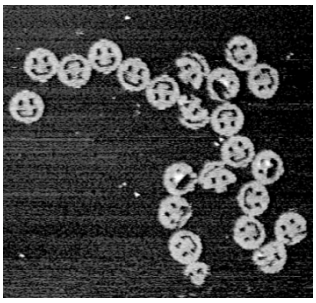
Davidson College

Department of Biology

Honors Thesis

Title:

Building a Modular Single-Stranded DNA Production Device Using Multi-Copy Single-Stranded DNA



Submitted by:

Pallavi Penumetcha

Class of 2011

Date of Submission:

11 April 2011

First Reader: Dr. A. Malcolm Campbell

Second Reader: Dr. Barbara Lom

Abstract

The goal of this project was to build a modular device for *in vivo* production of single-stranded DNA (ssDNA) using multi-copy single-stranded DNA (msDNA). ssDNA has many applications in the field of biology and one rapidly growing subfield is the use of ssDNA in DNA origami. DNA origami involves the use of ssDNA to build two or three dimensional nanostructures. These DNA structures have potential applications in a wide variety of fields ranging from medical drug delivery to device physics. Synthesis of these ssDNA structures to make DNA origami is time consuming and expensive. *In vivo* production of ssDNA would streamline large scale production of DNA origami. I proposed using the extrachromosomal ssDNA-RNA structure called msDNA that is produced from chromosomal structures called retrons to make ssDNA *in vivo*. The first goal of this project was to build a modular system for producing msDNA. I successfully built the portion of the msDNA production construct that produced the *msr/msd* regions (portions of the retron that produce the msDNA structure) of the EC83 and EC86 retrons (the two bacterial retrons I redesigned in my msDNA production system). One of the biggest challenges with building the msDNA production construct was cloning reverse transcriptase (RT) into a vector. Traditional methods for cloning the reverse transcriptase were not successful, so I designed a novel method for cloning a part into a vector that involved using three rounds of PCR to add a desired DNA sequence upstream of an existing sequence. I was able to successfully clone a part into a vector using my new method, but the part that was cloned had many mutations that conferred the reverse transcriptase non functional. I decided to use the reverse transcriptase that is naturally present in the *E. coli* B genome as a part of my msDNA production device. The second goal of this project was to design a novel method for detecting msDNA. Currently, polyacrylamide gels are the most common method used to detect ssDNA, and our lab attempted to detect msDNA on a polyacrylamide gel, but was

unsuccessful. Thus, I designed a new protocol for detecting msDNA using restriction digestion and PCR. A novel method for detecting msDNA will provide a quick screen to determine if anmsDNA producing construct is making its product. I worked on developing the protocol for this novel msDNA detection method and was able to overcome many of the barriers I encountered. Further protocol development is necessary to streamline the msDNA detection protocol. The next step of this project is to test themsDNA detection protocol in conjunction with the msDNA producing construct.

Introduction

DNA Origami

Recently, DNA origami has emerged in the field of nanotechnology as a way to capitalize onDNA's ability to self assemble (Rothemund, 2006). The complementary base pairing associated with single strands of DNA allows an investigator to design segments of DNA to bind to each other in a particular orientation. The ability of DNA molecules to self assemble into both linear structures and other patterns such as stars and squares has far exceeded the complexity associated with other forms of self assembly and is comparable to the resolution achieved by atomic force microscopy (Rothemund, 2006). DNA origami has many potential applications. Rothemund proposed the benefits of a "nanobreadboard," to study the assembly of different protein molecules based on a DNA origami template and the spatial organization between these protein molecules (Rothemund, 2006). The ability to construct stationary protein complexes with defined nanometer spacing has very important applications in the fields of proteomics, tissue engineering and medical diagnostics (Chhabra *et al.,* 2007; http://2008.igem.org/DNA-Origami). Formation of higher order DNA structures has important implications in the biomedical field. A DNA origami cage could be constructed to hold a drug for tissue specific delivery in a diseased

individual (Greenwood, 2009).DNA origami also has promising applications in the construction of electronic devices and chemical sensors, also known as nanoelectronics (Lin *et al.,*2006). The construction of nanostructures in the field of nanoelectronics, however, requires nanomaterials to be organized into precise patterns with defined interparticle spacing (Lin *et al.,* 2006). Rothemund (2006) demonstrated the ability of scaffold DNA origami to allow nanostructures to assemble in an ordered manner. Thus, DNA origami has the potential to have important applications in the area of device physics as well as biology.

Although the idea of DNA self-assembly proposed by Rothemund (2006) has many potential applications, *in vitro* production of ssDNA necessary to create DNA origami is very expensive and impractical to implement on a large scale. The long DNA scaffolding strand in combination with the short DNA strands must be purified and the strand concentrations have to be equimolar to make sure that the higher order structures are assembled with the right stoichiometry(Rothemund, 2006). Each of the "staple" strands used to fold the template strand have to be chemically synthesized*in vitro,*which is expensive and time consuming(Rothemund, 2006). I propose the use of extrachromosomalssDNA-RNA structures, called multi-copy single-stranded DNA (msDNA), to produce ssDNA*in vivo*. *In vivo* production of ssDNA is cheaper than *in vitro* ssDNA production, because there is less manpower and high tech equipmentrequired. *In vivo* production is more efficient that *in vitro* production because the cells are performing all the procedures that were originally done by hand. Consequently, *in vivo* production of ssDNA is an important step towards streamlining DNA origami production.

Retrons and msDNA

msDNA was first discovered in the gram-negative soil bacterium, *Myxococcusxanthus*and then isolated in *E. coli* and other species of bacteria (Inouye and Inouye, 1991). The biological significance of retrons is still unknown, but there is speculation about the

impacts retrons may have on modifying the host's genome (Lampson *et al.,* 2005). msDNA exists as a branched structure that is made up of a ssDNA molecule and a ssRNA molecule (Figure 1). These two molecules are linked by a 2',5'-phosphodiester linkage at a highly conserved guanine residue and base pairing between the RNA and DNA molecules at the 3' end of the structure (Inouye and Inouye, 1991).

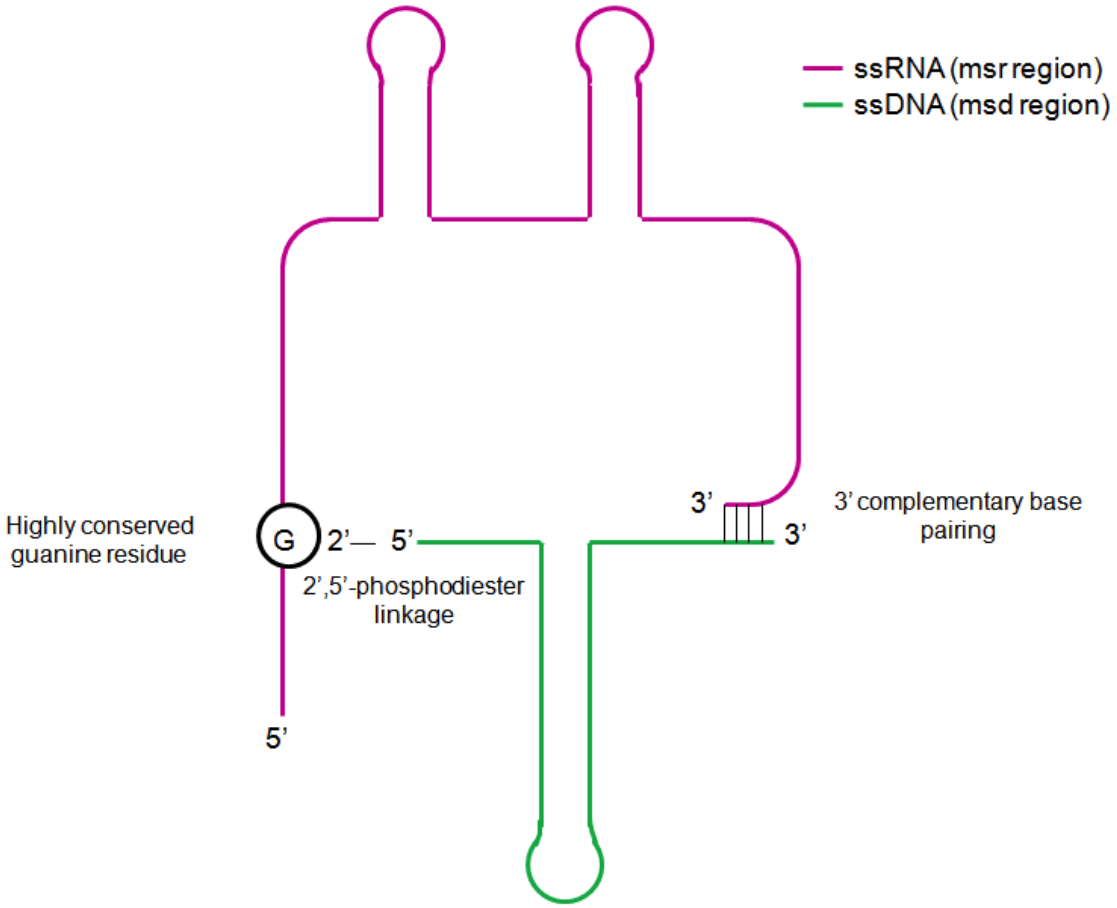


Figure 1.Structure of msDNA. The pink portion represents ssRNA and the green portion represents ssDNA. The highly conserved guanine residue and the 2',5'-phosphodiester linkage are highlighted as well as the complementary base pairing at the 3' end of the structure. The highly looped characteristic of the entire structurewas determined by extensive experimental evidence (Inouye and Inouye, 1991).

Within bacteria, msDNA is produced from a genetic element called a retron. The retron contains an *msd* region (that codes for msDNA), an *msr* (that codes for msdRNA) and a reverse transcriptase (RT; Inouye and Inouye, 1991; Figure 2). During the biosynthesis of msDNA, RNA

polymerase transcribes the entire retron (*msd, msr,* and RT). The RT transcript is translated into a reverse transcriptase protein. Inverted repeat sequences within the *msd* and *msr* regions of the RNA transcript fold upon each other to form a looped structure. In the looped structure, the highly conserved guanine residue is exposed at the end of the loop and serves as the primer for RT and allows for the formation of a ssDNA-RNA complex (Figure 2). In the *E. coli* EC83 retron (the *msd* region is 83 nucleotides long) the ssDNA produced in msDNA is cleaved from the RNA complex such that free ssDNA can be isolated (Kim *et al.,* 1997). This cleavage does not occur in the EC86 retron (Lampson *et al.,* 1990).

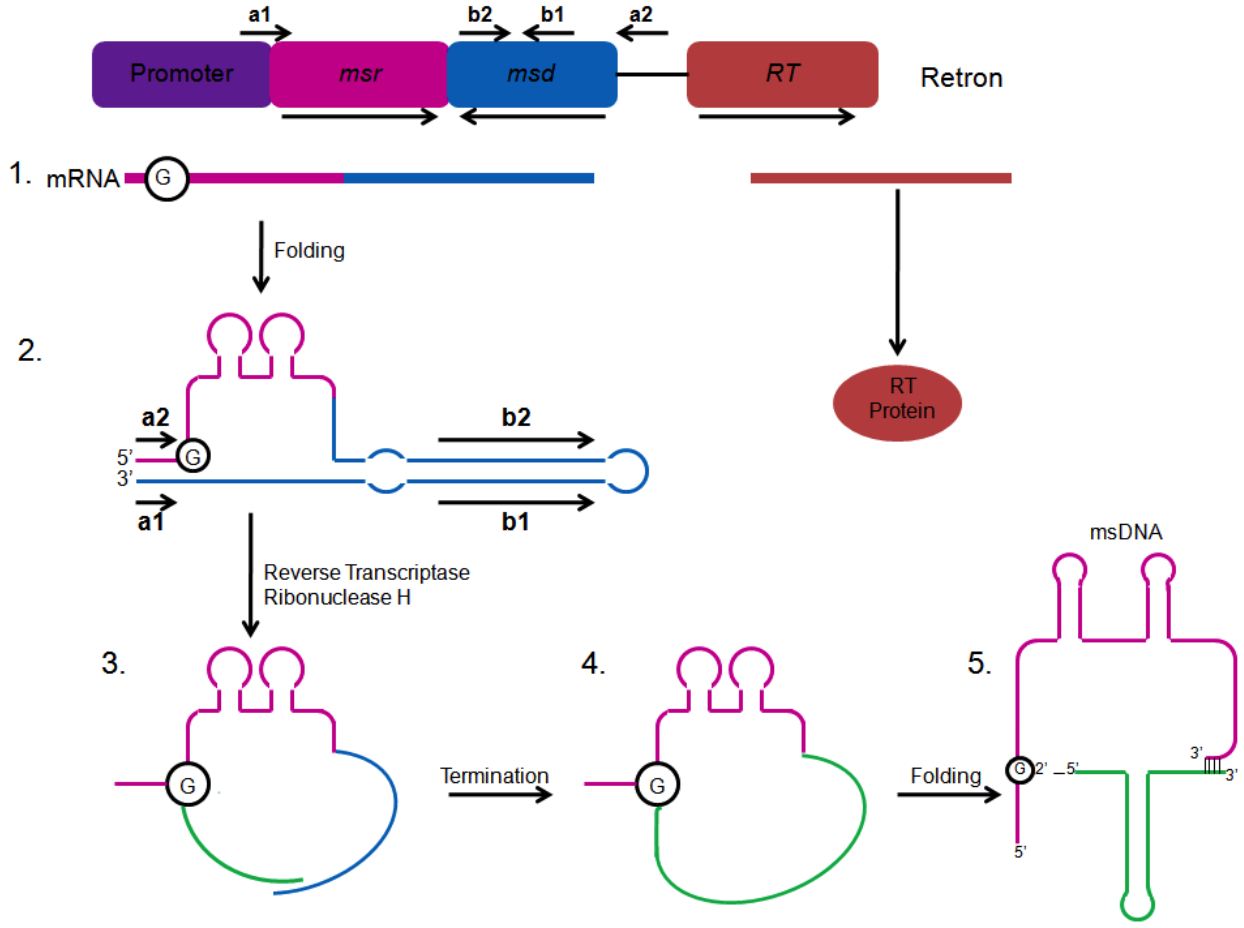


Figure 2.Biosynthesis of msDNA. The retron as arranged in the chromosome is depicted at the top. The black arrows above the retron represent complementary regions that allow formation of secondary structure shown in (5) (a1 binds to a2 and b1 binds to b2). The black arrows below the retron represent the orientation of the *msr* and *msd* regions.(1) The entire bacterial retron is transcribed from left to right. The reverse transcriptase portion of the retron is translated into protein. (2) Complementary regions in the *msr/msd* transcript base pair forming secondary

structure. (3) Reverse transcriptase producesassDNA copy of *msd* region of the RNA transcript starting at the conserved guanine residue. Ribonuclease H destroys the template RNA during ssDNA synthesis. (4) Synthesis of ssDNA terminates at the 3' end of the *msr* region of the RNA transcript. (5) Further folding of the molecule forms msDNA.

Re-designing and building the EC83 and EC86 retrons

Figure 3 shows the two basic constructs designed by Waters (2010) to produce msDNA*in vivo*. Once these constructs are built, the ultimate goal of the project is to design the *msr/msd*region of the construct to produce a desired piece of ssDNA. Different *msr/msd*regions would be inserted into different cells such that when the cells are lysed, the individual pieces of ssDNA will self assemble into the desired three-dimensional structure as opposed to purifying individual strands and combining equimolar concentrations of ssDNA plus template. Both msDNA production constructs use the EC86 RT to form the final msDNA structure. Waters (2010) chose to test the effects of the same RT (EC86) on two different *msr/msd* regions (EC86 and EC83) to determine if an RT from a different retron can act on multiple *msr/msd*regions. In addition, Waters (2010) had previously obtained a strain of bacteria that contained EC86 RT in its genome, and had amplified this region using PCR instead of synthesizing the entire EC83 RT sequence. Waters (2010)cloned EC86 RT into a vector, but the protein was not successfully cloned downstream of a promoter plus a ribosomal binding site (RBS). One of the goals of my thesiswas to clone RT downstream of a promoter and RBS, in order to complete the msDNA production cassette (Figure 3).

msDNA Detection

In addition to building EC86 RT downstream of a promoter and RBS, the other main goal of my project was to develop a better method for detecting msDNA. Previously, Mike was able to visualizeassDNA PCR product on a polyacrylamide gel (Waters, 2010), but not msDNA produced by the *E. coli* B cells (strain of *E. coli* cells that has the EC86 retron). Detection of

msDNA is very important in order to make sure that the msDNA production constructs are producing the molecule they were designed to produce.

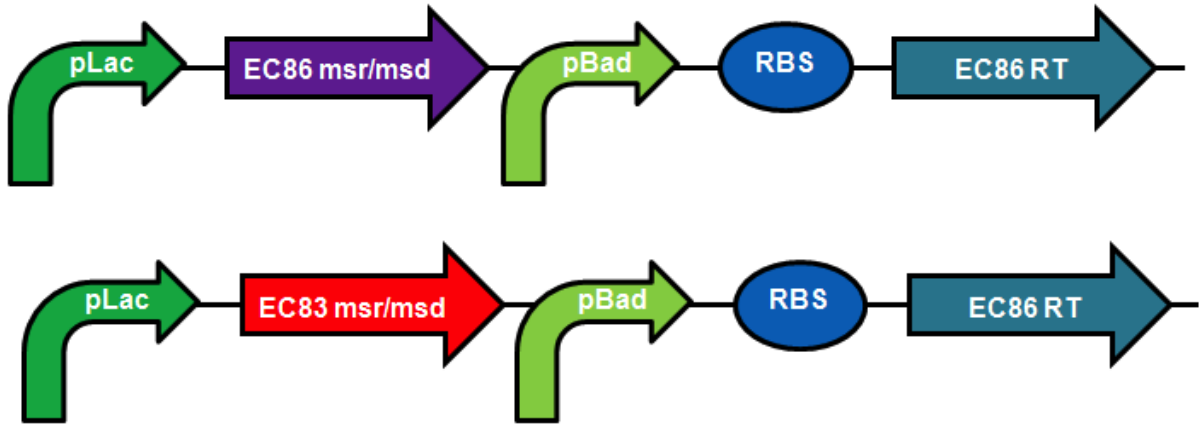


Figure 3. msDNA production constructs. These two testing constructs represent the modular systems designed to produce msDNA. The two different promoters increase modularity of the system so that each portion of the retron can be manipulated individually which controls the amount of RT that is produced in relation to the *msr/msd*region. Relative production of RT and *msr/msd* could have an impact on the formation of msDNA. Curved arrows represent promoters, straight arrows represent coding regions and circles represent ribosomal binding sites.

Methods

General laboratory procedures for methods can be found on the lab's wiki site that provides an extensive list of common protocols (These protocols can also be found in Appendix A): http://gcat.davidson.edu/GcatWiki/index.php/Davidson_Missouri_W/Davidson_Protocols

DNA Isolation and Purification: DNA was isolated using a standard mini-prep procedure. Both Promega and Zyppy mini preps were used. DNA was purified using the ethanol precipitation protocol.

Cloning*:* All cloning was performed according to BioBrick cloning procedures outlined by Knight (2003). All ligations were performed with Promega's rapid ligation buffer and enzymes. All transformations were done with Zyppy JM109 competent cells.

Gel Electrophoresis: Gels were run in 0.5X TBE or 1.0X TAE. All gel volumes were 60mL. The percent gel being used was determined using the Optimal Agarose Gel Concentration tool (http://gcat.davidson.edu/iGEM08/gelwebsite/gelwebsite.html).

Gel Purification: In order to increase the DNA yield from this procedure, DNA was eluted in a large volume (30µl-50µl), spun in a speed-vac for 10 minutes and then re-suspended in a small volume (4µl-6µl).
(http://www.bio.davidson.edu/courses/Molbio/Protocols/MN_gelpure.html).

Restriction Digestion: Digestions were performed with Promegaor New England BioLab enzymes. All Promega digestions were incubated for 1 hour in the appropriate Promega buffer for the enzymes being used. NEB digestions were incubated for 10-20 minutes in NEB buffer four (for digestions involved in ligation-transformation) or four to five hours (for digestions involved in msDNA detection).

Polymerase Chain Reaction (PCR): This procedure was used to screen for successful ligations. A modified version of this protocol was used in the msDNA detection protocol. The cycle times used are outlined here (http://www.bio.davidson.edu/courses/Molbio/Protocols/ColonyPCR_Screening.html). This procedure was also used to amplify desired sequences of DNA and to build particular constructs (http://www.bio.davidson.edu/courses/Molbio/Protocols/pcr.html).

Cloning reverse transcriptase:Traditional methods to clone reverse transcriptase were not successful. The figure below describes a novel method for building a part into a vector. This method involved three rounds of PCR to build pBad-RBS upstream of RT. The PCR products were size verified between each round of PCR and then gel purified to be used in the next round. This method definitively demonstratedif RT could be cloned into a vector downstream of a promoter and RBS. The primers used in this process are detailed in Table 1.

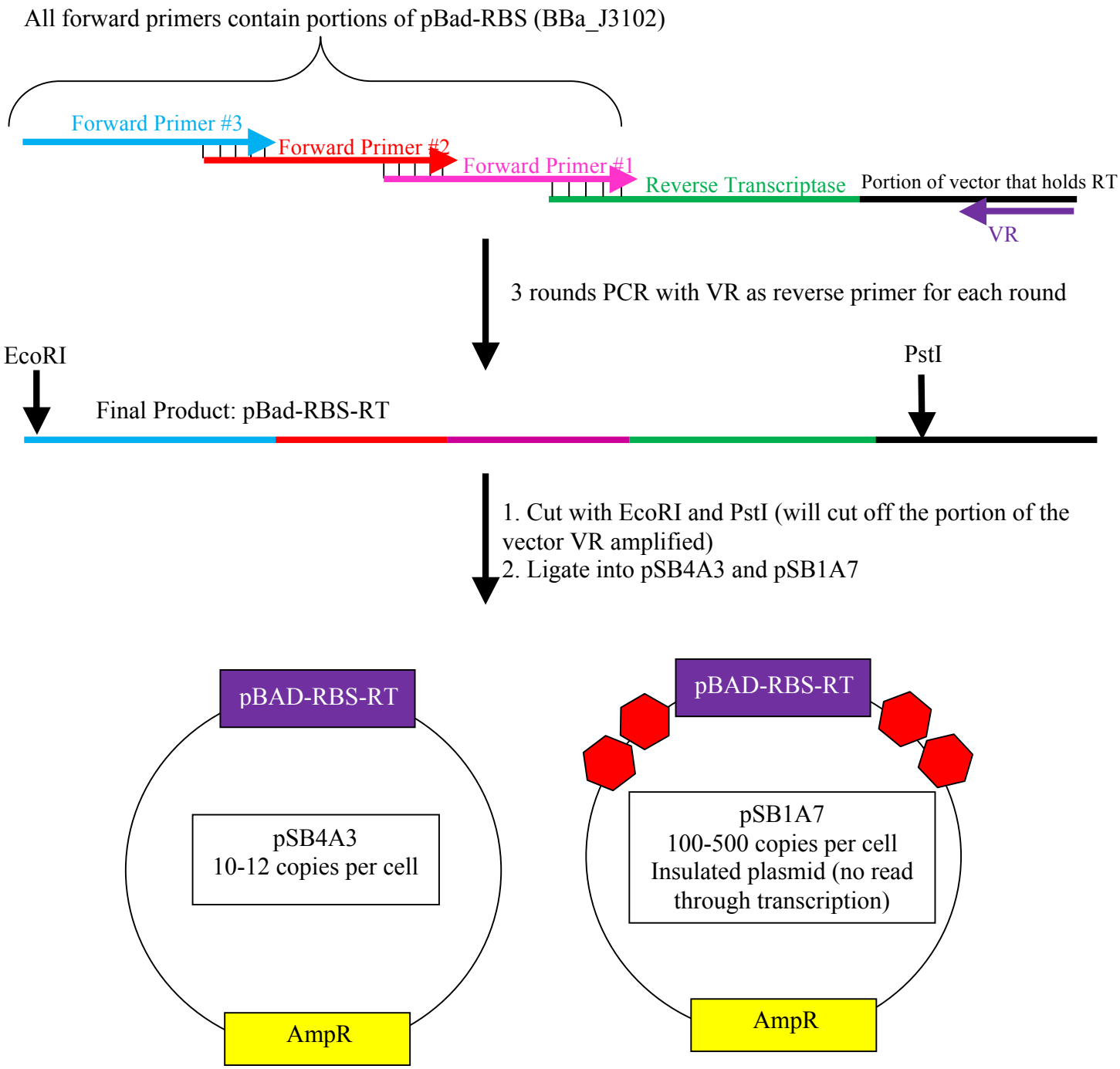


Figure 4. Building pBad-RBS-RT with PCR.
PCR was used to build pBad-RBS (part #BBa_J3102) in front of EC86 RT (part# BBa_I715074). Forward primers 1,2, and 3 each included portions of part #J3102 in them. Three successive rounds of PCR were performed using EC86 RT as the template. The elongation time (1 minute) for each round of PCR was kept short in order to minimize the possibility of amplifying mutations introduced by DNA polymerase. The PCR products were gel purified after each round of PCR. After the third round of PCR, the gel purified product (966bp) was ligated into pSB4A3 and pSB1A7. The yellow AmpR boxes in each vector represent the gene for

ampicillin resistance. The red hexagons that flank the part in pSB1A7 represent transcriptional terminators which prevent read through transcription.

Table 1: pBad-RBS-EC86RT Construction Primers. These three primers were used to build pBad-RBS in front of EC86 RT using PCR. Each successive round of PCR added another portion of the pBad-RBS sequence to the 5' end of the growing PCR product (Figure 4).

Primer Name	Primer Sequence (5'→3')	PCR Product Length
Forward Primer #1	ctctactgtttctccataccgtttttttgggctagctactagagattaaagaggagaaaatgaacaatttgcatgacatg	1046
Forward Primer #2	gccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctctactgtttctccatacc	1066
Forward Primer #3	gcatgaattcgcggccgcttctagagacattgattatttgcacggcgtcacactttgctatgccatagcatttttatccataag	1119

msDNA Detection:Current methods for purifying and utilizing ssDNA in DNA origami synthesis are time consuming and expensive. This thesis proposes a novel *in vivo* method for utilizing ssDNA in DNA origami production (Figure 5).In this novel method, ssDNA molecules do not need to be purified and ssDNA concentrations do not need to be equimolar. Thus, an *in vivo* approach eliminates multiple steps in DNA origami production. In conjunction with an *in vivo* method for producing msDNA, developing a more efficient protocol for msDNA detection is necessary to confirm that themsDNA production construct (see Figure 3) is producing the expected structure.The novel protocol I developed to detect msDNAutilizes the ability of restriction enzymes(which only cut double stranded DNA) to digest all double stranded DNA (dsDNA) so that,when regions of dsDNA are amplified in conjunction with msDNA, only the msDNA product will be visible on a gel (Figure 6). Table 2 indicates the primers used in my detection methods and figures 7 and 10 detail the protocols that were used. Three different enzymes were used in the msDNA detection protocol: dsDNase, HaeIII and CviKI-1. HaeIII and CviKI-1 had similar protocols, but the dsDNase protocol was slightly varied from the other two. dsDNase is a unique enzyme because it digests all double stranded DNA and does not have site-specific DNA digestion. Additionally, this enzyme was used first so a majority of the protocol development was done with dsDNase.

Table 2: msDNA Detection Primers. This table details the sequences of all the primers used for msDNA detection. Helicase primers were designed to ensureenzymes were cutting up dsDNA. The msDNA primers were designed to amplifymsDNA. msdRNA primers were designed to amplify the *msr* region and were used to determine how many times the EC86 retron occurred within the genome (see Results section). Detailed information about how these primers were designed can be found in Appendix B.

Primer Name	Primer Sequence (5'→3')	PCR Product Length
EC86 Helicase Forward	gtctgttttccttgttggaacggag	153bp
EC86 Helicase Reverse	gcgaccaaacgccagcaataaaatttg	153bp
EC86 msDNA Forward	gtctgttttccttgttggaacggag	75bp

EC86 msDNA Reserve	gtcagaaaaaacgggtttcctggttg	75bp
EC83 msDNA Forward	cttcggcgccttgtttgaaaaactaggcgttg	76bp
EC83 msDNA Reverse	gtaaatgtcaacgccttgtttgaaaaactcggc	76bp
EC86 msdRNA Forward	gcgcacccttagcgagaggtttatcattaag	75bp
EC86 msdRNA Reverse	ctcagattcaatgcaggatgccgaaacaacatcc	75bp
EC83 msdRNA Forward	cttcggcgccttgtttgaaaaactaggcgttg	76bp
EC83 msdRNA Reverse	gtaaatgtcaacgccttgtttgaaaaactcggc	76bp

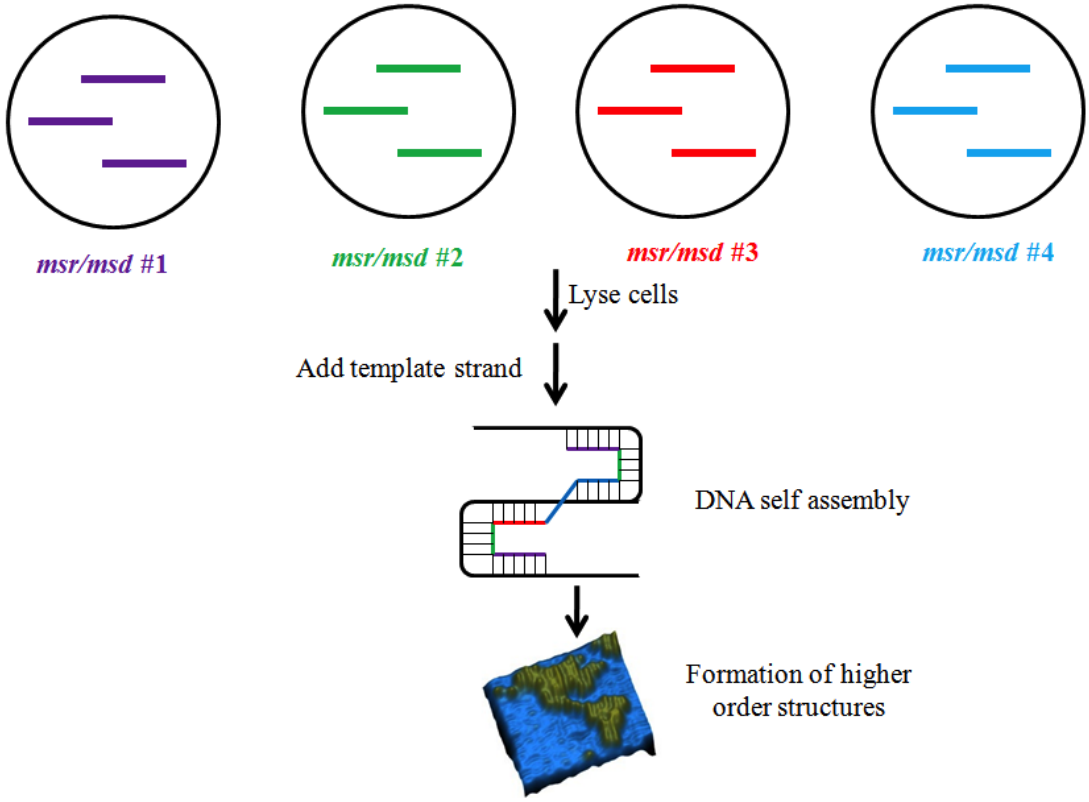


Figure 5. *In vivo* method for DNA origami production. In the *in vivo* approach for producing higher order DNA structures, *E. coli* cells will produce distinct *msr/msd* regions. The *E. coli* cells will be lysed and a template strand will be added. The ssDNA pieces will "staple" the template strand into a desired conformation based on the ability of DNA molecules to adhere to each other via complementary base pairing. The world map figure was taken from http://www.cbc.ca/news/health/story/2006/03/15/dna-origami060315.html.

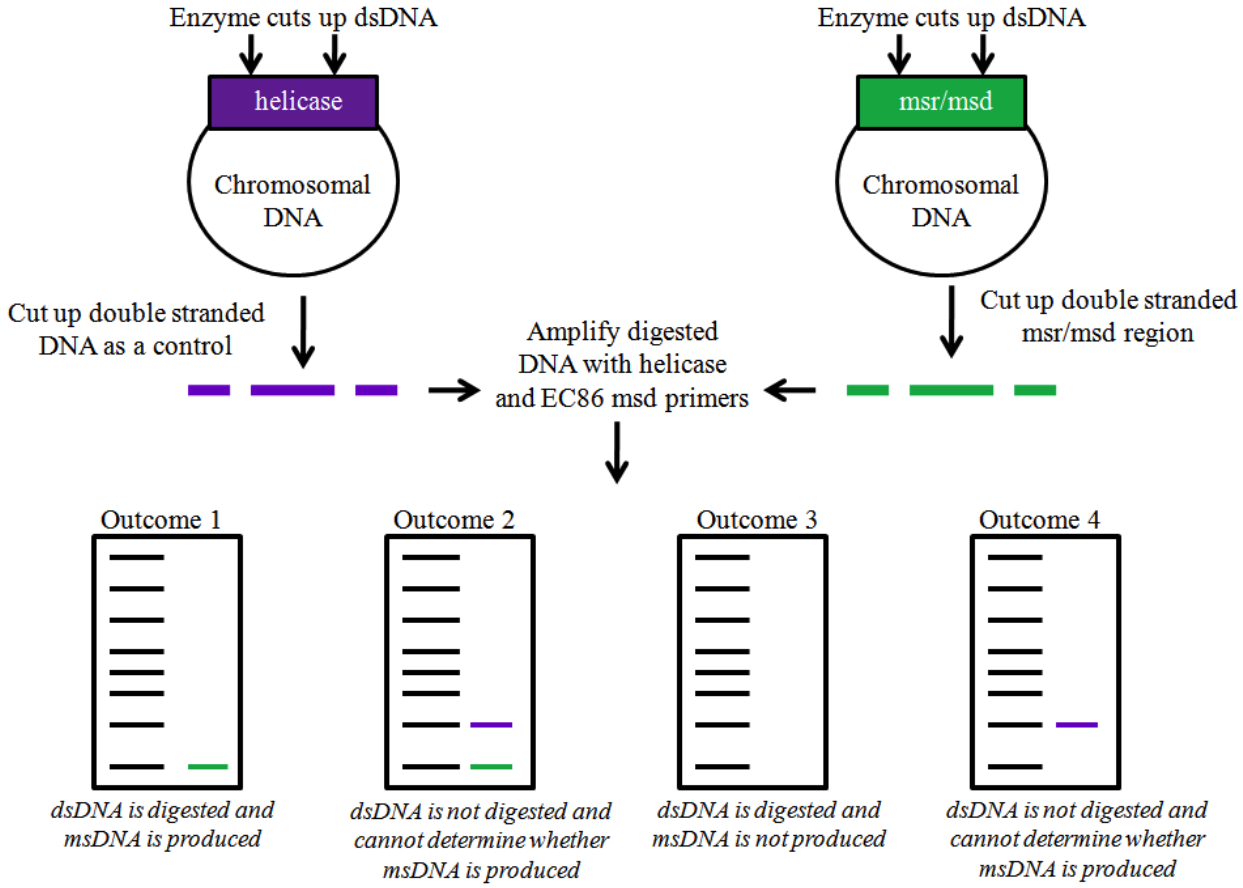


Figure 6. Rationale for msDNA detection method. This figure describes the rationale for the msDNA detection method. The black lines in each outcome box represent the molecular weight marker. If dsDNA is digested before PCR amplification, then no helicase band product should appear on the gel. The presence of an msDNA band in the absence of a helicase band implicates the formation of msDNA. The helicase serves as a negative control to make sure that dsDNA has been digested.

dsDNase Digestions*:* Shrimp recombinant dsDNase from USB (product # 78314) was used to determine if msDNA was being produced. According to this company's product literature, this enzyme is specific for dsDNA and does not cleave ssDNA. The following flow chart describes the overall procedure for using dsDNase.

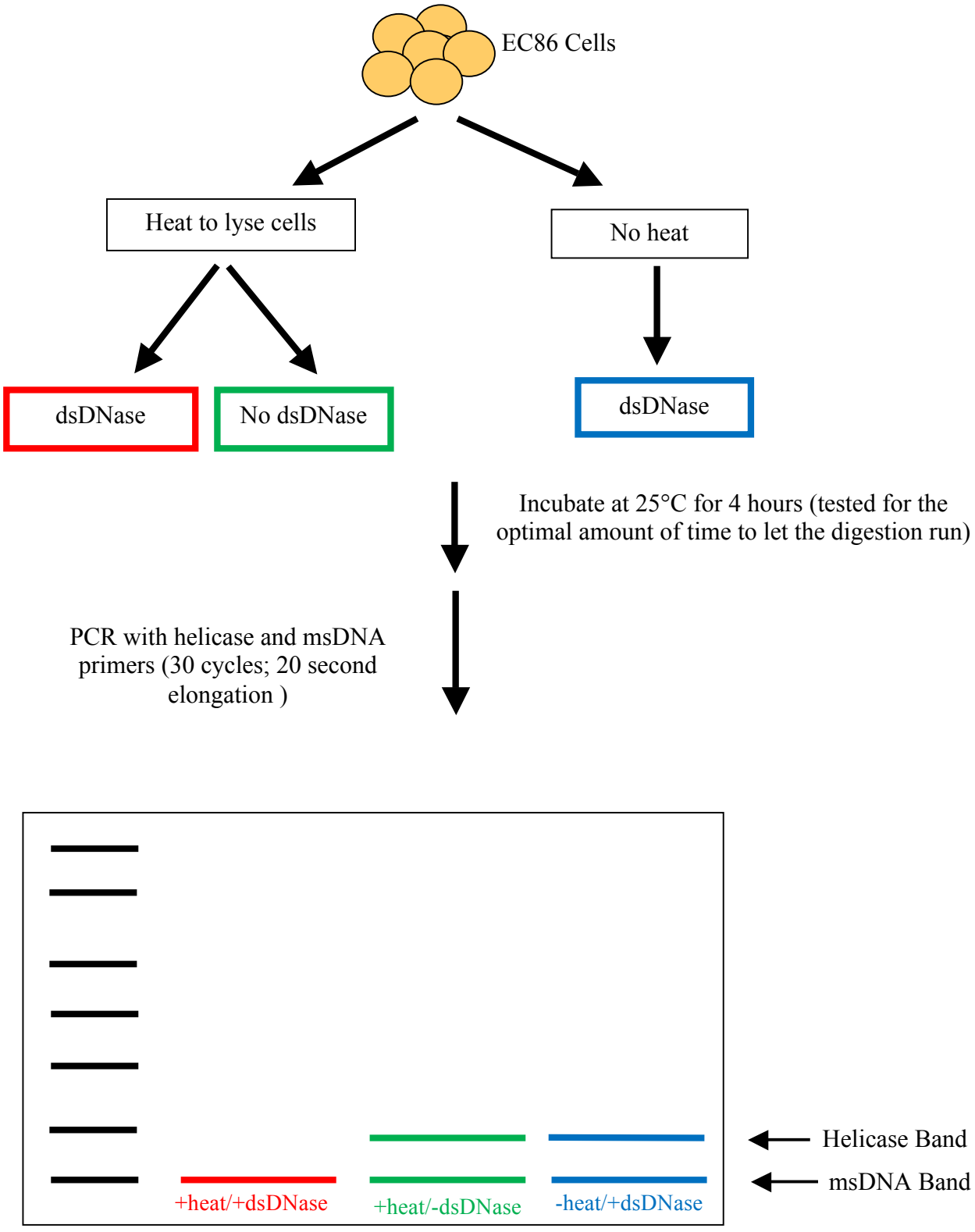


Figure 7. Protocol for dsDNase digestions. dsDNase digestions were performed to develop a novel protocol to detect msDNA. This protocol involved two controls. The first control (green) was cells that were lysed by heat, but not exposed to dsDNase, which served as a negative

control to ensure that dsDNase is the only enzyme cutting the chromosomal DNA. The second control (blue)was cells that were not lysed, but were exposed to dsDNase, which made sure that heat lysed the cells and that no external dsDNA PCR template was present. For all digestions, the enzyme was added to whole cell extracts. The gel at the bottom of the figure depicts an ideal banding pattern for this experiment. The bands of the control treatments (blue and green) should be of comparable intensity (helicase to helicase and msDNA to msDNA), because in both of these controls, the DNA should not be digested.

HaeIII Digestions*:* HaeIII, a traditional four-cutter restriction enzyme, was also used to develop a more efficient protocol to detect msDNA. The portion of the helicase gene that is amplified contains two HaeIII restriction sites and the *msd* portion of the EC86 retron contains no HaeIII restriction sites (Figure 8). HaeIII was used to develop a protocol that successfully destroyed all dsDNA, because this enzyme is less expensive than CviKI-1 (six-cutter), which has restriction sites in both the helicase gene and the EC86 *msd* region. The protocol for HaeIII mediated msDNA detection is detailed in Figure 10.

Helicase Gene

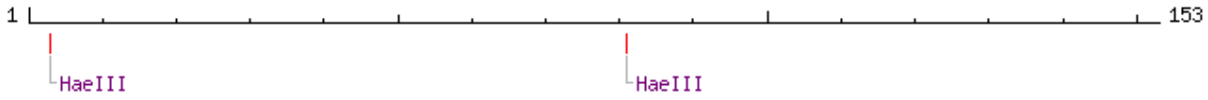
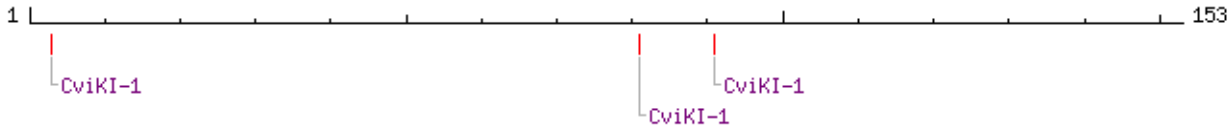


Figure 8. Helicase gene digested with HaeIII. Location of HaeIII restriction sites in the portion of the helicase gene that was amplified by the helicase primers. The location of the HaeIII sites was determined by using NEB cutter v2.0. There were no restriction sites in the *msd* region of *E. coli* B genomic DNA.

CviKI-1 Digestions*:* CviKI-1 was used to detect msDNA production in EC86 cells. CviKI-1 has three sites in the helicase gene and one restriction site in the *msd* portion of the *E. coli* B genome (Figure 9). The same protocol that was followed for the HaeIII digestions was used for this enzyme (Figure 10). The only difference between the HaeIII protocol and the CviKI-1 protocol is the CviKI-1 heat inactivation step.

Helicase Gene



msd Portion of EC86 retron

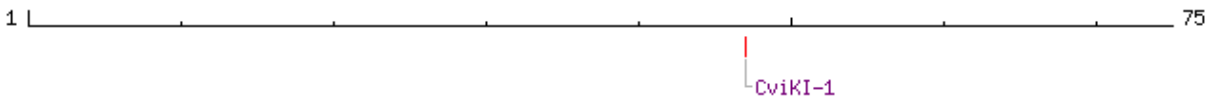


Figure 9. Helicase and *msd* region of the EC86 genome digested with CviKI-1. Location of CviKI-1 digestion sites in the helicase gene and in the msd portion of the EC86 retron. The first

restriction site in the helicase gene is where the forward helicase primer binds to amplify the gene, but is close enough to the beginning of this sequence that the primer should still be able to bind and amplify the region. The location of the CviKI-1 sites was determined by using NEB cutter v2.0.

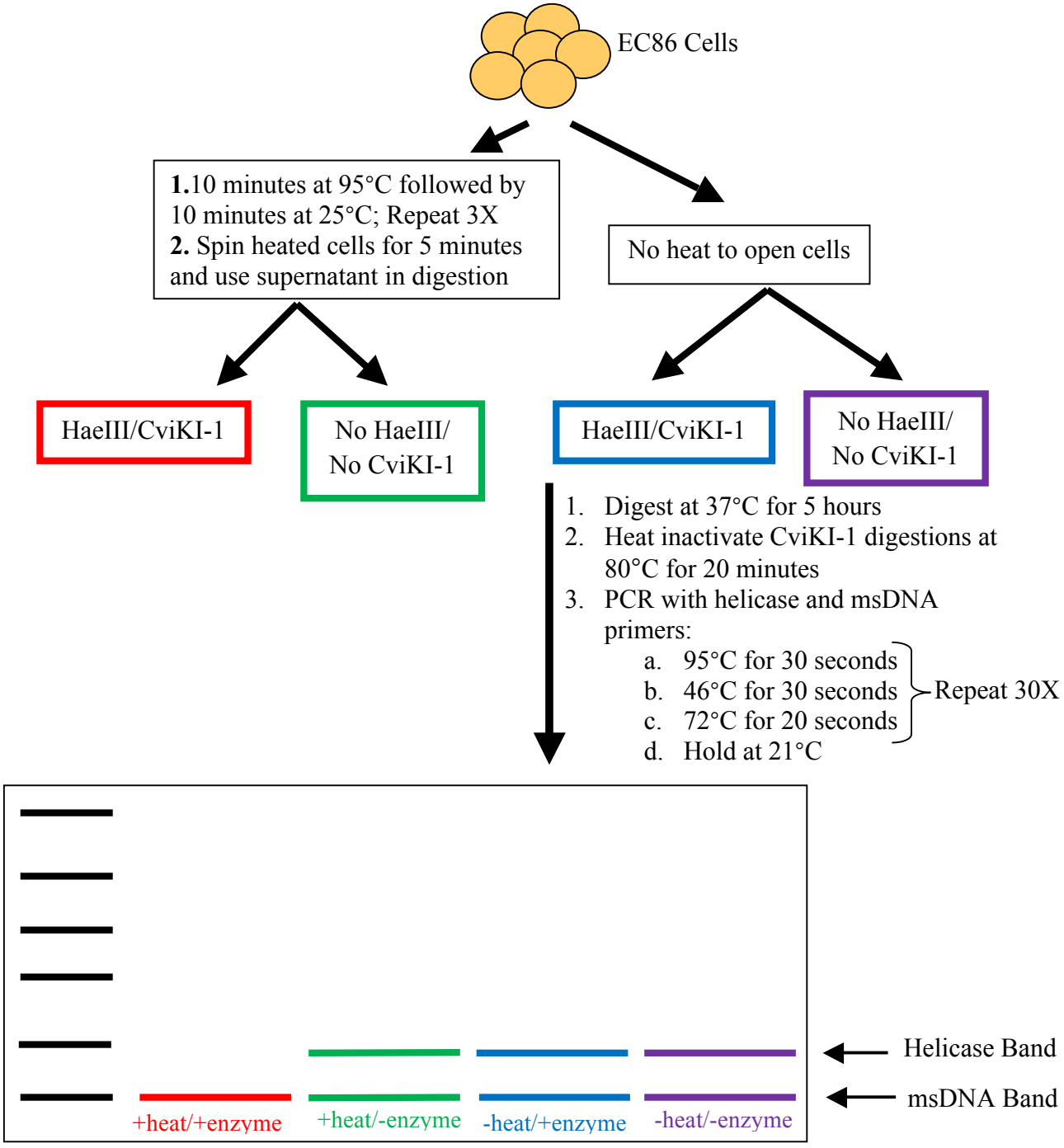


Figure 10. Protocol for HaeIII and CviKI-1 digestions. This protocol is very similar to the dsDNase protocols (Figure 7) with a few exceptions. Cells were opened by cycling them through a heat cool protocol in the heat block. Digestions were performed at 37°C. Heated cells (1µl of cells diluted 1:1)were spun and the supernatant was used in the digestion (1µl per

digestion).Whole cell extracts were used in digestions where the cells were not heated. Another control was added where the cells were not heated and no enzyme was added to determine the amount of PCR product amplified in the absence of heat (which could cause DNA degradation) and enzyme (which could digest the DNA during PCR). The green control determined the intensity of the PCR band in the absence of enzyme. The blue control demonstrated that cells were not being opened in the absence of heat. The purple control illustrated the banding intensity in the absence of enzyme and heat. The bands of the control treatments (blue, green and purple) should be of comparable intensity (helicase to helicase and msDNA to msDNA), because in all of these controls, the DNA should not be digested.

Results

Part 1: Building the msDNA expression cassette

Part 1A: Cloning Reverse Transcriptase

In order to overcome the difficulties Waters (2010) faced cloning reverse transcriptase I attempted to clone this protein into both pSB4A3 (a low copy plasmid) and pSB1A7 (a high copy, insulated vector; Haynes, 2008) using traditional BioBrick cloning methods. A low copy plasmid decreases the number of plasmids in a cell and, thus, will decrease the amount of RT in the cell. The insulated vector has transcriptional terminators on the outside of both BioBrick ends and prevents read through transcription, which reduces the number of RT molecules present in the cell. Attempts to clone the RT sequence into both of these vectors were unsuccessful (Figure 11). All pSB1A7 ligations failed to produce any positive results in the colony PCR reaction.pSB1A7 is a high copy plasmid and the failure of any cells in the pSB1A7 ligation plate to produce a positive result in the colony PCR suggests that cells are not able to handle an excess of RT. The pSB4A3 ligations produced some positive results in the colony PCR reaction, but digestion confirmation of these PCR produces shows an insert of about 200bp which was much smaller than the expected 1119bp insert (the desired product length for pBad-RBS-RT; Figure 11). I performed these ligations multiple times and these experiments yielded similar results every time. Therefore, ligating EC86 RT into either pSB1A7 or pSB4A3 was unsuccessful,

which suggests that RT is not clonable downstream of a weak promoter (pBad) into a low copy or insulated vector.

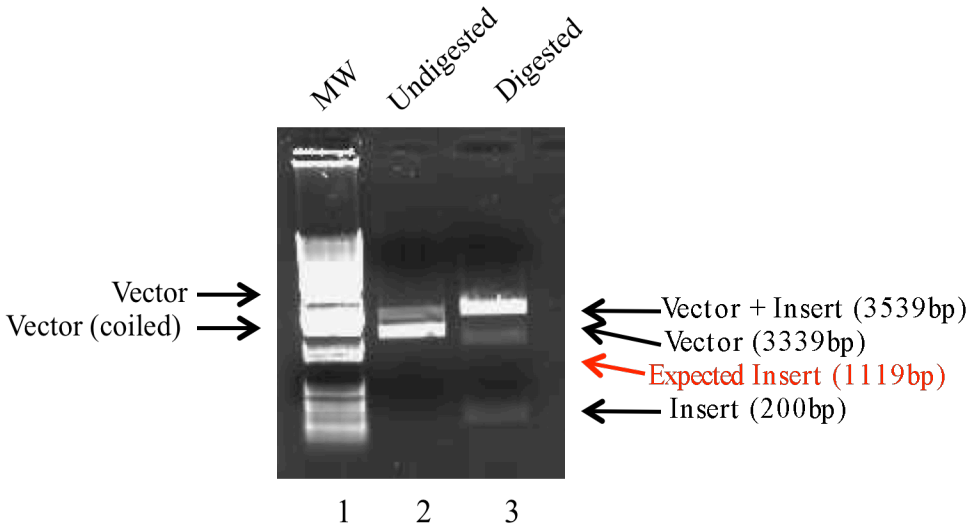


Figure 11. Digestion of pBad-RBS-EC86RT in pSB4A3 confirms that BioBrick cloning did not clone EC86 RT. This gel shows a clone that gave a positive result in the colony PCR reaction digested with EcoRI and PstI to confirm if the insert was the correct size. MW: molecular weight marker; lane 2: clone 2 from colony PCR that gave a positive result not digested with EcoRI and PstI; lane 3: clone 2 digested with EcoRI and PstI.

Traditional cloning methods for ligating RT downstream of a promoter-RBS construct were unsuccessful, so I devised an alternative method for accomplishing this goal. I used PCR to build pBad-RBS upstream of reverse transcriptase and clone the final PCR product into a vector (Figure 4). I performed three successive rounds of PCR with each round adding more of the pBad-RBS construct in front of the RT. Three rounds of PCR were necessary because the pBad-RBS part is too large to include in one primer. A primer that large would have required a very long elongation time, which would have increased the possibility for point mutations to be introduced into the sequence. The gels after each round of PCR show that the product of each reaction was the correct size (Figure 12).

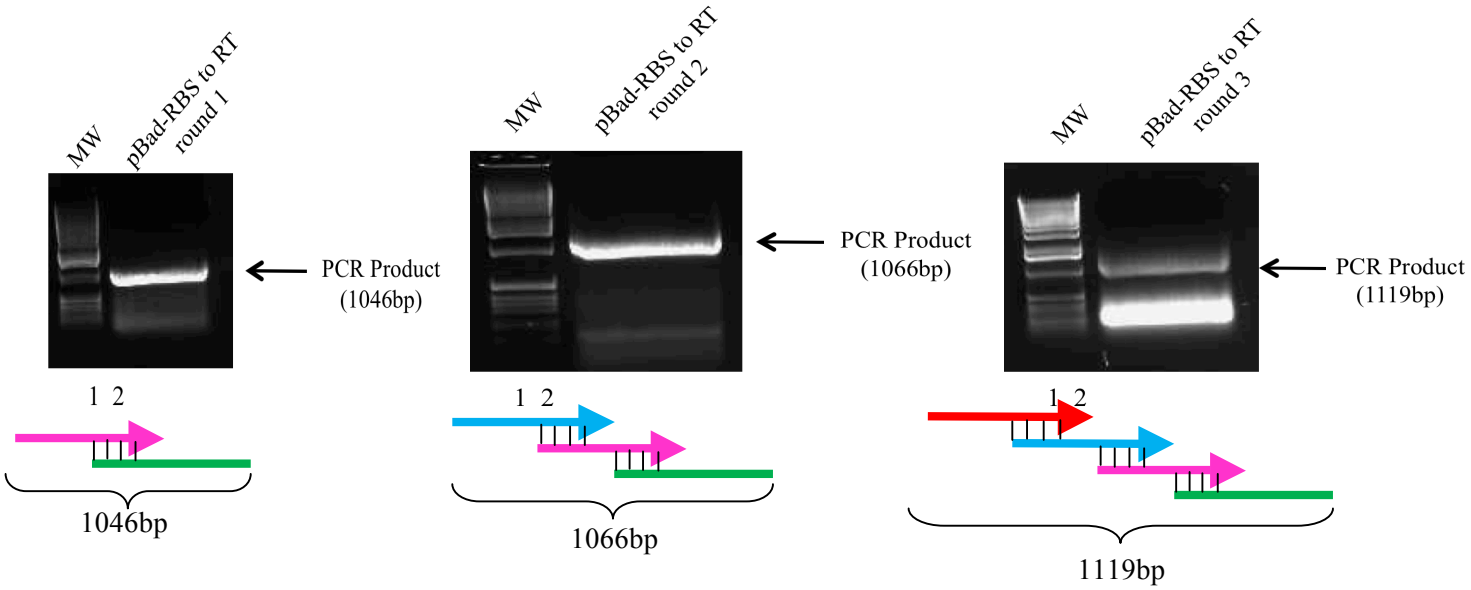


Figure 12. Three rounds of PCR to build pBad-RBS-RT reveal appropriately sized PCR products after each round. These three gels show the PCR products after each round of PCR for building pBad-RBS-RT. Each of the PCR products are between 1000-1119bp long. The diagrams below each gel depict the successive addition of pBad-RBS to EC86RT. Green: EC86RT; pink: forward primer #1; blue: forward primer #2; red: forward primer #3. The numbers below each diagram indicate the length the PCR product should be after that round of PCR.

I ligated the third round PCR product into pSB4A3 (Figure 13). I selected two controls for this experiment, because the registry part I took the pSB4A3 vector from had a green reporter gene (GFP) at the end (part# BBa_K199069). Having a green colony control in addition to a white colony control (the white colony control just shows a vector that has closed on itself with no insert)allowed me to differentiate between an insert that was the proper size for previous insert BBa_K199069 (1620bp)versus the desired insert (1119bp). The insert size for the colony of interest (colony 23) differed in size from the insert in the green colony control and was 1119bp, the correct size for the expected product (lanes 2 and 4; Figure 13).

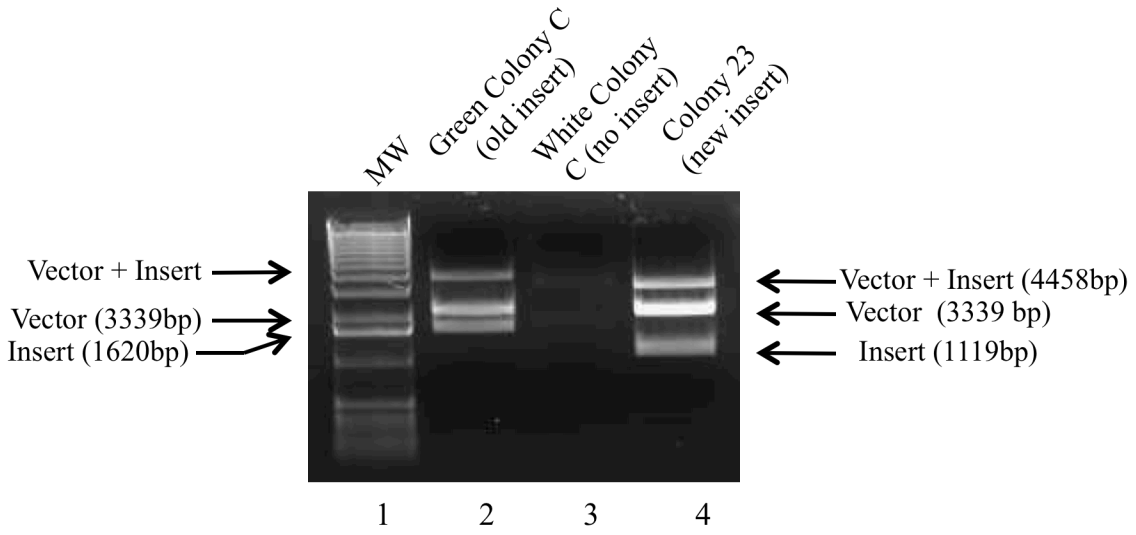


Figure 13. Digestion confirmation of pBad-RBS-EC86RT in pSB4A3 indicates a successful ligation. This gel shows both the controls and colony 23 digested with EcoRI and PstI after cloning pBad-RBS-EC86RT into pSB4A3. MW: molecular weight marker; lane 2: a green colony from the control plate; lane 3: a white colony from the control plate; lane 4: colony on ligation plate that gave a positive result in colony PCR.

Although I was able to size verify the insert, sequence verification revealed that the pBad-RBS-EC86RT sequence had several point mutations. Two of these mutations were in the RT sequence and one of the mutations was in the ribosomal binding site (RBS). The first mutation in the RT region was 3Asn→Ile and the second mutation was 107Thr→Ala meaning that both the mutations in the RT region resulted in amino acids that have different chemical properties. Asparagine and threnonine are both hydrophilic whereas isoleucine and alanine are hydrophobic. The change in chemical properties associated with these mutations could result in a protein has a different three dimensional structure, which could affect the ability of the protein to function properly.

In addition to mutations in the reverse transcriptase sequence, the insert I cloned into pSB4A3 had a mutation in the RBS. A mutation in this region is detrimental because this site is necessary for an mRNA transcript to bind to a ribosome for translation to occur. Thus, the mRNA may not have bound to the ribosome, which suggests that RT might not have been

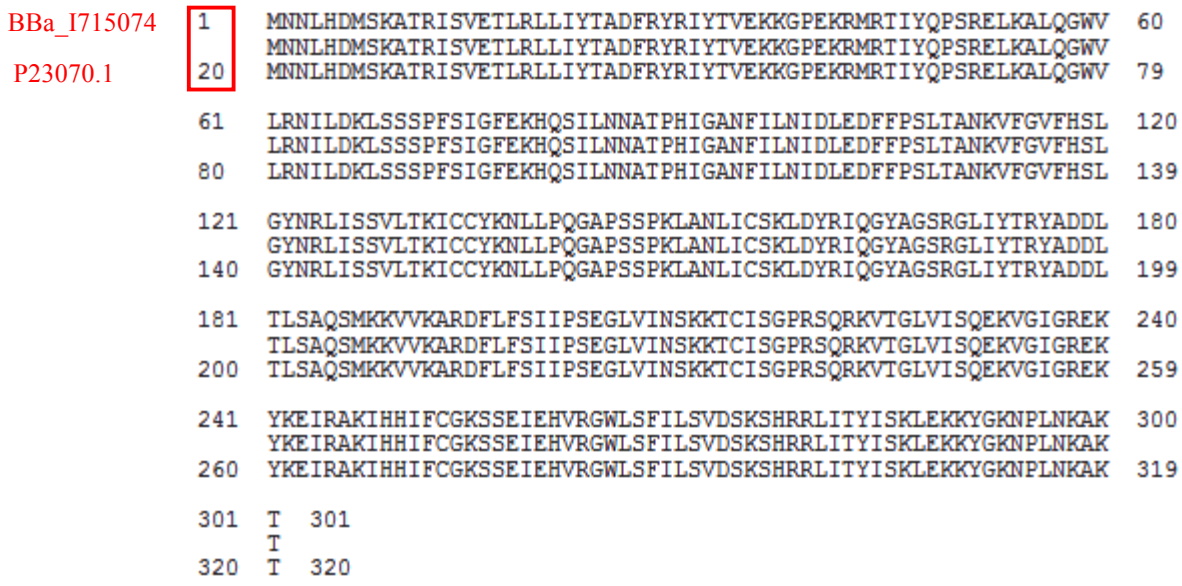
translated. The absence of a functional RT production cassette may have been the reason why the pBad-RBS-EC86RT was incorporated into the cell, because it is possible that multiple copies of RT are toxic to a cell. The mutation in the RBS region of the EC86 RT expression cassette (in addition to the mutations within the reverse transcriptase sequence itself) for the only successful clone in the pSB4A3 ligation suggests that RT is not clonable.

In analyzing the sequencing data from the positive clones, I noticed another discrepancy in the registry version of EC86 RT (part# BBa_I715074). BBa_I715074 was missing nineteen amino acids from the N-terminus of the protein (Figure 14A). Absence of these nineteen amino acids would not allow RT to fold into its three dimensional conformation. Thus, in addition to the mutations that resulted from the PCR, the absence of these nineteen amino acids in EC86 RT detrimentally affects the ability for our modular testing system (Figure 3) to produce msDNA. Although this finding does provide important information regarding the testing construct, the absence of these nineteen amino acids likely does not have a significant effect on the clonability of EC86 RT into a vector. Both Waters (2010) and I attempted to clone this protein multiple times with traditional cloning methods and were unable to clone part BBa_I715074 downstream of a promoter and RBS. It is doubtful that addition of nineteen amino acids will allow the protein to clone downstream of a promoter and RBS. Thus, I conclude that EC86 RT is not clonable and will use the naturally occurring EC86 RT in my testing system.

The absence of nineteen amino acids at the 5' end of BBa_I715074 affects the location of mutations I found in the clone pBad-RBS-RT construct. The mutation at amino acid three is located at position twenty-two and the mutation at amino acid one-hundred and nine is located at position one-hundred and twenty eight in the native EC86 RT. The 22Asn→Ile is located outside the seven highly conserved domains in reverse transcriptases and the 128 Thr→Ala mutation is located in the third highly conserved domain (Figure 14B). The occurrence of one of the two

mutations in a highly conserved region of RT further bolsters the idea that, had RT been translated, it would not have been functional.

A.



B.

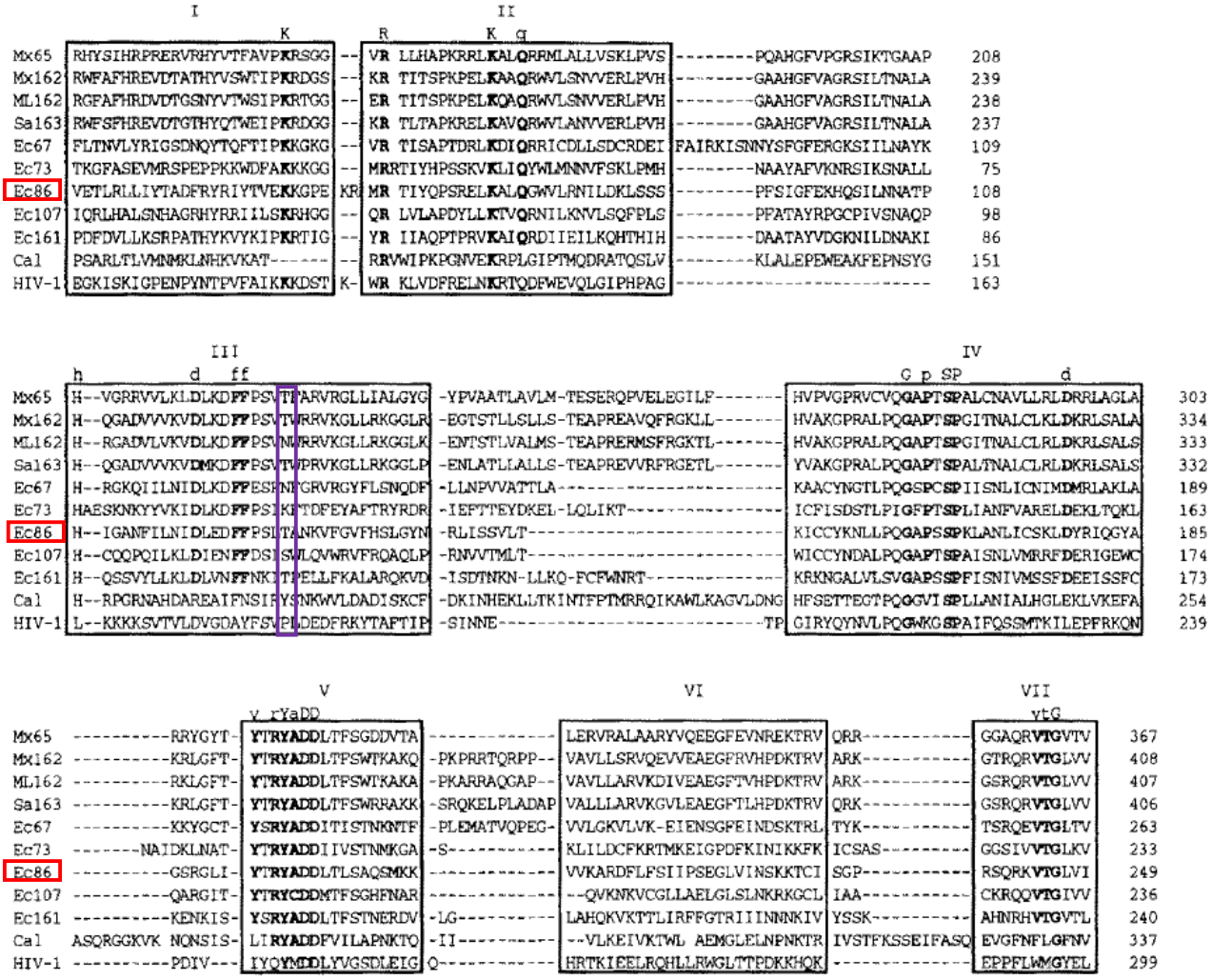


Figure 14. Comparison of registry reverse transcriptase (part #BBa_I715074) to published RT sequences reveals that part BBa_I715074 is missing 5' amino acids. A. The amino acid sequence for part BBa_ I715074 (EC86 RT) translated using Expasy's Translate tool aligned with EC86 RT from the NCBI database (accession #P23070.1). The sequence from NCBI aligns with the registry sequence starting at amino acid 20, which indicates that the first nineteen amino acids are missing in part BBa_I715074. B. Alignment taken from Rice and Lampson (1996). The purple box delineates the location of the 128Thr→Ala mutation. This mutation occurs inside one of the seven highly conserved regions within reverse transcriptases. The 22Asn→Ile mutation occurs before the first highly conserved domain in reverse transcriptases and is, thus, not located in a highly conserved region of the reverse transcriptase.

Part 1B: Cloning msr/msd regions downstream of a promoter

Although cloning reverse transcriptase was not successful, I built the 5' portion of the testing construct (Figure 3) to use in conjunction with the naturally occurring RT. The 5' portion of the cassette was originally built by Mike Waters, but the glycerol stocks failed to culture bacteria, so I had to rebuild these parts. I was able to successfully ligate pLac upstream of the EC83 *msr/msd* (Figure 15). The gel shows an insert that is approximately 365bp, which is the correct size for part# BBa_I715075 (pLac-EC83 *msr/msd*). This clone has been sequence verified, which means that the sequence contains no point mutations or deletions. I also ligated pLac upstream of the EC86*msr/msd*(part# BBa_I715076)region and sequence verified two colonies from this ligation (Figure 16). Colony three had multiple gaps and mutations in its sequence. Colony nine had no mutations or gaps except at position 402. This mutation (C→G) is located in the *msd* region of the construct and is not at a position that is involved in secondary structure formation (Figure 17). The mutation changed a cytosine to a guanine and, because this mutation occurs across from an adenine residue in the predicted 2D structure, there is no possibility for an extra complementary binding event to occur in this region. Because of the location of this mutation, I decided that this construct could still be utilized in manipulating msDNA production *in vivo.* Thus, I was able to successfully build the front half of both msDNA production constructs (Figure 3).

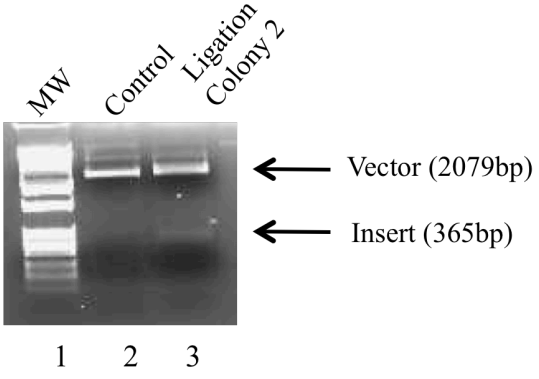


Figure 15. Digestion confirmation of pLac-EC83 *msr/msd* in pSB1A2 indicates a successful ligation. This figure shows the digestion confirmation after ligating pLac-EC83 *msr/msd* into pSB1A2. MW: molecular weight marker; lane 2: negative control (no insert included in ligation) digested with EcoRI and PstI; lane 3: colony 2 digested with EcoRI and PstI.

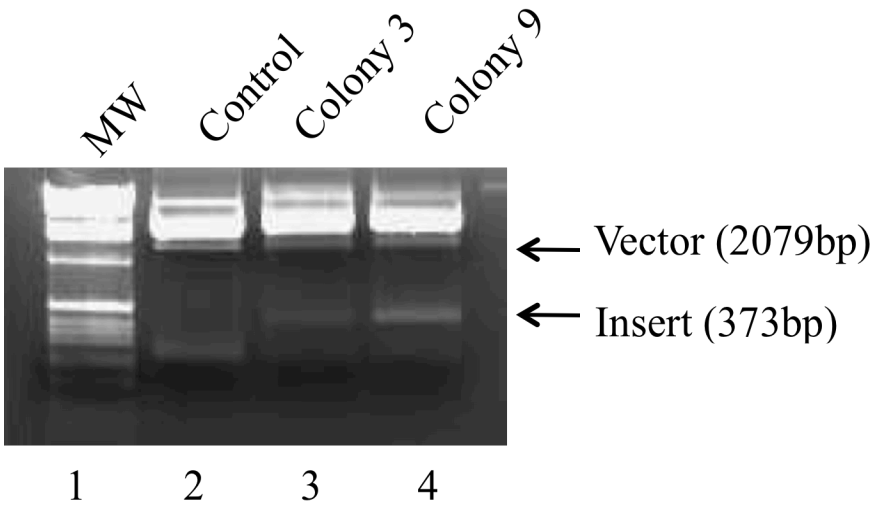


Figure 16. Digestion confirmation of pLac-EC86 *msr/msd* in pSB1A2 indicates a successful ligation. This gel shows the digestion confirmation after ligating pLac-EC86 *msr/msd* into pSB1A2.MW: molecular weight maker; Control: negative control (no insert included in ligation; the 200bp band in this lane is the pLac promoter with no *msr/msd* region downstream of it) digested with EcoRI and PstI; Colony 3 and Colony 9: ligation clones digested with EcoRI and PstI.

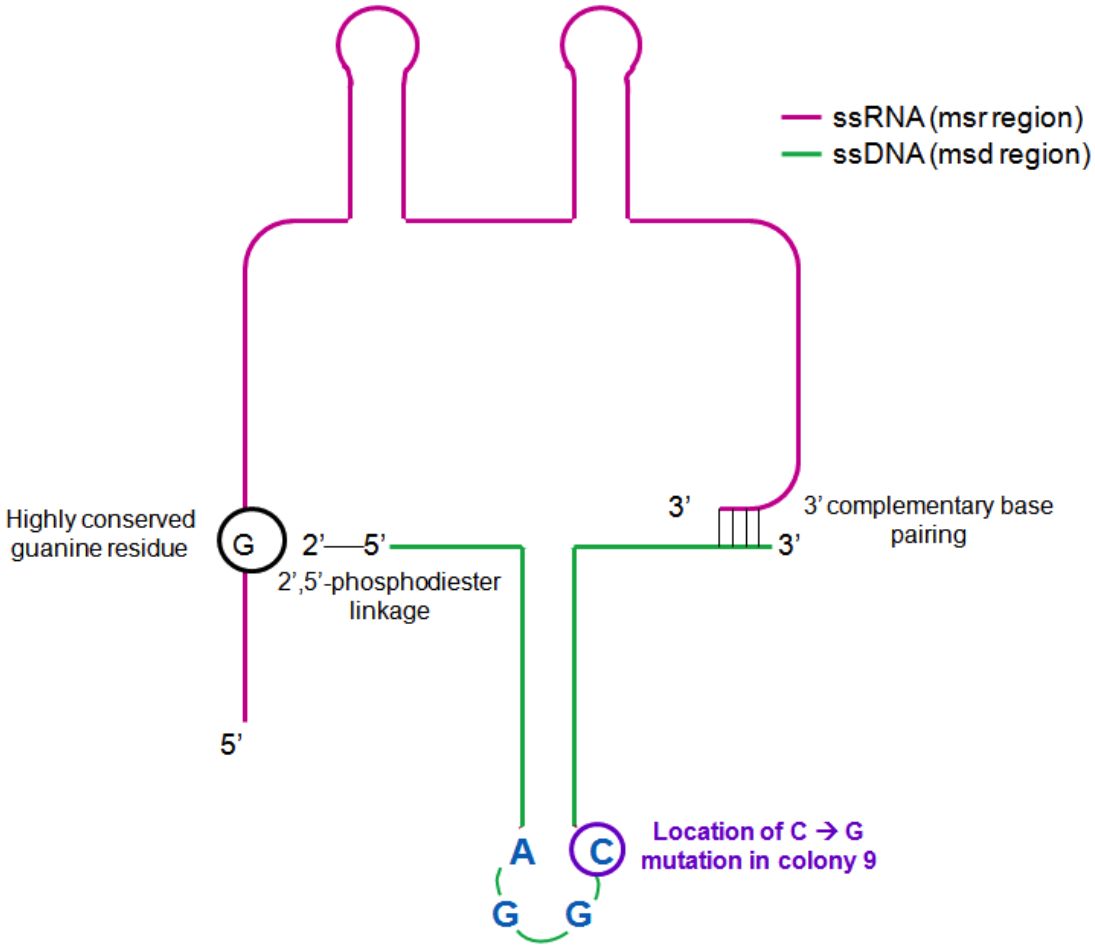


Figure 17. Location of mutation in pLac-EC86*msr/msd*. This figure shows the location of the one base mutation in colony nine of the pLac-EC86 *msr/msd* ligation. The purple circle indicates the location of the mutation (C→G) in colony nine. The adenine residue across from the mutated base is where an extra binding event could have occurred if the base mutated to a thymine instead of cytosine.

Part 1C: Investigating the Importance of the Spacer Region between msr/msd and Reverse Transcriptase

I investigated whether or not the DNA region in between the *msr/msd* region and the RT coding region in retrons of different bacteria was a highly conserved region. The spacer region is present in all retrons, which suggests that this region could have an important regulatory function in the retron. The significance of this spacer region, however, has never been investigated. If I found that the spacer region was highly conserved across different species of bacteria, then I would include it in my testing construct. To determine the importance of the spacer region, I

aligned spacer regions from several *E.coli* retrons as well as sequences from retrons in other species of msDNA-producing bacteria. I found that, not only was the length of the spacer region highly variable between species, but that sequences were also varied considerably (Figure 18A). A separate comparison of the spacer regions in *E. coli* retrons alone also revealed little sequence conservation (Figure 18B). These results suggest that although the spacer region is present all retrons, the region itself is not highly conserved. The absence of a highly conserved region in this area suggested that this region may not be important for proper retron function and was not incorporated into my design of the msDNA production construct.

A.

```
EC86        ------------------------------------------------------------GTAA
EC73        ----------------------------GTAGGTTTGGCTCTTTTAGTCCTCTACCATCA
Mx65        ---------------------------------------------GTCGTCCCCTGGCC
Mx162       GTGGTGCTTTCCCGGCCTCCCTCGACTGCTCGCGCCATGTC---CCGTCT-TCCATCGCC
Sal63       GTGGTGCCTTCG-GGCCTCCCTCGACCGCTCGCGCTCCGTCGCCCTGCCC-TGCCTCGCC
ML162       GTGGTGCCTTCCTGGCCTCCCTCGACCGCTCGGATTCGTCTCCTCCGTCAGCCCCTCCCT
EC107       -------------------------------------------------CCTCTTTTGGCC
EC83        GTAAAGAGCCTGGACCTAGGGATGGACAGGGAACTTAAAGCTCCTTGTTCATCTCTTGGT


EC86        GGGTGCGCAACTTTC--------------------
EC73        AGGTGCATAAGGATATTCTCG--------------
Mx65        TCTTCCGGAGCACC---------------------
Mx162       GCGCCCGCCCAAGGTGCAGAC--------------
Sal63       CCCCCCACCTTGCTCACCGGCGCCAGGAGCCGTC-
ML162       GAGCCAGCAGCGCCCGTCGGGTAACTCAG------
EC107       GCGGTGATGTGGAGAGTGGA---------------
EC83        ACTAGTTATTGACTGGTAAACGGAACAAGCTGGTA
```

B.

```
EC86        ------------------------------------------------------------
EC73        ------------------------------------------------GTAGGTTTGGCTCT
EC107       ------------------------------------------------------------
EC83        GTAAAGAGCCTGGACCTAGGGATGGACAGGGAACTTAAAGCTCCTTGTTCATCTCTTGGT


EC86        --------------GTAAGGGTGCGCAACTTTC------
EC73        TTTAGTCCTCTACCATCAAGGTGCATAAGGATATTCTCG
EC107       --CCTCTTTTGGCCGCGGTGATGTGGAGAGTGGA-----
EC83        ACTAGTTATTGACTGGTAAACGGAACAAGCTGGTA----
```

C.

Retron	Species	Accession Number
EC86	*Escherichia coli*	M24408
EC73	*Escherichia coli*	M64113
MX65	*Myxococcus xanthus*	J03763
Mx162	*Myxococcus xanthus*	M30609
Sal63	*Stigmatella aurantica*	M14945
ML162	*Melittangium lichenicola*	L36722
EC107	*Escherichia coli*	AF170088
EC83	*Escherichia coli*	Z12832

Figure 18. ClustalW multiple alignment of spacer region between *msr/msd* and RT domains of various bacterial retrons. A. Compares eight bacterial retrons from different subspecies of *E. coli* and other types of bacteria. No highly conserved region was found between *E. coli* and these other bacterial species. B. Compares four *E. coli* retrons and no similarity is found between the different subspecies of *E. coli* that would indicate a region of high conservation. C. A table that describes the retrons, their bacterial source and the retron accession number.

Part 2: Detection of msDNA

Part 2A: Primer Validation

The next step after building a functional msDNA production cassette is to design a method to detect msDNA. The most commonly used method for detecting ssDNA is polyacrylamide gels which are difficult to handle and manipulate. Additionally, our lab was not able to detect msDNA on a polyacrylamide gel. Because our ssDNA production system requires the production of msDNA intermediates, a definitive method to ensure that msDNA is being produced is necessary. In our model, the goal is for *E. coli* to produce the desired pieces of ssDNA and then be lysed open and allow the DNA to self assemble on its own without having to purify DNA strands from a gel. Thus, our model requires a method to detect msDNA (and eventually ssDNA) production so that we can confirm the formation of msDNA from the msDNA production construct.

Before performing the msDNA detection experiments, I verified that the primers I designed for msDNA detection (helicase and EC86 msDNA primers) functioned properly in isolation (used the primers for msDNA in one reaction and the primers for helicase in another)

and in a multiplex mixture(used both sets of primers in the same PCR reaction; Figure 19). The control lane shows that in the absence of cells, no PCR product is amplified. The multiplex reaction (helicase + EC86msDNA) produces two distinct bands of the appropriate sizes (helicase:150bp; msDNA: 75bp). When compared to the multiplex reactions (lane 3), the isolation reaction (lanes 4 and 5) bands appear to be of comparable intensity (Figure 19). These results suggest that the helicase and EC86msDNA primers can be used in a multiplex reaction in the msDNA detection protocol. Performing multiplex reactions increases the efficiency of the protocol and limits the amount of error that can be introduced into the experiment, because a fewer number of tubes are required for the PCR reactions.

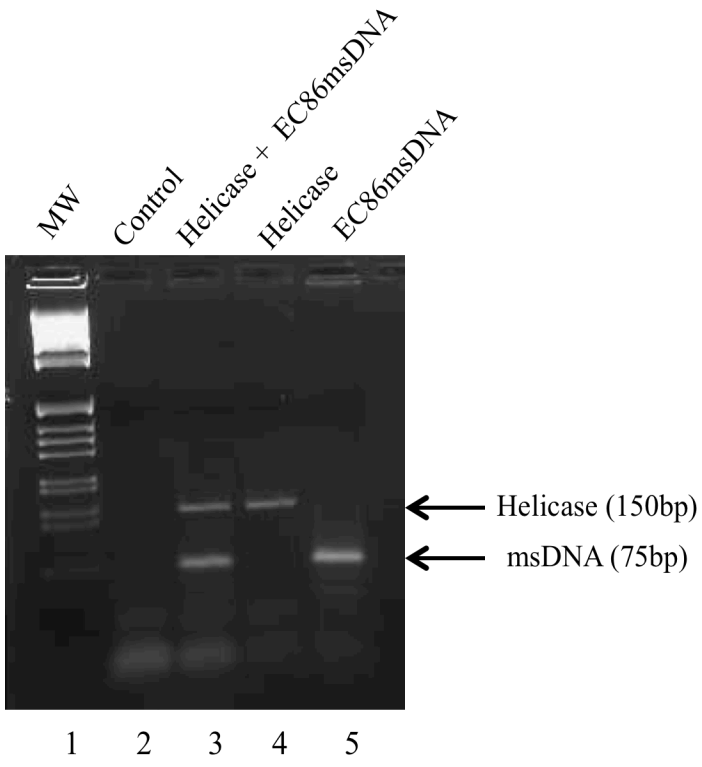


Figure 19. Helicase and EC86 msDNA primer validation indicates that multiplex PCR reactions are feasible with helicase and msDNA primers. This gel is the result of a PCR reaction that amplified the helicase and *msd* portions of the *E. coli* B genome. MW: molecular weight marker; lane 2: no cells; lane 3: both sets of primers used in the same reaction; lane 4: helicase primers alone; lane 5: msDNA primers alone.

Part 2B: dsDNase Experiments

After confirming the primers amplified the helicase and *msd* regions of the *E. coli* B genome, I used these primers in the msDNA detection protocol I developed (Figure 7). The first round of msDNA detection experiments involved the dsDNase enzyme. In the first dsDNase experiment I conducted, I determined the optimal amount of time to incubate the digestion. The DNA digestion time was important because all dsDNA needed to be digested to determine if msDNA was being produced. If any dsDNA remained intact in the helicase or *msd* regions of the genome, PCR would amplify these regions. Amplification of either of these regions would interfere with msDNA detection, because both the chromosomal EC86 *msd* region and EC86 msDNA can be amplified with the msDNA primers I designed.

I tested four different time points in the optimal digestion time experiment. I expected that the longer the digestions were incubated, a greater portion of the helicase band would disappear. The results that I saw contradicted these assumptions. Even at the longest digestion time(24 hours) the helicase band was not completely digested by dsDNase (Figure 20). Although there is decreased band intensity for the helicase band at the 24 hour time point, this decrease in intensity is also observed in the msDNA band (lane 7; Figure 20), which does not allow me to determine if msDNA is being produced. The decreased banding intensity could be due to DNA degradation that resulted from DNA being exposed to heat for such a prolonged period of time. At the 1, 5, and 12 hour digestions the msDNA band is much brighter than the helicase band, which would suggest that msDNA is being produced (lanes 4, 5, and 6; Figure 20). The helicase band does not, however, disappear completely at 1, 5, or 12 hour timepoints, which means that dsDNA has not been completely digested and does not allow me to determine if msDNA is being produced. Additionally, it would have been beneficial to perform the “no heat” and “no dsDNase” controls at each timepoint to evaluate the effect of time on digestion more effectively.

Overall, this experiment showed that increasing digestion time did not eliminate the intensity of the helicase band. I decided to use a four hour digestion for the rest of my experiments under the assumption that another portion of the protocol was causing the persistence of the helicase band.

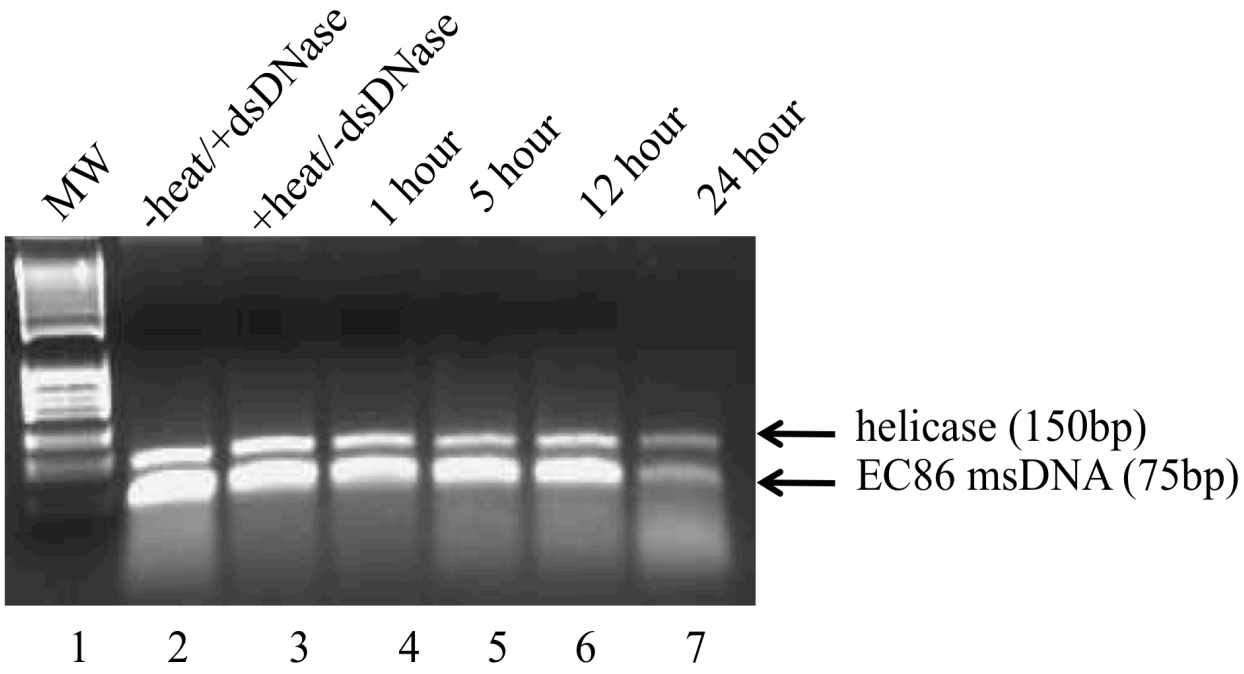


Figure 20. dsDNase digestion timing experiment does not reveal an optimal digestion incubation time. This gel shows the PCR reactions to determine the optimal amount of time to perform a dsDNase digestion. MW: molecular weight marker; lane 2: cells were not lysed open in the heat block; lane 3: no enzyme was added to these cells; lanes 4, 5, 6, and 7 indicate the amount of time the cells in these experimental groups were digested at room temperature. Cells in lanes 2 and 3 were digested for one hour.

After deciding that 4 hours was the most appropriate digestion time, I determined the optimal amount of time the cells needed to be placed in the heat block to lyse the cells. Optimizing the amount of time the cells spent in the heat block was necessary to make sure that the cells stayed in the heat long enough to open the cells so the enzyme could access DNA, but short enough that the heat did not start to degrade the DNA. For this experiment, I only used the helicase primers because I was interested in seeing the amount of time the cells needed to be exposed to heat and not the production of msDNA. I expected that the more time cells were

exposed to heat, the lighter the experimental bands (lysed cells that were digested with dsDNase) would be as compared to the control bands (lysed cells that were not digested with dsDNase). I hypothesized that the longer cells were exposed to heat, the more likely all cells would be lysed, which would decrease the possibility for cells to open during PCR and allow for amplification of a band when all the dsDNA should be digested. Contrary to my expectations, increasing the amount of time cells were exposed to heat did not consistently decrease the intensity of the helicase band. There was a larger decrease in the intensity of the band between the 30 second control and the 30 second experimental, but this pattern did not hold for all the time points (lanes 5 and 6; Figure 21). For example, the 60 second, 90 second and 120 second experimental reactions all have brighter band intensities than the 30 second experimental condition. An increase in band intensity for the 60 second, 90 second and 120 second experimental PCR reactions suggests that the decrease in band intensity observed between the 30 second control and experimental reactions could have resulted from pipetting errors, which would have influenced the number of cells in solution. Additionally, the longer the cells were heated, the more number of cells that could have opened, which would have resulted in larger volume of DNA that dsDNase needed to digest. dsDNase is only able to digest a certain amount of DNA before it loses its enzymatic activity. Thus, cells that spent a longer amount of time exposed to heat and released a larger overall volume of DNA would have a brighter experimental band that those cells that spent a shorter amount of time exposed to heat. Because this experiment did not provide conclusive results as to the optimal amount of time to place the cells in the heat block, I tried an alternative method to lyse the cells.

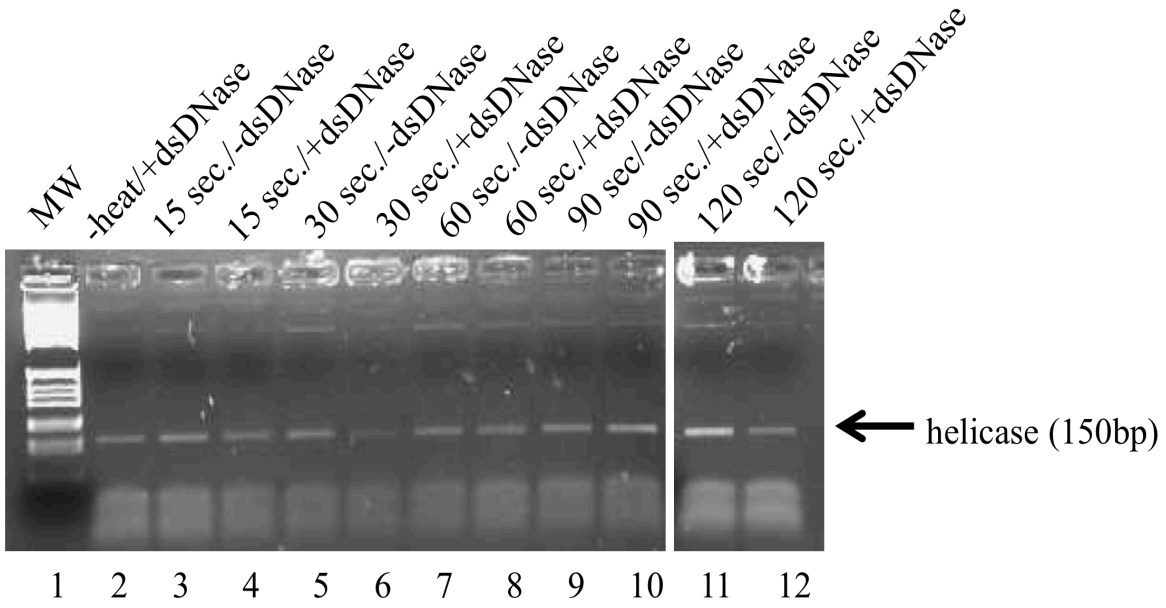


Figure 21. Increasing heat incubation does not consistently decrease the intensity of the helicase experimental band as compared to the control band. This gel shows PCR reactions that amplified the helicase gene in the optimal lyse time experiment. MW: molecular weight marker; lane 2: cells were not heated and dsDNase was added; the numbers above each lane indicate the amount of time cells were exposed to heat; -dsDNase indicates that enzyme was not added and +dsDNase indicated that enzyme was added. Only *E. coli* B helicase primers were used. All treatments were digested for four hours.

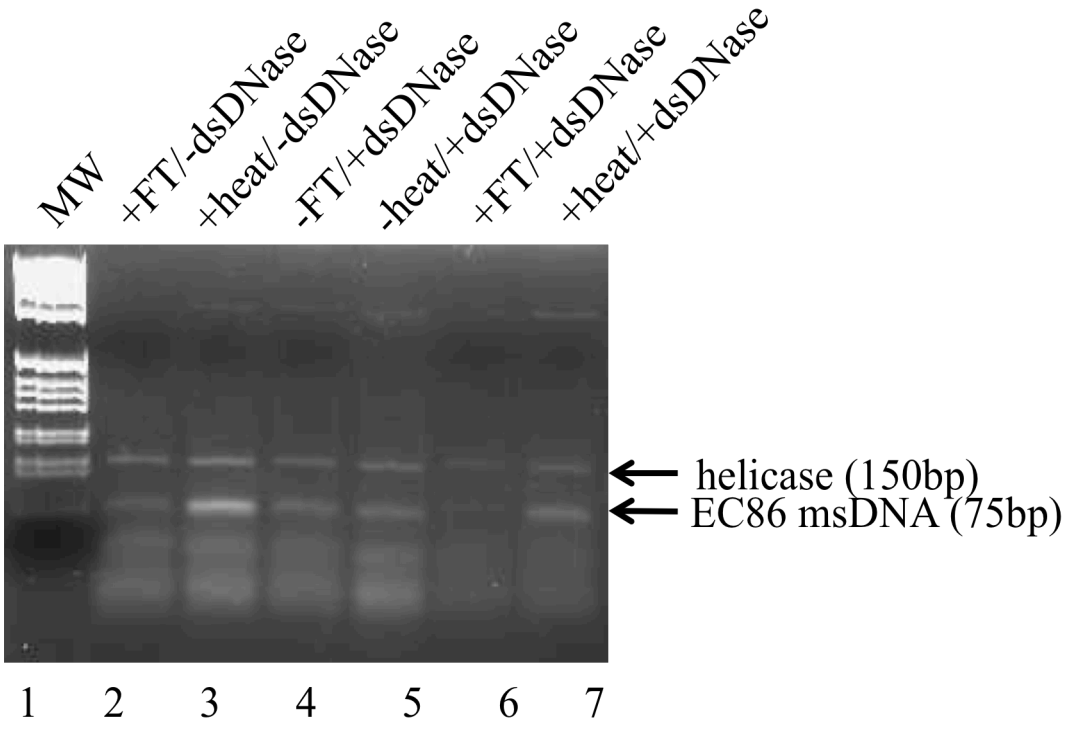
I chose to freeze thaw the cells as an alternative method for lysing the cells, because such drastic changes in temperature rupture the cell membrane. To freeze thaw the cells, I placed them in a -80°C freezer for 10 minutes and then allowed the cells to thaw completely (10-15 minutes at room temperature) and repeated this cycle three times. I expected the freeze thaw method to considerably decrease the intensity of the helicase band as compared to heat block method for lysing cells. The freeze thaw method did decrease the intensity of the helicase band as compared to the heat block method (lanes 6 and 7; Figure 22). The gel shows, however, that there is a less intense band for the msDNA primers as compared to the helicase primers in the experimental reaction for the freeze-thaw experiment (lane 6, figure 22A). The lower intensity of the msDNA band is contrary to what I expected, because the dsDNase should act exclusively on dsDNA and

leave msDNA intact to be amplified by primers. Several explanations could underlie this result: 1) msDNA is being produced, but is being digested by dsDNase or 2) this protocol does not effectively detect msDNA. The first explanation is a possibility because msDNA does have secondary structure formation which means that the dsDNase could digest this molecule. There are, however, supposed to be 200 copies of msDNA per cell and it seems unlikely that most (if not all) 200 copies would be digested before the chromosomal helicase band had been digested. The alternative explanation, however, is not feasible in this instance because the helicase band never disappears completely, which means that dsDNA is not being completely digested. If dsDNA is not completely digested, then the presence or absence of msDNA is indeterminate. Alternately, because the helicase band is also much less intense in lane 6 than in lane 7, it is possible that the freeze thaw method caused some of the DNA to degrade (Figure 22).

I used ImageJ's plot profile tool to quantify and compare the intensity of the helicase and msDNA bands (Figure 23B). The area under each peak in the plot profile is proportional to the amount of DNA present in each lane (Figure 23: this figure provides an example of the measurements and does not correlate with Figure 22). I performed this comparison in order to better visualize the difference in banding intensities between and within the different treatments. Lane three was of particular interest because the msDNA band appears to be brighter than the helicase band which could be an indication that msDNA is being produced and detected. The msDNA band appears to be approximately twice as bright as the helicase band (Figure 22B). This result could either mean that msDNA is being produced or that the EC86 *msd* region is present more than once in the *E. coli* B genome. msDNA, however, is supposed to be produced at a volume of approximately 200 copies per cell, which would suggest that the msDNA band should be much brighter than the chromosomal helicase band if msDNA is being detected, as helicase is only present once in the genome. Because the msDNA band is not considerably

brighter than the helicase band in lane three, the increased band brightness is better explained by multiple copies of the EC86 *msd* region (discussed further in results section 2E).The increased intensity in band brightness of the msDNA band as compared to the helicase band, however, is not consistent for all the controls (lanes 2, 4 and 5; Figure 22).If the msDNA detection protocol were detecting msDNA, the msDNA band should be consistently brighter than the helicase band. Thus, production of msDNA cannot be determined from this gel. The inconsistency in banding patterns between the controls led me to conclude that the freeze-thaw method was not the most effective way to lyse the cells.

A.



B.

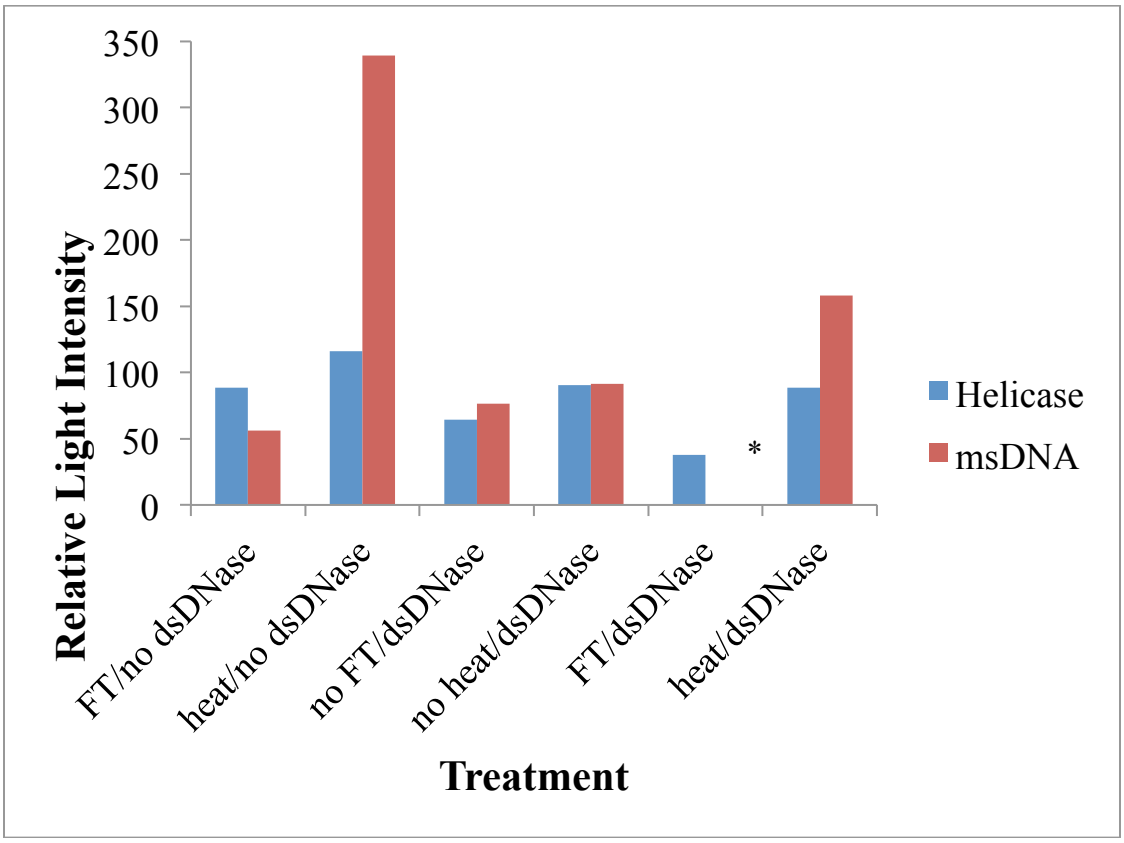
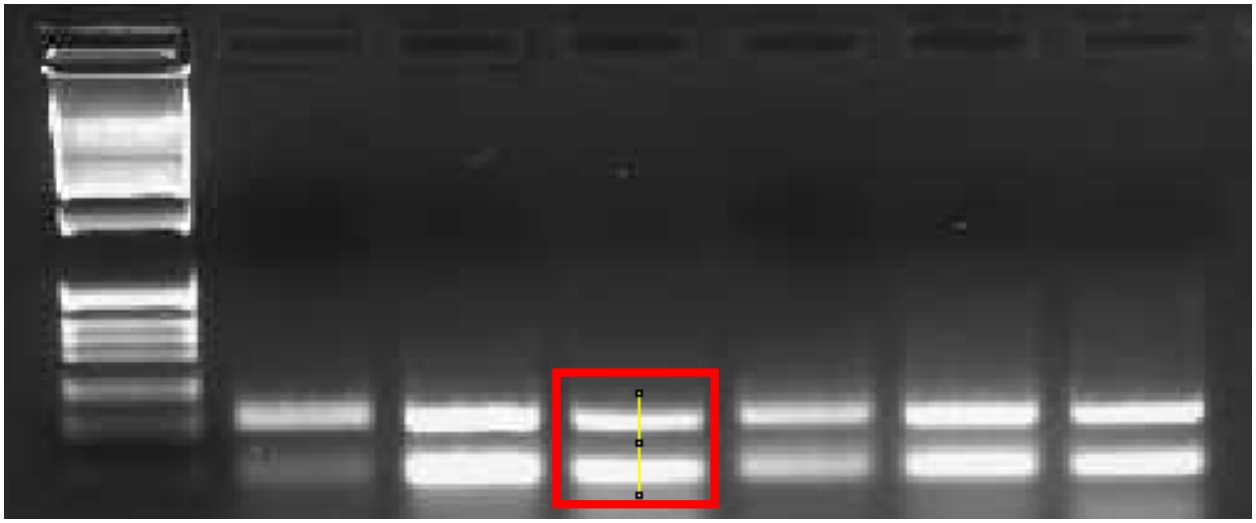


Figure 22. Freeze thaw is not a more effective method to lyse cells than heat block. A: This gel shows the results of the PCR reactions performed to test whether the heat block or freeze thaw was a more efficient method to open EC86 cells so dsDNase can digest all dsDNA. MW: molecular weight marker; FT: freeze-thaw; heat: heat block for 15 seconds. Both helicase and msDNA primers were used. B: Quantification of band intensity using ImageJ as described in Figure 23. The y-axis represents the area under the curve measured from the Plot Profile. The x-axis represents the treatment. *There was no peak for this band.

A.



B.

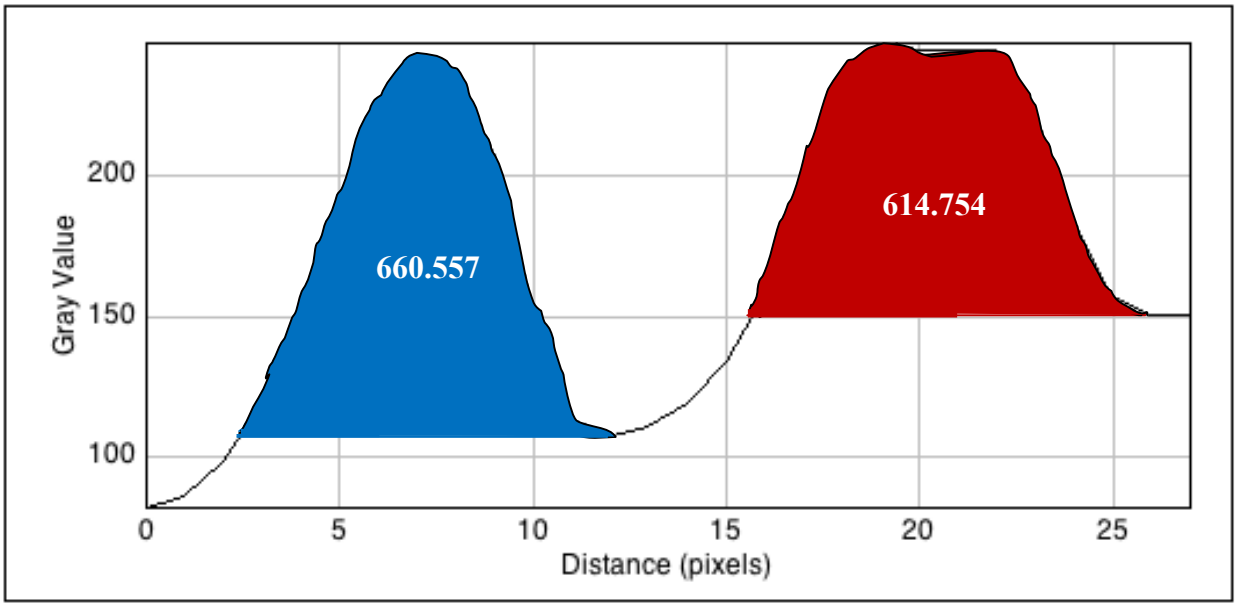
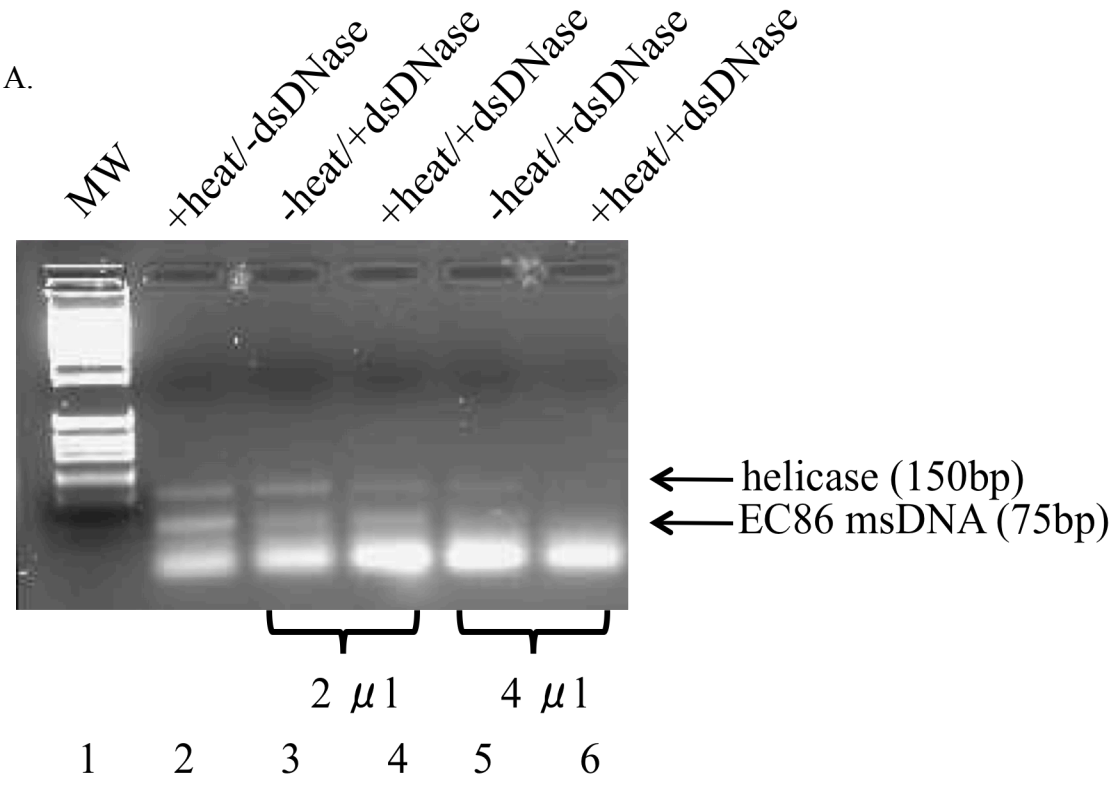


Figure 23. Method for Plot Profile Measurement in ImageJ. A. The red box indicates the bands measured. The yellow measurement indicates the line drawn to measure band intensity. B. Plot profile based on the measurement taken in part A. The first peak quantifies the helicase band and the second peak quantifies the msDNA band. I took the indicated polygon area measurements and plotted graphs based on these areas. The numbers under each peak represent the area measurements from ImageJ.

Because the freeze-thaw method for opening cells was unsuccessful, I decided to determine if different concentrations of dsDNase would better eliminate dsDNA (Figure 24). I hypothesized that increasing the amount of enzyme in the reaction mixture would digest dsDNA more efficiently in the given incubation time, because of the volume of dsDNA that needs to be digested after the cells are lysed. The gel I produced from this experiment shows that 4µl of dsDNase eliminates the helicase band more efficiently than 2µl of dsDNase (lane 4 and lane 6; Figure 24). In the 4 µl experimental reaction (lane 6; Figure 24), however, the helicase band is also absent in addition to the msDNA band. The absence of both bands suggests that msDNA is not being detected using this method because if msDNA were being detected a 75bp band would persist in the absence of the 150bp band. In lane six, however, determining the exact intensity of the msDNA band was difficult because of the brightness of the primer-dimer band (lane 6;

Figure 24).In addition, the helicase band is brighter than the msDNA band in lanes three, four and five, which also suggests that msDNA is not being detected (Figure 24B).The results of this experiment did, however, reveal that higher concentrations of dsDNase are effective for eliminating dsDNA. The absence of consistent banding patterns in the control treatments, however, suggests that increased concentration of dsDNase is not sufficient for creating an effective dsDNase mediated msDNA detection method.

A.



B.

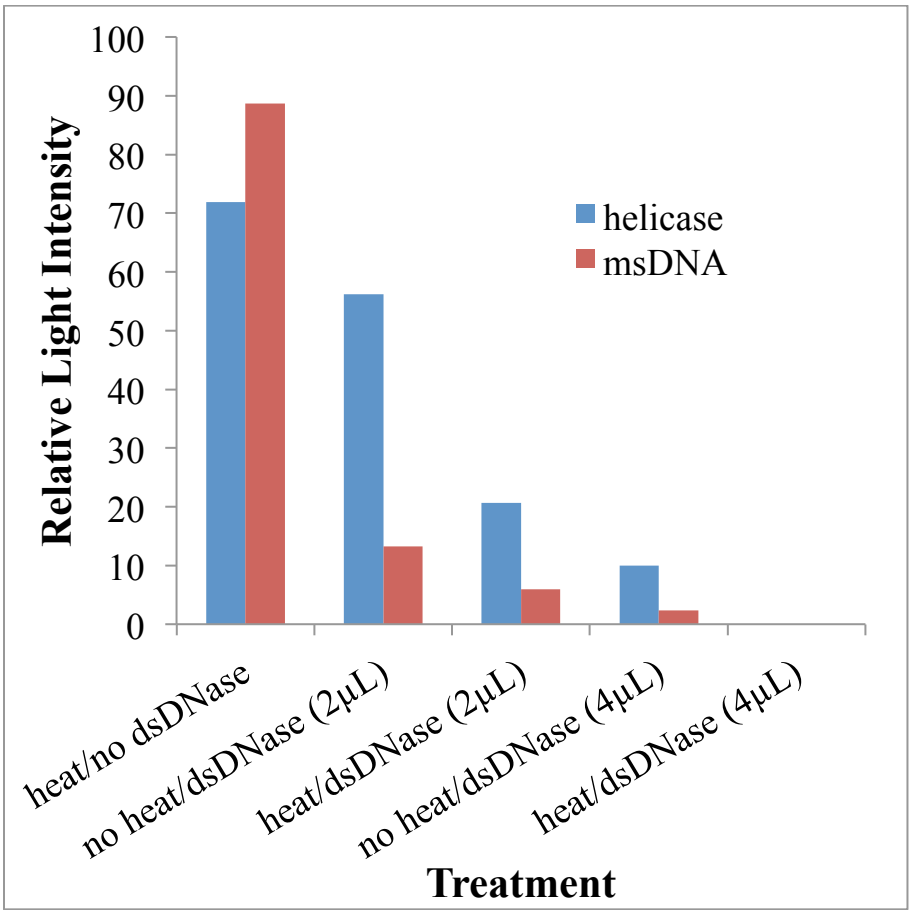


Figure 24. Increasing dsDNase concentration eliminates helicase and msDNA bands. A. This gel shows the PCR reactions from the experiment that tested different concentrations of dsDNase in the digestion reactions. MW: molecular weight marker; lanes 3 and 4 had 2µl of dsDNase added; lanes 5 and 6 had 4µl of dsDNase added. Cells were placed in the heat block for 15 seconds and digested for four hours. B. Quantification of band intensity using ImageJ. The y-axis represents the area under the curve measured from the Plot Profile. The x-axis represents the treatment.

Taken together, these results from the dsDNase experiments produced very inconsistent results, which made it difficult to determine if the protocol I developed detected msDNA production. The helicase band did not disappear in the experimental condition and the banding intensities in the control treatments were not consistent with the expected results (helicase bands should be of comparable intensity across the controls as should msDNA bands, because no DNA should be digested in any control treatments). I revised the msDNA detection protocol to utilize a more traditional restriction enzyme with a similar msDNA detection protocol.

Part 2C: HaeIII Digestions

I used the HaeIII enzyme (4-cutter) to revise the msDNA detection protocol. The HaeIII enzyme has two cut sites in the helicase gene and no cut sites in EC86 *msd* genomic region (Figure 8). I PCR amplified the helicase, EC86 *msr* and EC86 *msd* regions and then digested them with HaeIII to confirm that the *E. coli* B genome I was working with had HaeIII sites in the helicase gene but not in the *msd* and *msr* regions (Figure 25). This experiment confirmed that the HaeIII enzyme digested the helicase gene and not the *msr* or *msd* portions of the EC86 retron (Figure 25).The helicase PCR product that was digested with HaeIII was smaller than the helicase PCR product that was not digested with HaeIII (Figure 25A), whereas the msDNA and msdRNA PCR products were all the same size in the presence and absence of HaeIII (Figure 25B). I started with an enzyme that does not cut in EC86 *msd* region because the lab has had experience with the HaeIII enzyme and it was an economical alternative to the enzyme (CviKI-1) that had cut sites in both the helicase gene and the EC86*msd* region (the latter is necessary for msDNA detection). I used HaeIII to develop the revised msDNA detection protocol.

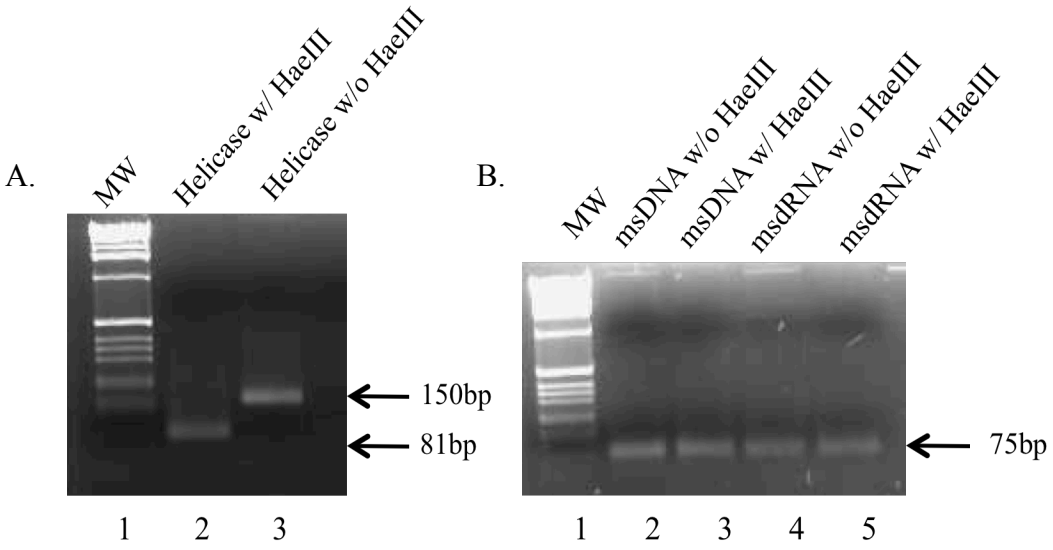


Figure 25. HaeIII digestion of helicase, EC86 *msd* and EC86 *msr* shows that HaeIII sites are present in helicase and not in EC86 *msd* or *msr*. A. This gel shows the helicase gene digested with HaeIII. In the absence of HaeIII, a band of the correct size for the PCR product (150bp) is detected on the gel. In the presence of HaeIII, the band is visibly shifted downwards, which suggests digestion of the PCR product with enzyme. B. This gel shows digestion of the EC86 msDNA and EC86 msdRNA regions with HaeIII. There are no band shifts when comparing the experimental groups with the control groups which shows that these portions of the genome do not have HaeIII sites.

The first round of HaeIII digestions had a similar protocol to the dsDNase digestions (Figure 7) with two important exceptions. Instead of placing the cells on the heat block for 15 seconds at 95°C, cells where lysed through a heat-cool cycle: 95°C for 10 minutes followed by 25°C for 10 minutes repeated three times. Additionally, the PCR protocol was modified to exclude the initial step of 94°C for 10 minutes to eliminate the possibility of lysing cells that were not originally exposed to heat and lysing more cells in the groups that were originally exposed to heat (Figure 10). These modifications did not, however, eliminate the helicase band in the experimental condition (lane 5; Figure 26). In the next round of HaeIII digestions, I took the supernatant from heated cells to use in the digestion. Taking the supernatant as opposed to whole cell extracts ensured that no new cells were being opened during the PCR reaction. If new cells were being opened in the PCR reaction, the helicase primers could have amplified an intact

helicase template in cells where the helicase gene should have been digested, which would explain the persistence of the helicase band in the experimental condition. Taking the supernatant from heated cells was an important protocol modification, because this change resulted in elimination of the helicase band in the experimental condition (lane 4; Figure 27). Disappearance of the helicase band was an important part of protocol development for a new screen to detect msDNA, because taking the supernatant from heated cells ensured that additional cells were not being opened during the PCR stage of the protocol (Figure 10; Figure 27).The results from the HaeIII digestions show that the msDNA protocol I developed effectively digests dsDNA and eliminates intact cells that confounded my earlier results (Figure 27). I used this final protocol for msDNA detection experiments using an enzyme (CviKI-1) that digests both the helicase gene and the *msd* portion of the EC86 retron.

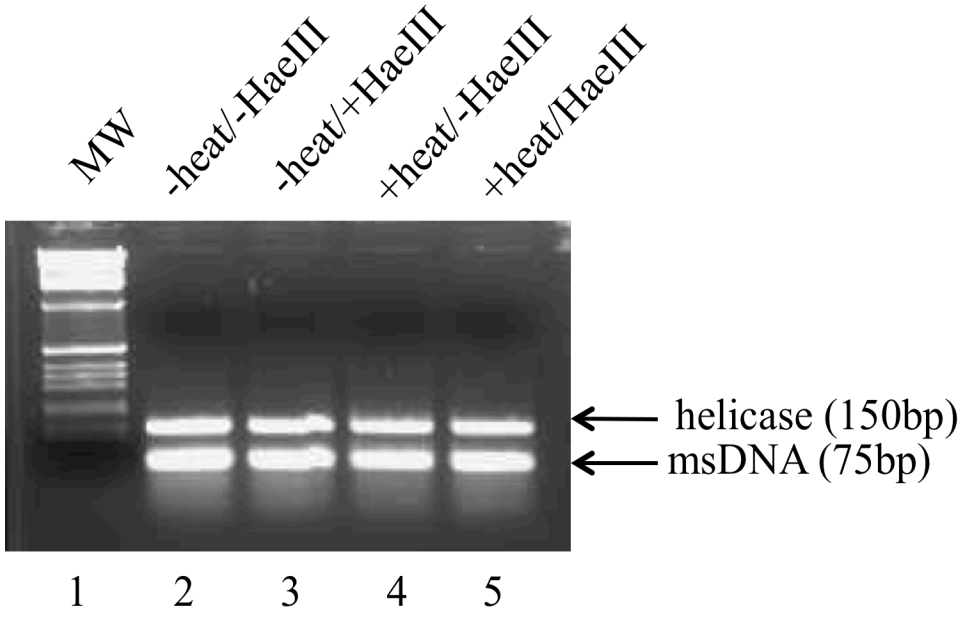


Figure 26. Whole cell extracts in HaeIII digestion do not eliminate the helicase band in the experimental treatment. This gel shows the PCR reactions from cells where whole cell extracts were used in all digestions. MW: molecular weight marker; lane 1: cells were not heated and not

digested with HaeIII; lane 3: cells were not heated and were digested with HaeIII; lane 4: cells were heated and not digested with HaeIII; lane 5: cells were heated and digested with HaeIII.

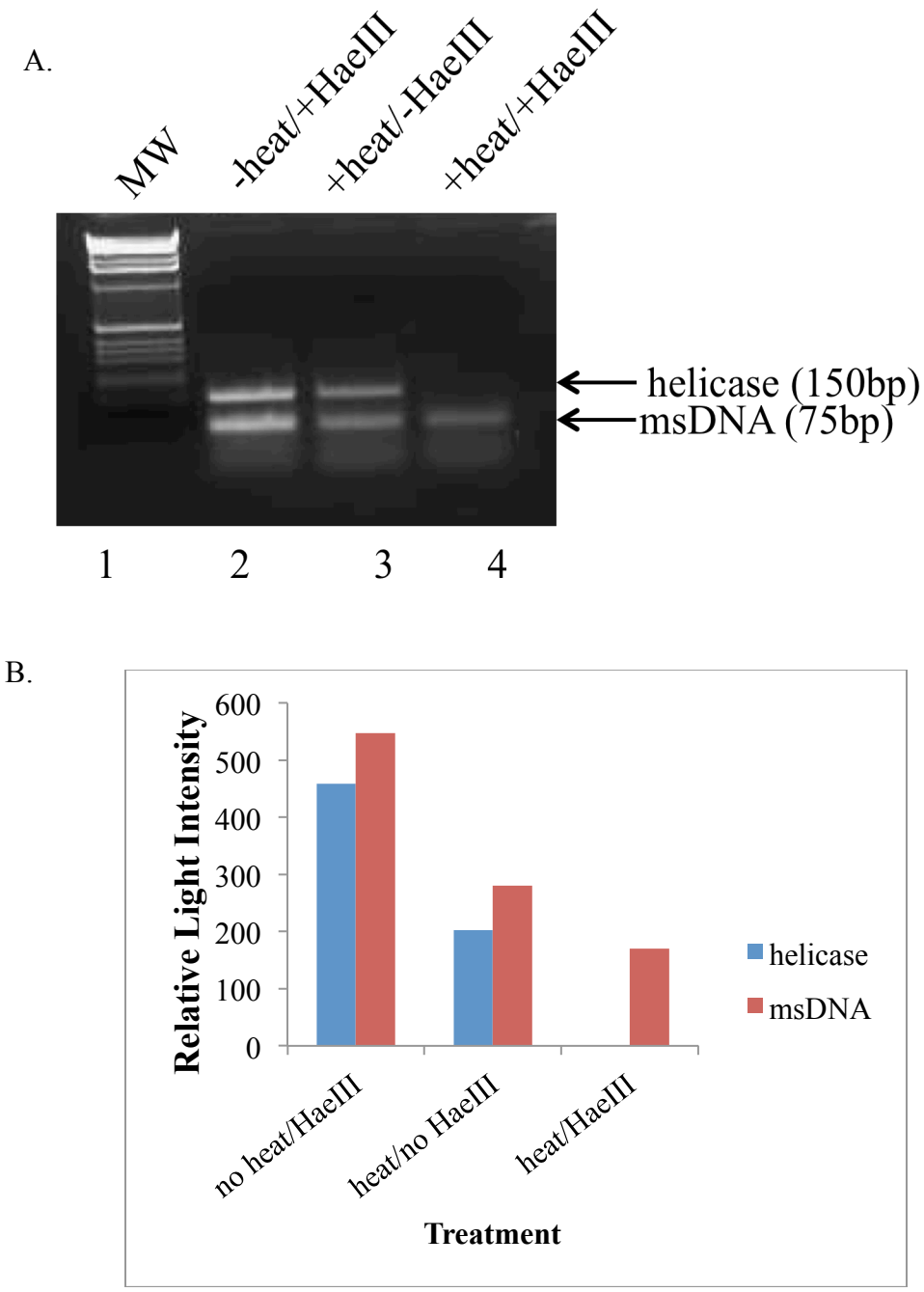


Figure 27. Taking the supernatant from heated cells successfully eliminates the helicase band in the experimental treatment. A. This gel shows the results of the PCR reactions when the supernatant from heated cells was used in the HaeIII digestion as opposed to whole cell extracts. MW: molecular weight marker; lane 2 shows cells that were not heated but digested with HaeIII; lane 3 shows cells that were heated but not digested with HaeIII; lane 4 shows cells

that were both heated and digested with HaeIII. B. Quantification of band intensity using ImageJ. The y-axis represents the area under the curve measured from the Plot Profile. The x-axis represents the treatment.

Part 2D: CviKI-1 Digestions

After protocol development using the HaeIII enzyme, I performed the same procedure using the CviKI-1 enzyme because this enzyme has recognition sites in both the helicase region of the *E. coli* B chromosome and the EC86*msd* region (Figure 9). Thus, if msDNA is being detected, cells that have been lysed and digested with CviKI-1 should not have a helicase band, but should have an msDNA band. To test this hypothesis, I used the same protocol I developed using the HaeIII enzyme and expected to see very similar results as the HaeIII digestion (see Figure 26) if msDNA was being detected. Initially, I performed the experiment in the absence of heat inactivation, but this experiment caused one of my controls to fail(Figure 28, lane 4).If the control in lane 4 had worked properly, both the helicase and msDNA bands should be present (whole cell extracts were used in digestions when the cells were not heated). Closer examination of lanes four and five does reveal that just above the primer-dimer bands, there appears to be a second band (Figure 28). These bands, however, appear to be shifted downwards and are much fainter when compared to the msDNA bands in lanes two and three which suggests that these are products that resulted from non-specific binding of the msDNA PCR primers (Figure 28). Overall, the failure of the control in lane four does not allow me to determine if msDNA was being detected.

The no heat negative control (lane 4; Figure 28) most likely failed in CviKI-1 mediated msDNA detection experiment because the CviKI-1 enzyme could have remained active in the PCR reaction and digested DNA even after the digestion incubation period. For this reason, I heat inactivated the enzyme (80°C for 20 minutes) before doing the PCR reaction to ensure that

CviKI-1 was not able to digest DNA in the PCR mixture. With heat inactivation, I would expect that in the absence of heat and presence of enzyme both the helicase and msDNA bands should be present because the cells should not be opened before digestion and heat inactivation should prevent digestion during PCR. Contrary to what I expected, heat inactivation of CviKI-1, still caused the no heat negative control to fail (lane 4; Figure 29). The presence of the CviKI-1 enzyme, even after heat inactivation, results in digestion of dsDNA. It is not possible to determine if msDNA is being produced because the no heat negative control did not produce the expected banding pattern (lane 4; Figure 29).

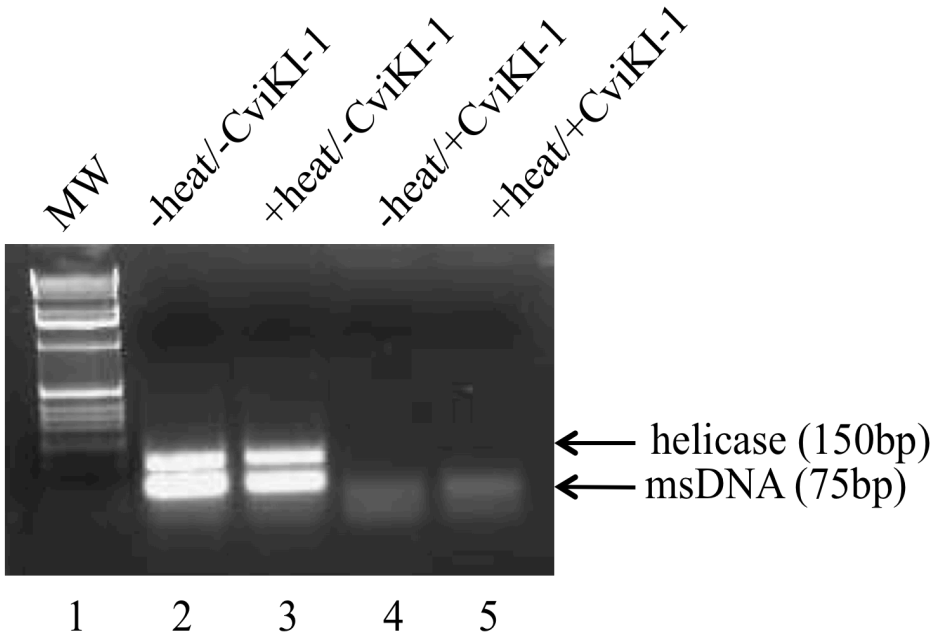


Figure 28. CviKI-1 digestions in the absence of heat inactivation cause "no heat" negative control to fail. This gel shows the PCR products from an msDNA experiment where CviKI-1 was not heat inactivated before the digestion run through a PCR reaction. MW: molecular weight marker; lane 2: cells that were not heated and not digested with CviKI-1; lane 3: cells that were heated and not digested with CviKI-1; lane 4: cells that were not heated and digested with CviKI-1; lane 5: cells that were heated and digested with CviKI-1.

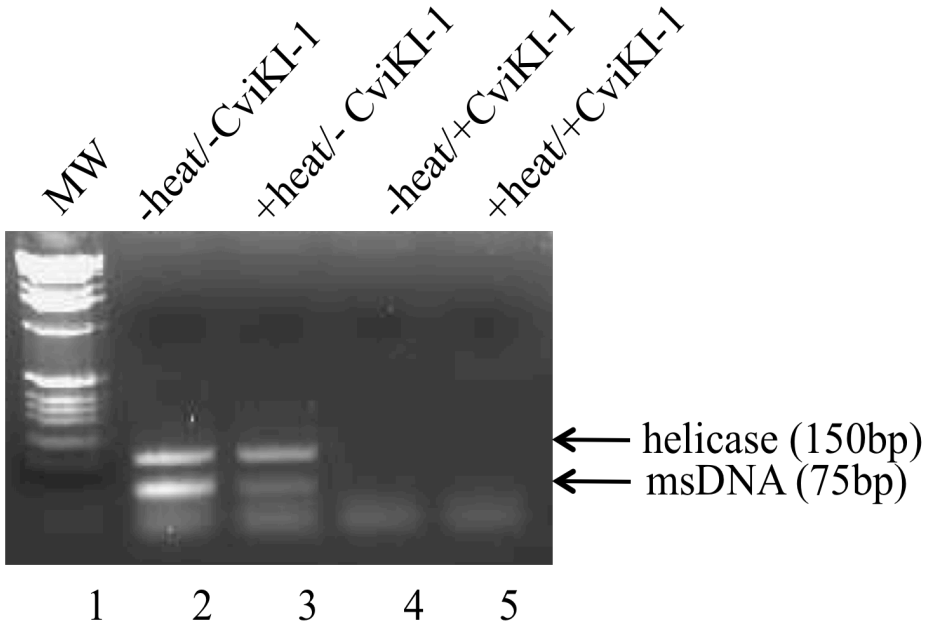


Figure 29. CviKI-1 digestions with heat inactivation cause "no heat" negative control to fail. This gel shows the PCR products from an msDNA experiment where CviKI-1 was heat inactivated before the digestion was included in a PCR reaction. MW: molecular weight marker; lane 2 shows cells that were not heated and not digested with CviKI-1; lane 3 shows cells that were heated and not digested with CviKI-1; lane 4 shows cells that were not heated and digested with CviKI-1; lane 5 shows cells that were heated and digested with CviKI-1.

The results from the CviKI-1 digestions did not allow me to make conclusive statements about the production of msDNA, because the controls that involved adding enzyme in the absence of heat both with and without CviKI-1 heat inactivation did not produce the predicted banding pattern. Thus, the CviKI-1 msDNA detection protocol requires further development and experimentation to produce a protocol that effectively detects msDNA

Part 2E: Testing for Multiplicity of EC86 Retron in the Genome

The next portion of my project involved determining if the method I used was detecting msDNA or multiple copies of the EC86*msd*region. There are two organizational possibilities for the presence of multiple copies of the *msd* region in the EC86 genome (Figure 30). The first

possibility involves the presence of multiple copies of the *msr/msd* region within the retron (Figure 30A). The second possibility suggests the presence of several whole retrons throughout the genome (Figure 30B). The first arrangement is unlikely because amplification of genomic DNA with EC86 *msd* specific primers would produce PCR products of multiple sizes, and this banding pattern was not observed in the msDNA detection experiments. To determine if the second arrangement was possible, I designed primers to the msdRNA portion of the EC86 retron to compare the relative banding intensities of the *msr* region versus the *msd* region. If the *msr* and *msd* bands have the same intensity, then the msDNA detection protocol does not detect msDNA. If the *msd* band is brighter than the *msr* band, then the msDNA detection protocol does detect msDNA.

A.

B.

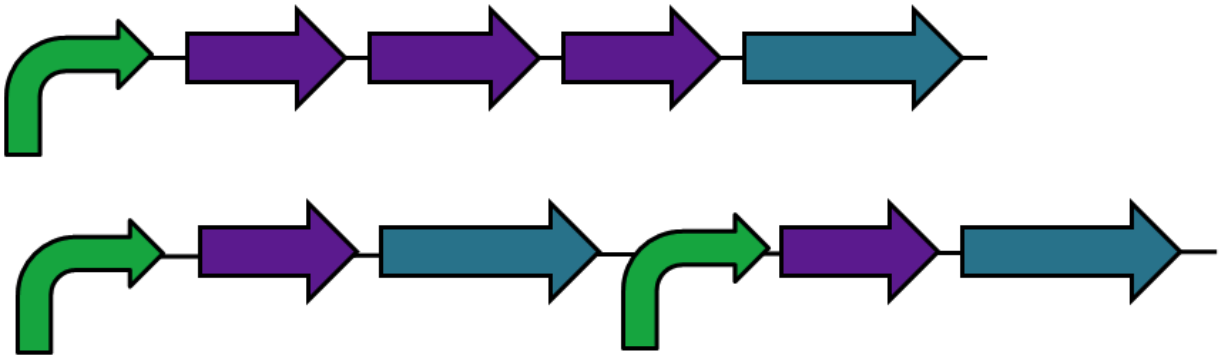


Figure 30. Two possibilities for the multiplicity of *msd* region in the *E. coli* B genome. The green arrows represent the promoter for the retron, the purple arrows represent the *msr/msd* region of the retron and the blue arrows represent RT. The top diagram suggests that multiple copies of the *msr/msd* region within a single retron. The bottom diagram suggests that multiple copies of the entire retron unit throughout the genome. The top diagram is unlikely, because with the primers that were designed for the msDNA portion of the retron, PCR products of multiple lengths would have been detected.

Before performing the experiment to determine the multiplicity of the *msd* region in the *E. coli* B genome, I tested the *msr* primers I designed to ensure they amplified a unique region of the genome (Figure 31). This reaction resulted in a single PCR product of the correct size (75bp), which suggests that these primers are amplifying the EC86 *msr* region.

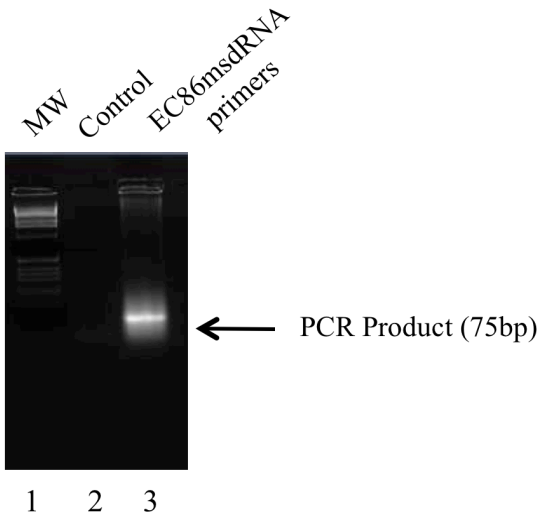


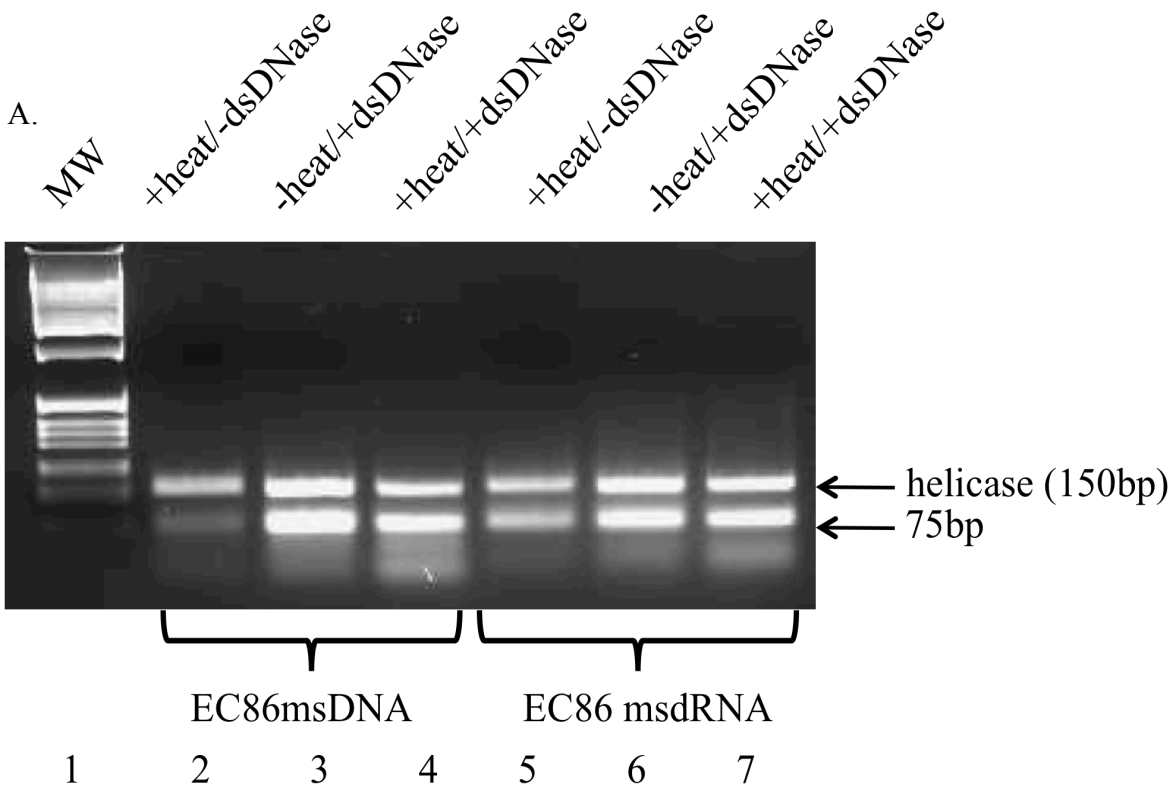
Figure 31. EC86msdRNA primers amplify a PCR product of the correct size. This gel is the product of a colony PCR reaction where the msdRNA portion of the EC86 retron was amplified. MW: molecular weight marker; lane 2: no cells; lane 3: cells + EC86msdRNA primers.

I used the EC86msdRNA primers in a dsDNase experiment. The EC86msdRNA primers were used to compare the intensity of the *msd* band to the *msr* band to determine if multiple copies of the EC86 *msd* are present in the EC86 genome. In all lanes, except for lane two (the discrepancy in this lane could have resulted from a variable amount of primer that was incorporated into this tube), the intensity of the *msr* or *msd* band is similar in intensity to the helicase band (Figure 32). Relatively equal banding intensity suggests that the *msr* and *msd* regions in the *E. coli* B genome are equimolar to the helicase gene. As the helicase gene is only present as a single copy in the *E. coli* B genome, this would suggest that the *msr* and *msd* regions are only present once in the genome, which is supported by Lampson *et al.* (1990) and others who have shown that, in general, retrons are present as a single copy within the chromosome. Alternatively, it is important to note, that the helicase band is twice as long as the msDNA band, which suggests that the helicase band should be twice as bright as the msDNA band if both helicase and msDNA are present once within the genome. The relatively equal intensity of the helicase and msDNA bands suggests that either msDNA is being produced or that the *msd* region

is present more than once within the genome. If msDNA being produced, however, the msDNA band should be much brighter than the helicase band because msDNA is present at approximately 200 copies per cell. Thus, it is possible that the EC86 *msd* region may have multiple copies within the genome. It is, however, difficult to definitely determine whether msDNA is being produced or multiple copies of the *msd* region are present in the genome because of the persistence of the helicase band in the experimental condition (lanes 4 and 7; Figure 32A).

Although this experiment provided insight to the multiplicity of the *msd* region in the *E. coli* B genome, the experiment did not provide conclusive data about msDNA detection. The *msr* and *msd* regions are of comparable to intensity to each other which would suggest that msDNA is not being detected with this protocol, or that it is not being produced (lanes 4 and 7; Figure 32 does not, however, allow me to definitely determine if msDNA is being produced because the helicase band does not disappear for the experimental condition in which the cells are exposed to heat and dsDNase is added to the digestion (lanes 4 and 7; Figure 32). Taken together, these results suggest that the EC86 *msd* region is only present once in the *E. coli* B genome, but do not determine if msDNA is being detected.

A.


MW
+heat/-dsDNase
-heat/+dsDNase
+heat/+dsDNase
+heat/-dsDNase
-heat/+dsDNase
+heat/+dsDNase
helicase (150bp)
75bp
EC86msDNA
EC86 msdRNA
1
2
3
4
5
6
7

B.

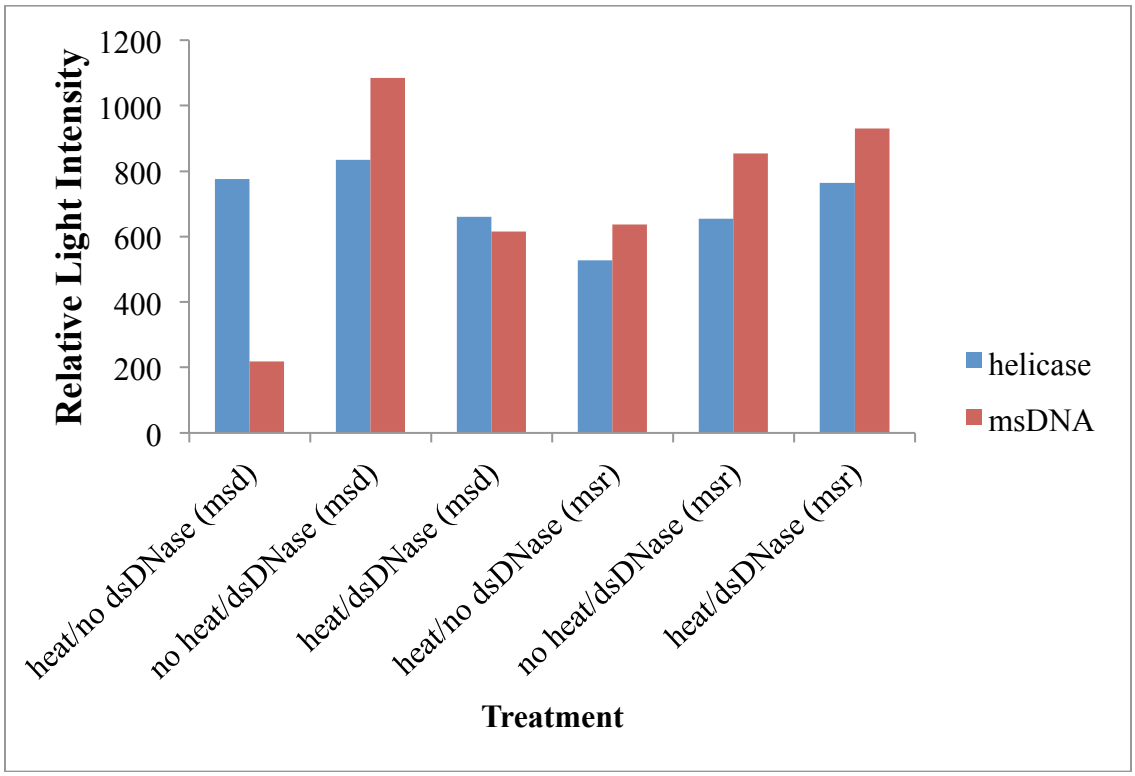


Figure 32. Testing for the multiplicity of the EC86 *msd* region indicates a single *msd* region in the *E. coli* B genome. A. This gel compares the *msd* and *msr* regions in the msDNA detection experiment.MW: molecular weight marker; the top band is EC86 helicase in all the lanes; lanes 2-4 had the msDNA primers and lanes 5-7 had the msdRNA primers. 4µl of dsDNase was used and cells that were exposed to heat were placed on the heat block for 15 seconds. B. Quantification of band intensity using ImageJ. The y-axis represents the area under the curve measured from the Plot Profile. The x-axis represents the treatment.

Discussion

1. Building the msDNA production construct

1A. msr/msd *and spacer regions*

My goal was to create a modular system to produce ssDNA *in vivo.* During the course of this thesis, I built a portion of the testing construct (Figure 3) that will be used in *in vivo* ssDNA synthesis. I built pLac upstream of EC83*msr/msd* and EC86*msr/msd* (Figure 15; Figure 16). Both of these constructs have been sequence verified. I also investigated the importance of the spacer region that exists in between the *msr/msd* region and the RT coding region, hypothesizing that this region has a regulatory function within the retron (Figure 18).I found that the spacer region

was not highly conserved among bacterial retrons. I concluded that this region may not have a regulatory function within retrons and did not incorporate the spacer region into my testing system construct.

1B. Difficulties cloning reverse transcriptase

Attempts to clone reverse transcriptase into a low copy vector and an insulated vector downstream of a weak promoter (pBad) were unsuccessful. I successfully executed a novel PCR method to clone EC86 RT into a vector, even though the part that was cloned into the vector had point mutations (Figure 4). The cloned insert had a mutation in the RBS region and two mutations in the protein sequence. These results suggest that even though the method I developed successfully built and cloned a part into a vector, reverse transcriptase itself is not clonable. Cloning RT into a plasmid, even if it is a low copy plasmid, results in a greater number of proteins than are normally present in the cell, which suggests that RT overexpression could be toxic to *E. coli* B cells. For this reason, I have decided to use the RT that naturally occurs in the *E. coli* B genome as part of the final testing construct (Figure 33). The *msr/msd* regions will be downstream of the pLac promoter on a plasmid. These constructs will have uniquely designed *msr/msd* sequences to produce desired pieces of ssDNA to be used in DNA origami designs. A high copy plasmid and a very strong promoter will allow the introduced *msr/msd* regions to overwhelm the endogenous *msr/msd* msDNA.

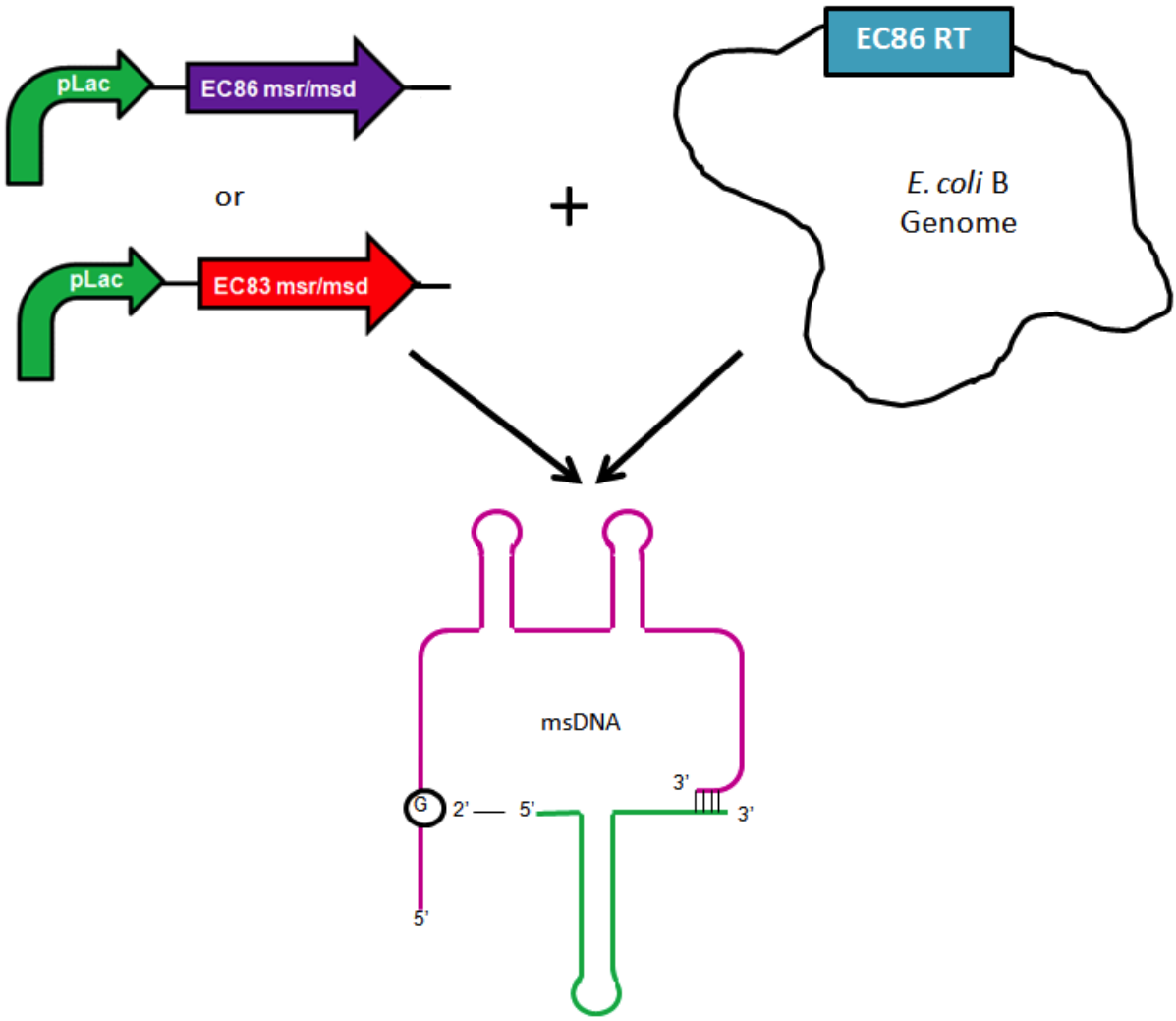


Figure 33. Revised model for production and detection of msDNA *in vivo.* Synthetically engineered *msr*/*msd* regions will be transformed into *E. coli* B cells and the naturally occurring RT in the *E. coli* B genome will be utilized to make the final msDNA structure.

1C. Overcoming the problem of decreased system modularity

Although the absence of reverse transcriptase on a vector reduces the modularity of the system, the presence of the *msr/msd* region downstream of pLac allows for a system that can be regulated. In the presence of the LacI protein, the pLac promoter is not able to transcribe DNA sequences (Penumetcha *et al.*, 2010). Other mutations in pLac and LacI produce promoters and proteins with varying levels of affinity for protein and DNA respectively, which alters the amount of protein that is transcribed (Penumetcha *et al*., 2010). Using the LacI-pLac repressor-

induction system permits control at the transcription level even if the modularity of the system has been reduced, because the amount of msDNA that is produced can be controlled by the strength of the promoter and/or repressor (Figure 34).

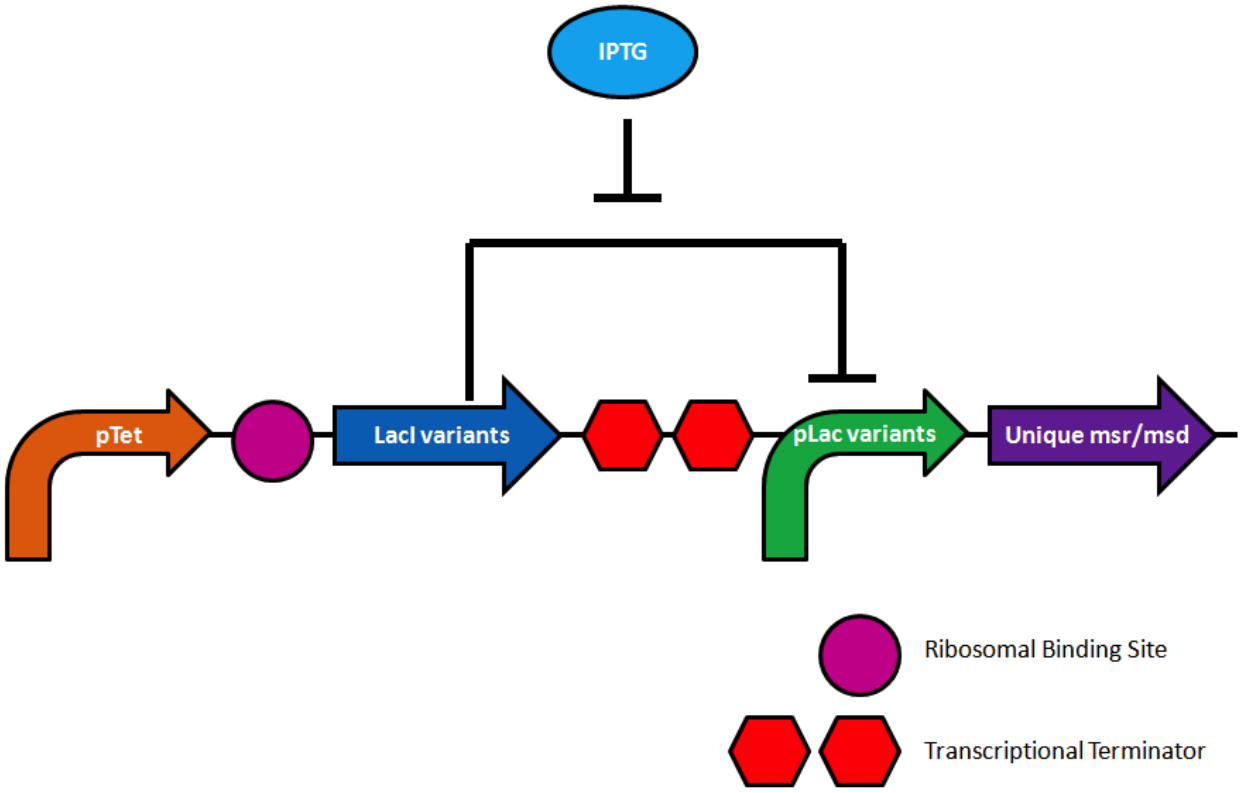


Figure 34. LacI-pLac testing construct for production of *msr*/*msd*. In this system, pTet controls the production of the LacI repressor protein. In the presence of the repressor, pLac cannot transcribe the *msr*/*msd* region. In the presence of IPTG, the repressor (LacI) is repressed and pLac is able to transcribe the *msr*/*msd* region. The *msr*/*msd* region is uniquely designed to include the desired ssDNA sequence. The RT that is present in the *E. coli* B genome will be used to make the final msDNA structure.

2. msDNA Detection

2A. dsDNase and HaeIII enzymes

In addition to building a modular ssDNA production system, I have also worked on a novel protocol to detect msDNA. Currently, polyacrylamide gels are the most common method to visualize ssDNA. Although gels achieve the desired goal of detecting ssDNA, polyacrylamide gels are difficult to work with and our lab was not able to visualize msDNA on a polyacrylamide

gel. Thus, I needed to develop a method for detecting msDNA to ensure that the msDNA production system produces the expected structures (Figure 33).I developed a novel protocol that uses PCR, enzymatic digestion and agarose gel electrophoresis to detect msDNA (Figure 4). Thus far, I have successfully designed and tested primers that amplify the, EC86 *msd* region, a portion of the EC86 helicase gene, and the EC86 *msr* region (Table 1; Figure 19; Figure 31). The first enzyme I used in the msDNA detection protocol was dsDNase, but the experiments performed with this enzyme did not produce consistent controls, which did not allow me to determine if msDNA was being detected. The dsDNase experiments did, however, allow me to make many modifications to the msDNA detection protocol which were important in the next stage of protocol development. I used the HaeIII enzyme to standardize a protocol that completely digests dsDNA (Figure 10; Figure 27). The protocol that was developed using this enzyme provides a template for using an enzyme that has recognition sites in both the helicase gene and the *msd* region of the EC86 retron, such that any amplification with EC86 msDNA primers would suggest detection of msDNA

2B. CviKI-1 Digestions

2Bi. CviKI-1 msDNA detection protocol

I used the enzyme CviKI-1, which digests both the helicase gene and the *msd* region of the EC86 retron. I performed a protocol similar to the one I developed using HaeIII with the exception that I heated inactivated CviKI-1 before performing PCR(Figure 10). The reason I added the heat inactivation step in the CviKI-1 protocol was because, in the absence of the heat inactivation step, helicase and msDNA primers did not amplify the expected PCR products (Figure 28). Even with heat inactivation, however, there were no bands when the cells were not opened and enzyme was added (Figure 29).This result suggests that the enzyme could be

digesting the DNA during the PCR reaction (during which unheated cells could have been opened), because the heat inactivation protocol may not have successfully deactivated the enzyme. To determine if CviKI-1 had actually been heat inactivated, I could PCR amplify the helicase region of the *E. coli* B genome and digest half of the PCR reaction with the CviKI-1 and digest the other half of the PCR reaction with the heat inactivated CviKI-1. If the enzyme had been heat inactivated, then I should only be able to see a single product on the gel. If the enzyme had not been heat inactivated, then I should be able to see two products on the gel.

If cells were not being opened during the PCR protocol, the cells originally not exposed to heat could have opened during the five hour digestion period. If non-heated cells were being opened during the digestion incubation, CviKI-1 would have digested DNA even in cells that were not originally heated, which would have prevented the helicase and msDNA primers from amplifying the expected PCR products. To determine if cells were being opened during the digestion incubation period, I could perform a control digestion wherein CviKI-1 is added to whole cell extracts of cells that have not been heated and incubate the digestion at room temperature (or another condition where cell lysis is known not to occur) instead of 37°C. In the control experiment, I would not expect to see the helicase or msDNA bands on the gel because the cells should not have lysed to allow the primers to access the DNA. If I did see helicase or msDNA bands on the gel, then something in the digestion mixture is causing the cells to lyse.

In comparing the CviKI-1 digestions to the HaeIII digestions, the HaeIII digestions also had both the msDNA and helicase bands present in the absence of heat and the presence of enzyme, which suggests that some cells were being lysed either at the PCR stage or the digestion stage in these digestions as well. Thus, in the case of the HaeIII digestions, if the cells were being opened during PCR or the digestion period, the HaeIII enzyme did not digest the helicase and msDNA bands nearly as efficiently as the CviKI-1 enzyme appears to have (Figure

26;Figure 28). This discrepancy could be explained by the fact that the CviKI-1 enzyme could be faster than HaeIII. According to New England Biolabs, approximately two times more HaeIII is necessary to digest a comparable amount of DNA as CviKI-1 (HaeIII: 10,000units/ml; CviKI-1 5,000units/ml). The higher efficiency of CviKI-1 is one possibility regarding the failure of cells that were not opened before digestion but had enzyme added to them, to amplify both a helicase band and an msDNA band.

Another possible explanation regarding the absence of helicase and msDNA bands in the "no heat" CviKI-1 negative controls is the presence of a CviKI-1 recognition site in the region where the helicase primers binds to the DNA (Figure 9). Digestion with CviKI-1 may have prevented this primer from binding and amplifying the helicase region of the *E. coli* B genome, which explains the absence of the helicase band in the CviKI-1 digestions (Figure 28; Figure 29). Lowering the annealing temperature in the PCR reaction (Figure 10) may solve this problem to ensure that the primer is able to bind DNA, because even after the CviKI-1 digests the helicase gene, there is still a considerable region of DNA left for the helicase forward primer to bind (Figure 9).

2Bii. Location of CviKI-1 recognition site

I investigated the location of the CviKI-1recognition site in the msDNA structure to determine ifCviKI-1 could digest the final msDNA product. If the CviKI-1 recognition site is located in a region of secondary structure formation, then the msDNA product will be digested and the msDNA primers cannot amplify the DNA region of the msDNA structure meaning that my method will not detect msDNA. I used palindrome (http://emboss.bioinformatics.nl/cgi-bin/emboss/palindrome) and einverted (http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted) to look for secondary structure formation in EC86 msDNA. After finding

the regions of secondary structure formation, I searched for the CviKI-1 site within these regions and found that the recognition site is located in one of the regions of secondary structure formation in msDNA (Figure 35). This presence of a CviKI-1 recognition site at this location suggests that msDNA may not be detected with the current procedure.

The most appropriate next step would be to find a new enzyme that has digestion sites in both the helicase gene and the *msd* region of the retron, but not in a location that is involved in msDNA secondary structure formation. Alternatively, I could use two different enzymes if I do not find one enzyme that fulfills both of these functions. If I did use two separate enzymes, I would test each enzyme separately to make sure that they cut the appropriate sequences. I would ideally want to find an enzyme cuts twice in the *msd* region and a separate enzyme that only cuts once in the helicase region. A greater number of recognition sites in the *msd* region than the helicase gene is important, because I would have to be able to conclude that if the helicase gene is digested, the *msd* portion of the EC86 retron has been digested as well. If the helicase is only cut once and this band disappears, then it is fair to conclude that the *msd* (which has two cut sites in it) has also been completely digested, because the enzyme is more likely to digest the region with two restriction sites (*msd*) than one restriction site (helicase). Confirming that the genomic region of *msd* has been digested is very important to make sure that only extrachromosomal msDNA is detected in the gel.

If I am not able to find an enzyme that meets the above criteria, a third option would be to use the dsDNase enzyme again with the revised protocol (Figure 10). This procedure is more likely to be successful because I will supernatant from heated cells in contrast to using whole cell extracts. Using dsDNase would eliminate the problem of finding an enzyme (or two) that have more cut sites in the *msd* region of the E. coli B genome than in the helicase gene, because dsDNase digests all dsDNA. The problem with dsDNase is that it may digest the msDNA

structure because of the regions of secondary structure formation within the molecule (Figure 1). Given the three possible strategies to circumvent the problems associated with CviKI-1, the most appropriate strategy would be find one (or two) enzymes that digest the helicase and *msd* regions of the EC86 retron in a location that does not result in secondary structure formation in the final msDNA structure.

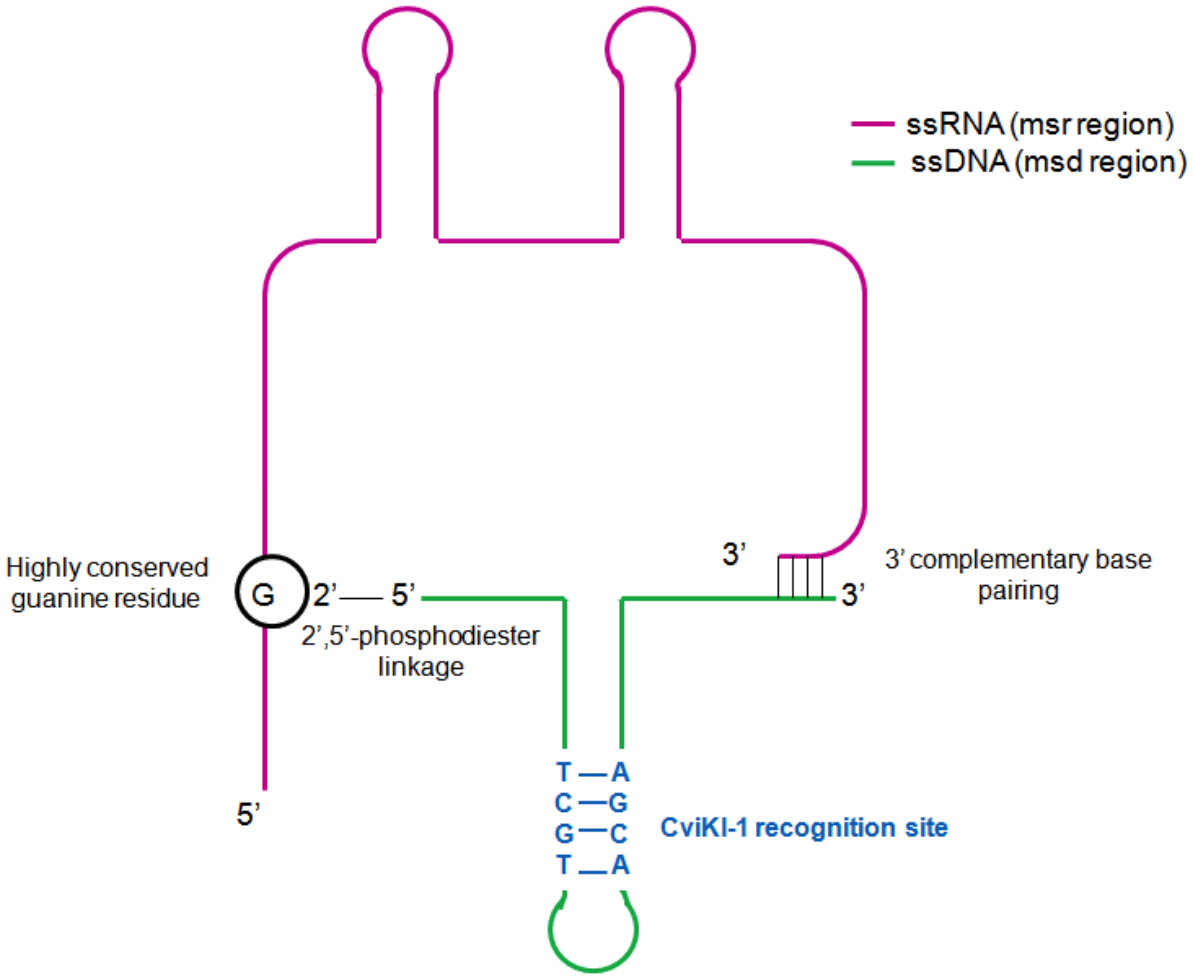


Figure 35. Location of CviKI-1 recognition site in EC86 msDNA. This figure shows the location of the CviKI-1 recognition site in the DNA portion of EC86 msDNA (delineated by the blue nucleotides). The digestion site occurs in a region where there is secondary structure formation.

Future Directions

When I have an enzyme (or two) that function properly in the msDNA detection protocol, I will attempt to simplify the protocol. Both the time it takes to open the cells and digestion time can likely be reduced so that the entire procedure can be performed on a single day. After the protocol has been streamlined, I will compare the detection of msDNA using my mature protocol with ssDNA detection in a polyacrylamide gel. Comparing my msDNA detection method to ssDNA detection in polyacrylamide gels will allow me to determine the efficacy of our novel detection method with the currently accepted method.

The next step will be to design unique *msr/msd* sequences that can be incorporated into a new testing construct (Figure 33), so that desired segments of ssDNA can be produced. Mao *et al* (1995) have demonstrated *in vivo* production of antisense ssDNA using msDNA in *E. coli* to regulate other genes on the chromosome. This group's approach will have important implications in my *in vivo* approach to produce higher order DNA structures. Other groups have demonstrated that changing the sequence of the *msd* region of a retron does not have a significant impact on msDNA production because RT recognizes the *msr* region to initiate reverse transcription and variations in the *msd* region do not affect this process (Rice and Lampson, 1996). Thus, production of unique ssDNA structures using msDNA has been successful as demonstrated by other groups. Additionally, because the EC86 retron does not automatically cleave the RNA portion of the msDNA structure, it is important the test the cross compatibility of the EC83 RT in forming msDNA structures with EC86 *msr/msd*, because the EC83 RT is able to cleave ssDNA from ssRNA in the msDNA structure (Kim *et al.,*1997). If the EC83 RT is able to successfully form the msDNA structure and then cleave the RNA portion of EC86 msDNA, then this RT would be the most appropriate reverse transcriptase for achieving this project's goals of producing ssDNA *in vivo.* If EC83 RT is not able to cleave ssDNA away from ssRNA in the

EC86 msDNA structure, Mao *et al* (1995) have demonstrated that designing enzymatic digestions sites into the *msd* region successfully isolates ssDNA.

Summary

Over the course of this project, I have overcome and modified many of the challenges associated with developing a method to produce ssDNA *in vivo.* The two goals of my project involved cloning RT and designing a protocol to detect msDNA. Although traditional cloning methods were not successful, I developed a novel method to clone reverse transcriptase into a vector and ultimately decided that using the natural RT was the most effective way to proceed. The protocol I developed to detect msDNA overcame a number of challenges associated with digesting dsDNA in whole cell extracts. This protocol provides a template for further development of a novel msDNA detection protocol.

Acknowledgements

I would like to thank all the members of the Campbell lab for giving me suggestions and comments throughout the course of my project. In particular, I would like to thank Mary Gearing, Katie Richeson, Anvi Raina, Brianna Pearson and Eric Sawyer who were the individuals in the lab during the time I worked on my thesis. I would also like to thank Dr. Laurie Heyer for her helpful questions and suggestions during lab meeting. This work would not have been possible without comments on the written portion of the work from Dr. Barbara Lom. Special thanks to the James G. Martin Genomics Program and the Davidson College Biology Department for funding this work. Finally, I would like to thank Dr. Malcolm Campbell for being my mentor during this project and providing excellent guidance throughout.

References

Chhabra R, Sharma J, Ke Y, Liu Y, Rinker S, Lindsay S and Yan H. Spatially addressable Multiproteinnanoarraystemplated by aptamer-tagged DNA nanoarchitectures.*Journal of the American Chemical Society,*129: 10304-10305, 2007

Greenwood V. With a few strands of nucleic acids and some ingenious programming, DNA origami is remaking nanotechnology, from drug delivery to chip design. *SEED Magazine*, 12 October 2009.

Haynes K, Broderick ML, Brown AD, Butner TL, Dickson J, Harden L, Heard L, Jessen E, Malloy K, Ogden K, Rosemond S, Simpson S, Zwack E, Campbell AM, Eckdahl R, Heyer L, Poet, J. Engineering bacteria to solve the burnt pancake problem. *Journal of Biological Engineering,* 2: 1-12, 2008.

Herzer PJ, Inouye S, and Inouye M. Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34bp intergenic sequence. *Molecular Microbiology,*6: 345-354, 1992.

Hsu M, Xu C, Inouye M, and Inouye S. Similarity between the *Myxococcus Xanthus* and *Stigmatella aurantiaca* reverse transcriptase genes associated with multicopy, single-stranded DNA. *Journal of Bacteriology,*174: 2384-2387, 1992.

Inouye M and Inouye S. msDNA and bacterial reverse transcriptase.*Annual Reviews: Microbiology,* 45:163-186, 1991.

Inouye M and Inouye S. Retrons and multicopy single-stranded DNA. *Journal of Bacteriology,* 174(8):2419-2424, 1992.

Inouye M, Ke H, Yashio A, Yamanaka K, Nariya H, Shimamoto T and Inouye S. Complex formation between a putative 66-residue thumb domain of bacterial reverse transcriptase RT-EC86 and the primer recognition RNA. *The Journal of Biological Chemistr,y*279: 50735-50742, 2004.

Inouye S, Herzer PJ, and Inouye M. Two independent retrons with highly diverse reverse transcriptases in *Myxococcusxanthus. Proceedings of the National Academy of Science,*87: 942-945, 1990.

Inouye S, Hsu M, Eagle S and Inouye M. Reverse transcriptase associated with the biosynthesis of the branches RNA-linked msDNA in *Myxococcus xanthus. Cell,*56: 709-717, 1989.

Inouye S, Hsu MY, Xu A and Inouye M. Highly specific recognition of primer RNA structures for 2'-OH priming reaction by bacterial reverse transcriptases. *The Journal of Biological Chemistry,*274: 31236-31244, 1999.

Kim K, Jeong J, Lim D. A mutation study of the site-specific cleavage of EC83, a multicopy single-stranded DNA (msDNA): nucleotides at the msDNA stem are important for its cleavage. *Journal of Bacteriology,* 179(20):6518-6521, 1997.

Knight T. Idempotent Vector Design for Standard Assembly of Biobricks. http://dspace.mit.edu/handle/1721.1/21168. 2003.

Lampson BB, Viswananthan M, Inouye M and Inouye S. Reverse transcriptase from *Escherichia coli*exists as a complex with msDNA and is able to synthesize double-stranded DNA. *The Journal of Biological Chemistry,*265: 8490-8495, 1990.

Lampson BC, Inouye M and Inouye S. Retrons, msDNA and the bacterial genome.*Cytogenetic and Genome Research,* 110: 491-499, 2005.

Lim D. Structure and biosynthesis of unbranchedmulticopy single-stranded DNA by reverse transcriptase in a clinical E. coli isolate. *Molecular Microbiology*, 7: 3531-3542, 1992.

Lim D. Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site. *Molecular Microbiology,* 5: 1863-1872, 1991.

Lim D and Maas WK. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B. *Cell,*56: 891-904, 1989.

Lin C, Liu Y, Rinker S and Yan H. DNA tile based self-assembly: building complex Nano-architectures. *ChemPhysChem,* 7: 1641-1647, 2006.

Mao JR, Shimada M, Inouye S and Inouye M. Gene regulation by antisense DNA production *in vivo.The Journal of Biology Chemistry,* 270: 19684-19687, 1995.

Marine Biochemicals. Double-Strand Specific DNase (dsDNase). 2009.

MIT Parts Registry. Constructing Novel BioBrick Parts. 2000. URL http://syntheticbiology.org/BioBricks/Part_fabrication.html

Miyata S, Ohshima A, Inouye S and Inouye M. *In vivo* production of a stable single-stranded cDNA in *Saccharomycescerevisiae* by means of a bacterial retron. *Proceedings of the National Academy of Science,* 89: 5735-5739, 1992.

Patil NS, Deshmukh SS, and Shankar V. Extracellular nuclease from a thermophile*Streptomyces thermonitrificans:* production, purification and partial characterization of-double strand preferential-deoxyribonuclease activity. *Process Biochemistry*, 40: 1271-1278, 2005.

Penumetcha P, Lau K, Zhu X, Davis K, Eckdahl TT and Campbell AM. Improving the lac system for synthetic biology. *BIOS,* 81: 7-15, 2010.

Rice SA and Lampson BC. Bacterial reverse transcriptase and msDNA.*Virus Genes,* 11: 94-104, 2005.

Rice SA and Lampson BC. Phylogenetic comparison of retron elements among the myxobacteria: evidence for vertical inheritance. *Journal of Bacteriology,* 177: 37-45.

Rothemund PWK. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440:297-302, 2006.

Sun J, Inouye M and Inouye S. Association of a retroelement with a P4-like cryptic prophage integraged into the selenocystyl tRNA gene of *Escherichia coli.Journal of Bacteriology,* 173: 4171-4181, 1991.

Viswanathan M, Dower KW and Lovett ST. Identification of a potent DNase activity associated with Rnase T of *Escherischia coli.The Journal of Biological Chemistry,* 273: 35126-35131, 1998

Waters Mike. Designing a modular single-stranded DNA production device using msDNA. Senior Honors Thesis. Biology Department, Davidson College, Davidson NC. 2010

Appendix A: Lab Protocols

DNA Isolation and Purification

Mini Preps

Promega Mini Prep

1. For each miniprep, grow 2 ml culture, 37° C, overnight (O/N) in appropriate medium and anitbiotic (usually ampicillin,but not always); shake at 400 RPM and slant tubes.

Next Day

2. Pour the contents O/N culture into one labeled tube. Replace the metal cap and save the culture tube at 4° C.
3. Spin the microfuge tube for 2 min.
4. Aspirate off the medium.
5. Resuspend pellet in 250 µl **Cell Resuspension Solution**. Resuspend cells very well by pipetting up and down.
6. Add 250 µl of **Cell Lysis Solution**. Mix by inverting the tube 4 times.
7. Add 10 µl **Alkaline Protease Solution**. Mix by inverting the tube 4 times. Incubate 3 minutes at room temperature (RT°).
8. Add 350µl **Neutralization Solution**. Mix by inverting the tube 4 times.
9. Spin full speed for 10 minutes at RT°.
10. Prepare Spin Column (with binding resin) by inserting into 2 mL collection tube.
11. Transfer supernatant to Spin Column.
12. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and reinsert Spin Column to collection tube.
13. Add 750 µl **Wash Solution** (with ethonal already added).

14. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and reinsert Spin Column to collection tube.
15. Add 250 µl **Wash Solution** (with ethonol already added).
16. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and transfer Spin Column to a clean 1.5 mL microfuge tube.
    Be sure to label the spin column and the microfuge tube.
17. Add 100 µl **Nuclease-Free Water** to the Spin Column.
18. Spin full speed for 1 minute at RT°. **SAVE THE LIQUID with your plasmid.** Discard the spin column. Store DNA at -20° C.
19. If you want to digest some DNA, 5-10 µl of the MP DNA in a final volume of 20 µl is a nice place to start.

Zyppy Mini Prep

1. For each miniprep, grow 2 ml culture, 37° C, overnight (O/N) in appropriate medium and anitbiotic (usually ampicillin,but not always); shake at 400 RPM and slant tubes.

Next Day

2. Add 600 µL of O/N culture to an appropriately labeled 1.5 mL microfuge tube. Save the rest of the O/N culture at +4 ° C and keep them sterile.
   You can increase your yield by pelleting up to 3 mL of O/N culture and resuspending pellet very well in 600 µL water or TE.
3. Add 100 µL **7X Lysis Buffer (blue color)** . Mix by inverting the tube 4-10 times. Solution should become clear instead of opaque.
   Proceed to the next step within 3 minutes.
4. Add 350 µL of **Neutralization Buffer (yellow color; RNase A already added)** and mix by inverting the tube until the entire solution and precipitate is yellow. This buffer is stored at +4 ° C.
5. Spin full speed for 2 minutes at RT°.
6. Prepare Zymo-Spin II column (with binding resin) by inserting into 2 mL collection tube. Be sure to label the spin column and the collection tube. Also label a 1.5 mL microfuge tube for use in step 14.
7. Transfer ~900 µL supernatant to Zymo-Spin II column. Do not transfer any of the solid precipitate.
8. Spin full speed for 15 seconds at RT°.
9. Discard liquid flowthrough and reinsert Zymo-Spin II column into same collection tube.
10. Add 200 µL **Endo-Wash Buffer** to the Zymo-Spin II column.
11. Spin full speed for 15 seconds at RT°. No need to empty flow through.
12. Add 400 µL **Zyppy Wash Buffer** (with ethonol already added).
13. Spin full speed for 30 seconds at RT°. Discard liquid flowthrough and and the 2 mL collection tube. The DNA is still in the spin column.
14. Transfer Zymo-Spin II column to a clean and appropriately labeled 1.5 mL microfuge tube (from step 6 above).
15. Add 30 µL **Zyppy Elution BUffer ot the center of the** to the Zymo-Spin II column. Let it stand for 1 minute to maximize yield.
16. Spin full speed for 30 seconds at RT°. **SAVE THE LIQUID with your plasmid.**
    Discard the spin column. Store DNA at -20° C.

17. If you want to digest some DNA, 5-10 µL of the MP DNA in a final volume of 20 µL is a nice place to start. You can NanoDrop the DNA if you need to know the concentration.

Ethanol Precipitation

1. If the volume of the DNA is less than 200 µl, bring the volume up to 200 µl with sterile dH2O.
2. Add 1/10 th volume of 3M sodium acetate to the DNA solution and mix.
3. Add 2 volumes of -20° C 100% ethanol (EtOH) and vortex for 10 seconds. Put the tube in a -20° C freezer overnight or a -70° C freezer for 30 minutes.
4. Spin in a microfuge for 10 minutes. Pour out the EtOH but **save the pellet**!!
5. Wash the pellet with 500 µl of 4° C 70% EtOH, gently roll the tube, then dump the EtOH, and speedvac the pellet. **SAVE THE PELLET!**
6. Resuspend DNA in appropriate volume of TE or water.

Cloning

Ligation

X ng of insert = (2) (bp insert) (50 ng linearized plasmid-) ÷ (size of plasmid in bp)

The 2 in the numerator takes into account the fact that you want to have a 2:1 ratio of insert to vector. You want to use 50 ng of plasmid for a typical ligation.

Thaw the frozen 2X buffer at room temperature. After the first thawing, it is best if you aliquot the 2X buffer in smaller volumes to avoid repeated freeze/thaw cycles. If a precipitate is present, vortex the solid DTT until it is back in solution. This usually takes 1 - 2 minutes.

Set up the following 10 µl ligation reaction (if your DNA is too dilute, you can scale up to 20 µL ligation volume and transform all 20 µL. But try to minimize the ligation volume.

Digested Vector (50 ng)	1 µl
Insert (2:1 molar ratio insert:vector)	(< 3) **x** µl
Sterile water	3µl- x µl
2X ligation buffer	5 µl
3 units T4 DNA ligase (keep cold)	1 µl
Total Volume	**10.0 µl**

Incubate for 5 minutes at room temperature. From this point, you may either freeze the ligation or go directly to transformation.

Transformation

1. Thaw the competent cells on ice for 5 minutes.
2. Very gently, aliquot cells into smaller volumes (we have used as low as 20 µL of cells with 5 µL of ligation) using chilled microfuge tubes.
3. Add 1 - 5 µL of ligation mixture (can go as high as 10 µL).
4. Incubate on ice for 5 minutes.
5. Add SOC media with no antibiotic to a final volume of 60 - 100 µL. Spread cells onto plates containing antibiotic. You may want to let this sit for 30 minutes if the antibiotic is not ampicillin (not tested, just rumored to help)

Gel Electrophoresis (Pouring a Gel)

1. Tape the ends of a gel mold and make sure some of the tape wraps around the bottom of the mold by 1 - 2 mm. Choose the appropriate comb(s) and make sure it looks clean.
2. Make 60 mL of the appropriate gel mixture in a 250 mL flask, cover it with Saran, and microwave for 1 minute and 20 seconds on high power (a good starting time).
3. Visually check to see that all the agarose has melted. Unmelted agarose looks like tiny refractive lenses floating around. If not completely melted, nuke it a little longer. Try 20 seconds.
4. Allow the gel to cool a bit; you may hasten this by running cold water over it but do NOT let it cool too much. Don't be startled (and drop the flask) by the popping sound of the saran wrap as the flask cools.
5. Add ethidium bromide (stock EtBr; 10 mg/mL) to the 60 mL of gel so the **final concentration is 0.2 µg/mL**. Then pour the gel into the mold.
EtBr is a known mutagen so wear gloves.
6. Allow this to cool until it turns slightly white. The gel is ready to run, as soon as you pull off the tape, remove the comb, and submerge it in 0.5X TBE that has the same concentration of EtBr. Unless you are told otherwise, our gel boxes hold 450 mL of buffer.

Gel Purification

1. Run the fragment(s) on a gel and photograph the gel. Remember to use a MW marker. Also, use as low a percentage gel as you can to resolve your bands.
2. Cut out the band(s) of interest using a razor blade and the UV box with the hinged plexiglass covering. Cut as close to the band as possible to minimize the volume of the gel piece. Use as low a level UV light for as little time as possible. Protect your eyes and skin from the UV.
3. Make sure Buffer NT3 has the appropriate volume of ethanol added to it before you go any further.
4. Weigh the gel slice in a colorless tube. Add 2 volumes of buffer Buffer NT to 1 volume of gel (100 mg gel = 200 µL NT). If your gel is > 2% agarose, use 400 µL Buffer NT.
5. Incubate in 50° C waterbath for 10 minutes, or until the gel slice has COMPLETELY dissolved. You can vortex every 2-3 minutes to speed up the dissolving process. Hold the tube up to the light and look for a translucent piece of undissolved gel.
6. Place a Nucleospin Extract II spin column in one of the provided 2 mL collection tubes.

7. To bind DNA, apply the DNA solution to the Nucleospin Extract II spin column and centrifuge as full speed for one minute. The maxiumum volume you can load at a time is 800 µL. If your volume is larger, then reload the same column and spin a second time.
8. Discard flow-through from step 8 and place the column back in the same collection tube.
9. To wash, add 700 µL of Buffer NT3 to the column and centrifuge for 1 minute at full speed.
10. Discard flow-through first and then spin again for an additional two minues to dry the column.
11. Place the QIAquick column in a clean, 1.5 mL centrifuge tube.
12. To elute the DNA, add 15 -50 µL of Buffer NE (30 µL is typical) to the CENTER of the QIAquick membrane. Let this sit for 1 minute at room temperature, then spin full speed for 1 minute. Your DNA is in this small volume. If your DNA is 5-10 kb or larger, preheat buffer NE to 70° C prior to adding to spin column.
13. Use 1 or 2 µL of this DNA to quantify using the NanDrop. You are ready to do a ligation now.

Restriction Digestion

1. Decide which buffer you must use and the correct incubation temperature.
2. Decide how much (in microliters) DNA you will digest.
3. Then construct a table similar to the one below:

- 3 µl DNA (volume depends on DNA concentration, 3 µl is good if using MP DNA)
- 14 µl water (up to desired volume)
- 2 µl 10X buffer (one tenth final volume)
- 1 µl restriction enzyme (never more than 10% final volume)
- 20 µl total volume

4. As you add each ingredient to a 500 µl microfuge tube, stir it in well with the pipet tip and
5. Make sure all of the liquid is in the bottom of the microfuge tube; spin if necessary.
6. Incubate at the appropriate temperature (typically 37° C) for at least 30 minutes.

Polymerase Chain Reaction

Colony PCR

1. Determine the number of colonies to be tested. Plan to conduct PCR on control plasmids with and without the insert. Assemble the following PCR mixture:

Per Reaction (you might want to make a cocktail, rather than multiple individual reactions)
1 µL forward primer (20 pmol = 0.2 µL of 100 µM oligo stock solution)
1 µL reverse primer (20 pmol = 0.2 µL of 100 µM oligo stock solution)
10 µL dH_2O
12 µL 2X Monster Mix (Green solution from Promega)
24 µL total volume

2. Use a micropipette tip to pick a single putative colony off a plate. Insert the tip into the PCR mixture and pipette up and down.

3. Reserve bacteria from each PCR mixture by removing 1 ul and placing into 100 ul of LB + Amp in a labeled tube and put in 37° C incubator.
4. Conduct PCR according to the following thermal profile:

94° C 10 minutes
20 cycles of:
94° C 15 seconds
46° C 15 seconds
72° C 30 seconds (time varies depending on the size of insert; rule of thumb is 1 minute per kb of DNA being amplified)
Hold at RT°

5. Run reaction on appropriate percentage agarose gel. If there is no insert, then the PCR product will be 258bp in size.
6. Add 1.9 mL of media to desired clones from reserved bacteria (step 3 above) for use in plasmid preps. Do your normal MiniPrep Procedure

PCR

Common Temperature Cycle (30 seconds per kb of DNA amplified):

- Step 1: 95° C 5 minutes (denature template)
- Step 2: 95° C 30 seconds (denature dsDNA)
- Step 3: 55° C 30 seconds (T_m minus 5 degrees)
- Step 4: 72° C 30 seconds (amplify about 1 kb per 30 seconds)
- Step 5: Repeat Steps 2 through 4, 29 more times
- Step 6: Store at RT°

REAGENT*	**VOLUME (µL)**	**FINAL CONC.**
water	50 - (X + Y + Z) µL	N/A
2X Master Mix (with buffer, Taq, dNTPs, $MgCl_2$**)	50 µL	1X
template DNA (~1 ng DNA)	X (≤ 1 µL)	1 ng
primer #1	Y µL	1 □M
primer #2	Z ML	1 □M
FINAL VOLUME	**100 µL**	

Appendix B: Primer Design

atggaactgaaagcgacaacgcttggaaaacgtctggcacagcacccttacgatcgggcggtgatcctcaatgccgggattaaagtctccggc
gatcgccacgaataccttattcctttcaatcaattactggcgattcactgtaagcgcggtctggtatggggcgagctggaatttgtactgccggacg
aaaaagtggtgcgtctgcacggcaccgaatggggcgagacgcagcgtttttaccatcatcttgatgctcactggcggcggtggagtggcgaga
tgagcgaaattgcgtctggtgttttacgccagcaactggatttgattgccacgcgcactggggaaaataaatggctgacgcgtgagcaaacctct

ggcgtgcagcaacaaatccgccaggctttgtcggcgttgccgttgccggttaaccgactggaagaattcgataactgccgtgaggcgtggcgt
aaatgtcaggcctggttgaaagatattgaaagcgctcggttgcagcataaccaggcgtataccgaagccatgcttaccgagtatgcggatttttc
cgccaggtcgagtcttcaccgctgaatccggcgcaggcccgggcagtcgttaatggcgagcattctctgttagtgctggcaggtgcaggaagc
ggaaaaacgtcggtgctggtggcccgtgcaggctggttgctggcgcgtggtgaagcgtcccctgagcaaattttattgctggcgtttggtcgca
aagccgctgaagagatggacgagcggattcgcgaacggctacataccgaagacattaccgcacgcacgtttcatgcgctggcgctgcatatta
ttcagcagggcagcaaaaaagttccgatagtcagcaaactggaaaatgataccgctgcccgtcatgaactctttattgctgagtggcgcaagca
atgcagcgaaaagaaagcgcaggcgaagggctggcggcaatggctgacggaagaaatgcagtggtcagtgccagaaggtaacttctggga
tgatgaaaaattacagcgtcgccttgcctctcgcctcgatcgctgggtaagtctgatgcggatgcacggtggtgcacaggcagaaatgattgcca
gtgcacccgaagagattcgcgatctgttcagtaaacgtatcaagttgatggccccgttattaaaagcctggaaaggtgcgctgaaggcagaaaa
cgctgtcgatttttcgggccttattcatcaggcgattgtgattctggagaaaggtcgctttatcagcccgtggaagcatattctggttgatgaatttca
ggatatctcgccgcagcgggcagcgttgttagcggcattacgcaagcaaaacagtcagacgacgttgttcgctgttggtgatgactggcaggc
gatttaccgattcagcggtgcgcaaatgtcgctcaccaccgctttccatgaaaactttggtgaaggcgaacgctgtgatttagacacgacttaccg
ttttaacagtcgtatcggtgaggtggcaaaccggtttattcagcagaacccaggccagctgaaaaagccgctaaacagcttaaccaatggagac
aaaaaagccgtcacgttattggatgagagtcaacttgacgctttgctggataagctctctggttatgccaaaccggaagagcgcattctgatcctg
gcgcgttaccatcacatgaggcctgccagcctggaaaaagcggcaacacgctggccgaagttgcaaatcgactttatgaccattcatgccagc
aaagggcaacaggcggattacgtcatcatcgttggcttgcaggagggaagtgatggttttccggctgcggcgcgggagtcgattatggaagag
gcgctactgccaccggttgaggatttcccggacgctgaagaacggcggttaatgtacgtggcgctgacccgggcacgccatcgggtatgggc
actgtttaacaaagagaatccctctccctttgtggaaatactgaaaaatctggatgtgccggtggcgagaaaaccg**taa**

Primer Set #1:The gene sequence is the DNA helicase IV gene from EC86 (NCBI Gene ID: 8175759). The green codon is the start codon and blue codon is the stop codon. The region highlighted in pink is where the forward primer will bind and the region highlighted in red is where the reverse primer will bind. The length of the PCR product is 153bp.

5'-tctgagttactgtctgttttccttgttggaacggagagcatcgcctgatgctctccgagccaaccaggaaacccgtttttctgac-3'

Primer Set #2*:* Primers to amplify msDNA region of the EC86 genome. Underlined region is the region of overlap between msDNA and msdRNA. The pink highlighted region is where the forward primer will bind and the red highlighted region is where the reverse primer will bind. The length of PCR product is 75bp. Sequence was obtained from Lim and Maas, 1989.

5'-aacttcggcgccttgtttgaaaaactaggcgttggatgacctaacgccgagtttttcaaacaaggcgttgacatttacattag-3'

Primer Set #3*:* Primers to amplify msDNA region of EC83. The pink highlighted region is where the forward primer will bind and the red highlighted region is where the reverse primer will bind. The PCR product length is 76bp. The underline portion indicates the area of overlap with the msdRNA region of EC83. Sequence obtained from Lim, 1992.

J3102 (pBad-RBS) DNA sequence (153bp)
acattgattatttgcacggcgtcacactttgctatgccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctcta
ctgtttctccataccgtttttttgggctagctactagagattaaagaggagaaa

*Purple highlighted portion: Forward Primer #1
*Red highlighted portion: Forward Primer #2
*Blue highlighted: Forward Primer #3
*Underlined: regions of overlap for each primer set

I715074 (EC86 Reverse Transcriptase) DNA Sequence (966bp)

atgaacaatttgcatgacatgtctaaggcgactcgcatatctgttgaaacacttcggttgttaatctatacagctgattttcgctataggatctacact
gtagaaaagaaaggcccagagaagagaatgagaaccatttaccaaccttctcgagaacttaaagccttacaaggatgggttctacgtaacattt
tagataaactgtcgtcatctccttttctattggatttgaaaagcaccaatctattttgaataatgctaccccgcatattggggcaaactttatactgaat
attgatttggaggatttttcccaagtttaactgctaacaaagtttttggagtgttccattctcttggttataatcgactaatatcttcagttttgacaaaaa
tatgttgttataaaaatctgctaccacaaggtgctccatcatcacctaaattagctaatctaatatgttctaaacttgattatcgtattcagggttatgca
ggtagtcggggcttgatatatacgagatatgccgatgatctcaccttatctgcacagtctatgaaaaaggttgttaaagcacgtgatttttttatttct
ataatcccaagtgaaggattggttattaactcaaaaaaaacttgtattagtgggcctcgtagtcagaggaaagttacaggtttagttatttcacaag
agaaagttgggataggtagagaaaaatataaagaaattagagcaaagatacatcatatattttgcggtaagtcttctgagatagaacacgttagg
ggatggttgtcatttatttaagtgtggattcaaaaagccataggagattaataacttatattagcaaattagaaaaaaaatatggaaagaaccctt
aaataaagcgaagacctaatggtcttcgttttaaaactaaagctcataggttgaaaaattgagcacttcttcgtccaac

*Underlined: region of overlap for Forward Primer #1

Forward Primer #1

5'-ctctactgtttctccataccgtttttttgggctagctactagagattaaagaggagaaaatgaacaatttgcatgacatg -3'

*Purple highlighted region is the annealing region

Forward Primer #2

5'-gccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctctactgtttctccatacc-3'

*Red highlight: region of annealing

Forward Primer #3

5'-gcat**gaattcgcggccgcttctagag**acattgattatttgcacggcgtcacactttgctatgccatagcatttttatccataag-3'

*Bolded: BioBrick Prefix

*Blue highlight: region of annealing

Primer Set #4: The first gene sequence (J3102) is pBad-RBS. The second gene sequence (I715074) is the reverse transcriptase The three primers are the three forward primers that were used to build pBad-RBS-RT. VR was used as the reverse primer for all three forward primers.

tgcgcacccttagcgagaggtttatcattaaggtcaacctctggatgttgtttcggcatcctgcattgaatctgagttact

Forward Primer:

5'-gcgcacccttagcgagaggtttatcattaag-3'

Reverse Primer:

5'-ctcagattcaatgcaggatgccgaaacaacatcc-3'

Primer Set #5: These are the EC86msdRNA primers.The underlined portion indicates the region of overlap with msDNA. The pink highlighted region indicates where the forward primer will bind and the red highlighted region indicates where the reverse primer will bind. The PCR product will be 75bp long. Sequence obtained from Lim and Maas 1989.

5'-aacttcggcgccttgtttgaaaaactaggcgttggatgacctaacgccgagtttttcaaacaaggcgttgacatttacattag-3'

Purple highlight: region where forward and reverse primers will bind.

Forward Primer:
5’-cttcggcgccttgtttgaaaaactaggcgttg-3’

Reverse Primer
5’-gtaaatgtcaacgccttgtttgaaaaactcggc-3’

Primer Set #6: These are theEC83msdRNA primers.The underlined portion indicates the region of overlap with msDNA. The pink highlighted region indicates the region where the forward primer will bind and the red highlighted region indicates where the reverse primer will bind. The PCR product will be 76bp long. Sequence obtained from Lim, 1992.