Interactions among Dam, SeqA and mismatch repair proteins in *Escherichia coli*

by

Sophia Tsai

B.Sc. (Biological Sciences), Simon Fraser University, 2014

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology & Biochemistry
Faculty of Science

© Sophia Tsai 2019
SIMON FRASER UNIVERSITY
Spring 2019

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name:	**Sophia Meng Pin Tsai**
Degree:	**Master of Science in Molecular Biology and Biochemistry**
Title:	**Interactions among Dam, SeqA and mismatch repair proteins in *Escherichia coli***
Examining Committee:	**Chair:** Jonathan Choy Associate Professor
	Claire Cupples Senior Supervisor Professor
	Edgar Young Supervisor Associate Professor
	Kathleen Fitzpatrick Supervisor Senior Lecturer Biological Sciences
	Julian Guttman Internal Examiner Professor Biological Sciences
Date Defended/Approved:	March 29, 2019

Abstract

The accuracy of DNA replication is very important, and organisms have several proofreading and repair systems to prevent mutations from occurring. Lesions can be introduced by errors during replication, chemical mutagens, UV or radiation. In *Escherichia coli*, mismatches are detected by MutS and MutL which together activate MutH to initiate repair. Repair is heavily dependent on GATC hemi-methylation signals on the DNA which is added by DNA adenosine methylase (Dam). SeqA acts as a regulator of the origin, sequestering it and preventing miscoordination of reinitiation. As such, we hypothesize Dam and SeqA are coordinated by MutL, and (2) Persistent mismatches caused by an error prone polymerase will increase mismatch repair activity. Results show that Dam binds to both SeqA and MutL, and no significant increase in mismatch repair activity was detected when the error prone polymerase was induced. These data show the importance of temporal coordination of methylation and/or constitutive binding of Dam and MutL in preparation for mismatch repair. We also conclude that our data is consistent with previous literature that shows mismatch repair primarily works on transitions and is inefficient at repairing transversions.

Keywords: DNA repair, mismatch repair, bacterial-2-hybrid, DNA adenosine methylase, DNA methylation

Dedication

To my family, for encouraging me to explore my greatest passions.

To Grae, for being there every step of the way with me.

Acknowledgements

Thank you to Dr. Claire Cupples who gave me this amazing opportunity to pursue my graduate studies with her. Thank you for accepting me into your lab even when I had almost no background in molecular biology lab techniques! The experience and knowledge I've gained here has been invaluable.

Thank you to both my committee members Dr. Edgar Young and Dr. Kathleen Fitzpatrick. Both your knowledge and expertise in molecular techniques and scientific writing has helped me so much! Thank you, Kathleen, for encouraging me to get my Master's when I first expressed interest in teaching.

My mother Sherry, my father Peter and my brother Phil: Thank you all for your constant support, advice and providing me with a lovely home to stay in so I did not have to stress about finances while I worked on my research.

My partner Grae, you are amazing! Thank you for your encouragement, positivity, believing in me and letting me cuddle Toby all the time! I would not be where I am today without you!

Dr. Scott Pownall, where would I be without your guidance in teaching me so many molecular lab techniques? Thank you for your patience and allowing me to do some of my work in your lab. You are a great mentor and I am so glad we met back in my first year of graduate school.

Steven Barrett, graduate school would have been so lonely without you! Who else would I ask to go and get watered down free coffee with ;)? So grateful we met in microbiology class and forever grateful for the opportunity to present my data at your departmental conference.

My CrossFit Lions fam, thank you for letting me come to the box everyday to throw some weights around and do what I love. I am so proud to be part of this community. Thank you Mina Calupitan for your excellence and attention to detail to coaching and challenging me everyday. I have become a better athlete because of you.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. DNA and maintenance of a stable genome

All living organisms contain genetic material in the form of deoxyribonucleic acid (DNA). Because DNA is the genetic code that holds the information needed for organisms to survive and reproduce, copying fidelity is extremely important. For a mutation to occur, an error or lesion has to first be created during the DNA replication. When this error is not detected and repaired, it becomes permanently incorporated into the genome and is now defined as a mutation.

Mutations in DNA can be caused by various mutagens such as chemical carcinogens and UV rays or can be spontaneous, e.g caused by errors in DNA replication. Mutations may be neutral, deleterious or advantageous, depending on the environment the organism is in. While generation of an advantageous mutations may be quite rare, they can be strongly selected for and so the prevalence of this type of mutation can increase exponentially. Deleterious mutations can lead to improper gene expression and/or malformed and malfunctioning RNAs and proteins, leading to cell death and, eventually, organismal death. Overall, the frequency and distribution of each type of mutation is complex, due to many contributing factors. Therefore, organisms have multiple systems in place to prevent such mutations, such as proofreading and mismatch repair. In the event that the damage is too great, the cells in multicellular eukaryotes will induce apoptosis to prevent the passing on of incorrect and/or incomplete genetic material.

1.2. Mutations can be advantageous

As mentioned above, mutations can confer an advantage, although this is rare. Bacteria growing in a nutrient rich and antibiotic free environment would be under low stress. If, however, some ampicillin was added to the media, the bacteria that have an ampicillin resistance mutation would be at a much greater fitness advantage. These bacteria would survive and would be able to pass on their genes to the next generation. Thus, whether a mutation is considered advantageous or not depends entirely on the

environment the cell is in. A mutation that might be considered deleterious in one environment could be advantageous in another.

1.3. *Escherichia coli* as a model organism

E. coli K-12 has been a popular model organism since the 1940's. It has several advantages. It is inexpensive to culture and maintain and has a rapid growth rate. Its genome has been fully sequenced and researchers have access to many genetic tools such P1 phage or plasmids. *E. coli* is haploid, therefore knockouts or genetic manipulations are made much easier as only one copy of the gene needs to be mutated to observe any phenotypic effects. Further, *E. coli* is a single celled organism and there are no ethical concerns as there would be with studying more complex organisms such as mice or chimpanzees. Being non-pathogenic it can be cultured under low levels of biocontainment. Most importantly, many genes found in *E. coli* (such as the mismatch repair genes we are studying) have human homologues so discoveries made using this bacterium can have important implications for human health.

1.4. Replication of DNA & Proofreading

DNA is replicated in a semiconservative fashion with DNA Polymerase III as the main enzyme involved in this process. DNA Polymerase III has an extremely fast replication rate, synthesizing approximately 1000 nucleotides/second in bacteria[1]. Maintenance of DNA fidelity during replication is extremely important and so there are three mechanisms used to achieve this: 1) base selection, 2) proofreading, and 3) mismatch repair [2]. Base selection is carried out by DNA Polymerase III and is the process by which it discerns correct versus incorrect bases during the insertion step of DNA synthesis. It is the first step in ensuring DNA replication is accurate. DNA Polymerase III has two subunits tightly bound together that work in conjunction to proofread; these are the α subunit (polymerase) encoded by *dnaE* and the ε subunit (exonuclease) encoded by *dnaQ*. Proofreading occurs simultaneously with replication. If an error is detected, the polymerase holoenzyme complex can reverse, excise the incorrect nucleotide and insert the correct one. Because this proofreading activity is not perfect there are multiple post-replication DNA repair pathways available. It has been shown that both base selection and proofreading discriminate 40-fold more strongly

against base pair mismatches that cause transversions than transitions; thus, transversions are more numerous in cells lacking proofreading. A transverion occurs when purines (two ringed based) are interchanged for pyrimidines (one ringed bases) or vice versa, while a transition involves an interchange where a purine is switched for a purine or a pyrimidine is switched for a pyrimidine. The more numerous transition-causing mismatches remaining in the replicated DNA are repaired by mismatch repair downstream [3]. Since mismatch repair more easily prevents transitions, the combination of base selection, proofreading and mismatch repair means that rates of both transversions and transitions are quite low in wild type bacteria. DNA repair intermediates left unguarded (unprotected by proteins or other molecules) can be as mutagenic as the original lesion so coordination in hand offs between repair pathways is paramount [4]. Hand-offs occur anytime DNA moves from one step to another; for instance, moving from replication to mismatch repair, or mismatch repair to nucleotide excision repair. Repair pathways must be coordinated to prevent unprotected DNA from being left vulnerable to other repair systems. A nick in the DNA that is left unattended can trigger many other repair pathways (e.g. strand invasion), so coordination of DNA handoff between enzymes is really important. Thus, it is possible that some protein-protein interactions can cause a conformational shift in an enzyme to cause it to have high affinity for the DNA or for a protein that replaces it at the next repair step (i.e. stimulating the hand off) [5].

1.5. Methyl directed Mismatch Repair System in Bacteria

The mispairing of bases during DNA replication is usually the result of errors made by Pol III. Error frequencies are quite low at only one error per 10^{10} bases [6]. With a genome size of approximately 4.6×10^9 base pairs, the possibility of a mismatch is less than one per replication cycle. The mismatch repair (MMR) system is able to catch mistakes that Pol III's proofreading did not. Along with repairing base-base mismatches, MMR also works to repair insertion/deletion loops [7]. The MMR pathway is evolutionarily conserved, thus mammalian MMR functions in much the same way as in bacteria. The conservation of this complex pathway shows how important it is. For instance, in humans, hereditary non-polyposis colon cancer (HNPCC) arises due to loss of function mutations in various MMR genes.

It is quite simple to recognize base mismatches and insertion/deletions based on changed topography of the DNA, but in *E. coli*, determining which strand is the parental (and thus the correct) strand requires the use of methyl tags. This idea of strand discrimination based on methylation was initially proposed by Wagner and Meselson [8]. The basis of MMR is the ability of the repair machinery to distinguish between the parental and the newly synthesized DNA. Immediately after DNA synthesis, nascent DNA is unmethylated unlike the parental DNA. The methylation of the new strand, at hemi-methylated GATC sites, is completed by DNA adenine methylase (Dam) after a short time lag, giving the MMR proteins time to complete the necessary mismatch repairs. The methylation is incredibly important, as *dam* strains of *E. coli* are hypermutagenic [9]. Overexpression of wildtype *dam* is also mutagenic (for further explanation on this, please see section 1.6.1). The MMR pathway can correctly repair a mismatch despite the nearest hemimethylated site being located as far as two kilobases away.

There are several components needed for MMR to function; these being all MMR proteins (MutS, MutL, MutH), DNA helicase II, single stranded DNA binding protein (SSB), DNA Polymerase III, exonuclease I, exonuclease VII, RecJ and DNA Ligase [10]. Immediately following DNA replication, a MutS homodimer scans the newly synthesized DNA strand for base-base mispairings, and insertion/deletion loops (IDLs). When a lesion is found, MutS will bind to the lesion. MutS has been shown to have different binding affinities to different types of mismatches or lesions [11]. It has the highest affinity for unpaired T's and GT mismatches and has the lowest affinity for Watson-Crick pairs, as would be expected. However, high affinity binding does not necessarily mean the lesion in question will be well repaired. For instance, of all types of mismatches, GT and are the best repaired despite the fact that MutH has a low affinity for AC mismatches due to a pH dependent wobble conformation [11]. This is probably because MutS bound to GT and AC mismatches is more efficient at recruiting MutL and/or to signalling downstream repair machinery [11]. Because MutS has ATPase activity, it can use a translocation mechanism to catalyze the creation of an α-shaped DNA loop containing the mismatch and the hemimethylated GATC [12]. In general, transition mutations are more efficiently repaired than transversion mutations.

After the MutS homodimer has bound to the mismatch site, it will recruit MutL, a Mg^{2+}-dependent ATPase, which helps to increase the size of the α-loops. The MutS-

MutL complex then activates MutH, an endonuclease responsible for cleaving the DNA at the GATC sequence closest to the mismatch. Normally, the endonuclease activity of MutH is very weak; however, when activated by the MutS-MutL complex, its activity increases by ~50 fold [13]. Interestingly, MutH homologues are only found in gram-negative bacteria. This suggests that other mechanisms of strand specification and cleavage are used in other organisms [13]. In fact, the need for MutH's endonuclease activity can be alleviated by the presence of a persistent nick in the DNA [14]. Usually DNA is methylated at GATC sites; however newly synthesized strands lack this methylation because the Dam protein lags behind DNA Polymerase III [7]. Thus, during this short time interval, MutH is able to differentiate between the two strands and cleave at the nearby unmethylated GATC site of the new strand to start the repair process. When the nick made by MutH is 3' to the mismatch, exonuclease I, exonuclease VII, or exonuclease X can degrade the DNA; however, when the nick is 5' to the mismatch, either exonuclease VII or RecJ are required [15].

DNA helicase II, assisted by MutL, unwinds the DNA from the nick created by MutH to the other side of the mismatch, allowing exonucleases to degrade the DNA. Next, single stranded binding proteins (SSB) help stabilize the parent strand. DNA Polymerase III resynthesizes the excised strand and the repair process ends with ligation by DNA ligase and methylation of the repaired strand by Dam. Thus, methylation is an incredibly important signal in the process of mismatch repair and thus under or over methylation can cause mutations. It is the maintenance of the hemi-methylation state of DNA that allows for mismatch repair; and thus the time lag in methylation of the nascent strand of DNA is crucial.

However, details of MMR protein-protein interactions beyond the initiation steps are still lacking.

1.5.1. Mismatch Repair in Eukaryotes

The methyl-directed mismatch repair system is highly conserved and is thus a very important pathway for maintaining DNA fidelity. Eukaryotes have multiple homologues for both MutS and MutL proteins. Many such homologues have been discovered in yeast, humans, and the model organisms *Xenopus laevis* and *Drosophila melanogaster* [16]. MutS homologues have also been discovered in *Arabidopsis* [17]. A

defect in mismatch repair in humans results in HNPCC. The prevalence of homologues in other species supports the idea that the mismatch repair pathway is evolutionarily conserved due to its importance. Homologues of the mismatch repair proteins are listed in Table 1.

In eukaryotes, the average mutation rate is 10^{-10} mutations per base pair per generation [18]. In contrast to MMR in *E. coli*, MMR in eukaryotes uses strand discontinuity rather methylation signals on GATC sequences to identify the non-parent strand [19]. The daughter strand in eukaryotes is recognized by its gaps due to the presence of Okazaki fragments [19]. MMR in eukaryotes is heavily dependent on MutSα, which is a heterodimer comprised of Msh2 and Msh6. This heterodimer's role is recognizing and binding the mismatch in the DNA strand. The hydrolysis of ATP in addition to mismatch binding will cause a conformational shift in MutSα, which then allows it to recruit MutLα (another heterodimer composed of Mlh1 and Pms2) [20]. As there is no eukaryotic homologue for MutH, MutLα takes on the role of nicking the discontinuous strand and EXO1 will remove a short segment of DNA. As mentioned in the previous section, MutH is not required for MMR to take place; and even just a persistent nick in the DNA will allow for repair. Following the nicking by MutLα, DNA Polymerase III will fill in the missing bases and DNA Ligase will seal the nick.

E. coli	Yeast (*Sacchromyces cerevisiae)*	Human
MutS	MSH2/MSH3	hMSH2/hMSH3/hMSH6/Duc-1
MutL	PMS1/PMS2/MLH1	hPMS1/hPMS2/hMLH1

Table 1: Mismatch Repair genes in E. coli and their homologues in Eukaryotes

A loss of function mutation in any MMR homologues in humans dramatically increases the risk of tumour formation. The vast majority of hereditary nonpolyposis colorectal cancer cases are a result of mutations in *MLH1* and/or *MSH2* [21]. It has also been found that mice with knockouts in *msh2, msh6, mlh1* and *mms2* that form the MutSα (Msh2-Msh6) and MutLα (Mlh1-Pms2) complexes which are critical to repair, show a mutator phenotype with a strong predisposition to developing cancer [15]. Mouse lines with mutations in genes that contribute to less important complexes in the MMR pathway such as MutSβ show a milder cancer phenotype [15]. Additionally, mouse lines

with homozygous mutations in either *msh2; p53, msh6; p53,* show greatly accelerated tumorigenesis and T-cell lymphoma [15].

In *E. coli*, strand differentiation is determined by the presence or absence of methylation where the time interval during which MMR takes place is extremely important as the nascent strand quickly gets methylated after replication. It is known that in eukaryotes mutation frequency can change as the cells cycle through different stages; this may be due to a similar timing issue that bacteria experience [18]. In eukaryotes there is a fluctuation in the availability of MMR proteins depending on the stage of the cell cycle [18], [22]. MMR proteins are first expressed in G1, with expression increasing in both S and G2 stages. The change in timing of MMR availability is further evidence that DNA repair is tightly linked to DNA replication and the cell cycle.

1.6. Role of *dam* and *seqA*

1.6.1. Dam

Dam methyltransferase plays a role in DNA replication in *E. coli* by methylating adenines at GATC sites. Dam is a monomer and transfers a methyl group from a donor, S-adenosylmethione, onto the adenine residue, creating a 6-methyladenine product [23]. Methylation at GATC sites is essential for correct mismatch repair and for the regulation of DNA replication. Newly synthesized DNA is unmethylated for several minutes [24]. This transient hemimethylated state helps enzymes distinguish between parental and daughter strands of DNA and acts as a block to prevent refiring of the origin of replication, which is a particular sequence in the genome where DNA replication is initiated. The methylation at GATC sites is thus critical to maintaining the fidelity of DNA and controlling the cell cycle.

Mutant cells containing a *dam* knock out show pleiotropic phenotypic alterations. Firstly, as expected, *dam*$^-$ strains are hypersensitive to chemical mutagens [9]. They also exhibit a higher number of spontaneous double strand breaks. Additionally, since methylation of promoter sequences affects transcription, > 200 genes are over-expressed in *dam* deficient cells [9]. Boye et. al demonstrated that *dam*$^-$ cells lack the ability to fire the origin in a controlled manner [25]. Initiation of DNA replication was uncoordinated and random in both *dam* knockouts and cells that overexpress *dam* [25].

Yet, overexpression of *dam* is also detrimental to the cell, making it difficult for mismatch repair to act [26]. Thus, about 130 molecules of Dam was determined to be optimal [24].

1.6.2. seqA

SeqA binds to hemimethylated GATC sites but its functionality depends on whether such sites are within or outside the origin. SeqA plays an important role as a regulator of DNA replication to prevent misfiring of the origin at incorrect times. SeqA binds to the high numbers of GATC sites in the replication origin, *oriC,* of newly replicated DNA that is still hemimethylated [27]. The hemimethylation is due to methylation on the parental strand by Dam methyltransferase at the GATC sequences as described in section 1.6.1 above. *In vitro* studies show that SeqA binding to DNA requires at least two GATC hemimethylated sites on a single DNA fragment [28]. Studies by Waldminghaus et. al show that SeqA has two distinct domains connected by a flexible linker [27]. The binding of the SeqA dimer on the newly replicated origin physically prevents the initiator protein, DnaA, from binding again and reinitiating replication. SeqA has been shown to bind as a dimer to pairs of hemi-methylated GATC sites on the DNA [28]. Besides binding at *oriC*, SeqA has also been found to bind hemi-methylated GATCs throughout the chromosome [29]. SeqA binding correlates with the frequency of GATC sites found throughout the genome. Less SeqA is bound in highly transcribed regions, and more SeqA is bound in the less transcribed regions [29]. This negative correlation between SeqA and transcription suggests that SeqA binding prevents transcription by somehow interfering with RNA Polymerase binding. At the same time, the interference could also hypothetically be caused by RNA Polymerase preventing SeqA from binding to GATC sites.

Typically, in wildtype cells, firing of the origins occurs asynchronously but in a controlled manner; those that are fully methylated are initiated while those that are hemimethylated are not [30]. Bacteria only have one replication origin on their chromosome, but they commonly carry plasmids which also carry an origin of replication that works by binding with DnaA. In cells with a mutant *seqA* gene, refiring of the origin is constant and dysregulated. This leads to high amounts of uncontrolled DNA replication. Further, increased DNA replication, leads to an abnormally high copy number for genes. In this case, *dnaA* (which promotes the initiation of DNA replication) is produced and transcribed at a higher rate due to lack of sequestration. Because of the

lack of SeqA blocking the origin, the now uncontrolled overproduction of DnaA results in even more replication initiation. In a wildtype cell, sequestration of the origin will keep the origin inactive until SeqA dissociates spontaneously, at which point the origin and the rest of the newly replicated DNA can be methylated by Dam methyltransferase.

1.6.3. SeqA & Mismatch Repair

In addition to the above mentioned role, SeqA also binds with high affinity at hemi-methylated GATC sites outside of the origi although the purpose of this is still unclear. MutH also binds to such hemi-methylated GATC sites and thus, their roles in mismatch repair and replication could be connected in some way although few studies have been done on their relationship with each other. Nevertheless, there are some suggestions about how SeqA and MutH might interact. SeqA may prolong the hemi-methylated state of DNA by blocking methylation by Dam methyltransferase through competition at the same sites [31]. This was suggested because in cells that are *seqA* deficient or *dam* overexpressed, the hemi-methylated state was drastically shorter. In Δ*seqA* strains, an increased mutation rate is observed, likely due to less accurate mismatch repair resulting from the shorter than usual time that the DNA is in a hemi-methylated state [32].

MutH is only able to cut at hemi-methylated GATC sites when activated by the presence of MutS and MutL. But, when SeqA is overexpressed, mismatch repair becomes faulty because now the overexpressed SeqA is able to outcompete MutH for GATC sites [33]. In fact, *seqA* has been identified as a mutator gene with strong mutagenic effects by Yang et. al [32]. It's likely that normal levels of SeqA helps prevent MutH from cleaving DNA until MutS and MutL activate MutH's catalytic endonuclease activity and help facilitate it binding to the DNA substrate. SeqA produced in excess will prevent Dam from being able to methylate the GATC sites, resulting a hypo-methylated parental strand which prevents MMR from functioning properly [32]. The balance between Dam and SeqA is essential to correct MMR activity.

1.7. Protein – Protein Interactions

Many proteins in pathways interact specifically with other proteins in the same pathway. An interaction is defined as some sort of purposeful binding where either a reaction is catalyzed, a conformation shift results etc. We have already discussed some types of interactions in the MMR pathway. There are several methods that can be used to detect such interactions: co-immunoprecipitation, pull-down assays, crosslinking and affinity chromatography. The major drawback to these methods is amount of time required to purify these proteins using antibodies. Additionally, these methods only allow for *in vitro* studies of protein interactions where co-factors may be lacking, and often do not include a measure of protein function. In such cases, proteins are often studied at much higher concentration than found normally in cells; such levels can promote protein – protein interactions between proteins that normally wouldn't interact.

A quick method to determine protein interactions is the two hybrid system originally developed by Fields & Song in yeast [34]. This system takes advantage of transcriptional activators which have a DNA binding domain and a transcriptional activation domain. Normally, both of these domains are on the same protein. However, when the protein is split into two separate domains, it will still be functional when they are brought close enough together to interact. Proteins that are suspected to interact are fused to either domain. If the proteins of interest interact, the transcription factor components are brought together in close proximity. The transcriptional activator is reconstituted, driving the expression of a reporter gene such as *lacZ*. The expression can then be qualitatively and quantitatively measured using a variety of assays.

A similar assay called the Bacterial-2-Hybrid (B2H) assay is available in bacteria. The B2H assay uses complementary protein fragments (T18 and T25) of adenylate cyclase from *Bordetella pertussis* that when brought together and reconstituted, will produce cAMP. Genes for these two T18 and T25 fragments are located on separate vectors with multi-cloning sites at both the N and C terminals to allow for cloning sequences of interest in at either end to produce a fusion protein. The assay effectively measures the interaction between two proteins of interest using the *lac* operon as a reporter gene (Figure 3). False positives arising from "sticky" proteins can be controlled for by simply pairing the "sticky" protein fusion with a leucine zipper fusion and seeing if

an interaction still arises. There are several reasons why the B2H assay based on adenylate cyclase reconstitution is the ideal assay for studying protein-protein interactions for our research in *E. coli*. The B2H assay was used to study interactions between MutS, MutL, Vsr and MutH and has proved its effectiveness [35]. Lesions in the DNA were induced using 2-aminopurine which mis-pairs with C during DNA replication. This causes transitions and frameshift mutations. Mismatch repair proteins will interact together at these transition and frameshift sites which will be detected by the B2H assay through the generation of cAMP as the adenylate cyclase fragments are reconstituted. This then allows causes the lac operon to be turned on and a signal to be produced. The dose response curve observed with 2AP mirrors that dose response curve of transitions and mutations (Figure 1, 2). In Figure 2, such GC-AT mutation were detected using a lac reversion test in a strain that had a single point mutation in *lacZ*; reversion of this mutation back to wildtype allowed for it to grow on minimal lactose media which allows for a visual method of mutation screening. In Figure 1, mutations were induced using 2AP and protein-protein interactions for vsr/MutL (gray bars) and MutH/MutL (black bars) were detected with the same dose response curve shape as Figure 1. This means that the B2H assay can effectively detect protein interactions from mismatches produced. Moreover, the B2H assay was able to detect all physical interactions that was detected by more traditional biochemistry methods such as pull downs, crosslinking and FRET (refer to Figure 3). The B2H assay proves to be just as effective but less time consuming than other methods used to assay the same types of interactions.

1.8. The Bacterial Two Hybrid (B2H) System

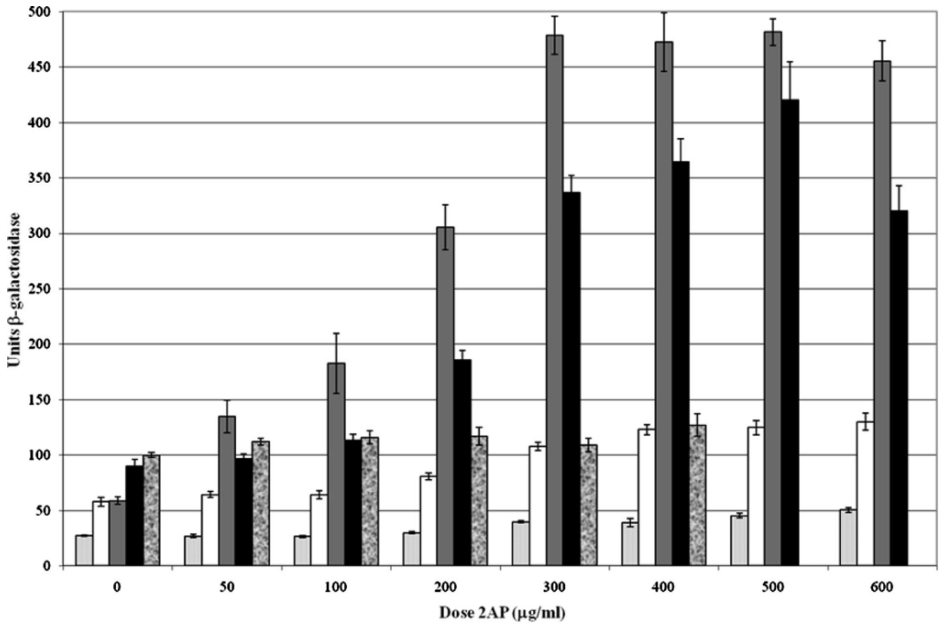

Figure 1: Effect of 2AP treatment on protein-protein interactions in the bacterial two-hybrid assay. Cells were cotransformed with pT18 and pT25 vectors (light gray bars), pT18-mutS and pT25-mutL (white bars), pT18-vsr and pT25-mutL (gray bars), pT18-mutH and pT25-mutL (black bars), or pT18-mutL and pT25-mutL (mottled bars). (NB: The dose-response curve for the pT18-mutS pT25-mutS transformants is similar to that of the pT18-mutL pT25-mutL transformants; it has been omitted for graphical clarity. [36]

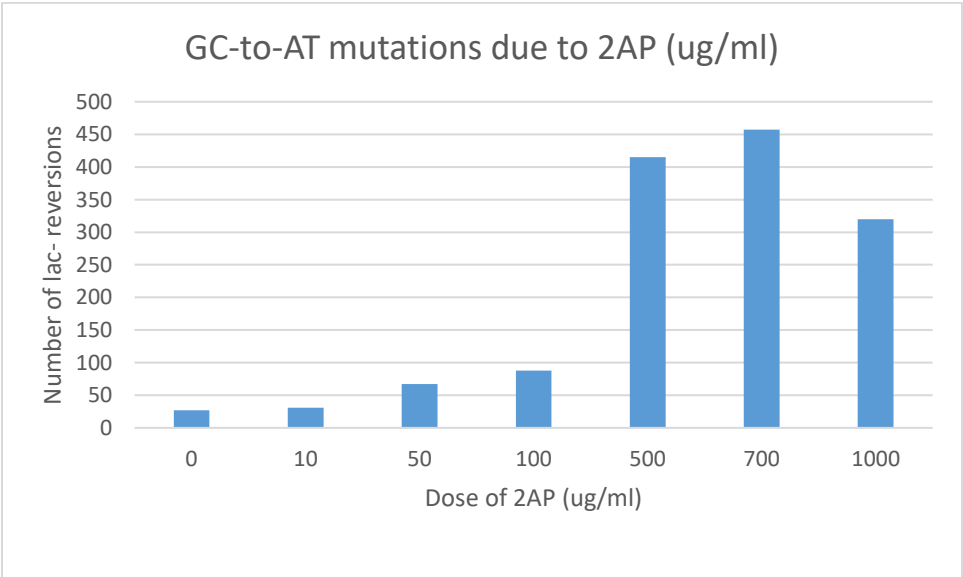

Figure 2 Number of lac- reversions in response to an increasing dose of 2AP. [37]

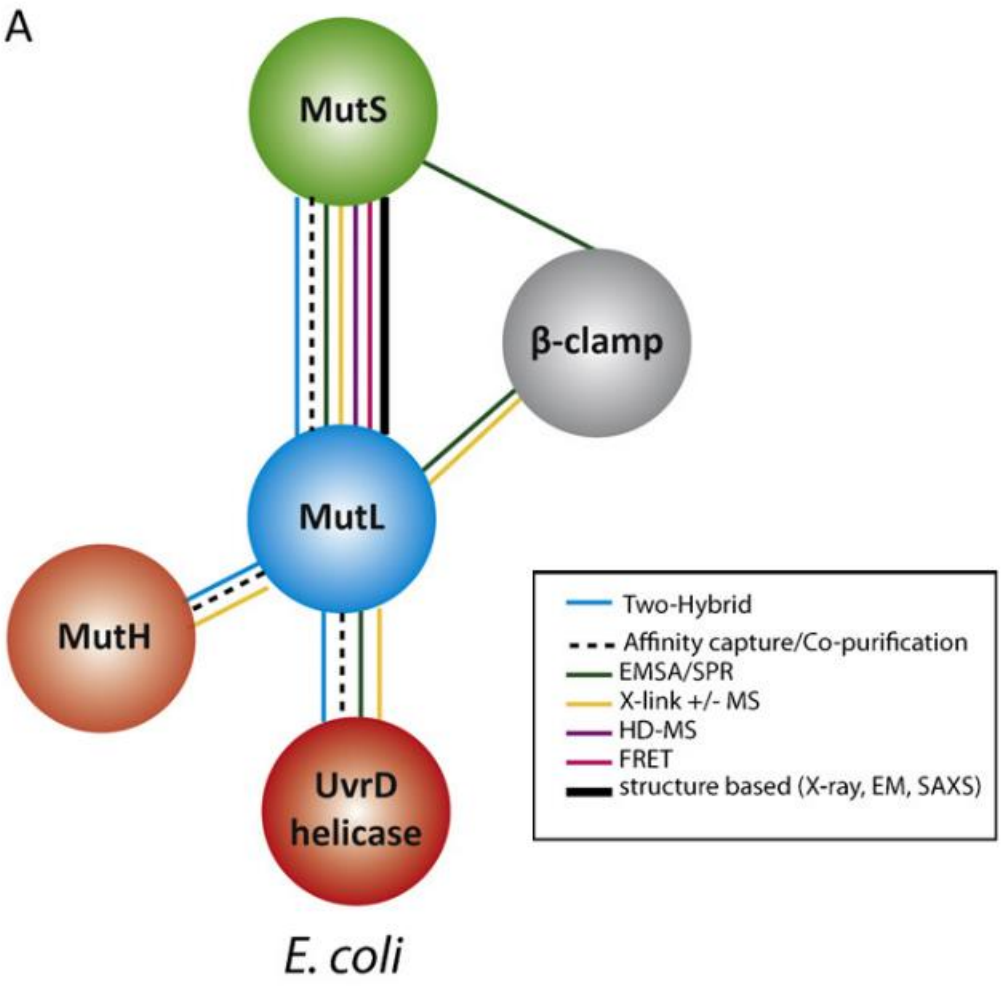

Figure 3: The physical interactions detected between MMR proteins in E. coli using different methods [38]. EMSA/SPR = Electrophoretic Mobility Shift Assay & Surface Plasmon Resonance. X-link +/- MS = Crosslinking +/- mass spectrometry. HD-MS = high definition mass spectrometry. FRET = Fluorescence Resonance Energy Transfer.

The interactions observed with the B2H assay closely mirror interactions observed using other methods of detection such as pull downs. For instance, the B2H assay detected interactions between MutS and MutL; this interaction was also detected using a gel shift analysis [15], [39]. There are several methods to detect such protein interactions, but the B2H assay has a few advantages over others. Namely, it is a very quick assay with only an overnight incubation needed and secondly, protein-protein interactions can be detected *in vivo* rather than *in vitro*. Further, because no protein

needs to be purified, no biochemical work is needed, which drastically reduces the workload. The techniques and methods used here are simple and quick, but are just as effective as some of the biochemical techniques used.

When comparing the exact same protein pairs in both Y2H and B2H, the Y2H failed to detect several strong interactions that was detected by the B2H assay which is likely due to limitations in the Y2H assay itself [40]. Further, there are fewer false positives and negatives with B2H, bacteria are faster growing and interactions anywhere in the cell can be detected with B2H. The assay uses the *lac* operon as a reporter gene which heavily relies on cAMP as a transcription enhancer. Adenylate cyclase when reconstituted via interacting proteins catalyzes the formation of cyclic AMP (cAMP) from ATP; this cAMP then binds the catabolite activating protein (CAP), forming the CAP/cAMP complex which can then bind to the *lacZ* promoter, stimulating binding of the RNAP. The catalytic domain of CyaA can be divided into two fragments: 1-224 (T25) and 225-399 (T18) [41]. The reconstitution of these fragments leads to successful production of cAMP which enhances the expression of the *lac* operon.

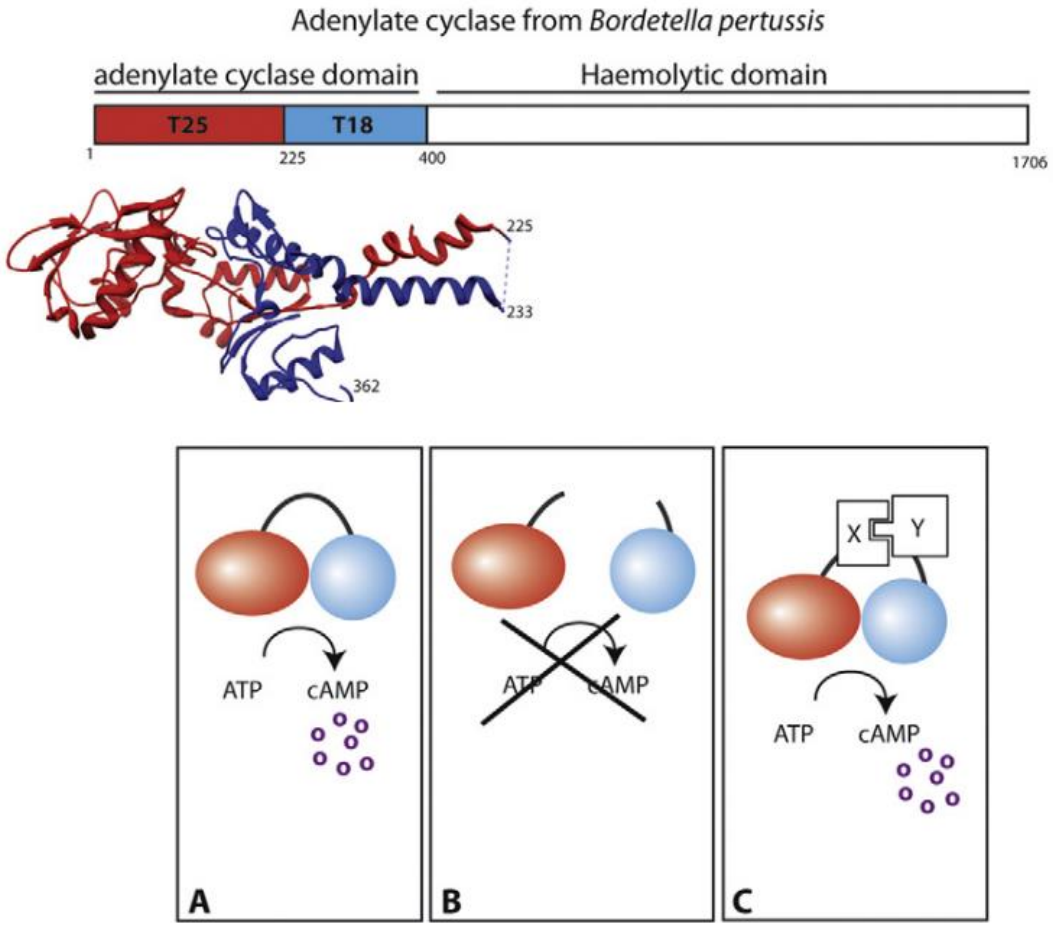

Figure 4. Schematic diagrams of the B2H technique and the two complementary fragments. A: When T25 (red) and T18 (blue) fragments are expressed and translated as one, cAMP is produced. B: When T25 and T18 are expressed separately, no cAMP is produced. C: When the two complementary fragments are brought close together due to being fused to hybrid proteins (X & Y) that interact, cAMP is produced. [42]

Simply expressing T18 and T25 independently in an *E. coli cyaA* strain results in no signal as the T18 and T25 fragment are unable to bind to each other (Figure 4). The two fragments must be each fused to proteins that interact to produce a signal. When the T18 and T25 are brought near enough to each other to interact, their catalytic activity is reconstituted, and cAMP is produced, turning on the *lac* operon. *lacZ* is transcribed and translated into the β-galactosidase enzyme when induced by the presence of IPTG (Isopropyl β-D-1-thiogalactopyranoside). When XGAL is in its native, uncleaved, form, it is colourless; when cleaved by β-galactosidase, insoluble blue compounds are formed, producing deep blue colonies on X-GAL agar plates. This allows for a quick and qualitative visualization of which colonies have successful protein-protein interaction. This assay can also be measured quantitatively using an O-Nitrophenyl-β-D-

galactopyranoside (ONPG assay) that uses a spectrophotometer to measure the yellow-coloured compound formed from cleavage of ONPG by β-galactosidase.

Because the B2H assay uses the *lac* operon as a reporter gene, detectable signal occurs as soon as the production of cAMP reaches a threshold; it does not increase even if the amount of damage per cell increases. Instead, an increase in signal within a culture indicates the approximate number of cells with DNA damage. With increasing 2AP concentrations for example, we see an increase in the amount of β-galactosidase activity due to the MMR proteins coming together to interact and thereby reconstituting CyaA. Therefore, the more cells that have enough mismatches to activate the B2H assay, the bluer the colony/patch would get and the higher the units of β-galactosidase activity on the liquid ONPG assay.

As the B2H assay uses the reconstitution of cAMP to measure protein interaction, an *E. coli* strain that is deficient in the *cyaA* gene is required; otherwise we would be measuring the cell's production of its own cAMP. The slight drawback here is that cells that are *cyaA*$^{-}$ also tend to grow very slowly in comparison to *cyaA*$^{+}$ cells. An advantage of the use of the *lac* operon as a reporter is that it allows one to positively select only those cells with protein interactions by plating on minimal media with lactose. Finally, in the B2H assay the direct consequence of a positive protein interaction is cAMP (a diffusible signal), so it can also be used to study localized membrane proteins that are suspected to interact. Overall, the B2H assay is an excellent choice for our methods as it is quick and accurate with a clear readout.

1.8.1. Other Uses of the B2H System

While most uses of B2H involves bacterial protein interactions understandably, this system has also been used for eukaryotic proteins. For example, the B2H assay was used in mice to detect interactions between variants of Lama1 proteins[43]. It has also been used to study HIV protease protein dimerizations [44]. This leads to the possibility of studying viral to host protein interactions and how to minimize viral infection. In addition, the B2H system is not just used for detection of an interaction, but it can also be used to help characterize which parts of the protein interact. This was done to characterize the interacting domains between Vsr and MutL for example [40]. With this method, one can more accurately predict at which domain the protein is

binding. It is also possible to study interactions where a third bridging protein may be involved by simply creating a mutation in that third protein then retesting the interaction again. This was shown in a previous study done in our lab where MutL was discovered to act as a bridge between MutH and MutS [40].

1.8.2. Known interactions in the MMR pathway

While much is known about each individual protein's role in the MMR pathway, the full extent of the network of protein-protein complexes formed is still unclear. Understanding the complexes formed, particularly in response to specific DNA lesions, can help us clarify how the pathway works in real time.

Several studies have shown how some of the MMR proteins interact. MMR is a complex set of protein-protein and protein-DNA interactions that are constantly changing. For instance, Hall & Matson discovered MutL interacts with MutH via the last 218 C terminal amino acids [39]. This would also imply that MutL's N terminus is responsible for binding to the DNA. Interaction of MutL with MutH also greatly increases MutH's endonuclease activity, suggesting that MutL acts as an activator. MutL is also known to associate with DNA helicase II, greatly stimulating the helicase activity. The nature of these interactions change through the repair process as MutL interacts with MutS near the start of the process, and later interacts with MutH. Coordination among all MMR proteins is essential to prevent nicks and single stranded DNA from being targeted and degraded by other enzymes. Accidental release of single stranded DNA can cause recruitment of other repair processes such as strand invasion, double strand breaks, single strand breaks and very short patch repair [5].

Previous studies in the Cupples lab, have demonstrated some additional novel interactions in the MMR pathway. When an error prone DNA Polymerase III with no proofreading function (*dnaQ49)* is expressed, the number of transversions and frameshift mutations increases, indicative of a high number of mismatches that saturate MMR. When MutL or MutH is overexpressed alongside the error prone polyermase, the mutation rate returns to near normal. However, overexpression of MutS has no effect on the mutation rate. Further, there is evidence of random variation of mutation rate

visualized with the B2H assay – some cells will have a transient mutator state causing colonies to be patchy white/blue (Figure 5).

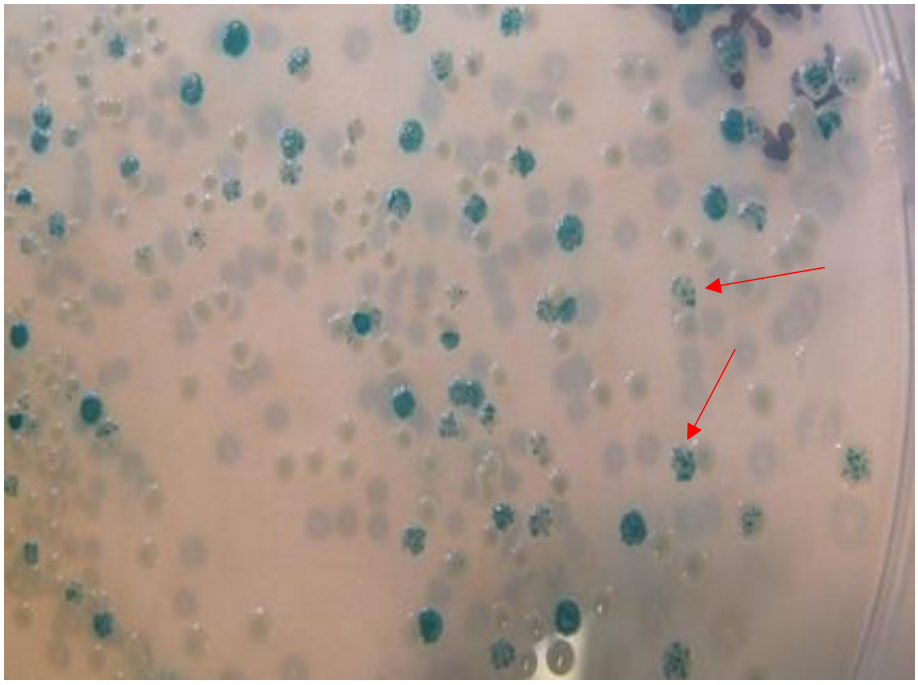

Figure 5: Mutator patchy white/blue colonies showing MutH/MutL/MutS activity is not uniform in cells with an error prone polymerase. (Data unpublished, 2011)

1.9. Mutagenic Plasmids

Mutagenic plasmids overexpress certain genes that will increase the rate of mutagenesis *in vivo*. Mutagenic plasmid 2 (MP2) overexpresses *dnaQ926*. *dnaQ926* is a mutant variant of *dnaQ*, which codes for the Ɛ subunit of DNA Polymerase III. DnaQ is essential for the 5' – 3' exonuclease activity used in proofreading during DNA replication. Without the 5' – 3' exonuclease activity, no proofreading can be done and *dnaQ926* cannot correct its own errors.

Mutagenic plasmid 3 (MP3) is a variant of MP2. It overexpresses both *dnaQ926* and *dam*. As described in section 1.6.1, *dam* methylates GATC sites to create hemi-methylated DNA, necessary to mismatch repair. With an overexpressed *dam*, we may expect high amounts of MutL/MutS interaction and a bottlenecking of MutH due to its inability to cleave at fully methylated sites.

1.9.1. Mutations caused by *dnaQ926* (MP2)

On the MP2 plasmid, *dnaQ926* causes a very strong, dominant mutator phenotype which suggests that it is still binding to the polymerase α subunit. The lesions caused by *dnaQ926* are primarily AT → TA transversions [45]. While AT -> GC transitions are also observed, AT → TA transversions were observed at a rate 50X higher than that of transitions. The mismatch repair system is designed to detect both types of mutations but is much better at repairing transitions. Too high a mutation rate could cause cell death by overwhelming the mismatch repair system, which in turn makes sense with the *mutL* plasmid supplement rescue in cells expressing *dnaQ* where mutation rates returned to normal following the addition of a wildtype *mutL* on a plasmid.

1.10. Hypotheses

We have previously discussed that Dam, SeqA and MutL all bind at hemimethylated GATC sequences. All three proteins compete for the same sequence so what determines which protein binds when? There are a few possibilities to consider: 1) Binding affinity of proteins for hemi-methylated GATC sequence 2) protein concentration within the cell 3) targeting of the protein to the hemi-methylated GATC sequences.

If targeting of proteins is what determines when proteins will act at these sequences, then it is most likely linked to mismatch repair. In order for repair to take place, DNA must be hemi-methylated so the machinery can distinguish the parental strand from the daughter strand. Following repair, DNA needs to be fully methylated for the next round of replication. We know that MutL acts as the coordinator for mismatch repair so we hypothesize that it could also interact with Dam, linking completion of repair to methylation of the DNA. So, we are testing two specific hypotheses.

1) **Correct mismatch repair requires regulated interactions among competing proteins. This means that Dam, SeqA and MutH could all be coordinated by MutL.**

    - To test this, I will use the B2H assay to test all pairwise interactions on a patch plate. I'd expect all interactions to be positive to some extent because they are all competing for the same sites and spatially close to each other. Positive interactions can be detected on an XGAL plate

by the presence of the colour blue indicating the XGAL has been cleaved by β-galactosidase.

2) **Persistent mismatches caused by failure of DNA Polymerase III to proofread will increase mismatch repair activity.**

- An error prone polymerase would cause more lesions in the DNA and therefore, I would expect to see an increase in the number of β-galactosidase due to more cells having mismatches and thus mismatch repair proteins interacting. This would be indicative of a higher amount of mismatch repair protein-protein interaction.

Chapter 2. Materials & Methods

2.1. Bacterial strains and media used

All strains used are listed in Table 1. Bacteria were cultured at 37ºC overnight, unless otherwise noted. A nutrient rich LB broth was used for most liquid cultures: 10g tryptone, 5g yeast extract, 10g NaCl per litre, aliquoted into smaller volumes and autoclaved. LB agar was prepared by adding 15g of agar per liter to the LB recipe. Filter sterilized antibiotics/supplements were added as required in the concentrations listed in Table 3.

Strain Name	Genotype
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1
LJ2809	F-, fruR11::Tn10, xyl-7, ΔcyaA854, ΔargH1
GM3819	F-, *thr*-1, *araC*14, *leu*B6(Am), Δ(*gpt-pro*A)62, *lacY*1, *tsx*-33, *qsr'*-0, *gln*X44(AS), *gal*K2(Oc)?, λ-, *Rac*-0, *his*G4(Oc), *rfb*C1, *mgl*-51, *rps*L31(strR), Δ*dam*-16::KanR, *kdg*K51, *xyl*A5, *mtl*-1, *arg*E3(Oc), *thi*E1
JW07674-1	F-, Δ(*araD-araB*)567, Δ*lacZ*4787(::*rrnB*-3), Δ*seqA*735::kan, λ-, *rph*-1, Δ(*rhaD-rhaB)568, hsdR514*
C2925I	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)* Tet^S *endA1 rspL136* (Str^R)*dam13::Tn9* (Cam^R) *xylA-5 mtl-1 thi-1 mcrB1 hsdR2*

Table 2 List of all *E. coli* strains used.

Media Type	Purpose
LB Lysogeny Broth	Culturing, preps, plating
SOB/SOC	Recovery after heatshock in transformations
Minimal media with lactose	Point mutation assays
GYT	Electroporation
Papillation media	Visual detection of Lac reversions in CC strains

Table 3 List of all media used.

Antibiotic	Working concentration
Ampicillin	100ug/ml
Carbenicillin	100ug/ml
Kanamycin	50ug/ml
Chloramphenicol	35ug/ml
Rifampicin	100ug/ml
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	1mM
Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)	200ug/ml
ONPG (ortho-Nitrophenyl-β-galactoside)	4mg/ml

Table 4 List of antibiotics and supplements used

2.2. Competent Cells

2.2.1. Chemical competency using $CaCl_2$ and $MgCl_2$

Cells were made competent using the calcium chloride method for heat shock using Protocol 25 described in Molecular Cloning: A Laboratory Manual Volume 1 [46]. A 250mL Erlenmeyer flask containing 50mL of LB medium with appropriate antibiotics if needed was inoculated with 0.75mL of overnight culture. This culture flask was then placed in a shaker incubator at 37ºC until the OD_{600} had reached a value of approximately 0.25 – 0.35. The culture was placed on ice for 10 minutes to halt growth and was then centrifuged for 8 minutes at 1800*g*. The supernatant was decanted, and the pellet resuspended in 30mL of ice cold 80mM $MgCl_2$ – 20mM $CaCl_2$. The resuspended cells were then centrifuged again for 8 minutes at 1800*g*. The supernatant was poured off and the pellet was resuspended in 2mL of ice cold 0.1M $CaCl_2$. The cells were then used for transformation.

2.2.2. Chemical competency using Transformation & Storage Solution (TSS)

TSS has been shown to be much more effective at inducing competency in cells in a simpler and quicker method so when putting two separate plasmids into one cell at once, the TSS method was used [15]. The addition of DMSO increases transformation efficiency to > 10^8 transformants per µg of plasmid DNA. TSS is composed of LB containing 10% (w/v) PEG 8000, 5% (v/v) DMSO and 50mM Mg^{2+} at pH 6.5 [15]. 50 ml of LB was inoculated with 0.75ml of overnight culture. Cells were grown up to early log phase (OD_{600} = 0.3-0.4), then transferred to 50ml conical tubes for centrifugation at

3000*g* for 10 minutes. The supernatant was decanted and the pellet was resuspended in 5ml of TSS Buffer. Aliquots of 100ul of competent cells were made which were then used for transformation.

2.2.3. Electrocompetency

Electroporation was also used in addition to chemical transformation because of higher transformation efficiency, which is critical when transforming multiple plasmids into a single cell. Cells were made electrocompetent using the method described in Protocol 26 of Molecular Cloning: A Laboratory Manual Volume 1 [46]. A 250mL Erlenmeyer flask containing 50mL of LB with appropriate antibiotics if needed was inoculated with 0.75mL of overnight culture and placed in a shaker incubator at 37ºC until the OD_{600} had reached a value of 0.4. The culture was then placed on ice for 10 minutes to halt growth. The culture was transferred to a 50mL conical tube and centrifuged at 1200*g* for 15 minutes at 4ºC. The supernatant was decanted and the pellet was resuspended 50mL of ice cold autoclaved H_2O. Cells were centrifuged at 1200*g* for 15 minutes, decanted and resuspended in 25mL of ice cold 10% glycerol; this step was repeated. The cells were centrifuged once more at 1200*g* for 15 minutes and resuspended gently by pipetting up and down in 1mL of ice cold GYT medium. 40µL aliquots of the cells were used for electroporation.

2.3. Transformation of Competent Cells with Plasmid DNA

2.3.1. Transformation of chemically competent cells & TSS competent cells

Plasmid DNA was introduced to chemically competent cells using the protocol indicated in Sambrook & Miller [46]. 200µl of competent cells and 2µl (~100ng) plasmid DNA were mixed together in a microcentrifuge tube and incubated on ice for 30 minutes. The tubes were then heat shocked in a DriBath at 42ºC for 90 seconds then placed back on ice for 2 minutes. LB or SOC media was added and the mixture was placed in a 37ºC waterbath for 45 minutes to recover to allow for antibiotic expression. 80µl - 100µl was plated on LB plates with the appropriate selective antibiotic and incubated overnight at 37ºC.

For TSS competent cells, 100ul of cells was mixed with 2ul of plasmid DNA in a microcentrifuge tube and incubated on ice for 60 minutes then heatshocked at 42ºC for 45 seconds. Following the heatshock, 900µl of LB was added and the tubes were placed at 37C for 45min for recovery. 80µl - 100µl was plated on LB plates with the appropriate selective antibiotic and incubated overnight at 37ºC.

2.3.2. Transformation of electrocompetent cells

Plasmid DNA was transformed into electrocompetent cells as per protocol 26 in Sambrook & Miller [46]. 40uL of electrocompetent cells was pipetted into an ice-cold sterile electroporation cuvette. 1-2uL (10pg to 25ng) of plasmid DNA was added to the electroporation cuvette. The electroporator was set to deliver a pulse of 25uF capacitance, 2.5kV over 4 milliseconds with a field strength of 12.5kV/cm. The cuvette was then removed from the electroporator and 800µl of LB at room temperature was added. The cells were transferred to a 1.5mL microcentrifuge tube and incubated in a 37ºC waterbath for 45min. 80µl of the cell mixture was plated on LB plates with the appropriate selective antibiotic. The plates were then incubated at 37ºC overnight.

2.4. Extraction of Plasmid DNA from Cells

Alkaline lysis was used to recover plasmid from cells. All steps were carried out at room temperature. The BioBasic EZ-10 Spin Column Plasmid DNA Miniprep Kit was used. 5-10mL of overnight culture was centrifuged at 6000rcf for 8 minutes. The supernatant was decanted and the pellet was resuspended in 100µl of Solution I with RNase A. The mixture was transferred to a 1.5mL microcentrifuge tube. 1µl of VisualLyse was added to the mixture to ensure sufficient alkalinity is reached. 200µl of Solution II was added to the mixture which turned blue. The tube was inverted 4-6 times and kept at room temperature for 1 minute. 350µl of Solution III was added and mixed by inversion 4-6 times. The sample was centrifuged at 12000rpm for 5 minutes and the supernatant was transferred to an EZ-10 column which was then centrifuged at 6000*g* for 2 minutes. The flow through was discarded and 750µl of wash solution was added and the centrifuged at 10000rpom for 2 minutes. This step was repeated. The column was transferred to a clean microcentrifuge tube and 50µl of Elution Buffer was added.

The tube was incubated at 37ºC for 2 minutes then centrifuged at 10000rpm for 2 minutes. The DNA was then stored at -20ºC.

2.5. Construct Design

The B2H Kit was ordered from Euromedex and the pKNT25 and pUT18 vectors were selected for cloning (refer to Figure 6). This is because having the adenylate cyclase T25 and T18 fragments on the C terminus of the fusion protein optimizes the protein-protein interactions [35].

PCR was done to amplify our genes of interest with suitable restriction sites for cloning. Restriction sites were chosen based on where they cut in the vector and the lack of cutting in the gene insert. Table 4 lists all primers used. All PCR was done using lysed DH5α as template. Vectors and inserts were cut using the chosen restriction enzymes and inserts were ligated in. Verification of the correct clone was done using restriction enzyme digest followed by gel electrophoresis.

2.5.1. *dam* (pDT25, pTD18) and *seqA* (pTSeq25, pTSeq18), *mutL* (pTLK25)

Amplification of *dam* was achieved through PCR with lysed DH5α cells as template. Oligonucleotide primers #101 and #102 were used to amplify *dam* with *Hind*III and *Bam*HI sites at either end to allow for insertion into the multi cloning sites of pKNT25 and pUT18. Ingredients needed for the PCR reactions as well as PCR program settings are described in Table 5. A 1% 1X TAE agarose gel was run to confirm that the correct amplified PCR products were obtained. Recovery and purification of the PCR product was carried out using the EZ-10 Spin Column Gel Extraction Kit (BioBasic).

Both vector and insert were first digested with *Hind*III and *Bam*HI. A 1% TAE gel confirmed that proper digestion had occurred and the digests were purified using the EZ-10 Spin Column Gel Extraction Kit (BioBasic). *dam* and pUT18 were ligated together using a 5:1 ratio at room temperature for 1 hour with 0.5µl T4 DNA ligase. The ligation mixtures were transformed into competent DH5α and single colonies were selected for overnight culture. Plasmids were recovered and clones were screened by restriction

digestion with appropriate restriction enzymes. Construction of pDT25 was essentially as above except the pKNT25 vector was used instead.

Construction of the remaining *SeqA* fusion genes was essentially the same as described with *dam* with the exception of the primers and restriction enzymes used.

pTLK25 construction – Subcloning

Construction of the *mutL*-T25 fusion gene was slightly more complex. The *mutL* gene was first amplified via *Taq* polymerase based PCR. A 1% TAE agarose gel confirmed the correct product was amplified. As PCR products amplified by *Taq* based polymerase have an additional deoxyadenosine at the ends, it is possible to ligate the fragment into a vector with overhanging T's. The Thermo Scientific InsTAclone PCR Cloning Kit was used for this purpose and the *mutL* product was ligated into the vector pTZ57R/T. Following this, our insert was cut out using the restriction enzymes *Kpn*I and *Bam*HI and gel purified. The insert was then ligated at 3:1 and 5:1 molar ratios into a prepped pKNT25 vector (also cut with *Kpn*I and *Bam*HI) via a 15 minute ligation reaction at room temperature. Ligations were then cut with *Sma*I to remove background vector religations. These were transformed into TSS competent DH5α cells and plated on appropriate selective antibiotic LB plates which were incubated overnight at 37C. Approximately 10 colonies were picked the next day for culturing, prepping and restriction digest analysis.

Primer Number	Primer Sequence	Purpose
101	GTT AAA GCT TAT GAA GAA AAA TCG CGC TTT TTTG	Insert *Hin*dIII site at 5' end of *dam*
102	CTT GAC TGC AGT TTT TTC GCG GGT GAA	Insert *Pst*I site and remove stop codon at the end of *dam.*
103	CTG GAA GCT TAT GAA AAC GAT TGA AGT TGA TGA TGA ACT C	Insert *Hin*dIII site at 5' end of *seqA*
104	CTG CGG ATC CGG GAT AGT TCC GCA AAC CTT CTC	Insert *Bam*HI site at 3' end of *seqA* and remove stop codon
105	TGC CAA ACT AAG GGA TCC TGA TGC CAA TTC AGG	Insert *Bam*HI site at 5' end of *mutL*
106	CTC GCC TTA CTG AGG TAC CTC ATC TTT CAG GG	Insert *Kpn*I site at 3' end of *mutL*

Table 5 List of primers used

Table 6 PCR reaction elements and program parameters

Goal	PCR reaction mixture (per one reaction)	Cycling conditions
Amplify *dam* with *Hind*III and *Pst*I sites	• 5ul 10X PCR Buffer (with $MgCl_2$) • 1ul 10mM dNTPs • 1ul Oligo 101 • 1ul Oligo 102 • 0.5ul Taq Polymerase • 1ul lysed DH5α cell mixture • 1.5ul nuclease free H_2O Total reaction volume: 50ul	Initial denaturation: 96C for 4 minutes Denaturation: 95C for 30 seconds Anneal: 50.7C for 30 seconds Extend: 68C for 1 min Final extension: 68C for 5 minutes Denaturation, extension and annealing steps are carried out for 30 cycles.
Amplify *seqA* with *Hind*III and *BamH*I sites	• 5ul 10X PCR Buffer (with $MgCl_2$) • 1ul 10mM dNTPs • 1ul Oligo 103 • 1ul Oligo 104 • 0.5ul Taq Polymerase • 1ul lysed DH5α cell mixture • 1.5ul nuclease free H_2O • Total reaction volume: 50ul	Initial denaturation: 96C for 4 minutes Denaturation: 95C for 30 seconds Anneal: 55.7C for 30 seconds Extend: 68C for 1 min Final extension: 68C for 5 minutes Denaturation, extension and annealing steps are carried out for 30 cycles.
Amply *mutL* with *BamH*I and *Kpn*I sites	• 5ul 10X PCR Buffer (with $MgCl_2$) • 1ul 10mM dNTPs • 1ul Oligo 105 • 1ul Oligo 106 • 0.5ul Taq Polymerase • 1ul lysed DH5α cell mixture • 1.5ul nuclease free H_2O • Total reaction volume: 50ul	Initial denaturation: 96C for 4 minutes Denaturation: 95C for 30 seconds Anneal: 56C for 30 seconds Extend: 68C for 1 min Final extension: 68C for 5 minutes Denaturation, extension and annealing steps are carried out for 30 cycles.

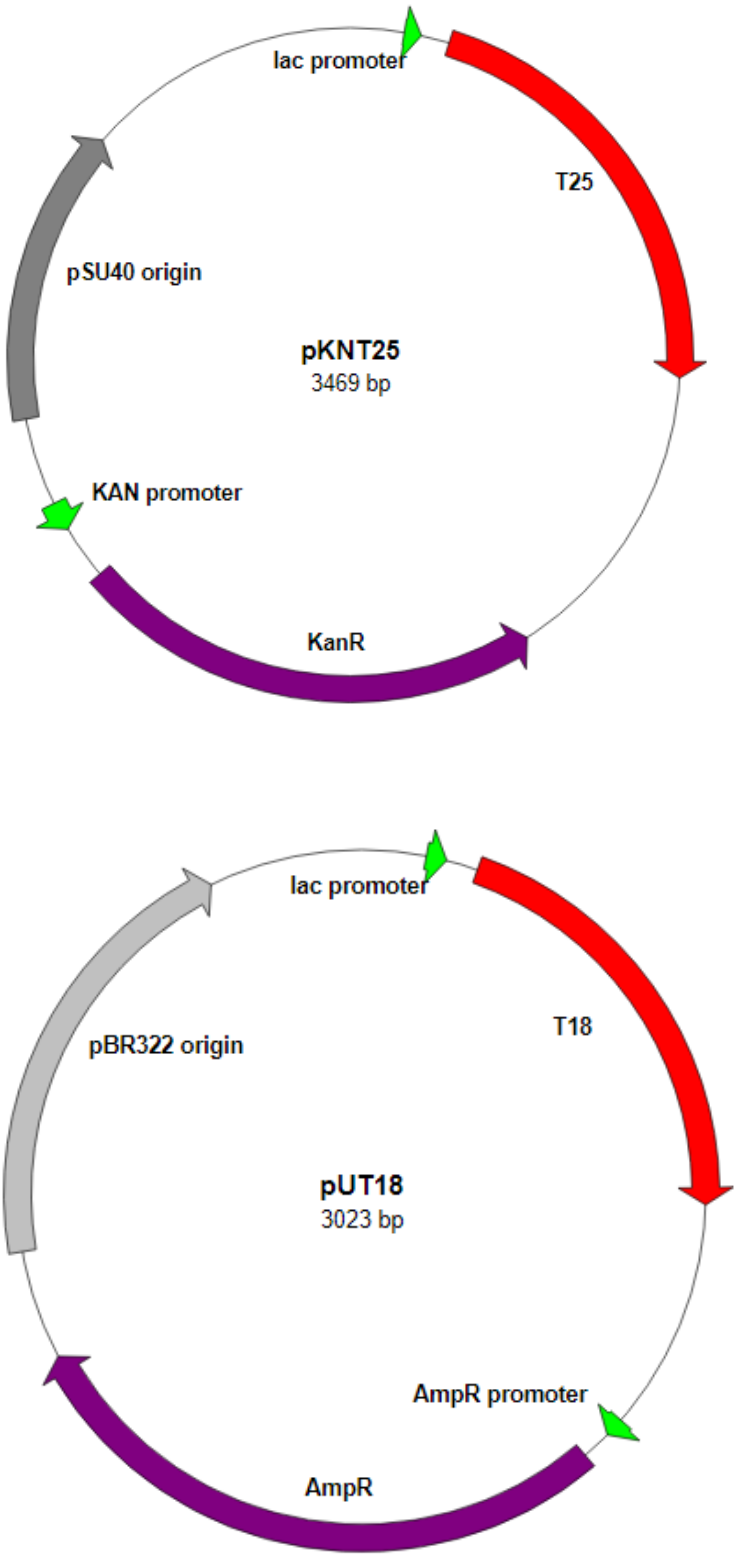

Figure 6: Diagram of pKNT25 and pUT18 cloning vectors ordered from EUROMEDEX. Multi-cloning sites are upstream of the T18/T25 segment.

2.6. Xgal spot plating - Qualitative Indications of protein-protein interactions

Potential interactions between the MMR proteins, Dam and SeqA were screened using the *E. coli* strain LJ2809. Constructs with *mutS*, *mutL*, *mutH* cloned into the T18 and T25 vectors were already made [35]. Pairwise combinations between the MMR proteins and Dam; and the MMR proteins and SeqA were transformed into LJ2809 competent cells. Single colonies were selected and cultured overnight at 37ºC with appropriate antibiotics. 8µl of each overnight culture was then spotted onto an LB plate containing 1mM IPTG and Xgal. A higher concentration of Xgal is used to increase blue intensity and reduce the number of ambiguous colonies that may need to be rescreened. Plates were incubated at 30ºC for overnight. Incubating at 30ºC allows for proteins to associate better. At 37ºC, the signal is significantly reduced due to heat-induced instability in the interactions between the proteins.

2.7. β-Galactosidase Liquid Assays: Quantification of *in vivo* protein-protein interactions

Since the spot plating only provides qualitative indications of protein-protein interactions, I opted to more accurately quantify such interactions through an ONPG assay. The ONPG substrate is a type of β-D-galactoside which can be cleaved by β-galactosidase. Upon cleavage, the colourless substrate turns yellow due to the presence of *o*-nitrophenol. Because the yellow colour produced is proportional to the amount of β-galactosidase activity, we can detect the enzyme's activity and thereby amount of protein-protein interaction using a colorimetric assay. The reaction is stopped by adding 1M Na_2CO_3 at pH12.3. The sodium bicarbonate increases the pH to denaturing the β-galactosidase enzyme and effectively preventing cleave of any further substrate.

Cultures of LJ2809 transformed with pairs of plasmids and MP2 were grown at 30ºC with appropriate antibiotics and IPTG and arabinose as an inducer. Subcultures into 5ml of fresh medium were prepared the next day and aerated continuously at 37ºC until the OD_{600} reached 0.28-0.70. 500µl of culture was added to 500ul of Z Buffer

(60mM $Na_2HPO_4 \cdot 7H_2O$, 40mM $NaHPO_4 \cdot H_2O$, 10mM KCl, 1mM $MgSO_4$, 50mM β-mercaptoethanol; pH 7.0) in a culture tube.

To lyse cells, 100ul of chloroform and 50ul of 0.1% SDS was added to the tubes. The tubes were then vortexed for 10 seconds and 0.2ml of ONPG (4mg/ml) was added to each tube and shaken to ensure even mixing. Tubes were then incubated in a 28ºC water bath until yellow colour (at minimum, a blonde shade similar to a Post-It note) had developed at which point the reaction was stopped and the reaction time was noted down. However, the exact shade of yellow is not crucial as the reaction is being timed, so stopping the reaction later at a deeper shade of yellow will not make a difference. Tubes that were cloudy were centrifuged first to pellet cell debris and reduce the amount of light scattering. The optical densities of the supernatant were then immediately recorded at 420nm and 550nm where the 420 measurement assays for amount of yellow and the 550 corrects for light scattering due to cell debris. β-galactosidase activity was calculated using Miller Units (MU) using the following equation:

$$\frac{1000 \times (OD420 - 1.75 \times OD550)}{t \times v \times \mathrm{OD600}} = \text{units of β-galactosidase}$$

Where t = time of the reaction (minutes) and v = volume of culture assayed (ml).

2.7.1. Determining arabinose concentration for induction of MP2 and MP3

MP2 was introduced into TSS competent LJ2809 cells. Resistance to rifampicin conferred by any one of a number of single nucleotide changes in *rpoB*, the beta subunit of RNA Polymerase. Tubes of 5ml LB, chloramphenicol, and arabinose at varying concentrations were set up and inoculated with transformed cells. Tubes were placed in the incubator shaker at 37C overnight. The next day, tubes of the exact same content were set up with fresh media and inoculated with 200ul of overnight culture then grown up to log phase. 100ul was plated directly onto LB rifampicin plates and a dilution of 10^{-7} was made for plating onto LB plates. Plates were incubated overnight at 37C. The next morning, colonies were counted to determine if MP2 was induced based upon whether the mutation rate increased. The assay was done several times in triplicate and I created

dose-response curves to ensure the dose I chose for further experiments was appropriate.

Chapter 3. Results

3.1. Functionality of *dam, seqA* fusion genes

The plasmids constructed for use in the B2H system contain either *dam, seqA* or *mutL* fused to a T18 or T25 adenylate cyclase catalytic domain. To test the function of the SeqA and Dam as part of fusion proteins, they were transformed into strains with defective *dam,* or *seqA.* (MutL had been tested previously.) We then tested to see if the proteins produced by the vectors complemented the mutant strains. *dam* constructs were transformed into the *E. coli* strain GM3819 which has no detectable Dam activity. A control plasmid (the parent vector using for cloning) was used as a negative control. We expect the *dam* clones to be able to rescue the mutant phenotype and thus methylate DNA. Plasmid DNA was then prepped from transformants and several single restriction enzyme digests set up with the following enzymes: *Dpn*I, *Sau*3AI, *Mbo*I. All three of these enzymes cleave at the same GATC sequences. However, *Dpn*I will only cleave at fully methylated GATC sites, while *Mbo*I is blocked by methylation at GATC sites (ie. MboI can only cleave at unmethylated GATC sites) and *Sau*3AI can cleave regardless of the methylation status. Using these three enzymes, we can see whether our *dam*-T18 and *dam*-T25 fusions are methylating the DNA. If our *dam* constructs are able to rescue the mutant, then we expect that *Dpn*I and *Sau*3AI to cleave while *Mbo*I should be blocked.

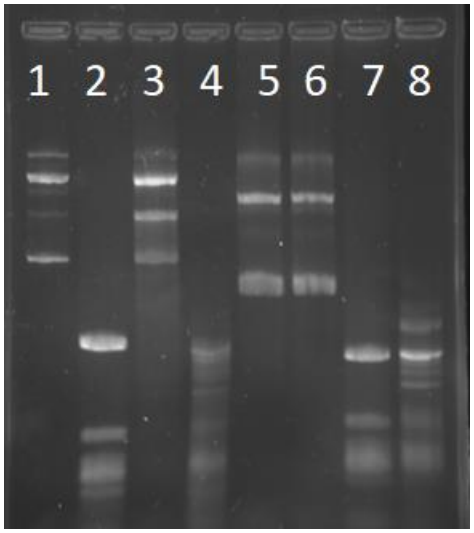

Figure 7 Restriction enzyme digests of pDT25 and pKNT25 (negative control) plasmids prepped from GM3819 E. coli strain. Lanes as follows: 1) pDT25 uncut 2) pDT25 digested with DpnI 3) pDT25 digested with MboI 4) pDT25 digested with Sau3AI 5) pKNT25 uncut 6) pKNT25 digested with DpnI 7) pKNT25 digested with MboI 8) pKNT25 digested with Sau3AI.

We can see that based on the agarose gel results, with the *dam* clone present, methylation is occurring because *Dpn*I and *Sau*3AI are able to cut but *Mbo*I is not. (Figure 7). For a successful rescue, we should expect linearization of the plasmid so approximately a 3.8kb. This is what we saw in both lanes 2 and 4 in which the plasmid was digested with *Dpn*I and *Sau*3AI respectively. With simply only the pKNT25 vector present, no methylation is occurring. This is indicative of a functional fusion protein.

To confirm functionality of the SeqA fusions, we used the JW01674 strain which has a deletion of the *seqA* gene. Our *seqA* fusion vectors were transformed into JW01674 cells. Transformants were selected then cultured for 3-4 hours to enter log phase. Cells were then fixed and stained with a methylene blue ethanol solution and visualized under at 100X magnification under a compound light microscope. Because *seqA* plays an important role in cell cycle regulation, cells that lack *seqA* do not complete cytokinesis and long strands of cells were observed in the uninduced negative control [47]. Cells that were induced with IPTG appeared normal (data not shown).

3.2. Qualitative interactions of *in vivo* protein-protein interactions in the Bacterial-2-Hybrid system

To identify possible interactions between Dam/Seq and the rest of the MMR proteins, the Dam & Seq clones along with previously made MMR clones were transformed in pairwise combinations into LJ2809. LJ2809 was the strain chosen as it has a *cyaA* deletion, which reduces any background activity that may be observed. We originally used a strain with a point mutation, but found that it easily reverts to wildtype, especially in cells with a higher than normal mutation rate, resulting in all of the patches having the uniformly dark blue colour diagnostic of a wild type cell. These revertants predominate because they have a growth advantage over *cyaA* strains. Potential interactions were first screened by spot plating onto LB plates containing Xgal and IPTG. The positive and negative controls were pT25zip/pT18zip and pKNT25/pUT18 respectively. The positive control contains leucine zippers fused to the T25 & T18 segments which interact very strongly.

Positive interaction between two proteins will result in the reconstitution of adenylate cyclase and turning on of the *lac* operon which will cleave Xgal resulting in a blue product on the LB indicator plate. I screened for 4 possible interactions (Figure 8):

- Dam18/SeqA25
- Dam18/MutL25
- Dam18/MutS25
- Dam18/MutH25

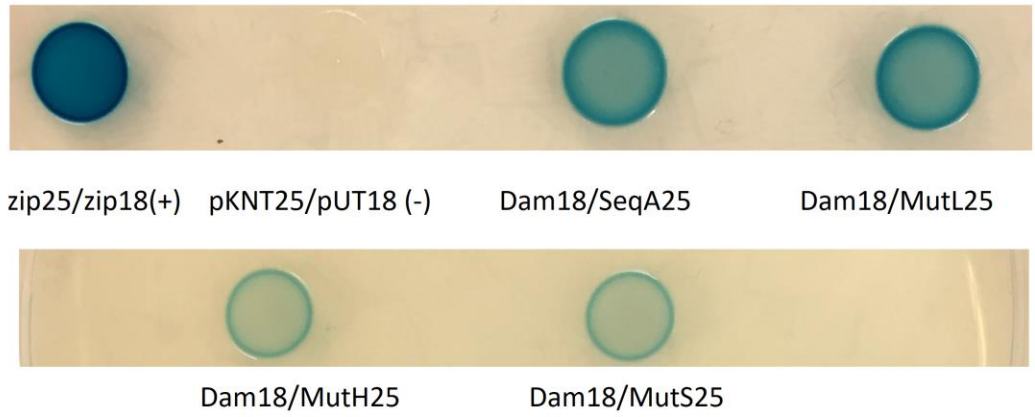

Figure 8: Qualitative interactions between Dam, Seq and MMR proteins.

This patch assay was done three times with different cultures of the same strain but only one spot per trial was patched. Both the positive and negative controls behaved as expected with the positive control giving a very strong blue readout and the negative completely white. Visually, Dam interacts with all proteins that we tested: SeqA, MutL, MutH and MutS. However, the strength of interaction varies. The interaction between Dam and SeqA was quite strong. The same can be said of the Dam and MutL interaction. Both these interactions are the strongest of the four tested experimentally, although not quite as deeply blue as the positive control.

There are also some weaker interactions that were observed. These are between Dam/MutH and Dam/MutS. Both of these patches are a light blue and much less intense than the positive control. Because of how weak these interactions are, it may indicate that these proteins normally do not associate, their interaction is very transient or that there is a third bridging molecule in between them. However, there are several factors contributing a high/low intensity in the patch colours which will be discussed in Chapter 4. Overall, these findings can help us determine at what stage Dam and Seq act in the mismatch repair cycle and their respective roles.

3.3. Quantitative interactions of *in vivo* protein-protein interactions in the Bacterial-2-Hybrid system

The same interactions tested in the qualitative Xgal patches, were also tested using a quantitative liquid β-galactosidase assay (Figure 6).

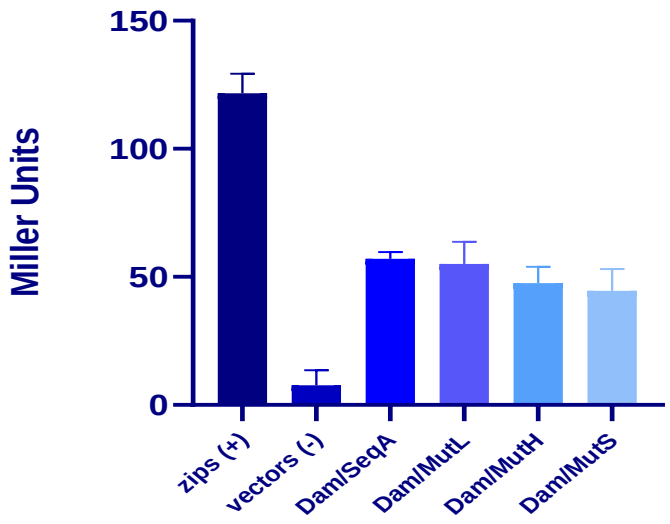

Figure 9: Quantification of in vivo interactions between Dam, SeqA and MMR Proteins. The first protein listed is always the T18 fusion and the second protein listed is the T25 fusion. Assay was done in triplicate and the error bars represent standard error.

.

The positive control and negative control gave values as expected with the zip pairs producing a much higher amount of Miller Units than the experimental pairs (refer to Figure 9). The positive control produces about two to three times as much β-galactosidase activity as the experimentals. The data seen in this quantitative ONPG assay is comparable to that seen in Figure 8 of qualitative interactions between the same pairs. Here, Dam18/SeqA25 and Dam18/L25 produce marginally higher Miller Units than Dam18/H25 and Dam18/S25. This is nearly the same result we saw in the qualitative patches as the latter two pairs were very faint.

Dam18/L25 has slightly higher the amount of Miller Units as Dam18/H25 as we might expect from the patches. One thing to note is that in this quantitative assay,

Dam18/S25 has a Miller Unit value of almost the same as Dam18/SeqA25, but Dam18/S25 appeared to be much fainter on the qualitative patch plates.

3.4. Arabinose Dosage to Induce MP2

Mutagenic plasmids were obtained from Badran et. al [48]. To determine the optimal concentration of arabinose for induction of dnaQ926 in MP2, several dose response curves were done before screening of protein-protein interactions (Figure 10). These were necessary as a dose too low could mean almost no mutation and a dose too high would mean lethality.

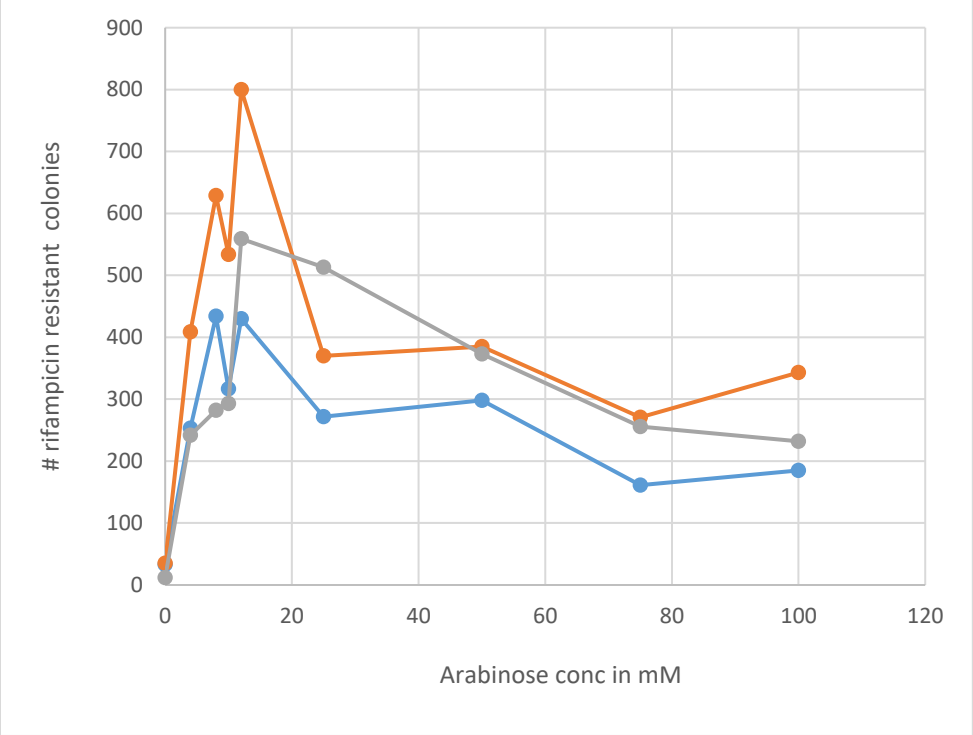

Figure 10: Triplicate dose response curve of arabinose dosage to induce MP2

All three curves appear to peak at around 8-12mM of arabinose before dropping off as the concentration increases at 25mM, 50mM, 75mM etc. The drop off is likely due to too much *dnaQ926* being overexpressed and causing lethality in the cells. Thus, the range between 8-12mM is likely to be the best for overnight induction allowing for some

expression of *dnaQ926* to cause mutations, but not an overwhelming amount. We chose 10mM as our concentration to use.

3.4.1. Mutation Rate Observed due to MP2

Determining the number of mutants as a function of the number of viable cells is important because it corrects for toxicity of the over-expressed mutant polymerase. It is a more accurate measurement of how much we are increasing the mutations by, particularly given the potential for overexpression of the mutant polymerase to make the cells weak.

Arabinose Concentration	# colonies on rifampicin plates	# colonies on LB plates (10^-7 diluted)	Total # of viable cells	Number of Rif^R mutants per 10^8 viable cells
0mM	1	712	7.12E+09	< 1.4
10mM	41	114	1.14E+09	36

Table 7: Mutation Rate of LJ2809 transformed with pTL25, pTH18, and MP2

Based on the data in Table 4,10mM arabinose overnight increase the number of mutants per 10^8 viable cells by well over 26-fold. Note, also, that viability of the arabinose-treated cells is considerably less than that of the untreated cells. Taken together, these results are indicative of 1) the arabinose induction working and 2) the *dnaQ926* creating mutations.

Induction of mutation was also checked because the production of cAMP is needed to turn on both the arabinose operon and the *lac* operon. Because *dnaQ926* is under the control of P_{BAD}, which requires cAMP-bound CAP and therefore active CyaA, this is essential in ensuring the induction is working correctly. The strain chosen is LJ2809 which is Δ*cyaA*, so the only source of cAMP the cell produces is from the fusion T18 and T25 proteins. Thus, based on Table 4, even though the cell has no working adenylate cyclase, the interaction of the fusion proteins is enough to produce enough cAMP to turn on the arabinose operon and therefore generate mutations.

3.5. Quantification of protein interactions under the effect of MP2

As the mismatch repair pathway detects transitions/transversions, we proposed that the induction of MP2 may increase the interactions between MutL/MutH and MutL/MutS. These proteins are already known to be interactors due to their role in the mismatch repair pathway; the number of cells with a positive B2H interactions increases in response to some types of mismatch (e.g. 2-aminopurine-C, Fig. 1) but not others (such as 8oxoG-A mismatches caused by loss of MutT) [36]. A quantitative β-galactosidase assay was run to measure the change in protein interactions. These results are in Figure 9.

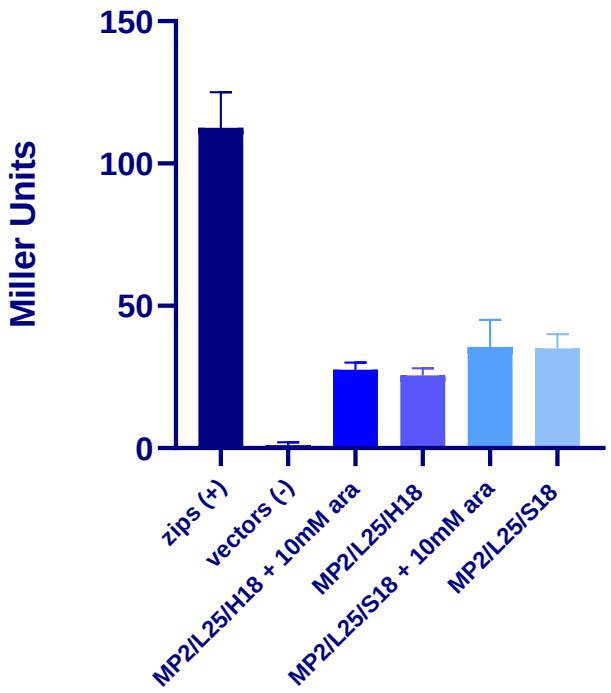

Figure 11: Quantification of mismatch repair protein interactions with increased mutagenesis from MP2. Assay was done in triplicate and the error bars represent standard error.

Both positive and negative controls behaved as expected with the zip18/zip25 positive control giving a very high Miller Units readout and the empty vectors producing

almost no Miller Units. For the MP2/L25/H18 assays, the culture induced with arabinose is marginally higher in Miller Units than the uninducd cells. For the MP2/L25/S18 assay, the culture that was uninduced produced somewhat more Miller Units than the culture that was induced. However, overall there does not appear to be a significant difference in the amount of protein-protein interaction observed between those samples with arabinose and those without.

3.6. Quantification of protein interactions under the effect of MP3

The same ONPG assay was done for cells carrying MP3 which contained both *dna926* and *dam* under $P_{BAD.}$ Both the positive and negative controls behaved as expected. Preliminary data shows that upon induction with 10mM arabinose, β-galactosidase activity has decreased for MutL/MutS as compared to the uninduced culture (Figure 12).

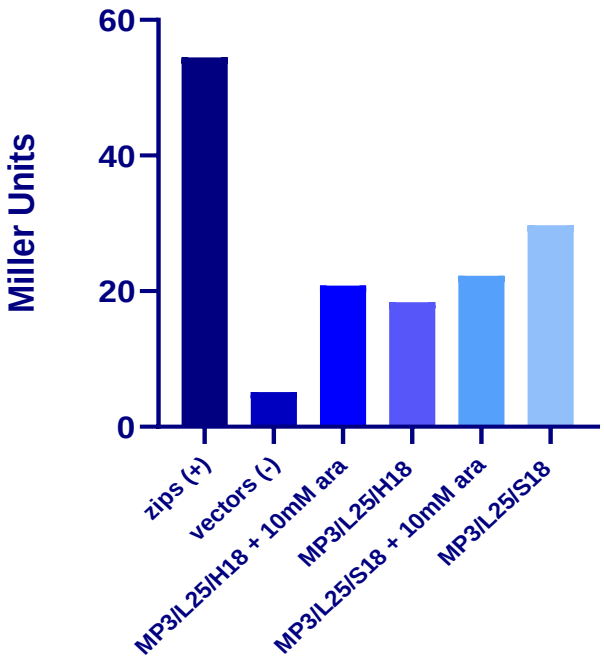

Figure 12: Quantification of protein interactions with increased mutagenesis and inhibition of mismatch repair by MP3

Chapter 4. Discussion

Although the basics of mismatch repair in *E. coli* are well-established there remains much to be uncovered in how this repair system interplays with other systems/cycles.

My study used the Bacterial-2-Hybrid (B2H) assay to determine how and if Dam interacts with MutL, MutH, MutS and SeqA. The B2H system is based on the fact that when two interacting proteins are fused to either fragment of the adenylate cyclase domain, the two adenylate cyclase fragments will be brought close enough together to be reconstituted and therefore functional to produce cAMP. Since the assay is done in a Δ*cyaA* strain, the reconstitution due to proteins interacting is the cell's only source of cAMP. The production of cAMP and formation of the CAP (Catabolite Activator Protein)-cAMP complex will then turn on a reporter gene, such as the *lac* operon in this case.

Hypothesis #2: persistent mismatches caused by failure of DNA Polymerase III proofreading will increase MMR repair activity, was failed to be upheld because of the lack of increase of β-galactosidase units as seen in the ONPG assays (Figure 11). There is undoubtedly an increase in the number of mutations due to the rifampicin assay that I did when determining the correct arabinose dosage (see Figure 10). The increase in the mutation rate seen in the rifampicin assay also shows there is enough cAMP being produced to induce the P_{BAD} promoter and thus there is also enough cAMP to turn on the *lac* operon. Thus, we know the B2H assay is working properly and the lack of increase isn't because insufficient number of mutations are being generated. It could likely be due to a failure to control some other factors such as culture density. While our assay did not show an increase in protein-protein interactions that were being tested, there could still have been increase that was undetectable (eg. increase in catalytic activity of MutH). A measure of K_{cat} would have be taken to see if MutH's endonuclease activity has increased and by how much.

Polymerase errors are known to include all 8 possible mismatches (G-A, G-T, G-G, A-C, A-A, C-C, C-T, T-T), and proofreading by the polymerase will preferentially repair those leading to transversions (G-A, G-G, A-A, C-C, C-T, T-T), likely because these are easier to detect due to a greater helix deformation [45]. Thus those causing transitions are left for the mismatch repair pathway to detect and repair. MMR defective

strains are shown to have an increase in transitions because of the lack of repair in these types of lesions while proofreading mutants have increases in transversions despite having a normal MMR. Our results show no significant increase in the amount of β-galactosidase produced which suggests that MMR does not repair mismatches leading to transversions. They may be repaired by other mismatch repair systems like MutY (G-A), MutT (G-A), MutM (8oxyG-C).

Hypothesis #1: Correct mismatch repair requires regulated interactions among proteins that potentially compete for access to the same binding site on the DNA. Alternatively, or in addition, these interactions may allow hand-off between proteins during MMR, protecting DNA intermediates (e.g. the nick formed by MutH) from non-specific processing (e.g. by DNA ligase). This hypothesis was supported by the results shown in Figures 8 and 9. The original prediction was that because coordination and regulation was needed, that all tested interactions should be positive. This was the case observed and because we hypothesized MutL as the coordinator of all such proteins, we expected the interaction between MutL and Dam to be strongest as compared to other tested pairs. This was supported by the quantitative ONPG β-galactosidase assay.

Nevertheless, there are many things that contribute to the intensity of the blue signal and when interpreting the data we had assumed many of these factors were consistent throughout. The first and most obvious is strength of binding affinity between the two proteins in question which is what we were really intending to test. Signal can also be produced via non-specific binding where perhaps one of our proteins is “sticky” and easily binds to any other protein. This could be controlled for by pairing the “sticky protein” with a zip fusion or a known combination that does not interact and seeing how much signal is produced. Third is the number of cells that are being patched on the plate. More cells in a patch would mean a stronger signal. This could easily be controlled for in future assays by testing the OD_{600} and ensuring all cultures are at roughly the same density before patching. Fourth is plasmid copy number where a higher amount of a protein can increase the amount of interactions. Thus, this can be controlled by testing interactions in both combinations: eg. Dam18/MutL25 and Dam25/MutL18. There could also be a bridging protein where perhaps MutL interacts with all proteins and thus gives a false positive readout; this could be controlled with the sticky protein control mentioned above. Lastly, there could also be cleavage of XGAL or ONPG by an enzyme other than

β-galatosidase, however this was controlled for as we saw minimal/no signal for the empty vector T25/T18.

The result observed between Dam and MutL is consistent with what we had observed between Dcm (DNA cytosine methyltransferase), Vsr, and the MMR proteins in studies conducted by previous graduate students in the Cupples Lab. Very short patch repair (VSP) is used to correct T/G mismatches caused by the deamination of 5-methylcytosine to thymine and is quite similar to MMR in the sense of cutting at a hemimethylated site and both endonucleases needing to be activated by MutL (Figures 13 & 14). The one key difference between VSP and MMR is that in VSP, the mismatch and the hemimethylated sites are both part of the same sequence whereas in MMR the hemimethylated site can be several base pairs upstream or downstream.

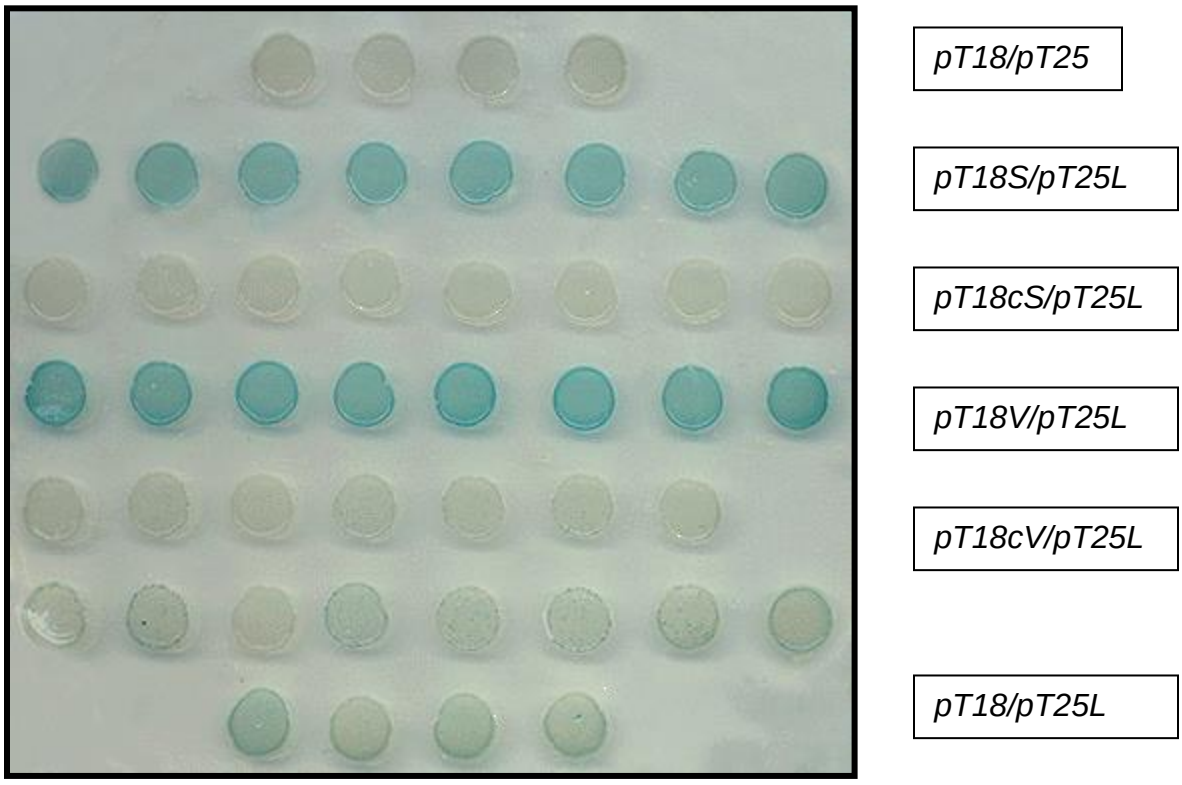

Figure 13: Assays for interactions between Vsr/MutL. Row 4 shows a strong interaction between Vsr and MutL (Mansour & Doiron, unpublished data)

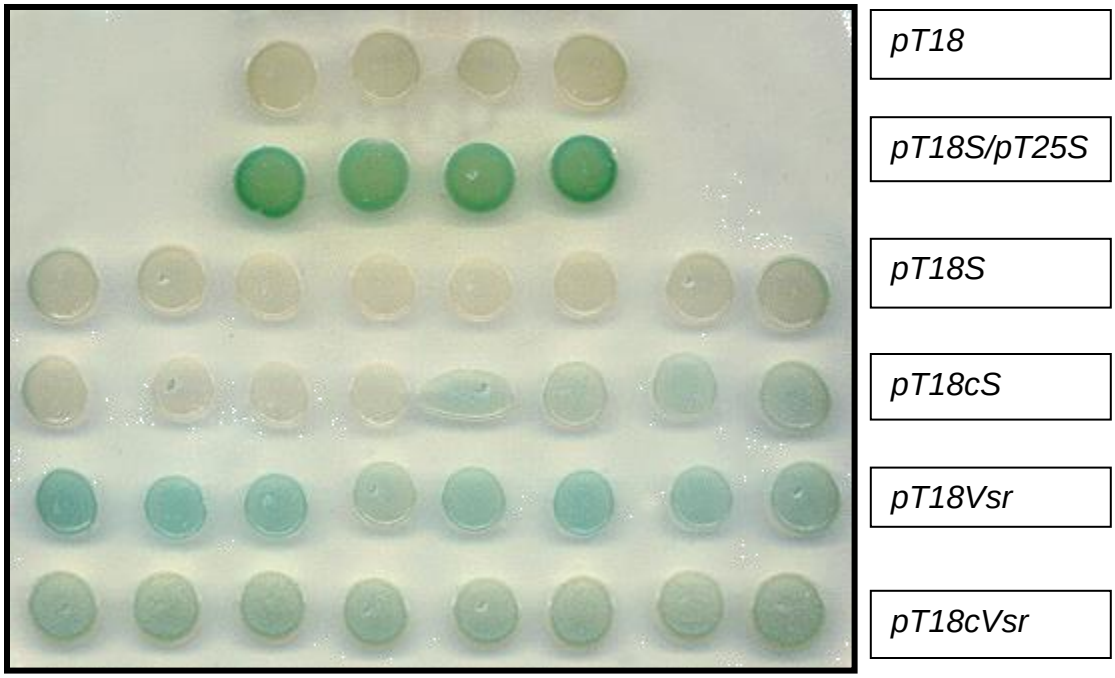

Figure 14: Assays for interactions between Vsr/Dcm. Row 5 shows an interaction between Dcm and Vsr proteins (Mansour & Doiron, unpublished data).

There are a few key enzymes in VSP. Dcm methylates the second C in CCWGG sequences, so Vsr's main job is the repair the T/G mismatches that occur in that sequence by deamination of 5MethylC to T. Cells lacking *vsr* have a high frequency of C-T mutations [49]. The first step of VSP is for Vsr to cleave 5' to the T mismatch. DNA Polymerase I will then remove a few bases including the mismatch and resynthesize the missing bases to restore the CG base pair. In this sense, Dcm and Vsr have to be produced proportionally to ensure that whenever Dcm is present Vsr is present in the same amount. Therefore, it is not a surprise that Dcm and Vsr are shown to interact, as a tight regulation on both is needed to minimize mutations in the genome (see Figure 14). Too much Dcm can produce too many T/G mismatches and too much Vsr produces a large number of transitions. MutL is also essential in VSP as it stimulates Vsr's endonuclease activity [36].

This is similar to what we saw with Dam and MutL. Dam is necessary for MMR because it creates the hemimethylation signals that tell the mismatch repair machinery what parental DNA is and what is daughter DNA. However, an overproduction of Dam will also cause an increase in mismatches, just like an overproduction of Vsr causes the

same MMR deficient phenotype. Similarly, Dam and the MMR proteins have to be kept in a balanced ratio, as do Vsr/Dcm/MMR. The interactions of the MMR proteins with Dcm and Dam show that the methylation takes place immediately after repair. This is especially important in the case of Dam because it ensures that the genome is fully methylated before replication can start over again.

The first set of qualitative screens done revealed that Dam interacts most strongly with MutL and SeqA, but to a weaker extent with MutH and MutS (Figure 5). Dam together with all of the mismatch repair proteins is not really a surprise. MutS's role is to travel down the nascent DNA strand looking for any errors. Once found, it will bind to the incorrect strand recruiting MutL and MutH. The MutL/MutS complex can then activate MutH endonuclease. Because this repair system is completely dependent on the presence and absence of methylation on the parental and daughter strands respectively, it is understandable why Dam may interact with these proteins.

Nevertheless, MutL and Dam show one of the strongest signals and this could be explained by hypothesizing that MutL binding Dam to effectively prevent it from methylating the DNA so quickly before the mismatch repair proteins have gotten a chance to complete their process. Another explanation is that if Dam is being overproduced, MutL could be responsible for binding the excess Dam. Overexpression of Dam has been shown to give the same phenotype as a mismatch repair deficient strain. This is because flooding of Dam in the cell causes too much methylation and possibly even methylation on the newly synthesized daughter strand before mismatch repair has been completed. This essentially prevents the MMR machinery from knowing what is correct and what is incorrect causing the entire repair process to fail. Additionally, Dam could be interacting with the other mismatch repair proteins through MutL; we have shown previously that interaction of MutH with MutS is dependent on MutL.

We know that MutL interacts with and/or activates the catalytic activity of all MMR proteins. The activation process requires the hydrolysis of ATP to change MutL's conformation. When a protein binds, another conformational shift occurs causing MutL to lose affinity for its current partner and recruit of a downstream protein, until Dam is reached. This allows MutL to stay on the DNA, protecting the lesion site as the repair process occurs. Because MutL binds to all MMR proteins, it thus controls the repair to

prevent these proteins from binding in the wrong order. Because of MutL's constant binding to the DNA, both the lesion on the nascent DNA and base on the methylated strand get restored with minimal risk of damage from other proteins, oxidative reactions and repair systems. Overall, MutL's role is to act as a protective hub for the DNA and controlled coordinator of all other mismatch repair proteins.

Dam is also seen to produce a fairly strongly signal with SeqA. This is seen in both the qualitative and quantitative assays (Figures 5 and 6). SeqA's primary function is to sequester the origin to prevent premature refiring by binding to hemimethylated GATC sequences [28], [31]. However, SeqA can also bind to other hemimethylated GATC sequences using cooperative binding with another protein for stabilization [28]. It is possible that SeqA is binding to Dam within the genome to further delay the methylation process and to give the mismatch repair machinery a longer chance to repair any errors. Dam could also be interacting with SeqA at the origin to induce some sort of conformational shift in SeqA causing it to lose affinity for the GATC sites. This would allow Dam to come in and methylate the origin which then permits refiring.

Here, we've demonstrated a strong likelihood that MutL interacts with Dam, indicating the importance and likelihood of MutL as an overall coordinator of mismatch repair. Having a protein that is "in command" is imperative for the accuracy of DNA repair and to ensure the organization of all other proteins. Additionally, MutL stays on the DNA to shield the lesion site from other repair systems that might accidentally get recruited. Our experiments are also consistent with the idea that mismatch repair is designed to repair transitions. This is evident with the error prone polymerase assays and the lack of significant protein interaction increase when MP2 was induced.

There are many biochemical methods that can be used to confirm these protein-protein interactions more precisely. Name, co-immunoprecipitation with Western Blotting is a conventional but more complex and lengthy technique. Fluorescence resonance energy transfer (FRET) could also be used to study known interactions between two proteins. Crosslinking is also a useful method when transient interactions need to be "fixed" before isolating the proteins.

MMR is a highly conserved system, which highlights its importance in multiple organisms. In humans, a defect in MMR causes a high predisposition to develop

HNPCC and possibly other cancers. These data together help us better understand the mismatch repair system as a whole and the implications of defective repair, inaccurate DNA replication or incorrect protein coordination in MMR.

References

[1] H. Maki, “DNA Polymerase III , Bacterial,” vol. 1, pp. 729–733, 2004.

[2] H. Echols and M. F. Goodman, “Fidelity Mechanisms in DNA Replication,” *Annu. Rev. Biochem.*, vol. 60, no. 1, pp. 477–511, 1991.

[3] R. M. Schaaper, “Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli,” *J. Biol. Chem.*, vol. 268, no. 32, pp. 23762–23765, 1993.

[4] J. J. P. Perry, E. Cotner-gohara, T. Ellenberger, and A. John, “Structural dynamics in DNA damage signaling and repair Tainer1,4,” *Curr Opin Struct Biol*, vol. 20, no. 3, pp. 283–294, 2011.

[5] S. Tsutakawa, J. Lafrance-Vanasse, and J. A. Trainer, “The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once,” *DNA Repair*, vol. 19, pp. 95–107, 2014.

[6] E. Cox, “Bacterial mutator genes and spontaneous mutations.” pp. 135–156, 1976.

[7] J. Jiricny, “The multifaceted mismatch-repair system,” *Nat. Rev. Mol. Cell Biol.*, vol. 7, no. 5, pp. 335–346, 2006.

[8] R. Wagner and M. Meselson, “Repair tracts in mismatched DNA heteroduplexes.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 73, no. 11, pp. 4135–9, 1976.

[9] J. L. Robbins-manke, Z. Z. Zdraveski, J. M. Essigmann, and M. Marinus, “Analysis of Global Gene Expression and Double-Strand-Break Formation in DNA Adenine Methyltransferase- and Mismatch Repair-Deficient Escherichia coli Analysis of Global Gene Expression and Double-Strand-Break Formation in DNA Adenine Methyltransferase- an,” vol. 187, no. 20, pp. 7027–7037, 2005.

[10] R. S. Lahue and P. M. Au, “DNA Mismatch Correction in a Defined System,” vol. 245, no. 4914, pp. 160–164, 1989.

[11] J. Brown, T. Brown, and K. R. Fox, “Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches,” *Biochem. J*, vol. 354, pp. 627–633, 2001.

[12] D. J. Allen *et al.*, “MutS mediates heteroduplex loop formation by atranslocation mechanism,” vol. 16, no. 14, pp. 4467–4476, 1997.

[13] W. Yang, “Structure and function of mismatch repair proteins.,” *Mutat. Res.*, vol. 460, no. 3–4, pp. 245–56, 2000.

[14] F. Längle-Rouault, G. Maenhaut-Michel, and M. Radman, “GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair.,” *EMBO J.*, vol. 6, no. 4, pp. 1121–7, 1987.

[15] A. Pluciennik and P. Modrich, “Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair,” *Proc. Natl. Acad. Sci.*, vol. 104, no. 31, pp. 12709–12713, 2007.

[16] P. Modrich and R. Lahue, “Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology,” *Annu. Rev. Biochem.*, vol. 65, no. 1, pp. 101–133, 1996.

[17] K. M. Culligan and J. B. Hays, “DNA mismatch repair in plants. An Arabidopsis thaliana gene that predicts a protein belonging to the MSH2 subfamily of eukaryotic MutS homologs.,” *Plant Physiol.*, vol. 115, no. 2, pp. 833–9, 1997.

[18] L. A. Loeb, “Mutator Phenotype May Be Required for Multistage Carcinogenesis Perspectivesin CancerResearch Mutator Phenotype May Be Required for Multistage Carcinogenesis1,” *Cancer Res.*, pp. 3075–3079, 1991.

[19] K. Fukui, “DNA Mismatch Repair in Eukaryotes and Bacteria,” vol. 2010, no. c, 2010.

[20] T. A. Kunkel and D. A. Erie, “Eukaryotic Mismatch Repair in Relation to DNA Replication,” *Annu. Rev. Genet.*, vol. 49, no. 1, pp. 291–313, 2015.

[21] C. Heinen, “Translating Mismatch Repair Mechanism into Cancer Care,” *Curr. Drug Targets*, vol. 15, no. 1, pp. 53–64, 2014.

[22] J. A. Cooper *et al.*, “Mismatch Repair, But Not Heteroduplex Rejection, Is Temporally Coupled to DNA Replication,” *Science (80-.).*, no. December, pp. 1713–1716, 2011.

[23] P. Hermann, Gail E. Modrich, “Escherichia coli dam Methylase,” *Mol. Biol. (Mosk).*, vol. 45, no. 4, pp. 642–51, 1982.

[24] E. Boye, M. G. Marinus, and A. Løbner-Olesen, “Quantitation of Dam methyltransferase in Escherichia coli.,” *J. Bacteriol.*, vol. 174, no. 5, pp. 1682–1685, 1992.

[25] E. Boye and A. Løbner-Olesen, “The role of dam methyltransferase in the control of DNA replication in E. coli,” *Cell*, vol. 62, no. 5, pp. 981–989, 1990.

[26] G. E. Herman and P. Modrich, “[i]Escherichia coli[/i] K-12 clones that overproduce [i]dam[/i] methylase are hypermutable.,” *J. Bacteriol.*, vol. 145, no. 1, pp. 644–646, 1981.

[27] T. Waldminghaus and K. Skarstad, “The Escherichia coli SeqA protein,” *Plasmid*, vol. 61, no. 3, pp. 141–150, 2009.

[28] T. Brendler *et al.*, “Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences,” *EMBO J.*, vol. 18, no. 8, pp. 2304–2310, 1999.

[29] M. A. Sánchez-Romero, S. J. W. Busby, N. P. Dyer, S. Ott, A. D. Millard, and D. C. Grainger, “Dynamic distribution of SeqA protein across the chromosome of Escherichia coli K-12,” *MBio*, vol. 1, no. 1, pp. 1–9, 2010.

[30] K. Skarstad, E. Boye, and H. B. Steen, “Timing of initiation of chromosome replication in individual Escherichia coli cells.,” *EMBO J.*, vol. 5, no. 7, pp. 1711–1717, 1986.

[31] T. Bach, M. A. Krekling, and K. Skarstad, “Excess SeqA prolongs sequestration of oriC and delays nucleoid segregation and cell division,” *EMBO J.*, vol. 22, no. 2, pp. 315–323, 2003.

[32] H. Yang, E. Wolff, M. Kim, A. Diep, and J. H. Miller, “Identification of mutator

genes and mutational pathways in Escherichia coli using a multicopy cloning approach," *Mol. Microbiol.*, vol. 53, no. 1, pp. 283–295, 2004.

[33] Y. L. Jae, J. Chang, N. Joseph, R. Ghirlando, D. N. Rao, and W. Yang, "MutH complexed with hemi- and unmethylated DNAs: Coupling base recognition and DNA cleavage," *Mol. Cell*, vol. 20, no. 1, pp. 155–166, 2005.

[34] S. Fields and O. Song, "A novel genetic system to detect protein-protein interactions. [Yeast two hybrid]," *Nature*, vol. 340, no. 6230, pp. 245–246, 1989.

[35] C. Mansour, "Shall we dance? Investigating Protein interactions implicated in Escherichia coli postreplicative mismatch repair using a bacterial two hybrid system," Concordia University, 2000.

[36] Y. Y. Polosina, J. Mui, P. Pitsikas, and C. G. Cupples, "The Escherichia coli mismatch repair protein MutL recruits the Vsr and muth endonucleases in response to DNA damage," *J. Bacteriol.*, vol. 191, no. 12, pp. 4041–4043, 2009.

[37] C. G. Cupples and J. H. Miller, "A Set of lacZ Mutations in Escherichia coli that Allow Rapid Detection of Each of the Six Base Substitutions," *Pnas*, vol. 86, no. 14, pp. 5345–5349, 1989.

[38] P. Friedhoff, P. Li, and J. Gotthardt, "Protein-protein interactions in DNA mismatch repair," *DNA Repair (Amst).*, vol. 38, pp. 50–57, 2016.

[39] M. C. Hall and S. W. Matson, "The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity," *J. Biol. Chem.*, vol. 274, no. 3, pp. 1306–1312, 1999.

[40] C. a Mansour, K. M. Doiron, and C. G. Cupples, "Characterization of functional interactions among the Escherichia coli mismatch repair proteins using a bacterial two-hybrid assay.," *Mutat. Res.*, vol. 485, no. 4, pp. 331–8, 2001.

[41] G. Karimova, J. Pidoux, A. Ullmann, and D. Ladant, "A bacterial two-hybrid system based on a reconstituted signal transduction pathway," *Proc. Natl. Acad. Sci.*, vol. 95, no. 10, pp. 5752–5756, 1998.

[42] A. Battesti and E. Bouveret, "The bacterial two-hybrid system based on adenylate

cyclase reconstitution in Escherichia coli," *Methods*, vol. 58, no. 4, pp. 325–334, 2012.

[43] M. M. Edwards *et al.*, "Mutations in Lama1 disrupt retinal vascular development and inner limiting membrane formation," *J. Biol. Chem.*, vol. 285, no. 10, pp. 7697–7711, 2010.

[44] N. Dautin, G. Karimova, and D. Ladant, "Human Immunodeficiency Virus (HIV) Type 1 Transframe Protein Can Restore Activity to a Dimerization-Deficient HIV Protease Variant," *J. Virol.*, vol. 77, no. 15, pp. 8216–8226, 2003.

[45] I. J. Fijalkowska and R. M. Schaaper, "Mutants in the Exo I motif of Escherichia coli dnaQ: Defective proofreading and inviability due to error catastrophe," *Proc. Natl. Acad. Sci.*, vol. 93, no. 1, pp. 2856–2861, 1995.

[46] J. Sambrook and D. Russell, *Molecular Cloning: A laboratory manual*, Third. New York: Cold Spring Harbor Laboratory Press, 2001.

[47] S. Fossum, S. Søreide, and K. Skarstad, "Lack of SeqA focus formation , specific DNA binding and proper protein multimerization in the Escherichia coli sequestration mutant seqA2," vol. 47, pp. 619–632, 2003.

[48] A. H. Badran and D. R. Liu, "Development of potent in vivo mutagenesis plasmids with broad mutational spectra," *Nat. Commun.*, vol. 6, pp. 1–10, 2015.

[49] M. Lieb, "Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair," *Genetics*, vol. 128, no. 1, pp. 23–27, 1991.

Appendix A.

Cloning of P_{DinB} into pGFP

pGFP was originally obtained to create fusion fluorescent proteins. The idea was to have GFP fused to *dinB,* which produces an error prone polymerase. When this polymerase was expressed under P_{DinB} , it could easily be detected using fluorescence microscopy.

The promoter was created by annealing of 4 oligos listed below:

PdinB1	AGCGGAAGAGCGGGCCCATGCTGAATCTTTACGCATTTCTCAAACCC
PdinB2	TTCAGGGTTTGAGAAATGCGTAAAGATTCAGCATGGGCCCGCTCTTCC
PdinB3	TGAAATCACTGTATACTTTACCAGTGTTGAGAGGTGAGCAGGTAC
PdinB4	CTGCTCACCTCTCAACACTGGTAAAGTATACAGTGAT

Oligos were mixed together in a 1:1:1:1 ratio, heated to 100ºC then allowed to slowly cool back down to room temperature for annealing. This promoter, now fully formed with SapI and KpnI sticky ends is now suitable for cloning into the pGFP plasmid. pGFP was digested with SapI and KpnI. The digest mixture was gel purified to remove any small bits of DNA. The promoter was then ligated in using T4 DNA ligase with a 10 minute incubation at room temperature. The ligation mixture was transformed into DH5α cells and plated on selective antibiotic plates. Restriction digest was used to confirm the correct clone.