increase in recombination based solely on the presence of LZD73, so the starting point for these transformations was yJD52 that had been previously transformed with either unmodified pCM173::eV or pCM173::LZD73. Both transformations also included the

KanMX donor plasmid. The yeast used were from yJD52 glycerol stocks already containing individual pCM173::LZD variants: pCM173::LZD can be stably maintained as long as doxycycline is provided at high enough concentrations to repress LZD expression. Therefore this experiment was a secondary transformation with the donor DNA only.

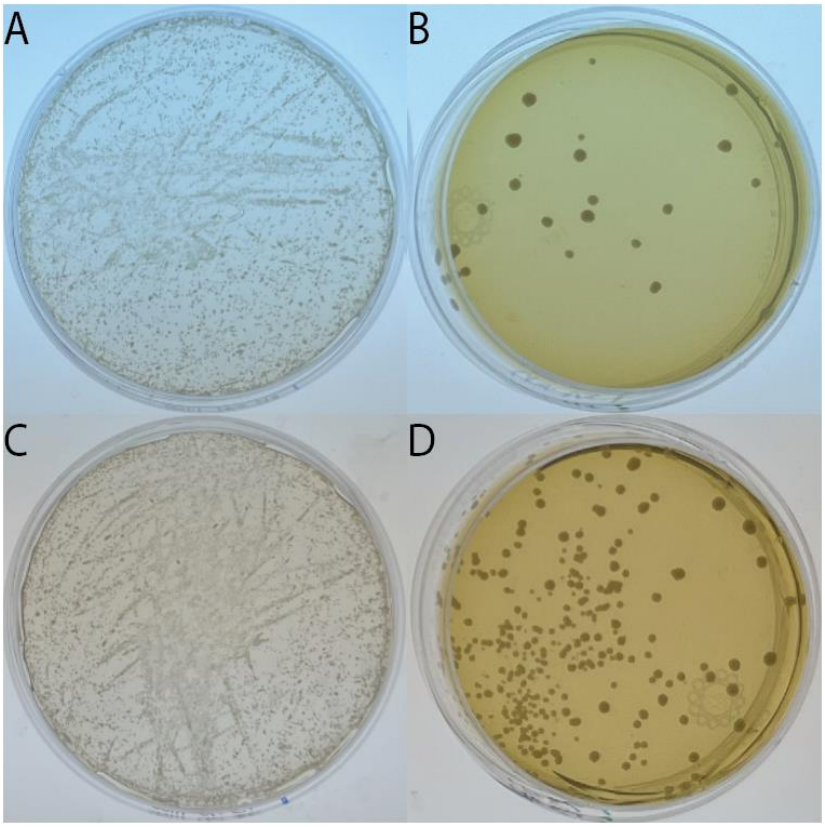

Figure 6.1 Transformation of pCM173::LZD73 and pUC57-agc1::KanMX: (A) pCM173::eV + pUC57-agc1::KanMX, plated on SD-Trp + 0.1 µg/mL doxycycline. (B) pCM173::eV + pUC57-agc1::KanMX, plated on YPAD + 200 µg/mL G418. (C) pCM173::LZD73 + pUC57-agc1::KanMX, plated on SD-Trp + 0.1 µg/mL doxycycline. (D) pCM173::LZD73 + pUC57-agc1::KanMX, plated on YPAD + 200 µg/mL G418. Plates A and C show a similar amount of growth, the only selection being the uptake of the pCM173 plasmid allowing growth in the absence of tryptophan. Plate D shows an increased number of colonies compared to plate B, indicating that LZD may play a role in enhancing recombination even in the absence of a targeted DSB. The image is enlarged for plates B and D, which are 4 cm in diameter as opposed to 10 cm for A and C.

As shown in Figure 6.1 above, there are similar numbers of colonies for empty vector vs. LZD73 vector on the SD-Trp + 0.1 µg/mL doxycycline plates (6.1A and 6.1C), indicating that the presence of this level of LZD does not have a large negative impact on the fitness of the yeast. Plates 6.1B and 6.1D (rich media plus 200 µg/mL G418), show that there is a substantial increase (~ 8-fold) in the number of colonies present when LZD73 is being expressed. This was an encouraging sign that LZD73 can enhance homologous recombination, albeit here in the absence of a targeted DSB.

The yeast pictured on the previous page were patched onto YPAA media to test their ability to grow on acetate. Unexpectedly, they grew without issue, indicating that either the AGC1 gene was still intact, or the knockout did not behave the way the literature described. This led to a change to the ADE2 target gene, as described in chapters 4 and 5.

Section 6.2.2: Combining LZD Expression, Cas9/sgRNA expression, and Donor DNA (Presented at the BPS Annual Meeting, February 2024)

After recognizing and resolving the problems with the AGC1 target and the pML107 Cas9/sgRNA delivery system as described in previous chapters, and after experiencing repeated episodes of contamination, requiring multiple deep cleanings of the laboratory space and new media and fresh plates for each round of transformations, the next set of experiments was performed. This set of transformations used the updated pG2 plasmid and a linear donor amplified from the pUC57::KanMX-ADE2 plasmid. The yeast strain is yJD100, using glycerol stocks that had been previously transformed with pCM173::LZD variants. Transformations were plated onto SD-Trp-Ura + 0.1 µg/mL doxycycline media. The results show a small number of colonies, but there are clear trends in the data in Table 6.1. When pG2 is transformed into yeast with and without LZD, the first column shows that the total number of colonies is similar (~160), inferring that LZD is not deleterious. The

number of red colonies is four times as large in the presence of LZD, again suggesting that LZD enhances Cas9 activity. In the second column, transformation with both pG2 and donor DNA, an ~8-fold decrease is seen in the total number of colonies grown for empty vector vs. the no-donor case or the +LZD; a result that is not understood. It may be an anomaly, or the donor may be deleterious due to initiating but not completing recombination with the defect being partially rescued by LZD. A larger percentage of colonies turned red in the presence of LZD than without.

Table 6.1 Colony Counts for Experiment 2

	pG2	pG2 + Donor	pG2 + Donor (Patched)	Desired Phenotype
pCM173::eV	163 total *(Figure 6.2A)*	18 total *(Figure 6.2B)*	2 total *(Figure 6.2C)*	0/18
	20 red (12.27%)	7 red (38.89%)	0 red (0%)	0%
pCM173::LZD73	160 total *(Figure 6.2D)*	106 total *(Figure 6.2E)*	9 total *(Figure 6.2F)*	9/106
	79 red (49.38%)	63 red (38.89%)	5 red (55.56%)	4.72%

Colony counts listed correspond to the plate in Figure 6.2 below. Total colony counts are provided as well as the number of colonies that grew red. Colonies growing red indicate a disrupted ade2 gene. The difference between red colonies and total colonies are those that grew white. White colonies indicate an intact ADE2 gene. All of the colonies from plates B and E were selected and then streaked (patched) onto plates C and F respectively for determination of G418 resistance. Ability to grow on G418-containing media (plates C and F) indicates recombination with the donor DNA. Red colonies that are also resistant to G418 have the desired phenotype.

All colonies from the pG2 plus donor plates were patched onto new plates with G418, and the number of colonies that grew on antibiotic are represented in the next column. Since G418 inhibits protein synthesis it was theorized that plating directly onto antibiotic media would prevent the successful production of Cas9, thus necessitating patching after the initial growth. None of the 7 red colonies produced in the absence of LZD grew on antibiotics, indicating that they resulted from Cas9 cleavage followed by NHEJ. In contrast,

5 of the 63 red colonies produced in the presence of LZD grew on antibiotics Indicating that although small, LZD may have a role in enhancing recombination.

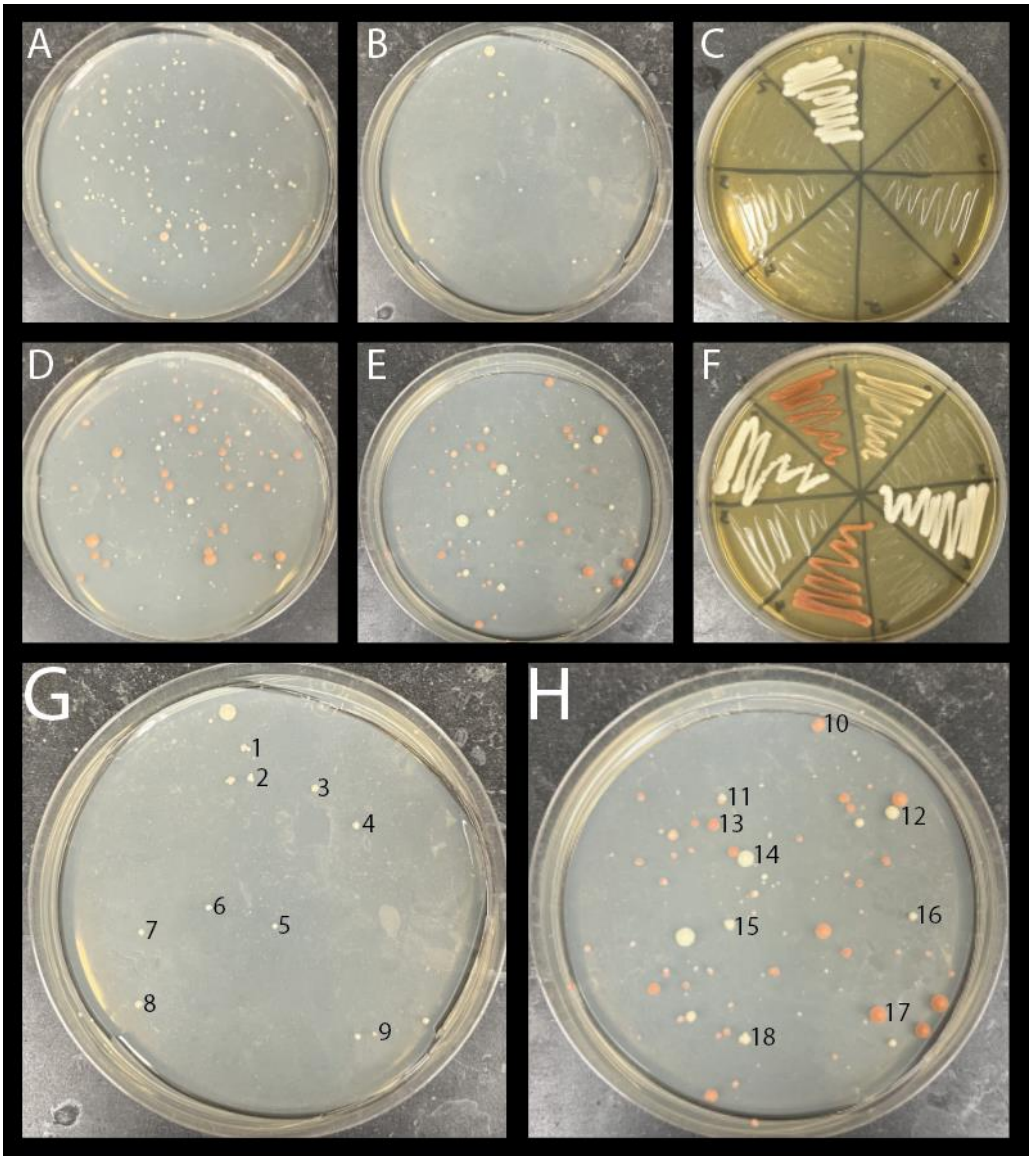

Figure 6.2 Transformation of pCM173::LZD73, pG2 and ade2::KanMX Linear Donor: All plates were transformed with pG2. (A) pCM173::eV + no donor, plated on SD-Trp-Ura + 0.1 µg/mL doxycycline. (B) pCM173::eV + linear donor, plated on SD-Trp-Ura + 0.1 µg/mL doxycycline. (C) Patched colonies from plate B, plated on YPD + 200 µg/mL G418. (D) pCM173::LZD73 + no donor, plated on SD-Trp-Ura + 0.1 µg/mL doxycycline. (E) pCM173::LZD73 + linear donor, plated on SD-Trp-Ura + 0.1 µg/mL doxycycline. (F) Patched colonies from plate E, plated on YPD + 200 µg/mL G418. Increased red colonies on plates where yeast were expressing LZD indicate that our protein may have an effect on Cas9 cleavage. The increased number of colonies on patched plates indicates that LZD could be

playing a role in enhancing recombination in yeast. (G) Enlarged image for (B) with numbered colonies used in the colony PCR shown in Figure 6.3 below. (H) Enlarged image for (E) with numbered colonies used in the colony PCR shown in Figure 6.3 below.

Yeast colonies that grow on G418 containing media and also turn red have the phenotype corresponding to targeted HDR, but they could also be the products of two simultaneous separate processes, off-target recombination of the donor DNA and a properly knocked out ADE2 gene repaired via NHEJ. Colony PCR was used to determine if the genome was modified via HDR or NHEJ. Figure 6.3 shows the results of 4 red colonies and 14 white colonies selected for colony PCR at the same time the patched plates were incubating. The first 9 colonies come from the plate shown in Figure 6.2G (no LZD) and include 8 white colonies and 1 red colony. The second 9 colonies come from the plate shown in Figure 6.2H (expressing LZD73) and include 6 white colonies and 3 red colonies. The first PCR, shown in Figure 6.3A, used primers that spanned from the 5' UTR of the ADE2 promoter to the other side of the DSB site, about 700 bp away. In this reaction there is a definitive band of higher molecular weight appearing for 3 of the yeast colonies. This band is approximately 2.1 bp in size and is the correct size to be the product of amplifying through the KanMX cassette inserted into the ADE2 gene through HDR. The 743 bp band at the bottom of the gel also seems to be less intense in at least two of the 3 lanes with larger molecular weight bands. A similar result is seen in the second gel (Figure 6.3B) where the same colonies were run with a different reverse primer that is further into the ADE2 gene 1760 bp away. The 3 red boxes indicate light or completely disappeared bands. The indicated lanes are the same in both gels, and correspond to yeast that expressed LZD73, and grew red on media containing G418. Based on the design of the PCR primers there should be a large molecular weight fragment of about 3.1 kb when the KanMX cassette is properly inserted. Although this larger molecular weight band is not seen it is

likely due to the PCR conditions not allowing for amplification of that large a fragment and thus only the disappearance of the intact ADE2 band is seen.

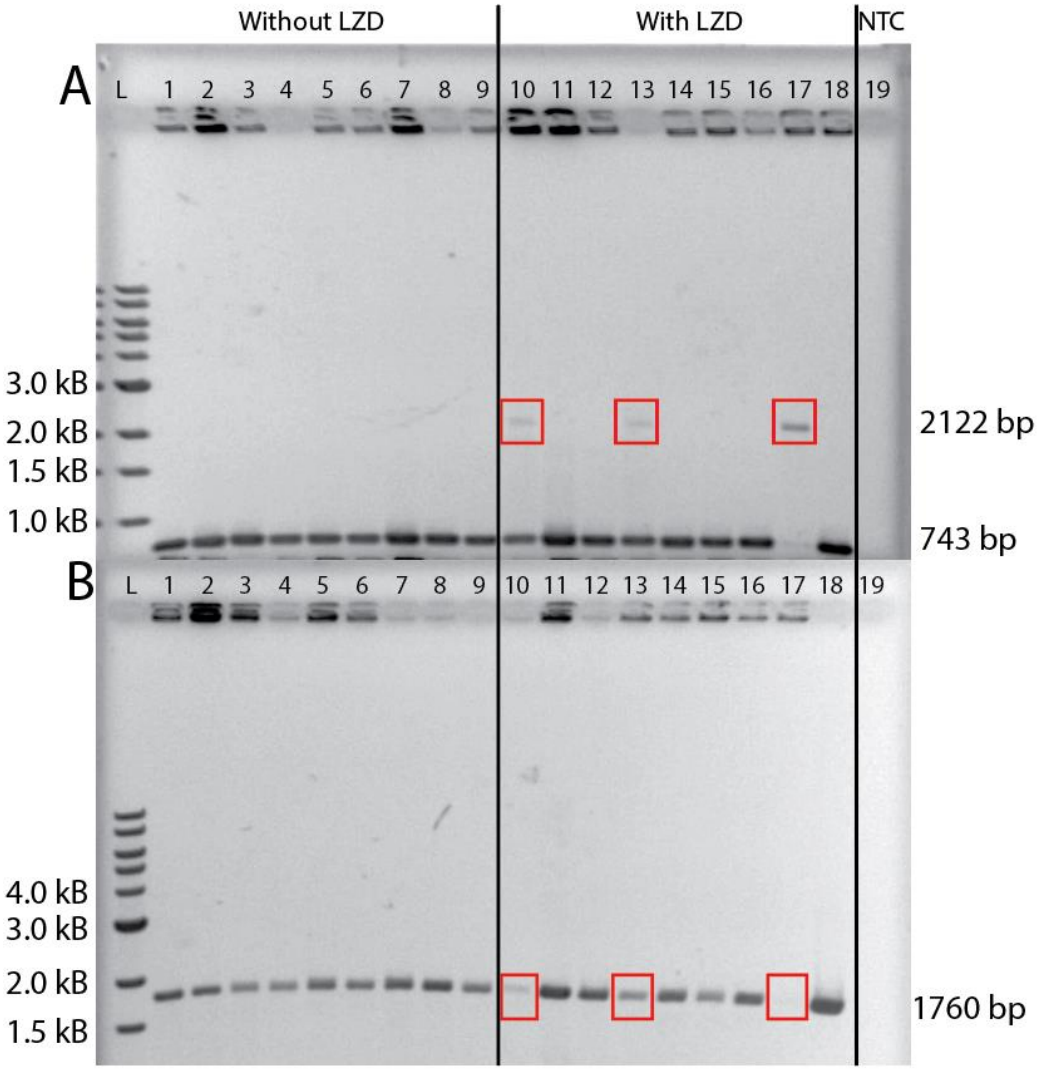

Figure 6.3 Results of Colony PCR: (A) Result of PCR using ExoADE2 forward and EndoADE2 reverse primers. Amplicon at 743 bp indicates an intact ADE2 gene. The presence of an amplicon at 2122 bp indicates a portion of the selected colony contains the desired mutation with KanMX interrupting the ADE2 gene. (B) Result of PCR using ExoADE2 forward and ExoADE2 reverse primers. The amplicon at 1760 bp indicates an intact ADE2 gene. The presence of an amplicon at 3139 bp would indicate the desired mutation with KanMX interrupting the ADE2 gene. All lanes are labeled according to the colony numbers assigned in Figure 6.2. Colonies 10, 13, and 17 light bands are seen for the intact ADE2 gene and the presence of a higher molecular weight amplicon in gel A. The same lanes show a light band for the intact ADE2 amplicon in gel B. The lack of a higher

molecular weight band in gel B is likely due to PCR conditions not being fully optimized. The ladder used in both gels is the NEB 1 kb DNA Ladder (Catalog # N3232L). The lane labeled NTC is a no template control intended to show primers will not amplify on their own.

Table 6.2 Colony PCR Expected Products			
Forward Primer	Reverse Primer	Expected Product Size (ADE2)	Expected Product Size (ade2::KanMX)
ExoADE2	ExoADE2	1760 bp	3139 bp
ExoADE2	EndoADE2	743 bp	2122 bp
ExoADE2	KanMX	N/A	1148 bp
EndoADE2	ExoADE2	532 bp	532 bp
EndoADE2	EndoADE2	N/A	N/A
EndoADE2	KanMX	N/A	N/A
KanMX	ExoADE2	N/A	1779 bp
KanMX	EndoADE2	N/A	762 bp
KanMX	KanMX	N/A	N/A

The continued detection of the smaller molecular weight bands in both gels is likely caused by incomplete conversion of the colonies to the desired genotype at the time of selection for colony PCR. In order to address this further in the future, samples to be used in colony PCR should be patched first and then selected so as to isolate single colonies that have completed the DNA repair process. Other changes that should be made would be to optimize the PCR conditions to ensure visibility of all expected bands regardless of size. This could involve extending the extension time during cycling to allow for the synthesis of longer amplicons.

Section 6.3: Modified System with LEU2 Donor

A third round of experiments were performed, with Cas9/sgRNA from pG2 targeting ADE2 or pEmpty-Control, but with the newly designed pUC57::LEU2 donor in place of the KanMX donor. The purpose of this experiment was to make replica plating or patching unnecessary and acquire the phenotypic results in one transformation. The transformations were split into two groups, one with a new yeast strain – FY1679-05A (S288C isogenic yeast

strain: *MATa HIS3 leu2Δ1, ura3-52 trp1-63 GAL2*) – and another set with pCM173::LZD variants in that same yeast strain.

Table 6.3 Colony Counts for Experiment 3

	Control Yeast (FY1679-05A)			
	With Leucine		Without Leucine	
	Total Colonies	Red Colonies	Total Colonies	Red Colonies
Plasmid Donor	Lawn	0	>1000	0
Linear Donor	Lawn	0	>1000	0
pEC	>1000	0	0	0
pG2	52	48	0	0
pEC + Donor	~500	0	40	0
pG2 + Donor	8	8	0	0
	pCM173::eV			
	With Leucine		Without Leucine	
	Total Colonies	Red Colonies	Total Colonies	Red Colonies
pEC	>500	0	0	0
pEC + Donor	400-500	0	28	0
pG2	8	6	0	0
pG2 + Donor	2	1	0	0
Donor Only	Lawn	0	Lawn	0
No DNA (-)	0	0		
	pCM173::PAAW			
	With Leucine		Without Leucine	
	Total Colonies	Red Colonies	Total Colonies	Red Colonies
pEC	>500	0	0	0
pEC + Donor	400-500	0	17	0
pG2	15	12	0	0
pG2 + Donor	2	2	0	0
Donor Only	Lawn	0	300-400	0
No DNA (-)	0	0		
	pCM173::LZD73			
	With Leucine		Without Leucine	
	Total Colonies	Red Colonies	Total Colonies	Red Colonies
pEC	>500	0	0	0
pEC + Donor	200-300	0	7	0
pG2	17	15	0	0
pG2 + Donor	1	1	0	0
Donor Only	Lawn	0	300-400	0
No DNA (-)	0	0		

Table 6.3, which shows colony counts for the different transformations, gave confusing results. Transforming the donor alone led to a large number of colonies growing in the absence of leucine in the media, suggesting that there is recombination occurring. The fact that none of the colonies were red, though, indicates that this is all off-target recombination. Additionally, a steep decline in the number of colonies that grow in the presence of pG2 compared to pEmpty-Control is seen. This may be due to inability to repair double strand break(s) due to the co-expression of the sgRNA, whereas pEmpty-Control only expresses Cas9. There is a high percentage of red colonies seen in the presence of pG2, suggesting some correct targeting of the ADE2 gene. Unfortunately, there are no colonies seen growing without leucine from the pG2 transformed cells.

This experiment has been done only once, with unexplained deleterious effects for the pEC plasmid that was previously found to be innocuous. It is unknown where the LEU2 marker integrated. It would require more experiments to work out what is going on.

Section 6.4: Materials and Methods

Yeast strains yJD52 and yJD100 were obtained from Dr. Dinman as previously stated. Yeast strain BY1679-05A was obtained from Euroscarf. All oligonucleotides were obtained from IDT. All enzymes including those used for PCR were obtained from NEB. All other reagents were obtained from Fisher Scientific. All yeast media formulations contain 0.001 % adenine hemisulfate unless otherwise indicated. PCRs were run in either an MJ Research PTC-200 Gradient Thermal Cycler or Applied Biosystems SimpliAmp Thermal Cycler with the permission of Dr. Myles Poulin.

Yeast transformations were performed using a different protocol than what was used previously. Single colonies of yeast were inoculated overnight in respective media (YPAD or SD-Trp + 0.1 µg/mL doxycycline depending on if pCM173::LZD is present or not)

shaking in a 30 °C water bath. The following morning cultures were diluted 1:40 in respective media and OD660 was observed. Undiluted suspensions were then diluted down to a final OD of 0.15 in 2 × media, again dependent on whether pCM173::LZD is present or not. The final volume of this secondary inoculation is calculated by multiplying the number of reactions to be done by 6 mL. This culture is shaken at 30 °C until the OD has reached 0.8, which takes approximately 6 hours. The cell suspensions are then collected via centrifugation at 5000 rpm for 3 minutes. The cells are then washed twice and resuspended in fresh water, first 5 mL per reaction and then 300 µL per reaction. The cell suspension is then aliquoted into Eppendorf tubes at 300 µL each, these tubes are spun one last time at 8000 rpm for one minute, and the supernatant is removed. Then 326 µL of PEG, Lithium Acetate, ssDNA (PLS) solution is added to each prepared tube and mixed well to resuspend the cell pellet. The PLS solution is 36.8 % PEG, 0.11 M LiOAc, and 0.31 mg/mL ssDNA. Finally, DNA is added to the transformation mixture, 500 ng for plasmid DNA or 1 ug of linear PCR product, before mixing well. The concentration of the plasmids is known from UV/vis spectrometry and the concentration of the PCR products is estimated based on comparing band intensities to a ladder purchased by NEB.

The transformation mixture is then placed in a 42 °C water bath to incubate for one hour. Every 20 minutes the samples should be inverted, or otherwise mixed to prevent the cells from clumping at the bottom of the tube. After an hour the tubes are spun in a centrifuge at 8000 rpm for 15 seconds to collect the cells. Supernatant is discarded and the cells are resuspended in 100 µL of fresh DI water. The final suspension of cells is split and plated onto 1 or 2 plates depending on conditions being tested, but the volume plated is always 50 µL to keep the quantity of cells as consistent as possible. Plate conditions for yeast cells containing only pG2, or pG2 and donor were plated onto SD-Ura media with or

without leucine. Yeast cells containing pCM173::LZD plus pG2, or pG2 and donor were plated onto SD-Trp-Ura + 0.1 µg/mL doxycycline with or without leucine. Plates were allowed to grow undisturbed for 4-5 days at 30 °C.

Once initial plates were grown colonies of interest are then patched to a plate with new conditions to determine antibiotic resistance to G418. To do so a small portion of a colony is collected on the end of a wooden toothpick and then spread onto a predetermined segment of a different plate. Secondary plates were rich media (YPD) with G418 at 200 µg/mL. These plates were also left undisturbed to grow for 4-5 days.

Colonies of interest that had desired phenotypes based on the plates they grew on were selected for colony PCR. To perform the colony PCR samples were collected from agar plates and resuspended in 20 µL of DI water. The reaction mixture is 1 × Q5 reaction buffer, 500 nM each forward and reverse primer, 500 µM dNTPs, and 1 unit of polymerase with 2 µL of dispersed colony to be analyzed. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 4 minutes

Cycle (35 repeats):

98 °C – 1 minute

55 °C – 1 minute

72 °C – 2 minutes

Final Extension: 72 °C – 10 minutes

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described before staining with EtBr and visualized on a UV transilluminator. Amplicon bands were compared to expected lengths based on SnapGene constructs of the yeast genome with the addition of our donor insert.

Figure 6.4 PCR Primer Locations for Colony PCR: Primers listed in purple can generate colony PCR amplicons for analysis of the genome. KanMX primers can participate in amplification only if the KanMX cassette is incorporated into the genome. Forward and reverse primers can be used in any combination to generate the potential results shown in Table 6.3.

Section 6.5: Discussion

To fully understand whether LZD has an effect on homologous recombination, the individually developed system components had to be tested individually and also combined into the fully assembled system. In addition, careful consideration had to be given to ensuring that the results seen were not solely influenced by the yeasts' natural ability to perform homologous recombination.

The first series of transformations addressed the effect of LZD on recombination in the absence of a targeted DSB, by omitting the Cas9 inducing plasmid. Using the yJD52 strain bearing pCM173::eV or pCM173::LZD73, the donor plasmid was transformed alone.

It was estimated that expression of LZD73 increased the number of antibiotic-resistant colonies 8-10-fold, compared to the yeast with pCM173::eV.

While promising, these results only indicate that the presence of LZD73 aids in the incorporation of the insert DNA somewhere in the yeast genome, and the G418-resistant yeast did not show the expected AGC1 knockout phenotype. Furthermore, it was at this point that we were unable to replicate our own preliminary results on the *agc1* phenotype, and literature was found stating that a knockout of the AGC1 gene does not prevent the growth of yeast on media containing acetate as a sole carbon source. It is possible that recombination occurred in our experiments, and at the appropriate site, but the lack of a scorable phenotype for appropriate insertion was disappointing. As described in chapter 4, the lack of an *agc1* phenotype could be ascribed to the YMC2 gene, and rather than knocking out YMC2 in all of the yeast strains we needed to use, the Cas9/sgRNA expression system and donor plasmids were redesigned to target the ADE2 gene. (In retrospect, it might be productive to return to this system and characterize recombinants at the genomic level without worrying about a growth phenotype.)

chapter 4 described the trajectory of achieving a working Cas9/sgRNA system targeting the yeast ADE2 gene. With the pG2 Cas9/sgRNA system, a new donor plasmid, and new yeast strain in hand, transformations began with all three components.

The transformations described in Figure 6.2 used yJD100 stocks including pCM173::eV or pCM173::LZD73 as the starting point. pG2 was transformed in, with and without the linear ADE2 donor from pUC57::ADE2. It was also noted that the number of red colonies was almost 4 times greater when LZD73 was present vs absent. These results were unexpected, and they have not been thoroughly reproduced. If this result is valid, it is possible the inclusion of a looping protein like LZD73 could rearrange the genomic DNA

and make specific DNA stretches more easily accessible or reactive to nucleases like Cas9 that bend and twist the DNA[68]. We have never observed LZD decreasing the efficiency of Cas9 cleavage, which is a good sign for the utility of the approach.

Additionally, in this transformation many colonies grew when donor and Cas9 were co-transformed into yeast expressing LZD73, and a much smaller number in the absence of LZD73. Large drop-offs had not been seen before in the absence of LZD, so this was not thought to be a systemic issue, and possibly just a one-off for this round of transformations. There is an approximate 50% increase in the relative ratio of red colonies to total colonies in the presence of LZD73. This alone is not indicative of homologous recombination, since knocking out the ADE2 gene is predominantly going to be driven by Cas9 activity.

Patching all of the yeast from this transformation onto G418-containing plates showed that about 10% of them grew. This amounted to only two colonies from the initial pool that had been transformed with the empty pCM173::eV vector, and both were white, indicating off-target recombination or some other accidental integration. We do not know why the level of background recombination is so low (i.e. undetectable). As for the cells that were initially expressing LZD73, 9 of the 106 colonies grew on media containing G418, and 5 of the 9 were red. These are still small numbers but is does indicate that the yeast expressing LZD73 presented the proper phenotype more often than the yeast that did not express any LZD. Colony PCR on three of the 5 red colonies, compared with white colonies as controls, suggested that the KanMX cassette had been correctly inserted at ADE2.

It should be noted that in the previous transformation there were a number of control plates also transformed to ensure consistency between the samples. In any transformation that used pG2, there was another transformation that used pEmpty-Control. In these experiments there was no indication of a drop off in colony counts as there

were no DSBs being introduced into the yeast. In most cases these transformations led to colony counts that were uncountable and well over 1000. It is also important to state that when pEmpty-Control was used in transformations no red colonies were observed, which is also good. It indicates that the only cause of red colonies is coming from the genomic breaks induced by pG2.

Other controls included a negative control, which entailed plating the yeast culture directly onto media lacking uracil without providing the pG2/pEmpty-Control plasmid. This transformation ensured that the only reason growth is seen on Ura$^-$ plates is because of the presence of pG2 or pEmpty-Control. The final control was transforming just donor DNA into the yeast cells in a similar manner to the first transformation experiment. This control should have given insight into the background level of recombination since the linear donor should be able to invade the genome even in the absence of a double strand break. No red colonies were observed and in fact only a lawn of white colonies grew. This may be due to the lack of auxotrophic control and the short homologous regions not being sufficient to initiate recombination on their own.

The necessity to replica or patch the yeast added a significant amount of time to all transformations and doubled the number of plates needed. If it could be simplified to just plating on a single media condition it would expedite the turnaround, similar to the effect that switching from AGC1 to ADE2 had. The instant feedback of seeing red yeast eliminated the need for two additional replica or patching plates and simplified it to a single additional plate. To streamline, two different methods were attempted. The first method was to allow the cell suspension time to grow in nonselective media for 2 hours after the transformation procedure was completed, right before the cells would regularly be plated. This was thought of as similar to doing a recovery step in bacterial transformations

prior to plating on antibiotics. Despite our best efforts, the yeast did not show any growth improvement under these conditions and there were still no colonies on media containing G418.

The other attempt involved an improvement to the system where instead of inserting a *kanR* gene into the yeast, the auxotrophic LEU2 gene is inserted that corrects for a knockout in the yeast strain. This involved an amount of reengineering the system, as a new yeast and a new donor were required. The yeast was easily obtained from Euroscarf (strain BY1679-05A), and the donor was generated by amplifying the LEU2 gene and its promoter sequence from the pML107 plasmid previously being used.

This set of results was inconclusive: it does not indicate that LZD73 aids in recombination, but it also doesn't show on-target recombination in the absence of LZD73. This system needs to be looked at more closely or modified to obtain results that would be more definitive.

Chapter 7: Conclusions and Future Directions

Section 7.1: Demonstration of Improved Homologous Recombination via Use of a DNA Looping Protein

Getting to this point took hard work and modifications to a system initially thought to be easy to design. Some of the various hurdles that had to be overcome were construction of plasmids, wrestling with gene targets, and general contamination.

Plasmid construction has mostly become trivial at this point with the use of NEBuilder® HiFi kits. Where initial attempts routinely took months to generate a single plasmid with the proper sequence, now it is possible to construct multiple plasmids in the span of a week using HiFi.

Moving from the AGC1 gene to ADE2 was also a huge step forward for the work being done. Though there was some amount of success using the initial system, it was clear from further literature searches that sticking with AGC1 would have led to lots of inconsistencies in the final data. Learning about the YMC2 gene certainly saved time that otherwise would have been spent trying to decipher results of an unreliable phenotype.

At this point we have presented data that does indicate that LZD73 has some positive effect on the rate of homologous recombination in our yeast system. This is still a work in progress and does need to be improved upon to generate the quantity and quality of data that would be necessary for responsibly moving forward into other organisms. The key take aways at this point are:

1. A system has been developed that allows for the controlled expression of LZD in a yeast system.
2. The system shows that the plasmid expressing LZD is stable in yeast, and we have evidence of biological activity in yeast.

3. The system confirms that knocking out the ADE2 gene in yeast produces a visible phenotype that can be visibly scored without the need for sequencing.
4. Expression of LZD73 sometimes coincides with an increase in Cas9 activity, and in every experiment that we think is interpretable it increases the efficiency of homologous recombination
5. Colony PCR is working well to assess the genotypes of recombinant and wild type yeast.

Section 7.2: Proposed Improvements and Modifications to the System

Throughout the course of this experiment many changes needed to be made to accommodate the struggles that were experienced. While the current system does work and yields results, there is potential to make it even better by changing additional factors. The first change that should be made is the inclusion of all other LZD proteins and determine how or if they affect recombination. Throughout these experiments only LZD73 and PAAW were used in experiments, PAAW being the N-terminal binding protein analogous to the bZip domain of the GCN4 protein in yeast.

The different LZD sequences theoretically orient the bound DNA strands in different angular orientations relative to each other based on the length of the coiled coil. Based on PyMol renderings, it appears that LZD73 holds the DNA with approximately a 60° angle between them. As the protein gets smaller the angle increases to almost 90° for LZD66. And when the protein gets larger the angle shrinks to 30° and then parallel for LZD80 and LZD87 respectively as seen in Figure 1.9 in chapter 1. This orientation could improve homologous recombination the more parallel or perpendicular it is, but at this time that information is not supported by experiments. Another aspect of the different proteins is the distance between the DNA binding domains. The difference between the smallest protein –

LZD66 – and the largest protein – LZD87 – is approximately 30 Å which could in theory make a large difference for recombination.

Section 7.2.1: Gene Targets

In addition to the use of all the LZD proteins it is possible to consider that different genes may be easier to screen for than the selected ones. The process of selecting a new gene target has already been experimented with when they system did not have the desired outcome in AGC1 and moved onto something that was better suited to the system's needs in ADE2. But using ADE2 did not solve all of our problems, and there was still an amount of troubleshooting with adenine concentrations to optimize the white to red color change of the yeast as well as the time allowed to let the yeast turn red.

In a similar manner the use of the *kanR* gene for the effector gene certainly caused some problems. Due to the mechanism of action for G418 it was nearly impossible to directly select for antibiotic resistance following a single transformation. As G418 inhibits protein synthesis it is unlikely Cas9 would be expressed in a large enough amount to cause genomic damage. Similarly, even if a DSB occurred and homologous recombination happened it is unlikely that the *kanR* gene would be expressed in the presence of G418. For this reason, all transformations had to be done onto non-selective media and then streaked or replica plated onto media containing G418.

Section 7.2.1.1: Knock Out Targets

When looking for alternative gene knock out targets it seems that there are three main groups that can be looked at for selection: Genes that change the yeast's color, genes that affect auxotrophy, and genes that change the shape of the yeast.

Modifying the ADE2 gene has shown some success making the yeast change color from white to red in the presence of suboptimal adenine concentrations. There are

additional genes that also affect the color of the yeast in a similar red/white manner. First is the ADE1 gene which catalyzes the step after ADE2 in the purine biosynthetic pathway. ADE1 knockout yeast also turn red in the presence of suboptimal adenine concentrations, but it is possible that a knockout of this gene would produce red colonies sooner than ADE2. Another member of the purine biosynthetic pathway is ADE8 which produces an enzyme that is used three steps before ADE2 in the pathway[69]. Literature shows that yeast with ADE1 or ADE2 knockouts can be reverted to white colonies by knocking out the ADE8 gene. This makes sense given that the colorimetric product will never be produced if the biosynthetic pathway never proceeds past the ADE8 catalyzed step. The good thing about using these ADE genes is that screening does not require additional plating or streaking steps. It can be directly observed if the yeast turns red or white and known from a single transformation if colonies are subject to further confirmation testing.

Another group of genes to potentially target are those that affect auxotrophy. While the ADE genes are technically auxotrophic markers, in most experiments they are used because they can affect the color of the yeast in suboptimal concentrations of adenine in media. If the adenine was removed entirely then the yeast would be unable to grow unless sufficient adenine is supplemented to the media. Almost all genes in their respective biosynthetic pathways for nonessential amino acids could be knocked out and used as selective agents when growing transformants. The downside to using these genes would be the phenotype on selective media would be death. This is something that is likely to be avoided, as it requires additional screening steps and looking for death as a phenotype is generally not ideal.

A final group of genes that could be used are those that affect the shape of the yeast. Genes like actin-related protein 8 (ARP8) are implicated in morphological changes of the

yeast as seen in Figure 7.2[70]. While these genes may be suitable targets for knocking out, they do tend to have some other deleterious effects on the fitness of the yeast which make them unlikely to be very useful. Additionally, observing the morphological changes in these yeast would require screening under a microscope which similar to the auxotrophic genes would require an additional layer of screening that may be more work than it is worth.

With any of the new potential targets, and even the currently used ADE2 gene, another factor to consider is modifying the sgRNA target sequence. There is current literature published indicating that not all sgRNAs target Cas9 to a gene with the same efficiency. Identifying the ideal target sequence could be done by analyzing a full gene sequence for all possible PAM sequences and then using modern bioinformatics to narrow down the selection to a manageable number of sequences. These potential sgRNA sequences could then be incorporated into a pG2-like plasmid and compared against each other to determine which sequences generate the greatest percent of knocked out genes.

Section 7.2.1.2: Knock In Targets

Aside from using *kanR* as a knock in target there are other options which may give better results and lead to less screening time. One option for knock-in targets are the genes used as auxotrophic markers. For example, if a yeast strain that is being used has URA3, TRP1 and LEU2 knocked out, one could attempt to insert a functional copy of the LEU2 gene in the DSB location. What would be required is the lack of the specific amino acid in the media, in this case leucine. Since the LEU2 gene is the only part of the native biosynthetic pathway knocked out and an intact copy of that gene is being introduced, the yeast that successfully undergo recombination will be able to grow in the absence of leucine in the media. The only nuance to this method would be making sure that the yeast in question has a full gene deletion for the marker that will be inserted in. Not all knockout

mutations are whole gene deletions, some are just partial deletions, frame shifting mutations, or SNPs that change a single amino acid into a stop codon. If any of these mutations are present in the gene being inserted, then it is possible to have reversion events where the knocked out genomic copy of the gene can become functional again. This process happens spontaneously at a fairly low rate but could also come to be from the donor being present acting as a homologous template for fixing the genetic defect.

This method was performed when the donor DNA was modified to use the LEU2 and its promoter. The goal was to avoid the need for replica plating or patching the yeast and directly transform onto media lacking leucine. The yeast biosynthetic pathway for leucine would be intact if the LEU2 gene was incorporated through recombination thus allowing the organisms to grow on media without leucine. This did work to an extent, but more work needs to be done to fully test the system, so it is still included here.

Another protein that could be of use is GFP, or any of the other colored fluorescent proteins. These proteins could be inserted into a cassette similar to the *kanR* gene and expressed with a specific promoter and terminator. The distinct advantage to these proteins over *kanR* and the auxotrophic genes, is the immediate knowledge of successful recombination. No special media is required to have the fluorescent proteins express and having GFP – or others – in the organism is not going to be deleterious to fitness. The color change should be visible with ambient light, but if not the grown colonies could be exposed to the optimal light source to produce colored colonies as seen in Figure 7.1. We did attempt a single round of transformations using a donor with GFP but did not observe any evidence of recombination.

Now a disadvantage to this method would be if it were used in conjunction with the current targeting of ADE2. The red color the yeast exhibits when ADE2 is knocked out is

strong and opaque. It might be difficult to detect additional color changes at the same time. So, if proceeding down this path it may be best to adjust the target gene to something that keeps the yeast white, so only a single-color change is detected. Perhaps comparing a GFP insert with a gene knockout that affects the morphological shape of the yeast would be the best path forward.

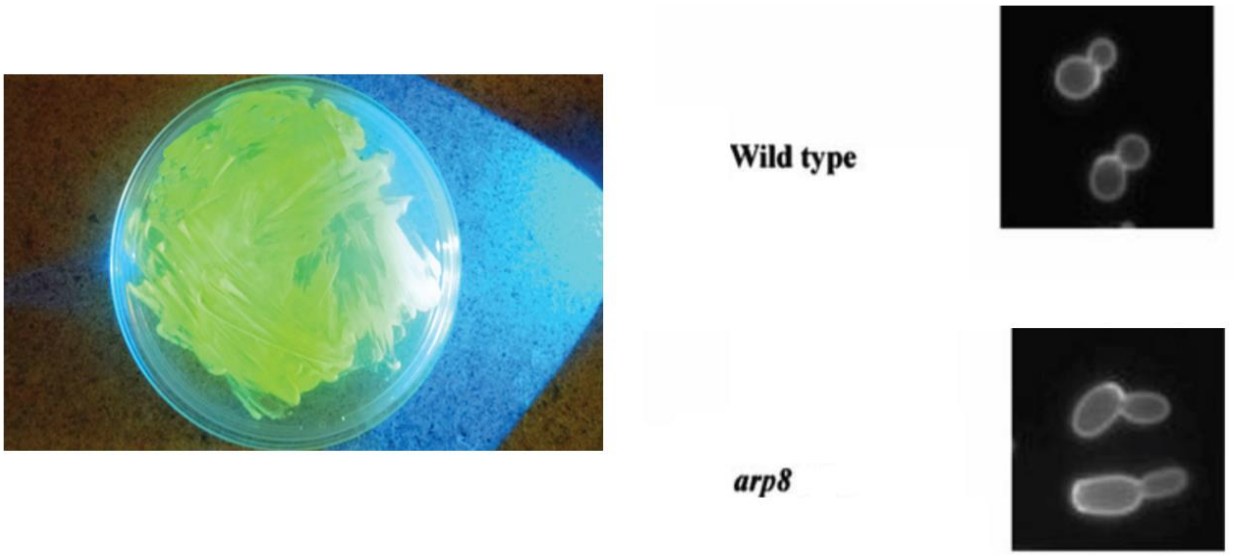

Figure 7.1 New Gene Target Effects: On the left yeast have been modified to express GFP. The yeast glow a distinct green color when introduced to blue light sources. On the right are microscope images of yeast that have a disrupted arp8 gene. This causes the budding yeast to appear more oblong as opposed to the wild type which are round.

Section 7.2.2: Further Donor Modifications

With the donor DNA being entirely flexible to sequences and features, there are many more changes that could be made in order to optimize the system. Homologous regions is a clear starting point, specifically generating a library of donors all with different length homologous regions. This could then be used to determine a more optimal window of homologous regions that would be acceptable for a given gene. Additionally, the relative position of the homologous regions to the cut site should be varied to determine what effect

that would have on recombination. This could also be performed using a library of different donors all with homologous regions that have relative positions to the desired DSB location. A final change to the donor is something that has already been planned for, which is the positional modification of the LZD binding site. The current system only uses a single INV-2 site in the donor, but there may be a benefit of having an INV-2 site on either end of the donor cassette. The distance between the homologous regions and the LZD binding site may also play a role in how available the donor is to undergo homologous recombination with the target sequence.

Section 7.2.3: Plasmid Targeting over Genome Targeting

A more drastic change to the system that could improve results is to target a plasmid in yeast instead of targeting the genome as depicted in Figure 7.2. This new system would involve designing a full plasmid that would be transformed and retained in the yeast under auxotrophic selection. The designed plasmid would also have a target gene like those listed in section 7.2.1.1. There would also be total control to design the N-terminal LZD binding site anywhere in the plasmid. The things to consider with this plasmid would be ensuring that the target gene is not present in the yeast being used, ideally through a whole gene deletion, or using a completely foreign target like GFP. This would ensure that our sgRNA sequence targets only the plasmid and not the genome. Alternatively, the system could target a random sequence in the plasmid, but this would require a larger amount of screening to determine Cas9 activity.

Another important factor for this genome-mimicking plasmid is that it would have to be a low, ideally single copy plasmid in the yeast. Designing the plasmid as a yeast centromere plasmid with a CEN/ARS sequence should address this, as these kinds of plasmids are typically retained as a single copy. Having a high copy number plasmid would

almost certainly lead to problems as not all the plasmids would necessarily be acted on by the Cas9 and DNA repair processes.

The sgRNA and Cas9 plasmid would be designed by analogy to pG2, using an HDV-ribozyme tagged sgRNA sequence to prevent degradation in the cells. And LZD could be expressed as it currently is. The donor to be used in this experiment would also likely pull from the possible knock in genes listed in section 7.2.1.2 and use homologous regions surrounding the target Cas9 cleavage site in the other plasmid.

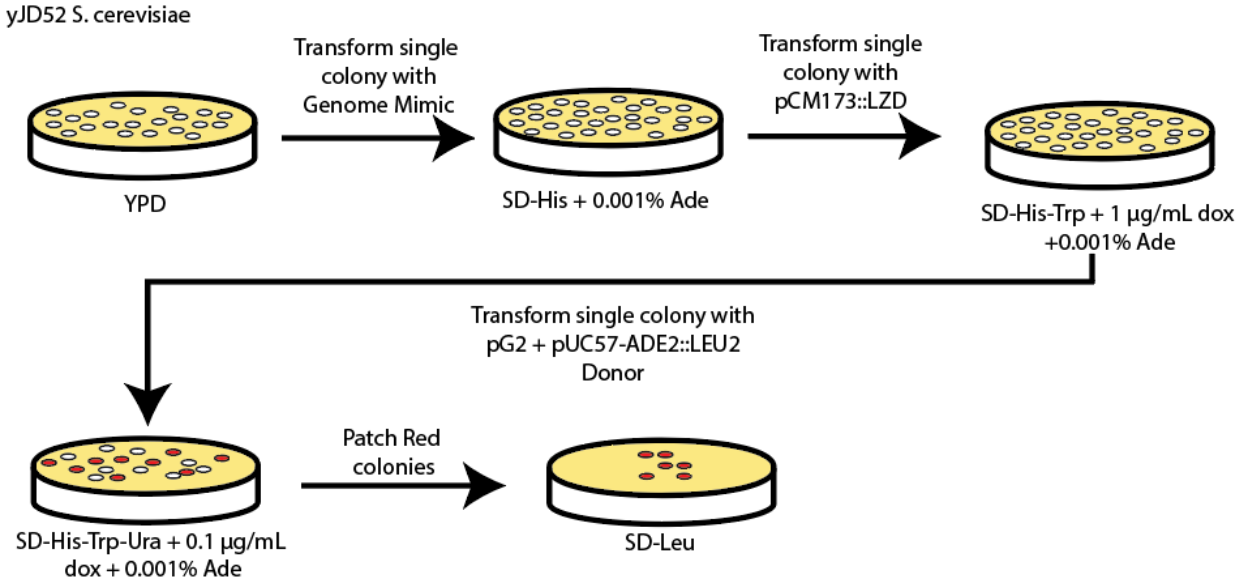

Figure 7.2 Potential Reimagination of Recombination Workflow: This modified scheme would introduce a plasmid that would mimic the genome – only in the sense that the target would be the plasmid instead of the actual genome. The initial transformation would bring in this "genome mimic" and then confirm proper activity. All subsequent steps would be in line with the current model.

To perform this experiment initial yeast strains would be transformed with the genome-mimicking plasmid to generate glycerol stocks for subsequent experiments. Similarly, these plasmid containing strains would then be transformed with the pCM173::LZD variants to generate secondary glycerol stocks. From here the experiment would proceed as it currently does, transforming in combinations of Cas9/sgRNA plasmid

and the donor DNA. Upon examination of the visible phenotype colonies of interest would be subjected to colony PCR to determine if the genome-mimic was modified in an expected way. Alternatively, the plasmid could be extracted through a yeast Midiprep kit, and the plasmid be analyzed via gel electrophoresis or sequencing to determine if changes occurred.

Section 7.2.4: LZD Expression

Another change to the system could be through the expression of LZD. Currently the system uses TetOff expression, where the presence of doxycycline is what tamps down the expression of the target protein. This requires a relatively large concentration of doxycycline to keep the expression system off when transformations are performed just to maintain the plasmid or in glycerol stocks. While there is no evidence to show that high doxycycline concentrations have a negative impact on the yeast it may be beneficial to eliminate the use of high concentrations if able. As an alternative the TetOn system could be looked at, where the absence of doxycycline turns expression of the proteins off. This system would be able to maintain LZD expression plasmids easily without the need for adding antibiotics. When expression is required add a low concentration to the media in order to introduce LZD at a controlled concentration.

Section 7.2.5: Model Organisms

Another way that this project can be expanded in the future is to branch into more organisms. Yeast is the lowest level of eukaryotic organism that can be used for this kind of experiment, and they historically are very good at homologous recombination. Moving forward into worms, fish, frogs, plants, and mammalian cells has an increased difficulty due to the systems all favoring NHEJ instead of HDR[71–73]. Forward progress is a must though if this kind of improvement on HDR is ever going to get out of research labs and into fully implementable treatment for human disease.

Section 7.3: Biological Implications

Enhancing HDR means science is moving in the right direction to one day perform this kind of work in human subjects. More improvements like this are what is needed to really drive the elimination of genetic diseases from the human population. This is not the final step, but a refinement of this process could be thing that launches HDR editing in humans to the next level.

Section 7.4: Final Conclusions

Despite the ever-changing scientific landscape, DNA repair through homologous recombination is still the future of treating and potentially curing some truly devastating genetic diseases. Current gene therapies that use NHEJ as a work around only do so because the ability to properly repair or restore function through HDR is inefficient and cumbersome.

When I took over this project I was determined to show that the inclusion of LZD in a CRISPR system would have a definitive impact on homologous recombination. Though the evidence we have at this point is indicative of an effect I do believe there is more to be done and the journey does not end here. There needs to be more quality data, at a larger quantity, in order to draw proper conclusions and perform useful statistical analysis.

I do believe the work that has been done lays the groundwork for those future experiments to occur. What I have done is designed a system in which LZD is expressed, Cas9 functions, and some amount of recombination occurs. This system is easily modifiable and modular to fit the needs of gene targets, and potential organisms. We also see that the inclusion of LZD alone does not negatively impact the otherwise normal CRISPR/Cas9 system. This is a huge benefit as it means our proteins could likely be implemented into other systems while not adding an additional burden. This is possibly the most important

revelation as the majority of genome engineering technologies could simply add LZD as an adjuvant to enhance their system without the need for additional modifications.

Appendix 1: Plasmid Sequences and Gene Fragments

Section A1.1: pRK793

ATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTT
CAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGG
CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATA
CGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTG
GGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGC
CCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGC
CGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAA
TGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGCACGGTGCACCA
ATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATT
CGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAA
TATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTCACGAGCACTTCACCAACAAGGACCATAGATTA
TGAAAACTGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGT
CGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAA
TTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCT
ACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTAC
CTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATT
TATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGA
AAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGC
TGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCT
GGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATT
ACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTC
CAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAA
CCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCC
TCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGC
GCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAA
GGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACG
CCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAA
CAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGGGAGAAAATCTTTATTTTCAAGGT**CATCAT**
CATCATCATCATCATGGAGAAAGCTTGTTTAAGGGGCCGCGTGATTACAACCCGATATCGAGCACCATTT
GTCATTTGACGAATGAATCTGATGGGCACACAACATCGTTGTATGGTATTGGATTTGGTCCCTTCATCAT
TACAAACAAGCACTTGTTTAGAAGAAATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAG
GTCAAGAACACCACGACTTTGCAACAACACCTCATTGATGGGAGGGACATGATAATTATTCGCATGCCTA
AGGATTTCCCACCATTTCCTCAAAAGCTGAAATTTAGAGAGCCACAAAGGGAAGAGCGCATATGTCTTGT
GACAACCAACTTCCAAACTAAGAGCATGTCTAGCATGGTGTCAGACACTAGTTGCACATTCCCTTCATCT
GATGGCATATTCTGGAAGCATTGGATTCAAACCAAGGATGGGCAGTGTGGCAGTCCATTAGTATCAACTA
GAGATGGGTTCATTGTTGGTATACACTCAGCATCGAATTTCACCAACACAAACAATTATTTCACAAGCGT
GCCGAAAAACTTCATGGAATTGTTGACAAATCAGGAGGCGCAGCAGTGGGTTAGTGGTTGGCGATTAAAT
GCTGACTCAGTATTGTGGGGGGGCCATAAAGTTTTCATGGTGAAACCTGAAGAGCCTTTTCAGCCAGTTA
AGGAAGCGACTCAACTCATGAATCGTCGTCGCCGTCGCTAATAAGGATCCTCTAGAGTCGACCTGCAGGC
AAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAAC
AGTTGCGCAGCCTGAATGGCGAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGATTTTCAGCCTGAT
ACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTC
CCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATG
CGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA

TCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAA
GCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCC
ATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATG
TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTC
AACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAAC
GCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAAC
AGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCA
GAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTA
TGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGA
AGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCT
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAA
CTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAG
TTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGA
GCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTAC
ACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA
AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAG
AAAAGCCCCAAAAACAGGAAGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAAT
TCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAA
ATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAAC
GTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCA
AATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAG
AGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGG
GCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGG
GCGCGTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTT
CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGT
AGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA
GATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT
AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGT
CCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT
GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT
TCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCC
TTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATA
GTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACAC
CCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGG
GAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA
GCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAA
GCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGC
CTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATA
CGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGC
GGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACA
GGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCC
AGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGC
AGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAG
CCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAATTCCG
ACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGT

GGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCC
CGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGC
TGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCAC
CTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGT
GCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTC
TCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGC
TGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTC
TCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGT
TAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAA

The red bolded sequence represents the coding sequence for the TEV Protease purified from this plasmid.

Section A1.1.1: TEV Protease Protein Sequence

GHHHHHHGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQSLHG
VFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTNFQTKSMSSMVSDTSCTF
PSSDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGW
RLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMN

The sequence includes TEV Protease cleavage site fragment and 7 × HIS tag, which are underlined for ease of viewing.

Section A1.2: LZD Coding Sequences

Exchanging the red bolded sequence in pRK793 with the following sequences generated the plasmid sequences for expressing LZD variants. Red underlined sequences indicate bases that were later mutated to convert a tyrosine into a tryptophan. The tyrosine residue is similarly red and underlined in the protein sequences.

Section A1.2.1: LZD66 Coding Sequences

GGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGTACCACCTGGAGAACGAAGTTGCGCGCCTGAAGAAGCT
GGTGGGTGAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGA
GCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.2.1.1: LZD66 Protein Sequences

GSDPAALKRARNTEAARRSRARKLQRMKQLEDKVYHLENEVARLKKLVGELQKLQRVKRARNTEAARRSR
ARKAALKG

Section A1.2.2: LZD73 Coding Sequences

GGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACT
GAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.2.2.1: LZD73 Protein Sequences

GSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGELQKLQRVKRARNT
EAARRSRARKAALKG

Section A1.2.3: LZD80 Coding Sequences

GGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAAAAGCTGGTGGAAGAACTGCTGAGCAAAGTGGGCGAACTGCAGAAGTTACAGCGG
GTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.2.3.1: LZD80 Protein Sequences

GSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELLSKVGELQKLQR
VKRARNTEAARRSRARKAALKG

Section A1.2.4: LZD87 Coding Sequences

GGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAAAAGCTGGTGGAAGAACTGCTGAGCAAAGTGCGTGCGCTGGCGGATTCTCTGGGC
GAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAA
AGGCTGCTCTGAAGGGATAG

Section A1.2.4.1: LZD87 Protein Sequences

GSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELLSKVRALADSLG
ELQKLQRVKRARNTEAARRSRARKAALKG

Section A1.2.5: PAA Coding Sequences

GGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACGGTAG

Section A1.2.5.1: PAA Protein Sequences

GSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER

Section A1.3: pCM173 (pCM173::eV)

GAATTCTTATTACGATCCTCGCGCCCCCTACCCACCGTACTCGTCAATTCCAAGGGCATCGGTAAACATC
TGCTCAAACTCGAAGTCGGCCATATCCAGAGCGCCGTAGGGGGCGGAGTCGTGGGGGGTAAATCCCGGAC
CCGGGGAATCCCCGTCCCCCAACATGTCCAGATCGAAATCGTCTAGCGCGTCGGCATGCGCCATCGCCAC

GTCCTCGCCGTCTAAGTGGAGCTCGTCCCCCAGGCTGACATCGGTCGGGGGGGCCGTCGACAGTCTGCGC
GTGTGTCCCGCGGGGAGAAAGGACAGGCGCGGAGCCGCCAGCCCCGCCTCTTCGGGGGCGTCGTCGTCCG
GGAGATCGAGCAGGCCCTCGATGGTAGACCCGTAATTGTTTTTCGTACGCGCGCGGCTGTACGCGGACCC
ACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGCCGAATAAGAAGGCTGGCTCT
GCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGTAGTAGGTGTTT
CCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAAAATGCCCCACAGC
GCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATTGATTTTCGAGA
GTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGCTCCATCGCGATGACTTAGTAAAG
CACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAAGTG
AGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTTTTACATGCCAA
TACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCT
CATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATATGAATTAATTCGGG
CCGCGGAGGCTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAACAGCGTGGATGGCGTCTCCAGGCGA
TCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGT
CAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTG
ACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCC
AAAACCGCATCACCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAA
GGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGC
ATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTAGCGTAAATACTCCACCCATTGACGTCAATGGA
AAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGG
TCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCATATATGGGCTATGAACTAATGAC
CCCGTAATTGATTACTATTAATAACTAGTCAATAATCAATGTCAACATGGCGGTAATGTTGGACATGAGC
CAATATAAATGTACATATTATGATATGGATACAACGTATGCAATGGGCCAAGCTCCTCGAGTAATTCGCG
CCACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAG
TGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTGTTCTTGAGTAACTCTTTCCTG
TAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACCGGCAGATCAA
TTCCTCGATCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTAT
CAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGT
TTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAA
AGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTAT
CAGTGATAGAGAAAAGTGAAAGTCGAGCTCGGTACCCTATGGCATGCATGTGCTCTGTATGTATATAAAA
CTCTTGTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTAGGTCCTTTGTAGCATAAATTACTAT
ACTTCTATAGACACGCAAACACAAATACACACACTAAATTACCGGATCAATTCGGGGGATCC**ATGACCAT**
GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT
CGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCC
AACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAG
CTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTAC
GATGCGCCCATCTACACCAACGTAACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATC
CGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTAT
TTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGT
CGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGC
TGCGTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGT
CTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTC
AGCCGCGCTGTACTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTT
CTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCG
TGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATC
CCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCG
ATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCG
AGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGAT
ATCCTGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGT
ACACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCC

AATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAG
CGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACG
ACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGA
CACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTG
CCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACG
CCCACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTT
ACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGG
TCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTG
CCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATC
CGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATG
GTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGT
TGATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCA
ACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAAC
CTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCA
TCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGA
TAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTA
AGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCG
AAGCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTG
GCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCG
ATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGG
CGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGC
CGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAAC
GGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACA
TCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCAC
ATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCG
CAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTCAAAAATAATAAGGATCGATCC
TCTAGAGTCGACCTGCAGGAGGGCCGCATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCC
CCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTAT
AGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTAC
GCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCC
AAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGT
TGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACG
CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT
TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
CAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCC
TAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG
GAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTA
AATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGT
GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGC
TCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAG

TTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCT
TCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCA
TAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
GTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGC
GTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC
TTCAAGAATTAATTCGGTCGAAAAAAGAAAAGGAGAGGGCCAAGAGGGAGGGCATTGGTGACTATTGAGC
ACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTATGTCTGTTATTAATTTCACAGGTAGTT
CTGGTCCATTGGTGAAAGTTTGCGGCTTGCAGAGCACAGAGGCCGCAGAATGTGCACTAGATTCCGATGC
TGACTTGCTGGGTATTATATGTGTGCCCAATAGAAAGAGAACAATTGACCCGGTTATTGCAAGGAAAATT
TCAAGTCTTGTAAAAGCATATAAAAATAGTTCAGGCACTCCGAAATACTTGGTTGGCGTGTTTCGTAATC
AACCTAAGGAGGATGTTTTGGCTCTGGTCAATGATTACGGCATTGATATCGTCCAACTGCATGGAGATGA
GTCGTGGCAAGAATACCAAGAGTTCCTCGGTTTGCCAGTTATTAAAAGACTCGTATTTCCAAAAGACTGC
AACATACTACTCAGTGCAGCTTCACAGAAACCTCATTCGTTTATTCCCTTGTTTGATTCAGAAGCAGGTG
GGACAGGTGAACTTTTGGATTGGAACTCGATTTCTGACTGGGTTGGAAGGCAAGAGAGCCCCGAGAGCTT
ACATTTTATGTTAGCTGGTGGACTGACGCCAGAAAATGTTGGTGATGCGCTTAGATTAAATGGCGTTATT
GGTGTTGATGTAAGCGGAGGTGTGGAGACAAATGGTGTAAAAGACTCTAACAAAATAGCAAATTTCGTCA
AAAATGCTAAGAAATAGGTTATTACTGAGTAGTATTTATTTAAGTATTGTTTGTGCACTTGCCTGCAAGC
CTTTTGAAAAGCAAGCATAAAAGATCTAAACATAAAATCTGTAAAATAACAAGATGTAAAGATAATGCTA
AATCATTTGGCTTTTTGATTGATTGTACAGGAAAATATACATCGCAGGGGGTTGACTTTTACCATTTCAC
CGCAATGGAATCAAACTTGTTGAAGAGAATGTTCACAGGCGCATACGCTACAATGACCCGATTCTTGCTA
GCCTTTTCTCGGTCTTGCAAACAACCGCCGGCAGCTTAGTATATAAATACACATGTACATACCTCTCTCC
GTATCCTCGTAATCATTTTCTTGTATTTATCGTCTTTTCGCTGTAAAAACTTTATCACACTTATCTCAAA
TACACTTATTAACCGCTTTTACTATTATCTTCTACGCTGACAGTAATATCAAACAGTGACACATATTAAA
CACAGTGGTTTCTTTGCATAAACACCATCAGCCTCAAGTCGTCAAGTAAAGATTTCGTGTTCATGCAGAT
AGATAACAATCTATATGTTGATAATTAGCGTTGCCTCATCAATGCGAGATCCGTTTAACCGGACCCTAGT
GCACTTACCCCACGTTCGGTCCACTGTGTGCCGAACATGCTCCTTCACTATTTTAACATGTGGAATTAAT
TCTCATGTTTGACAGCTTATCATCGAACTCTAAGAGGTGATACTTATTTACTGTAAAACTGTGACGATAA
AACCGGAAGGAAGAATAAGAAAACTCGAACTGATCTATAATGCCTATTTTCTGTAAAGAGTTTAAGCTAT
GAAAGCCTCGGCATTTTGGCCGCTCCTAGGTAGTGCTTTTTTTCCAAGGACAAAACAGTTTCTTTTTCTT
GAGCAGGTTTTATGTTTCGGTAATCATAAACAATAAATAAATTATTTCATTTATGTTTAAAAATAAAAAA
TAAAAAAGTATTTTAAATTTTTAAAAAAGTTGATTATAAGCATGTGACCTTTTGCAAGCAATTAAATTTT
GCAATTTGTGATTTTAGGCAAAAGTTACAATTTCTGGCTCGTGTAATATATGTATGCTAAAGTGAACTTT
TACAAAGTCGATATGGACTTAGTCAAAAGAAATTTTCTTAAAAATATATAGCACTAGCCAATTTAGCACT
TCTTTATGAGATATATTATAGACTTTATTAAGCCAGATTTGTGTATTATATGTATTTACCCGGCGAATCA
TGGACATACATTCTGAAATAGGTAATATTCTCTATGGTGAGACAGCATAGATAACCTAGGATACAAGTTA
AAAGCTAGTACTGTTTTGCAGTAATTTTTTTCTTTTTTTATAAGAATGTTACCACCTAAATAAGTTATAAA
GTCAATAGTTAAGTTTGATATTTGATTGTAAAATACCGTAATATATTTGCATGATCAAAAGGCTCAATGT
TGACTAGCCAGCATGTCAACCACTATATTGATCACCGATATATGGACTTCCACACCAACTAGTAATATGA
CAATAAATTCAAGATATTCTTCATGAGAATGGCCCAGCGATATATGCGGTGTGAAATACCGCACAGATGC
GTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGG
TGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAAC
GCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGT

The red and bolded sequence represents the lacZ gene present in pCM173::eV. This sequence is replaced with the following LZD coding sequences to generate the named plasmids.

Section A1.3.1: pCM173::LZD66

ATGACCATGCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGTACCACCTGGAGAACGAAGTTGCGCGCCTGAAGAAGCT
GGTGGGTGAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGA
GCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.3.2: pCM173::LZD73

ATGACCATGCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACT
GAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.3.3: pCM173::LZD80

ATGACCATGCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAAAAGCTGGTGGAAGAACTGCTGAGCAAAGTGGGCGAACTGCAGAAGTTACAGCGG
GTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCTCTGAAGGGATAG

Section A1.3.4: pCM173::LZD87

ATGACCATGCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGA
AGTTGCGCGCCTGAAAAAGCTGGTGGAAGAACTGCTGAGCAAAGTGCGTGCGCTGGCGGATTCTCTGGGC
GAACTGCAGAAGTTACAGCGGGTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAA
AGGCTGCTCTGAAGGGATAG

Section A1.3.5: pCM173::PAAW

ATGACCATGCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGC
TGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTGGCACCTGGAGAACGA
AGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACGGTAG

Section A1.4: pML107

AACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGG
ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGC
CAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGT
CAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC
TTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCT

TTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTA
ACGCGAATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC
GGTATTTCACACCGCATATCGACCGGTCGAGGAGAACTTCTAGTATATCTACATACCTAATATTATTGCC
TTATTAAAAATGGAATCCCAACAATTACATCAAAATCCACATTCTCTTCAAAATCAATTGTCCTGTACTT
CCTTGTTCATGTGTGTTCAAAAACGTTATATTTATAGGATAATTATACTCTATTTCTCAACAAGTAATTG
GTTGTTTGGCCGAGCGGTCTAAGGCGCCTGATTCAAGAAATATCTTGACCGCAGTTAACTGTGGGAATAC
TCAGGTATCGTAAGATGCAAGAGTTCGAATCTCTTAGCAACCATTATTTTTTTCCTCAACATAACGAGAA
CACACAGGGGCGCTATCGCACAGAATCAAATTCGATGACTGGAAATTTTTTGTTAATTTCAGAGGTCGCC
TGACGCATATACCTTTTTCAACTGAAAAATTGGGAGAAAAAGGAAAGGTGAGAGCGCCGGAACCGGCTTT
TCATATAGAATAGAGAAGCGTTCATGACTAAATGCTTGCATCACAATACTTGAAGTTGACAATATTATTT
AAGGACCTATTGTTTTTTCCAATAGGTGGTTAGCAATCGTCTTACTTTCTAACTTTTCTTACCTTTTACA
TTTCAGCAATATATATATATATTTCAAGGATATACCATTCTAATGTCTGCCCCTAAGAAGATCGTCGT
TTTGCCAGGTGACCACGTTGGTCAAGAAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGAT
GTTCGTTCCAATGTCAAGTTCGATTTCGAAAATCATTTAATTGGTGGTGCTGCTATAGATGCTACAGGTG
TTCCACTTCCAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGTGGGTGG
TCCTAAATGGGGTACAGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATCCGTAAAGAACTTCAATTG
TACGCCAACTTAAGACCATGTAACTTTGCATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCACAAT
TTGCTAAAGGTACTGACTTCGTTGTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGA
AGACGATGGTGATGGTGTCGCTTGGGATAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGA
ATGGCCGCTTTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCTTGGATAAAGCTAATGTTT
TGGCCTCTTCAAGATTATGGAGAAAAACTGTGGAGGAAACCATCAAGAACGAATTTCCTACATTGAAGGT
TCAACATCAATTGATTGATTCTGCCGCCATGATCCTAGTTAAGAACCCAACCCACCTAAATGGTATTATA
ATCACCAGCAACATGTTTGGTGATATCATCTCCGATGAAGCCTCCGTTATCCCAGGTTCCTTGGGTTTGT
TGCCATCTGCGTCCTTGGCCTCTTTGCCAGACAAGAACACCGCATTTGGTTTGTACGAACCATGCCACGG
TTCTGCTCCAGATTTGCCAAAGAATAAGGTCAACCCTATCGCCACTATCTTGTCTGCTGCAATGATGTTG
AAATTGTCATTGAACTTGCCTGAAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTTTTGGATGCAG
GTATCAGAACTGGTGATTTAGGTGGTTCCAACAGTACCACCGAAGTCGGTGATGCTGTCGCCGAAGAAGT
TAAGAAAATCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACATAAACTTTATAAATGAAATT
CATAATAGAAACGACACGAAATTACAAAATGGAATATGTTCATAGGGTAGACGAAACTATATACGCAATC
TACATACATTTATCAAGAAGGAGAAAAAGGAGGATGTAAAGGAATACAGGTAAGCAAATTGATACTAATG
GCTCAACGTGATAAGGAAAAAGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCACACAAA
AAGTTAGGTGTAACAGAAAATCATGAAACTATGATTCCTAATTTATATATTGGAGGATTTTCTCTAAAAA
AAAAAAAATACAACAAATAAAAAACACTCAATGACCTGACCATTTGATGGAGTTTAAGTCAATACCTTCT
TGAACCATTTCCCATAATGGTGAAAGTTCCCTCAAGAATTTTACTCTGTCAGAAACGGCCTTAACGACGT
AGTCGATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCC
AACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTC
TCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGA
TACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGTATGATCCAATATCAAAGGAAAT
GATAGCATTGAAGGATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGGGTAGTGCT
GAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCACAGGAGGTACTAGACTACCTTTCATCCT
ACATAAATAGACGCATATAAGTACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATA
TATATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTCGCGTTGCA
TTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGCTAGAA
AGTATAGGAACTTCAGAGCGCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAA
CGCACCGGACTGTAACGAGCTACTAAAATATTGCGAATACCGCTTCCACAAACATTGCTCAAAAGTATCT
CTTTGCTATATATCTCTGTGCTATATCCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCA
TCTAAACTCGACCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATTGTAGTAA
GAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAACATTTCCTATACGTAGTATATAGAGAC
AAAATAGAAGAAACCGTTCATAATTTTCTGACCAATGAAGAATCATCAACGCTATCACTTTCTGTTCACA
AAGTATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTGAAAAAATGCACCCGCA
GCTTCGCTAGTAATCAGTAAACGCGGGAAGTGGAGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCA

AAGTAGCCTTCTTCTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTATAGAGCGC
ACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATGCATTTTTGTAG
AACAAAAAAGAAGTATAGATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAA
TGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTTTACAAAAATGAAGCA
CAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCT
TTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTA
TCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGC
TGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG
CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA
TGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGA
ATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG
CAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGC
GTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC
GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAG
CATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTA
AAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT
CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG
GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAA
CGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGC
GCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAAC
GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT
TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGC
AATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGACAATCTTTGAAAAGATAATGTATG
ATTATGCTTTCACTCATATTTATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATG
TACGTTTGAAGTACAACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCA
CACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTC
GGCGTTCGAAACTTCTCCGCAGTGAAAGATAAAT**GATCAGAATCGATGCATTT**AAATAGCAAGTTAAAAT
AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTT
CGAGTCATGTAATTAGTTATAAGCATGTGAGCTAAGACACTGTAATTGCCAATCTAAACGATACCACGGC
CGCTCTAGAGAAATGGGGAGCGATTTGCAGGCATTTGCTCGGCATGCCGGTAGAGGTGTGGTCAATAAGA
GCGACCTCATGCTATACCTGAGAAAGCAACCTGACCTACAGGAAAGAGTTACTCAAGAATAAGAATTTTC
GTTTTAAAACCTAAGAGTCACTTTAAAATTTGTATACACTTATTTTTTTTATAACTTATTTAATAATAAA
AATCATAAATCATAAGAAATTCGCCTCGAGAGTCTAGGATCCGGAACTACCTACCTTGCGCTTTTTCTTG
GGAGCTCCCTCATCCCCTCCGAGCTGTGAGAGGTCAATCCGAGTTTCGTAAAGTCCGGTGATGCTCTGGT
GGATGAGGGTGGCGTCCAGCACTTCTTTGGTGGAGGTATAGCGTTTCCGGTCGATGGTGGTGTCGAAGTA
CTTGAATGCGGCAGGGGCACCAAGATTGGTGAGGGTGAACAGGTGGATGATATTCTCAGCCTGCTCCCGA

ATTGGTTTGTCTCTGTGCTTATTGTATGCGCTAAGGACTTTATCGAGGTTGGCATCAGCGAGAATGACGC
GTTTAGAAAACTCGGAAATCTGCTCAATGATCTCGTCCAGATAATGCTTATGCTGCTCCACAAACAGTTG
CTTTTGCTCGTTGTCCTCGGGGCTCCCCTTCAGCTTCTCGTAATGGCTTGCAAGGTAGAGGAAATTCACG
TACTTGGAGGGGAGAGCCAGCTCATTACCCTTCTGAAGTTCGCCAGCGGAGGCGAGCATCCGCTTCCGCC
CATTCTCCAGTTCGAACAGGCTATACTTTGGGAGCTTGATGATAAGGTCCTTCTTGACTTCCTTGTAACC
TTTGGCTTCGAGAAAGTCAATTGGGTTCTTCTCGAAGGAGCTGCGCTCCATAATGGTGATACCCAGCAGT
TCCTTCACGCTCTTGAGCTTCTTAGACTTTCCTTTTTCCACCTTAGCCACGACCAGGACAGAATAAGCCA
CAGTTGGTGAATCGAATCCGCCGTACTTCTTAGGGTCCCAATCCTTCTTCCTTGCAATGAGCTTGTCGGA
GTTGCGCTTGGGGAGGATTGATTCCTTAGAAAAGCCTCCGGTTTGCACCTCGGTTTTCTTCACGATGTTC
ACTTGAGGCATAGAAAGCACTTTGCGCACGGTAGCGAAATCTCTGCCCTTGTCCCACACGATTTCGCCAG
TCTCCCCATTAGTCTCGATGAGCGGCCTCTTCCGGATCTCACCGTTTGCAAGGGTGATTTCGGTCTTGAA
GAAATTCATGATGTTGCTGTAGAAGAAGTACTTTGCGGTGGCCTTTCCGATCTCCTGCTCAGATTTGGCA
ATCATCTTGCGCACGTCGTAGACCTTGTAGTCTCCGTACACGAACTCGCTTTCCAGCTTTGGGTACTTCT
TGATTAGGGCGGTTCCCACCACGGCATTCAGATATGCGTCATGAGCATGATGATAGTTATTGATCTCCCG
CACTTTGTAGAATTGAAAGTCCTTTCTGAAATCGCTGACCAGCTTGGACTTCAGGGTAATCACCTTCACC
TCGCGGATAAGTTTGTCGTTTTCGTCGTACTTGGTATTCATGCGTGAGTCAAGGATTTGGGCTACATGTT
TAGTAATCTGCCGAGTCTCCACAAGTTGCCGTTTAATGAATCCTGCTTTGTCCAGTTCGCTCAGTCCGCC
CCTCTCAGCCTTAGTGAGATTGTCGAACTTCCGCTGAGTGATTAGCTTTGCGTTCAGAAGCTGGCGCCAA
TAGTTCTTCATCTTTTTCACGACTTCCTCTGAGGGCACGTTATCTGACTTTCCCCGGTTCTTGTCTGAGC
GAGTCAGCACCTTATTGTCAATTGAGTCGTCTTTGAGGAAGCTCTGAGGGACGATGTGGTCCACGTCGTA
GTCTGAGAGCCTGTTGATATCAAGCTCTTGGTCCACATACATATCGCGTCCATTTTGAAGGTAGTACAGA
TAGAGCTTCTCATTCTGAAGCTGGGTGTTTTCCACCGGGTGCTCTTTAAGGATCTGTGACCCGAGCTCCT
TAATCCCTTCCTCAATTCTTTTCATCCGCTCCCTGCTGTTCTTCTGTCCCTTCTGGGTAGTCTGGTTTTC
CCTGGCCATCTCGATGACAATGTTCTCTGGTTTATGGCGACCCATCACCTTGACCAGCTCGTCCACGACT
TTCACGGTCTGGAGAATACCCTTCTTAATAGCGGGTGAACCAGCCAGATTTGCGATATGCTCGTGCAGTG
AGTCACCTTGTCCAGACACTTGGGCCTTCTGGATGTCCTCCTTAAAGGTCAGGGAATCATCATGGATAAG
TTGCATGAAGTTGCGATTAGCGAATCCGTCGCTCTTAAGGAAATCAAGAATGGTCTTTCCGCTCTGTTTG
TCCCGGATTCCGTTGATGAGTTTGCGGCTAAGGCGTCCCCATCCGGTATATCTTCTCCGTTTAAGTTGCT
TCATCACTTTATCGTCGAACAGATGGGCGTAGGTCTTAAGCCTCTCCTCGATCATTTCCCGGTCCTCGAA
CAGAGTGAGAGTCAGGACAATGTCCTCAAGGATGTCCTCTTTCTCCTCATTGTCCAGAAAATCCTTGTCC
TTGATGATCTTAAGCAGATCATGGTAGGTACCCAGGCTTGCGTTGAACCGATCTTCGACCCCGCTAATCT
CCACAGAGTCGAAGCATTCAATTTTCTTGAAATAGTCCTCCTTCAGCTGCTTCACGGTCACCTTTCTATT
GGTCTTGAACAGAAGGTCCACAATTGCCTTCTTTTGTTCGCCTGACAGGAATGCGGGCTTCCTCATCCCT
TCGGTGACATATTTCACCTTGGTCAGCTCGTTGTACACGGTAAAGTACTCGTAGAGCAGGGAATGTTTCG
GCAGGACCTTCTCATTGGGGAGGTTCTTATCGAAGTTAGTCATCCGTTCGATGAATGACTGAGCGCTGGC
TCCCTTGTCCACCACCTCTTCGAAGTTCCAGGGAGTGATAGTTTCTTCTGACTTTCTGGTCATCCAAGCA
AACCGGCTATTTCCTCTGGCGAGGGGTCCCACGTAGTAGGGGATGCGGAAAGTGAGAATTTTCTCAATCT
TCTCCCTGTTGTCCTTAAGAAAGGGGTAGAAGTCCTCTTGGCGCCGAAGGATGGCGTGAAGCTCCCCAAG
ATGGATCTGGTGGGGAATGCTCCCATTGTCGAAAGTGCGCTGCTTCCGCAGCAGATCCTCCCGATTCAGT
TTCACGAGGAGCTCTTCGGTGCCGTCCATCTTCTCAAGGATAGGTTTGATAAACTTATAAAATTCCTCTT
GGGAGGCACCTCCGTCGATGTAGCCGGCGTATCCATTCTTTGACTGGTCAAAGAAGATTTCCTTGTACTT
CTCCGGAAGCTGCTGCCGGACGAGTGCCTTAAGCAGGGTCAGGTCCTGGTGGTGCTCATCATACCGCTTA
ATCATTGAAGCTGAAAGAGGGGCCTTGGTGATTTCAGTGTTCACGCGGAGAATGTCAGACAGCAGGATGG
CGTCAGAAAGATTCTTAGCAGCAAGGAAAAGGTCGGCGTATTGGTCACCGATCTGGGCCAGGAGGTTGTC
GAGATCGTCATCATAGGTGTCCTTGGACAGTTGCAGCTTGGCATCCTCTGCCAGGTCGAAATTAGACTTG
AAATTGGGAGTCAGCCCAAGTGAAAGAGCAATCAGGTTCCCGAACAGTCCATTTTTCTTTTCACCCGGCA
GCTGAGCGATAAGATTCTCAAGTCTGCGGCTCTTTGACAGCCGGGCTGACAGAATGGCCTTGGCATCGAC
TCCGCTTGCATTGATGGGGTTCTCTTCAAAGAGCTGGTTGTAAGTCTGCACCAGCTGGATGAACAGTTTA
TCCACATCGCTATTGTCAGGATTGAGGTCGCCCTCGATCAGGAAGTGGCCTCTGAACTTAATCATGTGGG
CCAGGGCCAGGTAGATAAGCCGGAGGTCTGCCTTGTCGGTTGAGTCCACCAGCTTTTTGCGAAGATGGTA
GATGGTTGGGTATTTCTCGTGATAGGCCACCTCGTCGACGATATTACCGAAGATGGGATGCCGTTCATGC

TTCTTATCCTCTTCCACGAGGAATGATTCTTCCAGCCTATGGAAGAAGCTGTCGTCGACCTTTGCCATTT
CGTTGGAGAATATCTCTTGAAGGTAACAAATCCGGTTCTTCCTCCTGGTGTACCGCCGTCTAGCAGTCCG
CTTGAGGCGGGTTGCTTCAGCGGTCTCCCCTGAATCAAAGAGGAGAGCTCCGATCAGATTTTTCTTGATG
GAATGCCGGTCGGTGTTCCCGAGCACCTTGAACTTCTTAGAGGGCACCTTGTACTCGTCAGTGATGACGG
CCCACCCGACGCTATTAGTCCCGATGTCCAGTCCGATAGAATACTTCTTGTCCATGGTGGCGAGATCTAA
TTCGATCCATATCAGATCTTTTGTTTGTTTATGTGTGTTTATTCGAAACTAAGTTCTTGGTGTTTTAAAA
CTAAAAAAAAGACTAACTATAAAAGTAGAATTTAAGAAGTTTAAGAAATAGATTTACAGAATTACAATCA
ATACCTACCGTCTTTATATACTTATTAGTCAAGTAGGGGAATAATTTCAGGGAACTGGTTTCAACCTTTT
TTTTCAGCTTTTTCCAAATCAGAGAGAGCAGAAGGTAATAGAAGGTGTAAGAAAATGAGATAGATACATG
CGTGGGTCAATTGCCTTGTGTCATCATTTACTCCAGGCAGGTTGCATCACTCCATTGAGGTTGTGCCCGT
TTTTTGCCTGTTTGTGCCCCTGTTCTCTGTAGTTGCGCTAAGAGAATGGACCTATGAACTGATGGTTGGT
GAAGAAAACAATATTTTGGTGCTGGGATTCTTTTTTTTTCTGGATGCCAGCTTAAAAAGCGGGCTCCATT
ATATTTAGTGGATGCCAGGAATAAACTGTTCACCCAGACACCTACGATGTTATATATTCTGTGTAACCCG
CCCCCTATTTTGGGCATGTACGGGTTACAGCAGAATTAAAAGGCTAATTTTTTGACTAAATAAAGTTAGG
AAAATCACTACTATTAATTATTTACGTATTCTTTGAAATGGCAGTATTGATAATGATAAACTCGAACTGA
CTAGTAGACTGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTC
GCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTAC

The red and bolded sequence is replaced with the following sequences to make pML107 variants. The yellow highlighted sequence represents the GAP promoter. The MET25 promoter sequence is shown in section A1.4.4 and replaces this sequence in that construct.

Section A1.4.1: pML107::AGC1

GATCACTCAAATAGTCAAAAAATTGTTTTAGAGCTAG

Section A1.4.2: pML107::Random

GATCCAGACAAATACTCAATCAATGTTTTAGAGCTAG

Section A1.4.3: pML107::ADE2

GATCGAACAGTTGGTATATTAGGAGTTTTAGAGCTAG

Section A1.4.4: pML107::AGC1-MET25p

TGTATGGATGGGGGTAATAGAATTGTATCTATGTATCTGACGACCCTGTATTACACGAAGGAAATAGTAG
ATGAAAAGACAAGAGAGCAAGAAAAAGGAAAAACTTCCTTTCTAACAGACGCTTTACTTAACCTTATATA
TATCTTATTTTTTTCATCGAGCGTGTTCAATTGGACAAGGTGCCACCTTTTCGACACTTCGGTTATTATG
TTACACAGTTTTCATGAGGATGGCGCCTTGACTAACTTAATTAGTCATTTGCCAACCACCACAGTTCCCC
AATATCGACAGCTTCACGTGCCATTTGCCATTTTGCGATCATGTGACCTCAAGAGAAAAAGCAAAAAATA
A

Section A1.5: pG2

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGGA
TCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGT
TTATTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGA

AGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGAC
AAGAAAAGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGT
GTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAA
AAGGTAGTATTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTT
AATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTT
ATAGCACGTGATGAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT
TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTG
AAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTT
CCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGG
GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT
GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTC
GGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGG
ATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT
TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGC
CTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAG
CAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAT
AGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATT
GCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGC
CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGC
TGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATT
GATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAA
TCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCCTTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACTGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG
TTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAA
GGCGGACAGGTATCCGGTAAGCTGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAAC
GCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGT
CAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCC
TTTTGCTCTTAATTAAGTATATGTGTTATGTAGTATACTCTTTCTTCAACAATTAAATACTCTCGGTAGC
CAAGTTGGTTTAAGGCGCAAGACTGTAATTTATCACTACGAAATCTTGAGATCGGGCGTTCGACTCGCCC
CCGGGAGAGATGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACACCTTCGGGTGGCGA
ATGGGACTTTGGACAGTTGGTATATTAGGAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAA
ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTAACTCCGTTACTG
TGGGCTTGCCATATTTTTTGGAACTTTTCTGCCCTTTTTCTCAATGAGTAAGGAGGGCGTTTATCTTTTC
TATTTTATTTTTCTACAAATTTGATTCATGAGTCCATGGAACAAATATATAGAACTCATTTTCTTGTTTC
GGTATATATATTAACTCCATTGAAGCATTTGCTTTTATCCTGACTGGCTAGCTAGTCGTGCAATGTATGA
CTTTAAGCCGCGGTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAG
CTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCC
AATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGAC
TGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
ACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTA
TGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCATAGCTTCA
AAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAA
CACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAG
GTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATC
ACGTTTCTTTTTCTTGAAAATTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAG
TTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTT
CTTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGAGATGTCTAGAGCCCCAAAGAAG

AAGAGAAAAGTTAGACCCGGGATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCG
GTTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCG
CCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGG
CTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTA
GTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGA
TAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCA
ACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCG
CGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGA
TGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCA
TCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCG
CACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCC
CAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGAT
GATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGT
CAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAG
TATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTG
CCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAG
CAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCT
GGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAG
ATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACA
GGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTC
CAGATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGAT
AAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGG
TGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGT
CACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTC
AAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAAT
CATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCCTT
ACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAG
TCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGGT
ATGTGAACTGCAATATTAATAGCACGAGAAAATTGAGAGGAAGATAGATGGGAACTAGTAGAGTTGATAT
TGATGAGATACGAAAACCACACGTAAATAAACTATCCGACGACAAGAATAGTGCATTAAGGACTTGATTA
AGATATGGTGAGCAACGTAATTATCGGGCTCAACAGTTTATTAGCAATCGTTTTGATAGAAGCGTTGATG
CTGTGGAAGTTGTTCTTTTTACTACCAACAGCCATTAACAAATCCATTAGAGTGTTCGTTCGTTTTTCTG
CTCAGATTCAAGAAAATTATTCCATCCTCATTATACTTTTTTCTTCTATTTCGTGCTCCACGTCGAGGTA
TCAAGGAACATAGTTTACTAACATTAACGAATTCATCTCCTATGAATTTACTTTTTGTATAGATCCGAGA
CAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAG
TTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACA
GTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAA
GGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGA
GAGAACCAAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAA
AAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTA
CCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGAC
TACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAA
GATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTA
TTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGA
GGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCA
AGCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGA
GGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTG
AGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCA
AAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGAT
CGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTT

TTCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAA
CAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGT
GAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAAC
AGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAG
TCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGA
ACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAA
GGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACG
GCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATA
CGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAG
CAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAA
GAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCAT
CAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAG
TACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTC
ATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGA
CCCCAAGAAGAAGAGGAAGGTGTGATCTCTTCTCGAGTCATGTAATTAGTTATGTCACGCTTACATTCAC
GCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTA
TTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGA
CGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAA
TTTGCGGCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAA
CGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCT
GGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAC
GCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA
GCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCA
AGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTT
GATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGT
CCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTT
TGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
GCGAATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGG
TATTTCACACCGCATAGGGTAATAACTGATATAATTAAATTGAAGCTCTAATTTGTGAGTTTAGTATACA
TGCATTTACTTATAATACAGTTTTTTAGTTTTGCTGGCCGCATCTTCTCAAATATGCTTCCCAGCCTGCT
TTTCTGTAACGTTCACCCTCTACCTTAGCATCCCTTCCCTTTGCAAATAGTCCTCTTCCAACAATAATAA
TGTCAGATCCTGTAGAGACCACATCATCCACGGTTCTATACTGTTGACCCAATGCGTCTCCCTTGTCATC
TAAACCCACACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTTTGAGCA
ATAAAGCCGATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTCCAGTAG
ATAGGGAGCCCTTGCATGACAATTCTGCTAACATCAAAAGGCCTCTAGGTTCCTTTGTTACTTCTTCTGC
CGCCTGCTTCAAACCGCTAACAATACCTGGGCCCACCACACCGTGTGCATTCGTAATGTCTGCCCATTCT
GCTATTCTGTATACACCCGCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTT
CGAAGAGTAAAAAATTGTACTTGGCGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATGGAAAAATC
AGTCAAGATATCCACATGTGTTTTTAGTAAACAAATTTTGGGACCTAATGCTTCAACTAACTCCAGTAAT
TCCTTGGTGGTACGAACATCCAATGAAGCACACAAGTTTGTTTGCTTTTCGTGCATGATATTAAATAGCT
TGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATGATTTATCTTCG
TTTCCTGCAGGTTTTTGTTCTGTGCAGTTGGGTTAAGAATACTGGGCAATTTCATGTTTCTTCAACACTA
CATATGCGTATATATACCAATCTAAGTCTGTGCTCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATTACCG
AATCAAAAAAATTTCAAGGAAACCGAAATCAAAAAAAAGAATAAAAAAAAAATGATGAATTGAAAAGGTG
GTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACAC
CCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGG
GAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA

Section A1.5.1: pEmpty-Control

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGACGG
ATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTG
TTTATTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATG
AAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGA
CAAGAAAAGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCG
TGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATA
AAAGGTAGTATTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTT
TAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTT
TATAGCACGTGATGAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTA
TTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCT
TCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTG
GGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAA
TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACT
CGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG
GATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCG
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTA
GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAA
TAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG
CTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGAT
TGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA
ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG
ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTC
TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACG
ATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGA
ACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAA
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC
CTTTTGCTCTTAATTAAGTATATGTGTTATGTAGTATACTCTTTCTTCAACAATTAAATACTCTCGGTAG
CCAAGTTGGTTTAAGGCGCAAGACTGTAATTTATCACTACGAAATCTTGAGATCGGGCGTTCGACTCGCC
CCCGGGAGAGATGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACACCTTCGGGTGGCG
AATGGGACTTTGGCGGCCGCGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTAACTCCGTTACTGTGGGCTTGCCAT
ATTTTTTGGAACTTTTCTGCCCTTTTTCTCAATGAGTAAGGAGGGCGTTTATCTTTTCTATTTTATTTTT
CTACAAATTTGATTCATGAGTCCATGGAACAAATATATAGAACTCATTTTCTTGTTTCGGTATATATATT
AACTCCATTGAAGCATTTGCTTTTATCCTGACTGGCTAGCTAGTCGTGCAATGTATGACTTTAAGCCGCG
GTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTC
GCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACC
GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGC
AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTC
CGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTA
CGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCATAGCTTCAAAATGTTTCTAC
TCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAAGCACA
GCATACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGA
AAAAAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTT

CTTGAAAATTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGT
CTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGCTCATTAG
AAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGAGATGTCTAGAGCCCCAAAGAAGAAGAGAAAAGTT
AGACCCGGGATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCA
TTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAA
GAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACA
GCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGG
CTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGA
GCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCAT
CTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATA
TGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACT
CTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGAC
GCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTG
GGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATC
TAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAAT
CTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTC
TGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCG
CTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTAC
AAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGG
AATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAA
CAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGC
GAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTG
AGAAAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTG
GATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCT
GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAAC
ACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGAT
GAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCA
GCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAA
GGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAA
GATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGC
TCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGGTATGTGAACTGCA
ATATTAATAGCACGAGAAAATTGAGAGGAAGATAGATGGGAACTAGTAGAGTTGATATTGATGAGATACG
AAAACCACACGTAAATAAACTATCCGACGACAAGAATAGTGCATTAAGGACTTGATTAAGATATGGTGAG
CAACGTAATTATCGGGCTCAACAGTTTATTAGCAATCGTTTTGATAGAAGCGTTGATGCTGTGGAAGTTG
TTCTTTTTACTACCAACAGCCATTAACAAATCCATTAGAGTGTTCGTTCGTTTTTCTGCTCAGATTCAAG
AAAATTATTCCATCCTCATTATACTTTTTTCTTCTATTTCGTGCTCCACGTCGAGGTATCAAGGAACATA
GTTTACTAACATTAACGAATTCATCTCCTATGAATTTACTTTTTGTATAGATCCGAGACAAGCAGAGTGG
AAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGAT
GACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGC
ACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGA
ACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACT
ACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGT
CCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCT
GCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGAT
CATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAA
ATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCT
GCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCT
GAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCC
AAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTAT
TACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAAC

AATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCA
AGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGA
GCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAG
ATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCG
TGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAA
AAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTG
ATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTG
TACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCAT
CACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAG
GTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAA
TGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTT
GTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTG
GAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCG
CCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGC
AGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACC
ACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTA
CGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAA
GAGGAAGGTGTGATCTCTTCTCGAGTCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCA
CATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGT
TATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCA
TGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCCGGT
ACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGG
GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAA
ATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGG
TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTC
TTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT
TAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCAC
ACCGCATAGGGTAATAACTGATATAATTAAATTGAAGCTCTAATTTGTGAGTTTAGTATACATGCATTTA
CTTATAATACAGTTTTTTAGTTTTGCTGGCCGCATCTTCTCAAATATGCTTCCCAGCCTGCTTTTCTGTA
ACGTTCACCCTCTACCTTAGCATCCCTTCCCTTTGCAAATAGTCCTCTTCCAACAATAATAATGTCAGAT
CCTGTAGAGACCACATCATCCACGGTTCTATACTGTTGACCCAATGCGTCTCCCTTGTCATCTAAACCCA
CACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTTTGAGCAATAAAGCC
GATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTCCAGTAGATAGGGAG
CCCTTGCATGACAATTCTGCTAACATCAAAAGGCCTCTAGGTTCCTTTGTTACTTCTTCTGCCGCCTGCT
TCAAACCGCTAACAATACCTGGGCCCACCACACCGTGTGCATTCGTAATGTCTGCCCATTCTGCTATTCT
GTATACACCCGCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTTCGAAGAGT
AAAAAATTGTACTTGGCGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATGGAAAAATCAGTCAAGA
TATCCACATGTGTTTTTAGTAAACAAATTTTGGGACCTAATGCTTCAACTAACTCCAGTAATTCCTTGGT
GGTACGAACATCCAATGAAGCACACAAGTTTGTTTGCTTTTCGTGCATGATATTAAATAGCTTGGCAGCA
ACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATGATTTATCTTCGTTTCCTGC
AGGTTTTTGTTCTGTGCAGTTGGGTTAAGAATACTGGGCAATTTCATGTTTCTTCAACACTACATATGCG
TATATATACCAATCTAAGTCTGTGCTCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATTACCGAATCAAAA
AAATTTCAAAGAAACCGAAATCAAAAAAAAGAATAAAAAAAAAATGATGAATTGAATTGAAAAGCTGTGG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACC
CGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGG
AGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA

Section A1.6: pUC57-AGC1::KanMX

ATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG
CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCC
CGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGG
TGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGC
GCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGC
TGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAA
TTCGAGCTCGGTACCTCGCGAATGCATCTAGATATCGGGATCCGTCATATGACGCGCCAGATCTAGATAG
CTTGCCTCGTCCCC**GTACAATTTCTCCTTGGGTTCTATTGCGGGTTGTA**GCCGGGTCACCCGGCCAGCGA
CATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC
CCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGC
GAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAA
TTGTCCCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCAT
ATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAA
GGAAAAGACTCACGTTTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGG
GCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGT
TGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCT
GACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACC
ACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTG
ATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCG
CGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGAC
GAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATT
CAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTAT
TGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAG
TTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGC
AGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAAC
TTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTT
TTTTTCGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCG
TATGTGAATGCTGGTCGCTATACTGCCGTAGATTCGATACTAACGCCG**AAGCTGTCAAATTGCGAAAACA**
TTCATTGATTTGGCCGGGCCCGTCGACTGCAGAGGCCTGCATGCAAGCTTGGCGTAATCATGGTCATAGC
TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA
AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC
GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCA
CTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGC
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT
CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCG
TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA
ACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCT
CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAAT
CTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCG
ATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT
TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAAT

AAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGT
AAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG
TTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAA
TATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
AACAAATAGGGGTTCCGCGCAC

Homologous sequences are red and bolded. KanMX cassette is underlined between the homologous regions. Various constructs below will have different homologous sequences, or the KanMX cassette will be replaced with a different insert cassette.

Section A1.6.1: pUC57-ADE2::KanMX (Varied homologous sequences)

5′ Homologous Sequence TGTATAAATTGGTGCGTAAAATCGTTGGATCTCTC
3′ Homologous Sequence TTTCCCGGTTGTGGTATATTTGGTGTGGAAATGTT

Section A1.6.2: pUC57-ADE2::yEGFP (Varied insert cassette)

This construct uses the same homologous sequences as described in section A1.6.1.

GACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTC
GCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGC
GCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTG
AATTGTCCCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTC
ATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGTC
AGTAAGGGTGAAGAATTATTCACTGGTGTTGTTCCAATCTTGGTTGAATTGGATGGTGATGTTAACGGTC
ACAAGTTTTCTGTTTCTGGTGAAGGTGAAGGTGATGCTACTTATGGTAAATTGACCTTGAAGTTCATCTG
TACCACAGGTAAATTGCCAGTTCCATGGCCAACTTTGGTTACTACTTTGACTTATGGTGTCCAATGCTTC
TCTAGATACCCAGATCATATGAAGCAACACGACTTTTTCAAATCCGCTATGCCAGAAGGTTACGTTCAAG
AAAGAACCATCTTCTTCAAGGATGACGGTAACTACAAAACTAGAGCCGAAGTTAAGTTCGAAGGTGATAC
CTTGGTTAACAGAATCGAATTGAAGGGTATCGACTTCAAAGAAGATGGTAACATCTTGGGTCATAAGTTG
GAATACAACTACAACTCCCACAACGTTTACATTATGGCCGATAAGCAAAAGAACGGTATCAAGGTTAACT
TCAAGATCAGACACAACATCGAAGATGGTAGTGTTCAATTGGCTGATCACTACCAACAAAACACTCCAAT
TGGTGATGGTCCAGTTTTGTTGCCAGATAACCATTACTTGTCTACCCAATCTGCTTTGTCTAAGGACCCA
AACGAAAAAAGAGATCACATGGTCTTGTTGGAATTCGTTACTGCTGCTGGTATTACTTTGGGTATGGACG
AATTATACAAGTAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTT
TTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCA
TCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGC
TATACTGCCGTAGATTCGATACTAACGCCG

Section A1.6.3: pUC57-ADE2::KanMX v2 (Varied homologous sequences)

5′ Homologous sequence CAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGT

3′ Homologous sequence TAAGTTTATTGATATACTTGTACAGCAAATAATTATAAAA

Section A1.6.4: pUC57-ADE2::LEU2 (Varied insert cassette)

This construct uses the same homologous sequences as described in section A1.6.3.

GGCGCCTGATTCAAGAAATATCTTGACCGCAGTTAACTGTGGGAATACTCAGGTATCGTAAGATGCAAGA
GTTCGAATCTCTTAGCAACCATTATTTTTTTCCTCAACATAACGAGAACACACAGGGGCGCTATCGCACA
GAATCAAATTCGATGACTGGAAATTTTTTGTTAATTTCAGAGGTCGCCTGACGCATATACCTTTTTCAAC
TGAAAAATTGGGAGAAAAAGGAAAGGTGAGAGCGCCGGAACCGGCTTTTCATATAGAATAGAGAAGCGTT
CATGACTAAATGCTTGCATCACAATACTTGAAGTTGACAATATTATTTAAGGACCTATTGTTTTTTCCAA
TAGGTGGTTAGCAATCGTCTTACTTTCTAACTTTTCTTACCTTTTACATTTCAGCAATATATATATATAT
ATTTCAAGGATATACCATTCTAATGTCTGCCCCTAAGAAGATCGTCGTTTTGCCAGGTGACCACGTTGGT
CAAGAAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTCGTTCCAATGTCAAGTTCG
ATTTCGAAAATCATTTAATTGGTGGTGCTGCTATAGATGCTACAGGTGTTCCACTTCCAGATGAGGCGCT
GGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGTGGGTGGTCCTAAATGGGGTACAGGTAGT
GTTAGACCTGAACAAGGTTTACTAAAAATCCGTAAAGAACTTCAATTGTACGCCAACTTAAGACCATGTA
ACTTTGCATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTGACTTCGT
TGTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGAAGACGATGGTGATGGTGTCGCT
TGGGATAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGAATGGCCGCTTTCATGGCCCTAC
AACATGAGCCACCATTGCCTATTTGGTCCTTGGATAAAGCTAATGTTTTGGCCTCTTCAAGATTATGGAG
AAAAACTGTGGAGGAAACCATCAAGAACGAATTTCCTACATTGAAGGTTCAACATCAATTGATTGATTCT
GCCGCCATGATCCTAGTTAAGAACCCAACCCACCTAAATGGTATTATAATCACCAGCAACATGTTTGGTG
ATATCATCTCCGATGAAGCCTCCGTTATCCCAGGTTCCTTGGGTTTGTTGCCATCTGCGTCCTTGGCCTC
TTTGCCAGACAAGAACACCGCATTTGGTTTGTACGAACCATGCCACGGTTCTGCTCCAGATTTGCCAAAG
AATAAGGTCAACCCTATCGCCACTATCTTGTCTGCTGCAATGATGTTGAAATTGTCATTGAACTTGCCTG
AAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTTTTGGATGCAGGTATCAGAACTGGTGATTTAGG
TGGTTCCAACAGTACCACCGAAGTCGGTGATGCTGTCGCCGAAGAAGTTAAGAAAATCCTTGCTTAAAAA
GATTCTCTTTTTTTATGATATTTGTACATAAACTTTATAAATGAAATTCATAATAGAAACGACACGAAAT
TACAAAATGGAATATGTTCATAGGGTAGACGAA

Appendix 2: PCR Primers

Section A2.1: Generating LZD Coding Sequences

Section A2.1.1: SbfI and BamHI Tagged Ends

LZD_bamhI_for: CTGAAATAGGGGGATCCATGACCATGCCAGCTGCTC

LZD_sbfI_rev_v4: TCATTAGTATCCTGCAGGACGACGGCCAGTGCCAAGC

Section A2.2: Whole Plasmid Amplification of pML107

Section A2.2.1: AGC1 sgRNA

AGC1 PCR-F ITF: GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

AGC1 PCR-R ITF: AATTTTTTGACTATTTGAGTGATCATTTATCTTTCACTGC

Section A2.2.2: Random sgRNA

AGC1 PCR-F ITF: GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

Random sgRNA Reverse: ATTGATTGAGTATTTGTCTGGATCATTTATCTTTCACTGC

Section A2.2.3: ADE2 sgRNA

AGC1 PCR-F ITF: GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

ADE2 sgRNA Reverse: TCCTAATATACCAACTGTTCGATCATTTATCTTTCACTGC

Section A2.3: Donor PCRs

Section A2.3.1: Altered LZD Binding Sites

CREB Donor Forward: ATGACGTCATGCGCCAGATCTGTTTAGCTTGC

No LZD Donor Forward: AGAATCCGAGGCGCCAGATCTGTTTAGCTTGC

Donor Reverse: ATCCCAATACGCGTCAATTCACTGG

Section A2.3.2: Extended Homologous Regions

ADE2 Forward:

GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCCTAATATACCAACTGTTC

ADE2 Reverse:

CTTCTCCGCAGTGAAAGATAAATGATCGAACAGTTGGTATATTAGGAGTTTTAGAGCTAG

Section A2.3.3: Linear Donor Primers

5' INV2 F: GATCGGAATCGTCATATGACGCGCC

3' No LZD R: CACACAGGAAACAGCTATGACCATG

Section A2.3.4: Minimal Donor Primers

Primary PCR F:

CAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGTGATTCTAGAACAGTTGGTGA

Primary PCR R:

TTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTAGCGTTAGTATCGAATCTACG

Section A2.3.5: LZD Ends Primers

Donor CREB F:

ATTCTAAGAGATGACGTCATACGGACCTCTACTTGAGACACAATCAAGAAAAACAAGAAA

INV2 Donor F:

ATTCTAAGAGGTCATATGACACGGACCTCTACTTGAGACACAATCAAGAAAAACAAGAAA

Random Donor F:

ATTCTAAGAGGCACTCATGAACGGACCTCTACTTGAGACACAATCAAGAAAAACAAGAAA

CREB Donor R:

TATTGAGAATATGACGTCATCCACCTAGATAGCAATGAAGTTTTATAATTATTTGCTGTA

INV2 Donor R:

TATTGAGAATGTCATATGACCCACCTAGATAGCAATGAAGTTTTATAATTATTTGCTGTA

Random Donor R:

TATTGAGAATTTGTCGGAATCCACCTAGATAGCAATGAAGTTTTATAATTATTTGCTGTA

Section A2.3.6: LEU2 Donor Sequence

LEU2 Donor F:

CAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGTGGCGCCTGATTCAAGAAATA

LEU2 Donor R:

TTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTATTCGTCTACCCTATGAACAT

LEU2 HiFi F:

GCGAAAATCGGACAAAACAATCAAGTATGGATTCTAGAACCAATCAAGAAAAACAAGAAA

LEU2 HiFi R:

GCGCGCCATTGGGATCTCGGCGCGCCATTGGGATCCTAATTTTTATAATTATTTGCTGTA

Section A2.4: Colony PCRs

Section A2.4.1: AGC1 Primers

Consensus Forward: GCGAGTACTGGCTTGTTTGAATCTG

Consensus Reverse: ACCCGTTAATGCTTCTTAGGTGCTC

AGC1 Forward: AAATTATTTCTGGCGCTTCAGCTGG

ACG1 Reverse: ATTAAACATGCGGCTACACCATTGT

KanMX Forward: AATTGTCCTTTTAACAGCGATCGCG

KanMX Reverse: CTGTAACATCATTGGCAACGCTACC

Section A2.4.2: ADE2 Primers

ExoADE2 Forward: AACTTGACTAGCGCACTACCAGTAT

ExoADE2 Reverse: AAGTCCAACTTTTGAGCGACAGAGA

EndoADE2 Forward: ACTTATGTTATGCGCCTGCTAGAGT

EndoADE2 Reverse: TCAAGAGGATTGGAAAAGGAGCCAT

ADE2 primers are compatible with KanMX forward and reverse from section A2.4.1.

Section A2.5: LZD Mutagenesis

Section A2.5.1: LZD73 to PAA

LZD_PAA-For: CGGTAGTAGTTACAGCGGGTGAAGCGAGCT

LZD_PAA-Rev: TTCACCCACCAGCTTCTTCAGGCGCGCAAC

Section A2.5.2: Tyrosine to Tryptophan

Section A2.5.2.2: LZD66

LZD_Y35W-Y42W_For: TGGCACCTGGAGAACGAA

LZD_Y35W_Rev: CAGCTTGTCTTCCAGCTGCTT

Section A2.5.2.2: All other LZD Constructs and PAA

LZD_Y35W-Y42W_For: TGGCACCTGGAGAACGAA

LZD_Y42W_Rev: GTTCTTGCTCAGCAGTTCCTC

Appendix 3: Improving LZD Quantitation and Purification

Section A3.1: Rationale for Increasing Molar Absorptivity

When the LZD proteins were originally designed by Dr. Daniel Gowetski, they contained a tryptophan in the linker region between the 7 × HIS tag and the N-terminal DNA binding domain. When the expression system was changed to use a cleavable MBP tag, this tryptophan was no longer present in the final peptide. The only remaining way to quantitate the protein with UV spectroscopy became A_{280} using the single tyrosine present, or A_{214} with the peptide backbone. The molar absorptivity of the peptide bond is approximately 923 M^{-1} cm^{-1} at 214 nm[74] while tyrosine is 1480 M^{-1} cm^{-1} at 280 nm[75]. At the concentrations LZD has historically been purified at it is very difficult to determine an accurate concentration with the structure as it currently is. Placing a single tryptophan in the protein would increase the A_{280} signal of the protein almost 4-fold compared to having a single tyrosine present. It would also be beneficial to accurately determine the concentration of very dilute samples of LZD for the purpose of analytical ultracentrifugation and circular dichroism experiments that require low concentrations.

When deciding where the tryptophan should go in the protein the major concern was to not disrupt the leucine zipper, or the charged surface of the molecule. For these reasons it was decided to mutate the tyrosine into a tryptophan. Both residues are aromatic R groups and ideally the change would not impact the stability of the final protein. To generate this sequence, pMal-c2T-LZD plasmid was used as a PCR template and mutated the tyrosine codon (TAC) to a tryptophan codon (TGG). All pMal-c2T-LZD plasmids were modified this way through PCR, and then transformed into *E. coli* to be Miniprepped and sequenced for confirmation. PyMol renderings of the tryptophan variant can be found in Figure A3.1.

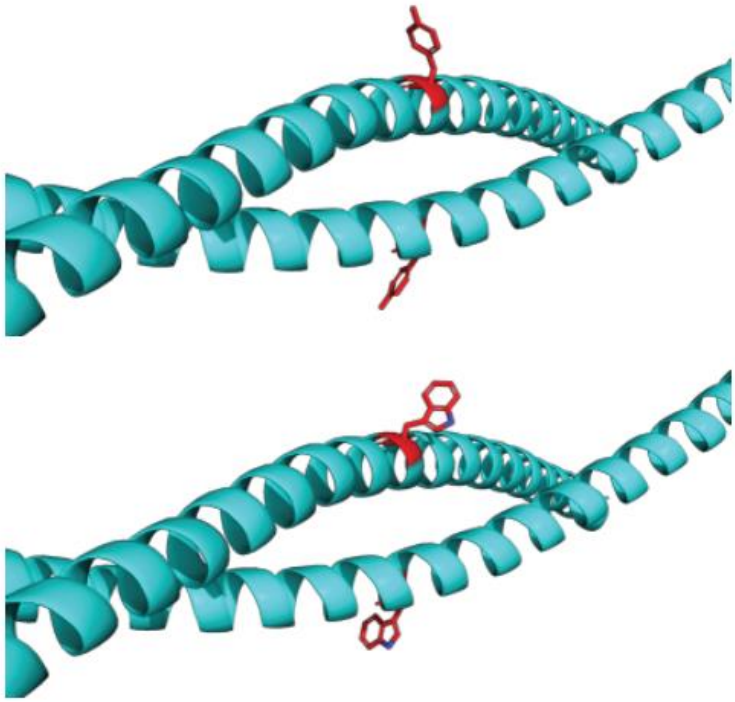

Figure A3.1 LZD Modeled with Tyrosine or Tryptophan: Red side chain residue is Tyrosine in the top and Tryptophan in the bottom. Molecules rendered in PyMol.

At this same time Dr. Jason Hustedt and myself designed another mutagenesis in order to generate a plasmid that would only express the bZip fragment of GCN4, essentially the LZD protein without a C-terminal DNA binding domain. This construct could then be purified for use in experiments as a control. The construct known in house as PAA would also go on to be used as a control in yeast recombination experiments. To generate this sequence, we began with pMal-c2T-LZD73 and mutated the sequence CTGCAGAAT (L-Q-K) to CGGTAGTAG (R-Stop-Stop). The double stop was not necessary but would ensure the protein would terminate at this point and not be read through. The LZD73 plasmid was modified this way through PCR, and then transformed into *E. coli* to be Miniprepped and sequenced for confirmation. Upon confirmation it subsequently underwent a second mutagenesis to generate a tryptophan variant as previously described.

Section A3.2: Materials and Methods

All enzymes used in reactions including PCR were obtained from NEB. Purification Kits were obtained from Qiagen. Oligonucleotides used as primers were synthesized and obtained from IDT. XL1-Blue Electrocompetent *E. coli* were purchased from Agilent. All other reagents used were purchased from Fisher Scientific. PCRs were run in an MJ research PTC-200 Gradient Thermal Cycler.

The pMal-c2T::LZD vectors used as templates were prepared from Midiprep aliquots that were diluted down to appropriate concentrations for PCR. The reaction mixtures were 50 µL and included 1 × Q5 reaction buffer, 200 µM dNTPs, 500 nM each primer (forward and reverse), 1 unit of Q5 DNA polymerase, and 10 ng of plasmid template. The cycling conditions for the reaction are as follows:

Initial Denaturation: 98 °C – 30 seconds

Cycle (30 repeats):

98 °C – 15 seconds

67 °C – 30 seconds

72 °C – 120 seconds

Final Extension: 72 °C – 4 minutes

Successful PCR reactions were confirmed by electrophoresis on 0.7 % agarose gels as previously described. Once proper size was confirmed reactions were digested with DpnI restriction enzyme to remove the methylated plasmid template. Samples were then treated with T4 PNK to phosphorylate 5' ends for ligation. T4 PNK reaction mixtures used 1 × T4 DNA ligase buffer (with ATP), 10 units of T4 PNK and 2 µL of PCR product. The reaction incubated at 37 °C for 30 minutes. The enzyme was then heat inactivated by incubating at 65 °C for 20 minutes. After cooling to room temperature 5 µL of the reaction mixture was

mixed with an additional 1 µL of T4 DNA Ligase buffer (with ATP) and 400 units of T4 DNA ligase and left to incubate at 16 °C overnight. Bacterial transformations were carried out using the electrocompetent cell method described previously. The transformation mixtures were plated onto LB-amp agar and incubated overnight at 37 °C. Plasmids were Miniprepped and then confirmed by sequencing before subsequently being Midiprepped for storage at higher concentrations.

Section A3.3: Modifications to Protein Purification

When LZD switched from a His tagged protein to an MBP tagged protein the purification system changed slightly to address changes in the system. By using an amylose affinity column (MBPTrap) to isolate the fusion and subsequent proteolytic reaction with TEV protease the MBP fraction could be removed by a second pass on the MBPTrap. To remove the TEV protease from solution it needed to be pushed across an IMAC column in a separate FPLC run. This posed some problems as the LZD protein bound to the IMAC column as well, leading to no separation of the two components. To address this issue, a complex purification scheme was developed that would address many concerns and lead to the purest protein product.

The modifications began with the workup of the lysed cells after expression of the fusion protein. A large amount of DNA was bound when trying to quantify the MBP-LZD fusion protein after eluting from the MBPTrap column. This makes sense since LZD is a DNA binding protein, but it was strange that the DNA would carry over through the FPLC run. To address this, the NaCl concentration was increased from 200 mM to 1 M in the MBP Elution and Equilibration buffers. The increased ionic strength of the solution should reduce the protein DNA interaction. Additionally, a polyethyleneimine (PEI) precipitation was added after the cell lysis step. PEI is well known to remove nucleic acids from solution

at certain concentrations. This addition helped clear samples and lead to overall less viscous lysates when it came to push filtering the solution.

After confirming that the high salt buffers and PEI did not interfere with the MBPTrap column and removed almost all the nucleic acids we proceeded to troubleshoot separating the post-TEV protease reaction mixture. This solution contained cleaved MBP, free LZD, and TEV protease. In order to accurately perform CD, AUC, or even AFM pure LZD samples were needed, free of any remaining MBP and TEV protease. To do this an ion exchange column was used, because the expected isoelectric points of the three components were at least one pI unit different from each other as can be seen in Table A3.1. Separating the three components, and only purifying LZD was necessary so the use a cation exchange HiTrap SP HP column made the most sense. The tris buffer used was at pH 7.5 which meant that the MBP in solution should not even bind the column, and just wash out in the flow through. Separating the TEV protease and LZD, which are about 2 pI units apart, was then possible by gradient elution with NaCl, starting at 10 mM and ending at 1 M. In our experience LZD has eluted last, and typically around the 800-850 mM NaCl concentration.

Table A3.1 Molecular Weight and pI of MBP, TEV, and LZD Proteins

Protein	Molecular Weight	pI
MBP	42821.30	5.02
TEV Protease	28617.50	9.59
PAA	6758.77	10.41
PAAW	6781.81	10.62
LZD66	9004.49	11.73
LZD66W	9027.53	11.84
LZD73	9818.40	11.47
LZD73W	9841.43	11.60
LZD80	10617.33	11.15
LZD80W	10640.37	11.31
LZD87	11344.16	11.13
LZD87W	11367.20	11.29

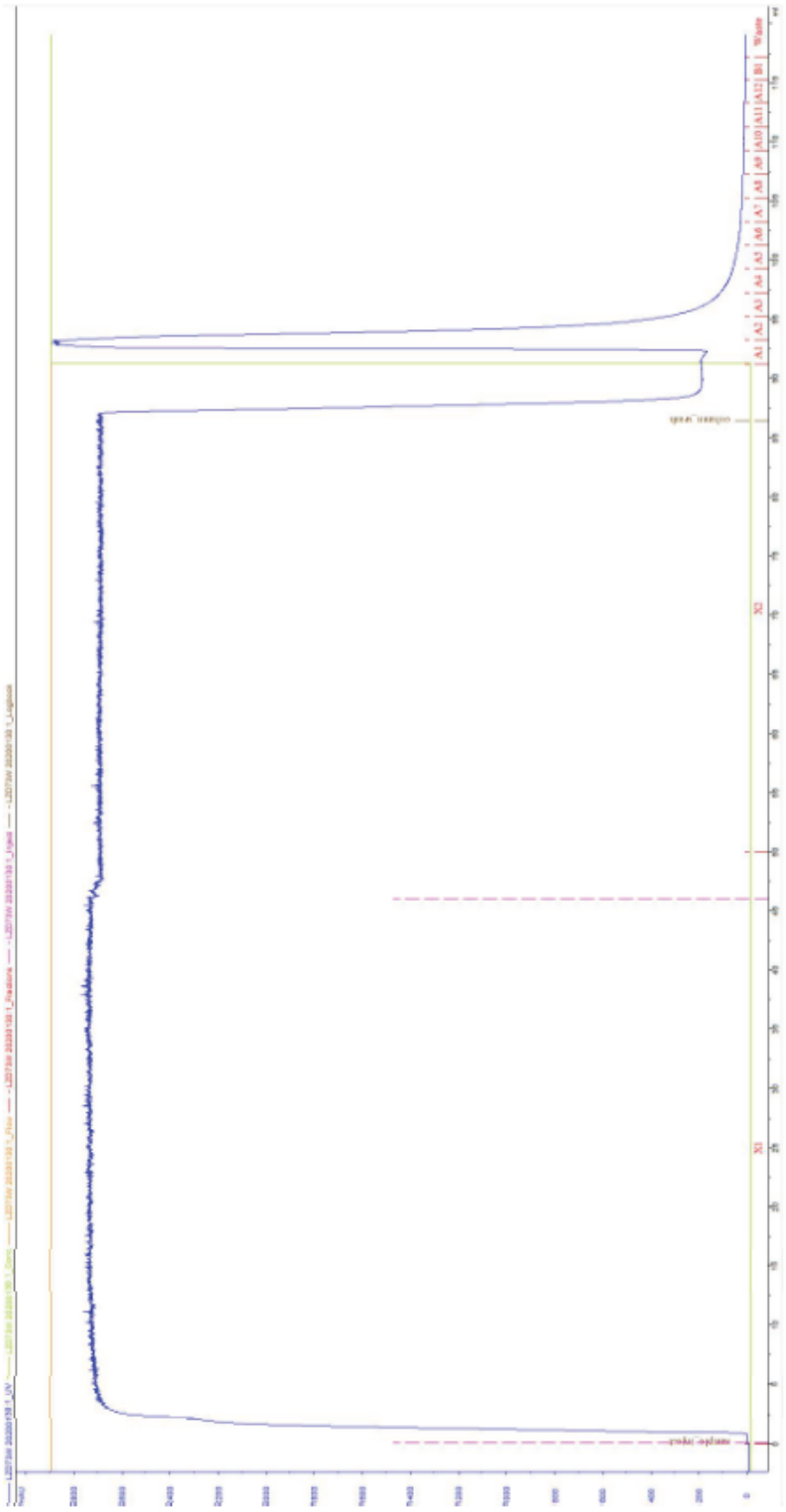

Figure A3.2 MBPTrap FPLC Trace: The blue line indicates the absorbance of the protein. There is a single elution peak that appears at A2-A4 and is the MBP-LZD fusion protein.

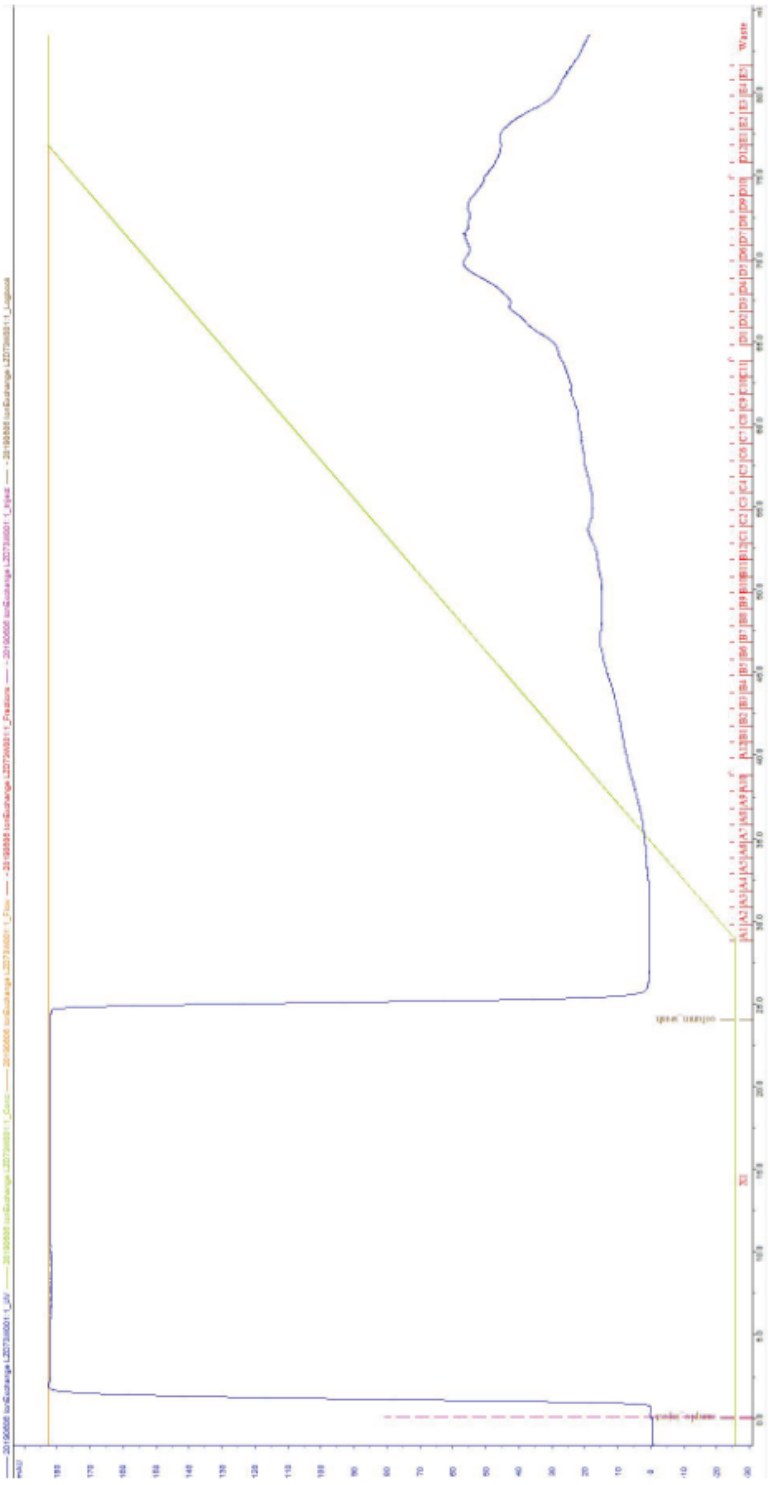

Figure A3.3 HiTrap HP SP FPLC Trace: The blue line indicates absorbance of the protein. Elution from the ion exchange column is not as clean as the MBPTrap and requires additional screening by SDS-PAGE to determine exactly where the desired LZD protein is.

Appendix 4: Protein Purification

Section A4.1: Purification Protocol

Day 0

Transform pMal-c2T-LZD into BL21 *E. coli* and allow it to grow overnight at 37 °C.

Day 1

- Morning - Inoculate single colony in 5 mL of LB with 5 µL of Ampicillin stock and grow while shaking for 6 hours at 37 °C.
- Morning – Prepare induction media (LB, Autoinduction buffer A and B, Ampicillin) and leave at 37 °C.
- Evening – Inoculate pre-warmed induction media with 1 mL of initial culture. Grow overnight (16 hours) with shaking at 37 °C. Keep the remainder for SDS-PAGE.

Day 2

1. Place cells on ice to cool for 20 minutes. *Do not keep the cells off ice for longer than necessary until after lysis*
2. Collect cells via centrifugation, 20 minutes at 5k rpm. Reserve 1 mL of culture for SDS-PAGE
3. Discard the supernatant. Collect cell pellets using a spatula and combine them into a tared 50 mL Falcon tube.
4. Spin down the pellet in a tabletop centrifuge, 5 minutes at 5k rpm.
5. Discard the supernatant and obtain the mass of the cell pellet.
6. Add 1.5-2 g of cell pellet to a fresh 50 mL Falcon tube and resuspend with 40 mL of MBP equilibration buffer. Add 400 µL of PMSF solution.
7. Use a probe sonicator to lyse cells with the following conditions: 30% amplitude, 15s on, 15s off, 10 minutes.

8. Repeat steps 6 and 7 until the full cell pellet has been sonicated.
9. Combine all lysed samples into a clean Erlenmeyer flask with a stir bar. Place the flask in an ice bath on a stir plate and begin stirring.
10. Add 1/100th volume of PEI solution. Allow the mixture to spin for 30 minutes.
11. Spin down the treated lysates for 20 minutes at 15k rpm. Supernatant contains the protein of interest. Keep an amount of the pellet for SDS-PAGE.
12. Filter the supernatant with a 60 mL syringe fitted with a 0.45 µm luer-lock filter.
13. Load the clarified lysate to the FPLC super-loop. Based on total volume this may need to be repeated. Keep 1 mL for SDS-PAGE
14. Run the FPLC on the MBPTrap protocol. Upon completion keep 1 mL of the waste fraction for SDS-PAGE. Keep 50 µL of eluted fraction for SDS-PAGE.
15. Transfer the remainder of the eluted fraction to a pre-washed cut of dialysis tubing. Seal both ends and place in large beaker with 4 L of TEV reaction buffer. Seal the beaker with plastic wrap and aluminum foil. Place in cold room and allow to equilibrate overnight (16 hours)

<u>Day 3</u>

16. Transfer the contents of the dialysis bag to a 15 mL Falcon tube. Obtain an OD for the eluted fraction.
17. Add TEV protease to the Falcon tube. 1 OD of TEV protease should be sufficient to cleave up to 100 OD of MBP-tagged protein.
18. Seal the Falcon tube with parafilm, label it, and place in 37 °C water bath overnight.

<u>Day 4</u>

19. Load the digested sample to the FPLC super-loop. Keep 50 µL for SDS-PAGE.

20. Run the FPLC on the IEX protocol. Upon completion keep 1 mL of the waste fraction for SDS-PAGE.
21. Pool the peak of interest (should elute around 800 mM NaCl) and put 50 µL aside for SDS-PAGE.
22. Buffer exchange the pooled sample with LZD storage buffer using Amicon® Ultra Centrifugal Filter, 3 kDa.
23. Determine concentration of the sample and dilute with excess LZD storage buffer if necessary for your application. Store aliquots at -80 °C.
24. Run SDS-PAGE with all nine collected samples plus an amount of the final sample in LZD storage buffer.

Section A4.2: AKTA FPLC Operating Protocols

Section A4.2.1: MBPTrap Column

Variables:
InjectionVolume: 56
Main Method:
    Main
    0.00 Base CV 0.96 {ml}HiTrap_Chelating_HP_1_ml
    0.00 Block starting_conditions
        (starting_conditions)
        0.00 Base SameAsMain
        0.00 AutozeroUV
        0.00 Alarm_Pressure Enabled 0.500 {MPa} 0.000 {MPa}
        0.00 Flow 1 {ml/min}
        0.00 ColumnPosition Position3
        0.00 InjectionValve Load
        0.00 End_Block
    0.00 Block column_equilibration
        (column_equilibration)
        0.00 Base SameAsMain
        0.00 Flow 1.00 {ml/min}
        0.00 ColumnPosition Position3
        8.00 AutozeroUV
        8.00 End_Block
    0.00 Block sample_inject
        (sample_inject)
        0.00 Base SameAsMain
        0.00 ColumnPosition Position3
        0.00 InjectionValve Inject

0.00 Fractionation 30mm 50.00 {ml}, TubeNumber [X,1], Volume
0.00 Flow 1.00 {ml/min}
0.00 Set_Mark “sample_inject”
(56.00) #InjectionVolume InjectionValve Load
56.00 End_Block

0.00 Block column_wash

(column_wash)
0.00 Base SameAsMain
0.00 InjectionValve Load
0.00 Flow 1.00 {ml/min}
0.00 Set_Mark “column_wash”
5.00 FractionationStop
5.00 End_Block

0.00 Block elution

(elution)
0.00 Base SameAsMain
0.00 Fractionation 18mm 2.000 {ml} TubeNumber[A.1] Volume
0.00 Gradient 100 {%B}, 0.00 {base}
0.00 Gradient 100 {%B}, 20 {base}
27.00 FractionationStop
29.00 End_Block

0.00 End_Method

Section A4.2.2: IEX SP HP Column

Variables:
InjectionVolume: 25
Gradient_Percent_B: 100
Gradient_length_cv: 100
Main Method:

Main
0.00 Base CV 0.96 {ml}HiTrap_SP_HP_1_ml

0.00 Block starting_conditions

(starting_conditions)
0.00 Base SameAsMain
0.00 AutozeroUV
0.00 Alarm_Pressure Enabled 0.300 {MPa} 0.000 {MPa}
0.00 Flow 1 {ml/min}
0.00 ColumnPosition Position3
0.00 InjectionValve Load
0.00 End_Block

0.00 Block column_equilibration

(column_equilibration)
0.00 Base SameAsMain
0.00 Flow 1.00 {ml/min}
0.00 ColumnPosition Position3
8.00 AutozeroUV
8.00 End_Block

0.00 Block sample_inject

(sample_inject)
0.00 Base SameAsMain
0.00 ColumnPosition Position3
0.00 InjectionValve Inject

0.00 Fractionation 30mm 50.00 {ml}, TubeNumber [X,1], Volume
0.00 Flow 1.00 {ml/min}
0.00 Set_Mark “sample_inject”
(25.00) #InjectionVolume InjectionValve Load
25.00 End_Block

0.00 Block column_wash

(column_wash)
0.00 Base SameAsMain
0.00 InjectionValve Load
0.00 Flow 1.00 {ml/min}
0.00 Set_Mark “column_wash”
5.00 FractionationStop
5.00 End_Block

0.00 Block elution

(elution)
0.00 Base SameAsMain
0.00 Fractionation 18mm 1.000 {ml} TubeNumber[A.1] Volume
0.00 Gradient (100)# Gradient_Percent_B {%B}, (100)#gradient_length_cv {base}
105.00 FractionationStop
107.00 End_Block

0.00 End_Method

Appendix 5: Setting Up a Yeast Lab

While Dr. Dinman was more than generous in allowing us to use his lab space for yeast work, it was a goal to set up a separate space in the biochemistry building in order to perform experiments in the same building as the plasmids would be stored. This would also have the benefit of not needing reagents transported from one building to another across campus. To do so room 2505, which was not in use, was provided by the department. After the previous occupants' supplies were cleared out the room was bleached and decontaminated.

A small combination refrigerator and freezer was left behind and cleaned as well, this would become storage for and temperature sensitive reagents like antibiotics and plates, as well as DNA stocks so they were not stored with the lab's bacterial reagents. Two benches were kept and used, one for a sterile cloning bench and the other for a tabletop centrifuge and heating element. The sterile bench was equipped with pipettes, and a Bunsen burner.

An incubator and water bath were acquired from the remains of Dr. Dorothy Beckett's equipment. The incubator was constantly set to 30 °C and used for storing yeast plates while in the process of growing. The water bath would be turned on when needed, set to 30 °C and used for growing overnight cultures and other small inoculations of yeast cell suspensions for transformations.

To ensure that the room was up to code with ESSR a hands-free eye was station was installed to the faucet of the sink in the room.

Appendix 6: Commonly Used Buffers

Section A6.1: Bacterial Media

LB – Liquid Media (1 L; autoclaved)
5 g Yeast Extract
10 g Tryptone
10 g NaCl

LB – Agar Plates (1 L; autoclaved)
5 g Yeast Extract
10 g Tryptone
10 g NaCl
7.5 g Agar

Add antibiotic solution if needed while cooling, mix well before pouring.

SOC (100 mL; autoclaved before addition of $MgSO_4$ and glucose)
5 g Yeast Extract
20 g Tryptone
0.5 g NaCl
0.186 g KCl
4.8 g $MgSO_4$ (separately filter sterilized, added after autoclaving)
3.6 g Glucose (separately filter sterilized, added after autoclaving)

Ampicillin Stock (10 mL; filter sterilized)
1 g dissolved in water (final 100 mg/mL; 1,000 × working concentration)

Section A6.2: Yeast Media

YPAD (1 L; autoclaved)
10 g Yeast Extract
20 g Peptone
20 g Dextrose
0.4 g Adenine Sulfate
20 g Agar (if making plates – omit for liquid media)

YPAA (1 L; autoclaved)
10 g Yeast Extract
20 g Peptone
20 g Sodium Acetate
0.4 g Adenine Sulfate
20 g Agar (if making plates – omit for liquid media)

Amino acid mix (1 L; autoclaved)
0.5 g Uracil
0.5 g L-tryptophan

0.5 g L-histidine
0.5 g L-arginine
0.5 g L-methionine
0.7 g L-tyrosine
0.7 g L-isoleucine
0.7 g L-lysine
1.2 g L-phenylalanine
3.0 g L-valine

pH of amino acid mix should be above 7, if not adjust with 1 M NaOH. Omit appropriate amino acids from the mix to account for auxotrophic selection. For SD media include only the amino acids necessary based on the strain genotype except those omitted for auxotrophic control. For SC media include all amino acids necessary based on the strain genotype.

Synthetic Dropout / Synthetic Complete Media (1 L; autoclaved)
1.7 g Yeast Nitrogenous Base (without Amino Acids or Ammonium Sulfate)
5 g Ammonium Sulfate **OR** 5 g Monosodium Glutamate (if using G418)
20 g Dextrose
1 mL 1% (w/v) Adenine Hemisulfate (final 0.001% w/v)

If making plates:
Adjust pH to 5.6 with 1 M NaOH (dropwise)

20 g Agar

After autoclaving liquid or solid media while cooling:
5 mL 5% L threonine
5 mL 2% Sodium aspartate
5 mL 1% L-leucine

Omit any or all of the amino acids if not required for auxotrophy based on yeast strain genotype. Add antibiotic solution if needed while cooling, mix well before pouring.

Doxycycline stock (20 mL; filter sterilized and aliquoted)
0.2 g dissolved in water (final 10 mg/mL; 1,000 × – 10,000 × working concentration)

G418 Stock (10 mL; filter sterilized and aliquoted)
2 g dissolved in water (final 200 mg/mL; 1,000 × working concentration)

Section A6.3: General Lab Reagents

Gel Extraction Buffer pH 7 (200 mL; filter sterilized)
0.82 g NaOAc (final 50 mM)
400 µL 500 mM EDTA (final 1 mM)
Adjust final pH to 7

5 × TBE (500 mL; autoclaved)

27 g Tris base (5 × is 445 mM, 1 × is 89 mM)
13.75 g Boric acid (5 × is 445 mM, 1 × is 89 mM)
20 mL 0.5 M EDTA pH 8 (5 × is 10 mM, 1 × is 2 mM)

Autoinduction A (250 mL; filter sterilized)
12.5 g lactose (5% w/v)
3.125 g glucose (1.25% w/v)
31.25 mL glycerol (12.5% v/v)

Autoinduction B (250 mL; filter sterilized)
19.8 g $(NH_4)SO_4$ (final 0.6 M)
42.5 g KH_2PO_4 (final 0.72 M)
44.375 g Na_2HPO_4 (final 1.48 M)

Autoinduction solution for culture
1 L LB, 40 mL Autoinduction A, 40 mL Autoinduction B, 1 mL 100 mg/mL Amp

Maltose Equilibration Buffer (500 mL; filter sterilized)
5 mL 2 M Tris HCl pH 7.5 (final 20 mM)
29.22 g NaCl (final 1 M)
0.146 g EDTA (final 1 mM)

Maltose Elution Buffer (500 mL; filter sterilized)
5 mL 2 M Tris HCl pH 7.5 (final 20 mM)
29.22 g NaCl (final 1 M)
0.146 g EDTA (final 1 mM)
1.802 g Maltose (final 10 mM)

TEV Reaction Buffer (200 mL; filter sterilized)
200 µL 0.5 M EDTA pH 8 (final 500 µM)
0.031 g DTT (final 1 mM)
5 mL 2 M Tris pH 8 (final 50 mM)

IEX Equilibration Buffer (500 mL; filter sterilized)
12.5 mL 2 M Tris HCl pH 7.5 (final 50 mM)
0.29 g NaCl (final 10 mM)

IEX Elution Buffer (500 mL; filter sterilized)
12.5 mL 2 M Tris HCl pH 7.5 (final 50 mM)
29.22 g NaCl (final 1 M)

LZD Storage Buffer (100 mL; filter sterilized)
1 mL 2 M Tris HCl pH 7.5 (final 20 mM)
0.88 g NaCl (final 150 mM)
20 mL Glycerol (20% v/v)

Polyethyleneimine (10 mL)
2 mL 50 % (w/v) dissolved in water (final 10% w/v; 100 × working concentration)

PMSF (10 mL)
0.174 g dissolved in isopropanol (final 100 mM; 100 × working concentration)

Bibliography

(1) Crick, F. (1970) Central dogma of molecular biology. *Nature 227*, 561–563.

(2) Goldberg, A. L., and Wittes, R. E. (1966) Genetic code: aspects of organization. *Science 153*, 420–424.

(3) Long, C., McAnally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R., and Olson, E. N. (2014) Prevention of muscular dystrophy in mice by CRISPR/Cas9–mediated editing of germline DNA. *Science 345*, 1184–1188.

(4) Nørremølle, A., Riess, O., Epplen, J. T., Fenger, K., Hasholt, L., and Sørensen, S. A. (1993) Trinucleotide repeat elongation in the huntingtin gene in Huntington disease patients from 71 Danish families. *Hum. Mol. Genet. 2*, 1475–1476.

(5) Bank, A., Mears, J. G., Ramirez, F., Baird, M., and Feldenzer, J. (1981) Organization of normal and abnormal human globin genes by restriction enzyme analysis, in *The molecular basis of mutant hemoglobin dysfunction* (Sigler, P. B., Ed.), pp 51–62.

(6) Feldenzer, J., Mears, J. G., Burns, A. L., Natta, C., and Bank, A. (1979) Heterogeneity of DNA fragments associated with the sickle-globin gene. *J. Clin. Investig. 64*, 751–755.

(7) Greenaway, P. J. (1980) Basic steps in genetic engineering. *Int. J. Environ. Stud. 15*, 11–24.

(8) Mach, B. (1978) Genetic engineering. *Bull. Schweiz. Akad. Med. Wiss. 34*, 401–410.

(9) Levinthal, C. (1956) The mechanism of DNA replication and genetic recombination in phage*. *Proc. Natl. Acad. Sci. U.S.A. 42*, 394–404.

(10) Poteete, A. R. (2001) What makes the bacteriophage λ red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol. Lett. 201*, 9–14.

(11) Wannier, T. M., Ciaccia, P. N., Ellington, A. D., Filsinger, G. T., Isaacs, F. J., Javanmardi, K., Jones, M. A., Kunjapur, A. M., Nyerges, A., Pal, C., Schubert, M. G., and Church, G. M. (2021) Recombineering and MAGE. *Nat. Rev. Methods Primers 1*, 1–55.

(12) Davis, L., and Maizels, N. (2011) DNA nicks promote efficient and safe targeted gene correction. *PLoS One 6*, 1–7.

(13) Ramirez, C. L., Certo, M. T., Mussolino, C., Goodwin, M. J., Cradick, T. J., McCaffrey, A. P., Cathomen, T., Scharenberg, A. M., and Joung, J. K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res. 40*, 5560 5568.

(14) Bibikova, M., Golic, M., Golic, K. G., and Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics 161*, 1169–1175.

(15) Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol. 31*, 397–405.

(16) Chandrasegaran, S., and Carroll, D. (2016) Origins of programmable nucleases for genome engineering. *J. Mol. Biol. 428*, 963–989.

(17) Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods 12*, 939–942.

(18) Charpentier, E., and Doudna, J. A. (2013) Rewriting a genome. *Nature 495*, 50–51.

(19) Doudna, J. A., and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science 346*, 1–9.

(20) Donohoue, P. D., Pacesa, M., Lau, E., Vidal, B., Irby, M. J., Nyer, D. B., Rotstein, T., Banh, L., Toh, M. S., Gibson, J., Kohrs, B., Baek, K., Owen, A. L. G., Slorach, E. M., Overbeek, M. van, Fuller, C. K., May, A. P., Jinek, M., and Cameron, P. (2021) Conformational control of Cas9 by CRISPR hybrid RNA-DNA guides mitigates off-target activity in T cells. *Mol. Cell 81*, 3637–3649.

(21) Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol. 38*, 824–844.

(22) Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., and Liu, D. R. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature 576*, 149–157.

(23) Nowak, C. M., Lawson, S., Zerez, M., and Bleris, L. (2016) Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res. 44*, 9555–9564.

(24) Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2014) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle 7*, 2902–2906.

(25) Labhart, P. (1999) Nonhomologous DNA end joining in cell-free systems. *Eur. J. Biochem. 265*, 849–861.

(26) Walker, J. R., Corpina, R. A., and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature 412*, 607–614.

(27) Wilson, T. E., Grawunder, U., and Lieber, M. R. (1997) Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature 388*, 495–498.

(28) Orr-Weaver, T. L., Szostak, J. W., and Rothstein, R. J. (1981) Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U.S.A. 78*, 6354–6358.

(29) Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell 33*, 25–35.

(30) Symington, L. S., and Gautier, J. (2011) Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet. 45*, 247–271.

(31) Bonilla, B., Hengel, S. R., Grundy, M. K., and Bernstein, K. A. (2020) RAD51 gene family structure and function. *Annu. Rev. Genet. 54*, 25–46.

(32) Scully, R., Panday, A., Elango, R., and Willis, N. A. (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol. 20*, 698–714.

(33) Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol. 18*, 495–506.

(34) Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Rivera, R. M. C., Madhavan, S., Pan, X., Ran, F. A., Yan, W. X., Asokan, A., Zhang, F., Duan, D., and Gersbach, C. A. (2016) In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science 351*, 403–407.

(35) Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G.-C., Zhang, F., Orkin, S. H., and Bauer, D. E. (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature 527*, 192–197.

(36) Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., Fuente, J. de la, Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., Montalembert, M. de, Rondelli, D., Sharma, A., Sheth, S., Soni, S., Steinberg, M. H., Wall, D., Yen, A., and Corbacioglu, S. (2020) CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *N. Engl. J. Med. 384*, 252–260.

(37) Chang, K.-H., Smith, S. E., Sullivan, T., Chen, K., Zhou, Q., West, J. A., Liu, M., Liu, Y., Vieira, B. F., Sun, C., Hong, V. P., Zhang, M., Yang, X., Reik, A., Urnov, F. D., Rebar, E. J., Holmes, M. C., Danos, O., Jiang, H., and Tan, S. (2017) Long-term engraftment and fetal globin induction upon BCL11A gene editing in bone-marrow-derived CD34+ hematopoietic stem and progenitor cells. *Mol. Ther. Methods Clin. Dev. 4*, 137–148.

(38) Urnov, F. D. (2021) The Cas9 hammer - and sickle: a challenge for genome editors. *CRISPR J. 4*, 6–13.

(39) Newman, A., Starrs, L., and Burgio, G. (2020) Cas9 cuts and consequences; detecting, predicting, and mitigating CRISPR/Cas9 on- and off-target damage. *BioEssays 42*, 1–13.

(40) Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell 154*, 1380–1389.

(41) Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2019) Methodologies for improving HDR efficiency. *Front. Genet. 9*, 1–9.

(42) Savic, N., Ringnalda, F. C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., Neri, D., Robinson, M. D., Ciaudo, C., Hall, J., Jinek, M., and Schwank, G. (2018) Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife 7*, 1–18.

(43) Aird, E. J., Lovendahl, K. N., Martin, A. St., Harris, R. S., and Gordon, W. R. (2018) Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol. 54*, 1–6.

(44) Roy, K. R., Smith, J. D., Vonesch, S. C., Lin, G., Tu, C. S., Lederer, A. R., Chu, A., Suresh, S., Nguyen, M., Horecka, J., Tripathi, A., Burnett, W. T., Morgan, M. A., Schulz, J., Orsley, K. M., Wei, W., Aiyar, R. S., Davis, R. W., Bankaitis, V. A., Haber, J. E., Salit, M. L., St.Onge, R. P., and Steinmetz, L. M. (2018) Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol. 36*, 512–520.

(45) Gowetski, D. B., Kodis, E. J., and Kahn, J. D. (2013) Rationally designed coiled-coil DNA looping peptides control DNA topology. *Nucleic Acids Res. 41*, 8253–8265.

(46) Hollenbeck, J. J., Gurnon, D. G., Fazio, G. C., Carlson, J. J., and Oakley, M. G. (2001) A GCN4 variant with a C-terminal basic region binds to DNA with wild-type affinity. *Biochem. 40*, 13833–13839.

(47) Hinnen, A., Hicks, J. B., and Fink, G. R. (1978) Transformation of yeast. *Proc. Natl. Acad. Sci. U.S.A. 75*, 1929–1933.

(48) Tycko, J., Myer, V. E., and Hsu, P. D. (2016) Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell 63*, 355–370.

(49) Laughery, M. F., Hunter, T., Brown, A., Hoopes, J., Ostbye, T., Shumaker, T., and Wyrick, J. J. (2015) New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. *Yeast 32*, 711–720.

(50) Webster, T. D., and Dickson, R. C. (1983) Direct selection of Saccharomyces cerevisiae resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin-resistance gene of Tn903. *Gene 26*, 243–252.

(51) Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast 13*, 837–848.

(52) Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science 268*, 1766–1769.

(53) Wishart, J. A., Hayes, A., Wardleworth, L., Zhang, N., and Oliver, S. G. (2005) Doxycycline, the drug used to control the tet-regulatable promoter system, has no effect on global gene expression in Saccharomyces cerevisiae. *Yeast 22*, 565–569.

(54) Gietz, R. D., and Schiestl, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc. 2*, 31–34.

(55) Gietz, R. D., and Woods, R. A. (2002) Transformation of yeast by lithium acetate/single-stranded Carrier DNA/polyethylene glycol method. *Methods Enzymol. 350*, 87–96.

(56) Hu, G., Luo, S., Rao, H., Cheng, H., and Gan, X. (2018) A simple PCR-based strategy for the introduction of point mutations in the yeast Saccharomyces cerevisiae via CRISPR/Cas9. *Biochem. Mol. Biol. J. 4*, 1–3.

(57) Sangsoda, S., Cherest, H., and Surdin-Kerjan, Y. (1985) The expression of the MET25 gene of Saccharomyces cerevisiae is regulated transcriptionally. *Mol. Gen. Genet. 200*, 407–14.

(58) Cavero, S., Vozza, A., Arco, A. D., Palmieri, L., Villa, A., Blanco, E., Runswick, M. J., Walker, J. E., Cerdán, S., Palmieri, F., and Satrústegui, J. (2003) Identification and metabolic role of the mitochondrial aspartate-glutamate transporter in Saccharomyces cerevisiae: yeast mitochondrial aspartate-glutamate carrier. *Mol. Microbiol. 50*, 1257–1269.

(59) Porcelli, V., Vozza, A., Calcagnile, V., Gorgoglione, R., Arrigoni, R., Fontanesi, F., Marobbio, C. M. T., Castegna, A., Palmieri, F., and Palmieri, L. (2018) Molecular identification and functional characterization of a novel glutamate transporter in yeast and plant mitochondria. *BBA 1859*, 1249–1258.

(60) Bharathi, V., Girdhar, A., Prasad, A., Verma, M., Taneja, V., and Patel, B. K. (2016) Use of ade1 and ade2 mutations for development of a versatile red/white colour assay of amyloid-induced oxidative stress in saccharomyces cerevisiae. *Yeast 33*, 607–620.

(61) Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., Lee, M. E., DeLoache, W., Dueber, J. E., Arkin, A. P., and Cate, J. H. (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife 3*, 1–15.

(62) Jimenez, A., and Davies, J. (1980) Expression of a transposable antibiotic resistance element in Saccharomyces. *Nature 287*, 869–871.

(63) Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., Xu, J., Chun, N., Yuan, W., Cheng, T., and Zhang, X.-B. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol. 18*, 1–18.

(64) Schubert, M. S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M. S., and Rettig, G. R. (2021) Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci. Rep. 11*, 1–15.

(65) Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A. R., Falkow, S., and Brown, A. J. P. (1997) Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. *Microbiology 143*, 303–311.

(66) Song, F., and Stieger, K. (2017) Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Mol. Ther. Nucleic Acids 7*, 53–60.

(67) Hutchins, G. G., and Key, S. C. S. (2008) Gene knockout/ gene therapy in yeast using homologous recombination. *ABLE 30*, 49–62.

(68) Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E., and Doudna, J. A. (2022) CRISPR–Cas9 bends and twists DNA to read its sequence. *Nat. Struct. Mol. Biol. 29*, 395–402.

(69) Dorfman, B. Z. (1969) The isolation of adenylosuccinate synthetase mutants in yeast by selection for constitutive behavior in pigmented strains*. *Genetics 61*, 377 389.

(70) Watanabe, M., Watanabe, D., Nogami, S., Morishita, S., and Ohya, Y. (2009) Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. *Curr. Genet. 55*, 365–380.

(71) Shi, Z., Xin, H., Tian, D., Lian, J., Wang, J., Liu, G., Ran, R., Shi, S., Zhang, Z., Shi, Y., Deng, Y., Hou, C., and Chen, Y. (2019) Modeling human point mutation diseases in Xenopus tropicalis with a modified CRISPR/Cas9 system. *FASEB J. 33*, 6962–6968.

(72) Paix, A., Folkmann, A., Rasoloson, D., and Seydoux, G. (2015) High efficiency, homology-directed genome editing in Caenorhabditis elegans using CRISPR-Cas9 ribonucleoprotein complexes. *Genetics 201*, 47–54.

(73) Chen, H., Neubauer, M., and Wang, J. P. (2022) Enhancing HR frequency for precise genome editing in plants. *Front. Plant Sci. 13*, 883421.

(74) Kuipers, B. J. H., and Gruppen, H. (2007) Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography−mass spectrometry analysis. *J. Agric. Food Chem. 55*, 5445–5451.

(75) Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Anal. Biochem. 200*, 74–80.