Purification of Soluble and Enzymatically Active T4 RNA Ligase 2

Jenny Y. Yoo

Colorado College

17 APR 2019

Primary Reader _____

Secondary Reader _____

ABSTRACT

In the past decade, RNA sequencing has become a notable method for identifying transcript alterations that correlate to disease development and progression. RNA-seq allows researchers to pinpoint mutated sequence elements that hinder transcriptional regulation, ultimately altering cellular functioning, and specify research questions around disease advancement. To use RNA-seq as a reliable molecular tool for mapping RNAs in diseased cells, researchers have focused on optimizing the step in which an adaptor is ligated to the RNA to be sequenced, prior to cDNA development. Although wildtype T4 RNA Ligase 2 (Rnl2) has been previously utilized for this step, its usage develops a mixture of ligated products and circularized RNA, making it an unreliable tool. C-terminus truncated T4 RNA Ligase 2 (1-249) with double mutations K227Q and R55K (DM Rnl2trunc) has been identified as an efficient tool for ligating known pre-adenylated adaptors to any RNA with reduced amounts of unwanted side products. In this study, we present the experimental parameters utilized to successfully purify a high yield of DM Rnl2trunc that is both soluble and enzymatically active from *E. coli* cultures. Growth media volume was identified as the largest contributing factor for solubilizing DM Rnl2trunc. Further, SDS-PAGE analysis suggests the recombinant protein was eluted from nickel columns using 100mM imidazole. Most importantly, our purification protocol yielded abundant DM Rnl2trunc with enzymatic activity comparable to commercially available product. Bioanalyzer data indicates both homemade and commercial DM Rnl2trunc is capable of ligating desired RNA sequences and pre-adenylated adaptors with very high efficiency and unwanted side products.

INTRODUCTION

RNA sequencing as a tool for determining the state of cell functioning

Studying changes in the type and quantity of transcriptional products that are produced as cells are exposed to shifting cellular environments can give insight into how gene expression and regulatory mechanisms are switching. One methodology that has become extremely important for studying gene expression is RNA sequencing (RNA-seq). RNA-seq is a molecular tool that can determine the nucleic acid sequence and abundance of RNAs present within a sample. Thus, RNA-seq provides information about which and how much of a specific gene is being transcribed. With this information gathered about the RNA from cells subject to varying conditions deviant from the normal cellular environment, researchers can study how gene expression has been altered, based on environmental circumstances.

RNA Ligases are critical for RNA-seq to yield high resolution sequence information about 5' and 3' mRNA ends

RNA-seq is comprised of three obligatory steps. First, RNA is extracted from the cell and the desired RNA type is selected for. For example, researchers interested in only studying mRNA can pull down mRNA with oligo-T beads and enrich the sample for transcripts with poly-A tails. Second, RNAs of interest are reverse transcribed and PCR amplified to develop cDNA libraries. Third, the resulting cDNA library is submitted for high throughput, next generation sequencing. Sequencing results provides short sequences reads of the RNA in the sample. Random 6,7,9-mers can be used to generate cDNA libraries and will give information about random portions of the RNA sequence, since these random primers are complementary to multiple regions of the RNA. Therefore, random 6,7,9-mers cannot be used to determine the sequence at the ends of the RNA. To generate detailed information about the sequence at the ends of the RNA of interest, researchers can use an RNA ligase to ligate an adaptor, a small RNA of known sequence, to the 3'-OH ends of the RNA of interest. The adaptor provides a binding site for a complementary primer that serves as a template for Reverse Transcriptase. Utilizing an adaptor for cDNA library generation ensures 3' or 5' ends will be sequenced with high detail. Researchers can then compare the resulting nucleic acid sequences to identify discrepancies in the 5' and 3' ends of RNA transcripts and determine how gene expression has been affected in different samples.

Analyzing 5' and 3' RNA ends can complement disease progression studies

Small changes in the DNA sequence can alter how the transcriptional machinery interacts with the gene and can result in the mis-regulation of gene expression. For example, transcription start site (TSS) recognition and polyadenylation site selection rely on specific consensus sequences to determine where the 5' and 3' RNA ends will form. Thus, mutations in regulatory sequences or genes involved in recognizing these consensus sequences can lead to sequence differences at 5' and 3' RNA ends. Alteration of the sequences at 5' or 3' RNA ends can change downstream regulation of the RNA and lead to disease. Sequence mutations that affect where polyadenylation factors bind can results in alternatively polyadenylated RNA isoforms with elongated or truncated 3' UTRs. There is evidence that in estrogen receptor positive (ER+) and triple negative breast cancer (BC) lines, uncontrolled proliferation results from bypassing miRNA negative regulation through alternative polyadenylation (APA)-mediated 3' UTR shortening (An et al., 2013). In fact, higher quantities of shorter isoforms in comparison to longer isoforms is considered one indicator of cancer development (Jain et al., 2018). Akman et al (2012)

demonstrated that in ER+ BC cells, estrogen-induced transcript shortening of a key factor in DNA replication, CDC6, led to an increase in CDC6 protein levels. As a result of CDC6 upregulation, more cells entered S- phase, leading to uncontrolled growth. Since alterations in 5' and 3' RNA sequence ends can promote downstream consequences that affect how cellular pathways are operating, it is important to analyze mRNA sequence changes at the 5' and 3' ends with RNA-seq to better grasp how diseases may be progressing.

T4 RNA Ligase 2 is an essential tool for studying 3' and 5' RNA ends

Since ligating an adaptor of known sequence to RNA ends is essential for creating cDNA libraries for RNA sequencing that can accurately map the 5' and 3' RNA ends, the enzyme that performs the ligation, RNA ligase, has become extensively studied. Wildtype (WT) Rnl2 has been noted for its use in developing RNA sequence libraries to map the 5' and 3' ends of RNA. Rnl2 originates from T4 bacteriophage and is involved in RNA editing, splicing and repair. The enzyme covalently joins the 5'-PO_4 group of one RNA sequence to the 3'-OH group of another transcript through a three-step enzymatic process using ATP (Figure 1) (Ho et al., 2004). Rnl2 first interacts with ATP to generate a lysyl-N bond between the lysine in the ligase's nucleotidyl transferase motif and AMP. The formation of this covalent bond releases pyrophosphate. In the second enzymatic step, the ligase transfers AMP to the 5'-PO_4 end of an RNA, creating the second intermediate, a 5' adenylated RNA. In the last enzymatic step, Rnl2 catalyzes the attack of a 3'-OH group to the 5' adenylated RNA and releases AMP, resulting in the final ligated product. Rnl2, along with DNA ligases and RNA capping enzymes, is categorized as an RNA Editing Ligase (REL). The conserved nucleotidyl transferase motif allows RELs to transfer a nucleotide, in this an adenine, to the 5'-PO_4 using an enzyme-(lysyl-N)-NMP intermediate. Rnl2's unique specificity for adenine results from interactions between the purine and the polypeptide backbone (Ho et al., 2004). Rnl2 strays structurally from other RELs in its C-terminus, which is essential for the second catalytic step, the transfer of AMP to the 5'-PO_4 (Viollet et al., 2011). On the other hand, ligase 5'-adenylation (step one) and phosphodiester bond formation at 5' adenylated ends (step three), are controlled by the N-terminus.

Although Rnl2's enzymatic activity is separated into three well defined steps, it is difficult to manipulate the activity of Rnl2 for the purposes of RNA-seq since each step is not necessary for RNA ligation and various steps are reversible. Therefore, it is difficult to control the ligation products that will result since Rnl2 can de-adenylate RNAs. During de-adenylation, the reverse of the second enzymatic step, AMP is removed from 5'-adenylated RNAs and transferred to the next available 5'-PO_4. This creates a reaction environment where both the adaptors and the RNAs

to be sequenced are adenylated at their 5' ends. Therefore, a mixture of ligation products results: the desired product where the adaptor has been ligated to the 3' end of the transcript, a second product where the adaptor is ligated to the 5' end rather than the 3' end, and circularized RNA products without the adaptor. Circularization may be attributed to the enzyme using the 3'-OH of the 5' adenylated RNA to attack its own adenylated 5'-PO_4 because of proximity. However, circularization is avoided in the presence of ATP, where high ATP concentrations allow Rnl2 to interact with another available ATP first, instead of immediately catalyzing phosphodiester bond formation between its ends (Ho et al., 2004). Since Rnl2 can create these undesired ligation products, Rnl2 is not considered an effective tool for ligating adaptors to RNAs for RNA-seq library preparation.

C-terminus truncated RNA Ligase 2 increases control over reaction environment

To solve this problem, Ho et al. (2004) demonstrated the benefits of using a truncated version of Rnl2, with intact amino acids 1-249, for increasing the amount of the correct product, RNA transcripts with the adaptor sequence ligated to its 3' end. By excluding the C-terminus, Rnl2 is expected to be incapable of catalyzing the transfer of AMP to the to the 5'-PO_4 end and creating the 5'-adenylated RNAs. However, this truncated form of Rnl2 retains its ligation activity because the N-terminus comprises a functioning adenyl transferase domain that operates independently of the C-terminus (Ho et al., 2004). Since truncated Rnl2 cannot the adenylate 5'-ends of RNAs, it can only perform RNA ligation reactions when provided a pre-adenylated RNA (AppRNA) that is supplied by the researcher. The AppRNA occupies the AMP binding pocket, which positions the AppRNA adaptor for attack by the 3'-OH end of the RNA of interest. Therefore, truncated Rnl2 (Rnl2trunc) was initially considered a superior ligation tool than WT Rnl2 since researchers could bypass the first two enzymatic steps and achieve the correct RNA ligation product, the RNA of interest with an adaptor ligated to the 3' end. Additionally, the RNA of interest, AppRNA adaptor and the ligase are the only components necessary in cDNA synthesis for RNA-seq. Since Rnl2trunc requires AppRNA, ATP is unnecessary. By restricting the number of components and steps, a more controlled *in vitro* reaction environment is created. Additionally, truncated Rnl2's sequence specificity is less than 1% WT Rnl2's specificity and is 10-fold more reactive than the wildtype enzyme (Ho et al., 2004). This makes Rnl2trunc a more flexible and efficient tool for ligating a variety of sequences and opens the potential to explore any RNA of interest without having to consider whether the sequence is compatible with the enzyme.

Problems with C-terminus truncated RNA Ligase 2

Although C-terminal truncation decreases the number of steps and reactants for ligation activity, Rnl2trunc is still able to de-adenylate AppRNA adaptors. Once AppRNA adaptors are de-adenylated, a ligase-(lysyl-N)-AMP intermediate is regenerated. AMP could then be transferred to adenylate the 5'-ends of the adaptor sequence or the RNA of interest. If the RNA of interest is adenylated, a ligation product with the adaptor sequence at the 5' end would result. Rnl2trunc retention of its de-adenylation activity suggested that more than just the C-terminus of Rnl2 is involved in adenylating/de-adenylating the 5' ends of RNA transcripts (Ho et al., 2004).

Double-Mutated Rnl2trunc is a reliable and efficient tool for ligating RNA adaptors

In 2011, Viollet et al (2011) introduced single- and double-point mutation combinations in Rnl2trunc's N-terminus to finetune and eliminate de-adenylation activity. The authors report that mutations at K227 and R55 optimize the ligation reaction by reducing 5'-adenylation activity, and thus decreasing the formation of unwanted byproducts (Viollet et al., 2011). A previous study had already suggested that K227 plays a pivotal role in Rnl2's 5'-adenylation activity by stabilizing ATP's α-phosphorus and AMP's phosphorus (Ho et al, 2004). Viollet et al (2011) further identified K227 as a key residue in de-adenylation, even in the absence of the C-terminus. A point mutation at K227 completely prevented de-adenylation activity, resulting in nearly 100% conversion to the expected product with significantly reduced circularization. Additionally, Rnl2trunc with double point mutations at K227 and R55 (DM Rnl2trunc) has comparable enzymatic activity to Rnl2trunc and is higher by 11% when compared to ligases with individual K227 or R55 mutation. DM Rnl2trunc also produces ligation products at a higher rate when compared to RNA ligases with alternative mutation combination and continues to ligate after 12 hours of reaction time. These attributes make DM Rnl2trunc an attractive tool for increasing RNA ligation efficiency, product accuracy and yield during RNA library sequence production. DM Rnl2trunc has become an extremely desirable molecular tool for RNA sequencing library preparation because of its high fidelity for generating the correct RNA ligation product with user supplied AppRNA adaptors.

Development of a DM Rnl2trunc purification protocol that solubilizes the enzyme and retains improved ligation activity

Though the enzyme is available commercially, we sought to determine the experimental conditions that are necessary for purifying soluble and enzymatically active DM Rnl2trunc recombinant proteins. Our procedure was inspired by a similar Rnl2 purification completed by Nandakumar et al (2004). We cloned DNA that encoded for Rnl2's amino acids 1-249 containing point mutations K227Q and R55K into a pEt28a plasmid, such that an N-terminus 6-HIS tag was

introduced along with controllable expression using an inducible T7 promoter. The plasmid was transformed into DH5α *E. coli* competent cells and resulting clones were confirmed by PCR and sequencing to harbor the correct plasmid. The plasmid was then transformed into Rosetta *E. coli* competent cells, a system optimal for eukaryotic protein expression. We induced the expression of the recombinant protein with IPTG and purified DM Rnl2trunc using a Nickel-affinity column. Enzymatic activity of the recombinant protein was tested and characterized.

Several problems were encountered during 6-HIS tagged DM Rnl2trunc purification, including protein solubility and retention of enzymatic activity. Sometimes, protein surplus due to overexpression is interpreted by the cell as waste and controlled by packaging proteins into inclusion bodies. Once the protein is in inclusion bodies, it is no longer soluble and cannot be purified through chromatography. Rather, the protein has to be first removed from the insoluble fraction and then induced to refold. This process can vastly reduce the chances that the protein will retain enzymatic activity, as it may refold incorrectly. We determined that the most essential factor for solubilizing DM Rnl2trunc is media volume. We demonstrate that additional factors for achieving DM Rnl2trunc solubility included induction time, media type, inoculating cell concentration, temperature control between inoculation and induction and ethanol addition. DM Rnl2trunc purified following our procedure is not only enzymatically active but also comparable to the commercially available product. Ultimately, we collected a high yield of ~1.2 mg/ml functioning recombinant protein.

RESULTS

PCR confirms DM Rnl2trunc DNA was cloned into the pET28a vector correctly. We sought to purify a high yield of enzymatically active DM Rnl2trunc to produce stocks available for laboratory use in preparing RNA-seq libraries. pUC57 plasmid with an insert encoding for DM Rnl2trunc was cut with NdeI and BamHI restriction enzymes to release DM Rnl2trunc DNA to enable subcloning into pET28a expression vectors. The 6-HIS tag, inducible T7 promoter and kanamycin (KAN) resistance made pET28a a favorable choice for induction, purification and selection steps. Ligating DM Rnl2trunc DNA into NdeI and BamHI sites of pET28a would maintain the pET28a 6-histidine (HIS) tag at the N-terminus of the induced protein and the translational frame of the enzyme (Figure 2). For the ligations, digested pET28a was dephosphorylated to prevent re-circularization and treated with T4 DNA ligase with and without the DM Rnl2trunc insert. The ligation reactions were transformed into DH5α *E. coli* competent cells and streaked onto LB+KAN plates to select for bacteria that had taken up the successful transformants. Colonies were not

expected to grow without the insert since KAN resistance should be encoded on a linearized plasmid and not expected to be propagated. No colony growth on –insert transformation alongside positive colony growth on + insert plates was the first indication that the desired DM pET28a+Rnl2trunc had been successfully subcloned (Figure 3). Sixteen colonies were chosen from +insert plates. PCR analysis tested whether colonies possessed the desired DM Rnl2trunc insert (Figure 4). T7 promoter and termination primers were used to identify if the recombinant DNA was present, which would result in a 1024 base pair band. An empty pET28a vector was used as a positive control for the PCR and was expected to produce a 314 base pair band. A single, bright band ~1000 bp was visualized for all sixteen colonies tested. The control, which lacked the 750 bp insert, produced a smaller ~300 bp band and verified that the primers had amplified the region where Rnl2 was inserted, instead of a random portion the vector. Thus, these results collectively indicated that all colonies tested contained the desired insert. Two colonies, #3 and #8, were sent for sequencing. A corresponding amino acid sequence based on the sequencing results, generated using ApE software, was compared manually to the expected amino acid sequence. This analysis further validated that clones #3 and #8 contained the pET28a+DM Rnl2trunc plasmid and that DM Rnl2trunc was successfully sub-cloned.

Ethanol addition and temperature control before DM Rnl2trunc induction slightly increases protein solubility. Clones were transformed into Rosetta *E. coli* competent cells, a system that optimizes eukaryotic protein expression, to multiply the clone and induce protein expression in preparation for purification. Cultures were inoculated and grown to an OD_{600}=0.60 in 500 mL growth media and induced with IPTG overnight. DM Rnl2trunc ligase presence can be tracked by a ~25 kDa band with SDS-PAGE, which correlates to its expected size of 24.9 kDa. SDS-PAGE with Coomassie staining analysis suggested that DM Rnl2trunc was trapped in insoluble, inclusion bodies under these growth conditions (Figure 5). DM Rnl2trunc trapped within insoluble bodies was problematic since its availability in the soluble fraction was essential for proceeding to purification. Previous studies demonstrated the benefits of adding 3% ethanol (EtOH) and controlling temperature during induction to enhance protein solubility by stabilizing native protein states (Chhetri et al., 2015; Dako et al., 2012). Cultures were inoculated in smaller volumes of growth media (30 mL) to test if controlling temperature and ethanol presence before induction could affect solubility. Cultures were also inoculated and grown to a reduced OD_{600}=0.45. After inoculation, temperature was reduced from 37°C to 0°C by allowing cultures to rest in ice water. Varying EtOH concentrations were added (0%, 1%, 2%, 3%) and protein was induced overnight. SDS-PAGE results showed that more DM Rnl2trunc was present in the soluble fraction than in

the insoluble fraction under all ethanol concentrations tested when compared to original experimental conditions (Figure 5, 6A-D). Surprisingly, there was little difference in the amount of soluble protein across differing ethanol concentrations (Figure 6A-D). Since the ultimate goal was to purify a high yield of DM Rnl2trunc to utilize in RNA sequence library development, the procedure was reproduced in large-scale induction (500 mL). However, DM Rnl2trunc solubility only increased slightly with 2% EtOH when compared to induction with no ethanol (Figure 7). Most of DM Rnl2trunc was still trapped in the insoluble fraction, indicating that EtOH and temperature control did not contribute to increasing the solubility of the ligase in large cultures.

Media volume is the most influential factor in producing soluble recombinant DM Rnl2trunc. The discrepancy in the amount of soluble fraction ligase in small-scale versus large-scale growth in the presence of EtOH results, suggests that media volume may affect protein solubility as more DM Rnl2trunc was present in the soluble fraction when the protein was inoculated and induced in a smaller volume (30 mL). To test if volume affected protein solubility and determine the most optimal volume for generating the most soluble DM Rnl2trunc, cultures were grown in 100 mL and 200 mL growth media. The media was also changed from 2X YT to LB and induction time was reduced to 5 hours to control for overexpression. Both 100 mL and 200 mL growth media volumes resulted in more DMRnl2trunc ligase in the soluble fraction than in the insoluble fraction (Figure 8). The proportion of soluble to insoluble protein in the 200 mL fraction was slightly greater than in the 100 mL fraction. Based on these results, 200 mL of LB is the optimal volume for solubilizing the most DM Rnl2trunc.

Recombinant DM Rnl2trunc is successfully eluted in 100mM imidazole. Once most DM Rnl2trunc had been successfully solubilized during the growth and induction steps, we were ready to purify the enzyme from the whole cell extract. Five 200mL cultures were inoculated and successfully induced with IPTG (Figure 9A). Cultures were combined, spun down, and the pellet was sonicated to break up cell membranes. Lysate was spun down to separate proteins from heavier cellular components. Soluble and insoluble fractions analyzed by SDS-PAGE suggested a greater proportion of the enzyme appeared in the soluble fraction of this large-scale experiment (1 liter, combined) and replicated results from the small-scale experiment (200 mL), when compared to results from previous experimental conditions (Figure 9B). Next, the soluble fraction was nutated with nickel (Ni) beads to allow the 6-HIS tag to bind nickel. Beads were washed in increasing concentrations of imidazole, which competes with histidine for Ni interactions. Lower concentrations of imidazole would compete off weak, background protein interactions to Ni beads,

while higher concentrations compete off DM Rnl2trunc. To test if DM Rnl2trunc was eluted, we assayed the fractions that were eluted with 50mM, 100mM and 200mM imidazole by Bradford reagent testing. Our results support that DM Rnl2trunc was eluted in 100mM imidazole. 100mM imidazole fractions 1-7 changed the color of Bradford reagent the most. Elution with 100mM imidazole is further supported by the absence of color change for any of the fractions in the 200mM imidazole wash, which followed the 100mM imidazole wash (Figure 10). Color change indicates the presence of protein but does not reveal protein type or fraction purity. Fraction 3 from the 100mM imidazole elution, which had the greatest Bradford reagent color shift, was run on a poly-acrylamide gel. A strong band ~25 kDa is visualized with minimal contaminating bands (Figure 9C). This results strongly supports that DM Rnl2trunc is not only eluted in 100mM imidazole but also successfully purified. Protein purification yield was established with a BSA standard curve to determine our average protein concentration as 1.175 mg/mL, yielding ~3.5 mg of DM Rnl2trunc per liter of culture.

Enzymatic activity of homemade DM Rnl2trunc is dose dependent. Ultimately, we hoped to utilize DM Rnl2trunc as a molecular tool for developing RNA sequence libraries. Thus, it was essential the protein retained its enzymatic properties after purification. To determine if purified DM Rnl2trunc was enzymatically active, an RNA ligation assay was designed (Figure 11A). If the homemade enzyme was active, it would ligate a synthetic GAPDH RNA 3'-OH end to the pre-adenylated adaptor RNA. A primer complementary to the pre-adenylated (App) adaptor would allow for specific cDNA synthesis with Reverse Transcriptase. cDNA would then be amplified with GAPDH 5' and App adaptor primers by PCR. PCR products could be visualized with agarose gel electrophoresis. Enzymatic activity of the ligase purified above was compared to the commercially available RNA ligase from *New England BioLabs*. Two concentrations of the purified enzyme, the commercially recommended concentration and a much higher concentration, were tested against the recommended concentration of commercial ligase (Figure 11B). We observed more PCR product in the reaction with the homemade ligase reaction when the same amount of commercial protein was utilized (0.025 μg/μL), suggesting that the homemade enzyme ligates more RNA effectively in comparison to the commercially available ligase. However, when a greater concentration is used (0.12 μg/μL), almost no ligation product is observed (Figure 11B). Enzyme activity seems to depend on a specific protein concentration, beyond which activity is inhibited.

Homemade enzyme activity is comparable to the commercially available product. Considering the discrepancy in ligation activity based on enzyme concentration, another ligation assay was

conducted to further characterize enzymatic activity of the homemade ligase. Five concentrations of homemade DM Rnl2trunc were tested from 0.12 μg/μL to 0.0075 μg/μL. Concentrations were halved between each reaction. The same ligation procedure mentioned above was completed to visualize the amount of resulting product from each reaction. PCR results suggest that the peak of homemade ligase enzymatic activity occurs at a concentration of 0.03 μg/μL ligase reaction with 20nmol of RNA and 5pmol of pre-adenylated RNA (20X molar excess). PCR products of similar intensity were observed for reactions with 0.3 μg of the homemade ligase and 0.25 μg of the commercial enzyme (Figure 12). Thus, we determined that the enzymatic activity of homemade DM Rnl2trunc to commercial ligase is nearly 1:1.

DM Rnl2trunc converts nearly all reactants into ligation products. Although previous ligation assays compared DM Rnl2trunc enzymatic activity to commercial ligase, it did not tell us about the efficiency of the purified ligase. To better characterize DM Rnl2trunc efficiency, a more sensitive system for detecting RNA size separation was used to determine how much reactant was converted into product. We used a bioanalyzer to evaluate the RNA after the ligation reaction (Figure 13). The no ligase lane has a single band, which likely correlates to un-ligated synthetic GAPDH RNA. App adaptor is not detected because its size is below the detection limits of the bioanalyzer. Both commercial and DM Rnl2trunc lanes have the same single band that is slightly higher than the no ligase band. This larger size suggests that these bands correspond to ligated synthetic GAPDH+ App adaptor. More significantly, there are no other bands visualized. Since there is no similar sized band that correlates to un-ligated RNA, all synthetic GAPDH RNA in the reaction is ligated to the App adaptor RNA with both the commercial and homemade ligases. Further, the presence of a single band suggests that both ligases do not produce unwanted side products.

MATERIALS AND METHODS

pET28a+ DM Rnl2trunc clone development. 5 μg of pET28a and pUC57 were digested with NdeI and BamHI in a 37°C water bath. Digested pET28a vector was isolated using phenol chloroform and de-phosphorylated with phosphatase in a 37°C water bath. De-phosphorylated pET28a vector was purified using phenol chloroform. pET28a vector and DM Rnl2trunc DNA from digested pUC57 were run on a 1% 1X TBE agarose gel and gel purified using the *Invitrogen Life Technologies "PureLink" ™* kit. 1μL of pET28a vector to 3 μL of DM Rnl2trunc DNA was ligated using T4 DNA ligase at 14.8°C cold water bath, overnight.

Verifying expected clone development. Ligation reactions were transformed into DH5α *E. coli* competent cells. Transformed cells were plated onto LB+ KAN plates and grown overnight at 37°C. Colonies were streaked onto a new LB+KAN plate and used to start cultures in LB+ KAN media. Both were grown at 37°C. Streaked colonies were confirmed by PCR amplification using 0.5μM T7 promoter and termination primers. The following cycling conditions were set: Initial denaturation- 94°C for 2 minutes (1 cycle), Denaturation- 94°C for 40 seconds (29 cycles), Annealing- 48°C for 40 seconds (29 cycles), Extension- 72°C for 1 minute (29 cycles), Final extension- 72°C for 10 minutes (1 cycle), Hold- 12°C for 10 minutes (1 cycle). PCR products were run on a 1% 1X TBE agarose gel. *Sigma-Aldrich MiniPrep* kit was used to purify the plasmid from overnight cultures. 40μM of each miniprep was submitted for *Quintara Bio* sequencing with 5μM T7 termination primer sequence.

Glycerol preparation for DH5α and Rosetta competent cells. DH5α overnight cultures for colonies were plated onto LB+KAN plates and grown overnight at 37°C. One colony from each Rosetta transformation plates were resuspended in LB media and plated onto LB+KAN plates. Plates were incubated at 37°C, overnight. 2X YT was added to DH5α and Rosetta overnight plates and bacteria was transferred with a glass rod into the appropriate glycerol tubes and glycerol was added to 30% total volume. Glycerol tubes were stored in the -80°C freezer.

Rosetta culture inoculation and DM Rnl2trunc induction. Procedure was adapted from Nandakumar et al (2004). Plasmid DMRnl2trunc_8 was transformed into Rosetta *E. coli* competent cells and plated onto LB+KAN plates. Plates were incubated overnight at 37°C. One colony was chosen from the Rosetta plates, resuspended in LB and plated onto a new LB+KAN plate and grown overnight at 37°C. For successive purification trials, the Rosetta glycerol stock was resuspended into LB, plated onto a LB+KAN plate and allowed to grow overnight at 37°C. LB was added to the overnight Rosetta plate and the bacteria was scraped up with a glass rod. 1 mL LB+ Rosetta was transferred into 200 mL LB+ 1X 30mg/ml KAN. Cultures were inoculated in a 37°C shaking incubator to OD_{600}= 0.45. Temperature was adjusted to 0°C by allowing beakers to rest on ice water for 30 minutes. Ethanol was added to 2%. Cultures were induced for five hours with 0.1mM IPTG in 17°C shaking incubator.

DM Rnl2trunc Purification. Procedure was adapted from Nandakumar et al (2004). Induced cultures were pelleted at 4000 rpm for 15 minutes. Pellet was resuspended in lysis buffer (50mM

Tris HCl pH 7.5, 0.25M NaCl, 15mM Imidazole, 10% glycerol) + protease inhibitors (1mM PMSF, 1 μg/ml Pepstatin, 1 μg/ml Leupeptin, 1 μg/ml Aprotinin, 1mM Benzamidine) and freeze dried with dry ice. Cell pellet was thawed in a 37°C water bath for 1-2 minutes, then 1mM PMSF and 0.2mM benzamidine was added. Cells were sonicated using a sonicator (program: power 8 duty cycle 40% with probe on ice, for 6 minutes total (6 cycles, 30 seconds on, 30 seconds off)). Sonication was repeated with five minutes rest on ice in between. 1% TritonX100 was added to sonicated cells and nutated at 4°C for 10 minutes. Lysate was spun down for 15 minutes at 15,000 rpm, 4°C. Soluble fraction was separated from the insoluble fraction. Insoluble fraction was stored in -80°C freezer for further analysis. 5% glycerol was added to soluble fraction. Beads were recharged with nickel according to Qiagen manufacturer instructions. Ni beads were washed with lysis buffer and centrifuged down for 5 minutes at 1600 rpm. Wash was repeated two additional times. Soluble fraction was added to Ni beads and nutated at 4°C for 1 hour. Beads were spun down for 5 minutes at 1000 rpm and washed twice in lysis buffer with the same centrifuge settings. Ni beads were resuspended in lysis buffer and poured into the BioRad column. Two additional wash steps were implemented with lysis buffer and Buffer B (50mM Tris HCl pH 7.5, 0.25M NaCl, 10% glycerol, 0.05% TritonX100) + 50mM imidazole. DM Rnl2trunc was eluted in Buffer B+100mM imidazole and 1mL fractions were collected. Peak of protein elution was tested with Bradford Reagent. Protein was concentrated by dialyzing overnight in dialysis buffer (50% glycerol, 50mM NaCl, 0.1mM EDTA, 0.1mM DTT, 10mM Tris HCl pH 7.5) at 4°C.

SDS-PAGE and Coomassie Blue Staining. Samples were resuspended in 1X Laemmeli Buffer with β-mercaptoethanol. Samples were heated in 90°C sand for 5 minutes and loaded onto a 10% Polyacrylamide gel. Empty lanes were filled with 1X Laemmeli Buffer. Gel was run in 1X Running Buffer (2.5mM Tris HCl pH 7.5, 19.2mM glycine, 1% SDS) at 120V. Surface of the gel was covered completely with Coomassie Blue Stain and rocked for 30 minutes. Stain was washed with water. De-stain was added with sponges. Gel was allowed to de-stain overnight.

BSA Standard Curve and DM Rnl2trunc Protein Concentration. 0 μg, 0.5 μg, 1 μg, 2 μg, 4 μg and 8 μg BSA with 1X Bradford Reagent were used to generate the BSA standard curve. Varying sample volumes in 1X Bradford Reagent were tested against the standard curve using a spectrometer.

Ligation assay for comparing enzymatic activity. Ligation reactions containing 20 ng synthetic GAPDH RNA and 5pmol pre-adenylated RNA (20X molar excess of GAPDH RNA), 15%

PEG8000 were setup with 0.25 μg of commercial enzyme (New England BioLabs, 50 units) or varying amounts of DM Rnl2trunc and left for 3 hours at 25°C. Excess AppRNA was removed from the ligations with *ThermoFisher* RNA Clean and Concentration -5 Kit. cDNA was synthesized with 20μM reverse primer, complementary to the App adaptor, and homemade Reverse Transcriptase (RT) "SuperScript". Reactions proceeded for 60 minutes at 50°C. RT products were amplified with 10μM 5' synthetic GAPDH forward primer and 10μM App adaptor reverse primer using Taq polymerase. The following cycling conditions were set: Initial denaturation- 94°C for 2 minutes (1 cycle), Denaturation- 94°C for 30 seconds (30 cycles), Annealing- 53°C for 40 seconds (30 cycles), Extension- 72°C for 1 minute (30 cycles), Final extension- 72°C for 1 minute (1 cycle), Hold- 12°C. PCR products were run on a 1.2% 1X TBE agarose gel and visualized under UV light.

Bioanalyzer. RNA ligation products and Agilent RNA buffer were added 2:1 in volume, respectively. Samples were vortexed to mix, spun down, heated at 70°C for three minutes and cooled in ice water for 2 minutes. Ligation products were loaded onto the RNA chip and analyzed on the Agilent Bioanalyzer.

DISCUSSION

Purifying enzymatically active DM Rnl2trunc in the lab makes efficient RNA-seq more accessible. Since DM Rnl2trunc will ligate any RNA strand of interest to a pre-adenylated adaptor, the range of RNA types that can be analyzed becomes limitless. In this study, we determined the conditions for solubilizing and purifying enzymatically active DM Rnl2trunc that is comparable in activity to the commercially available ligase. In our hands, protein overexpression was successfully limited by lowering OD_{600}, reducing IPTG concentration and switching to a less rich growth media. We successfully solubilized DM Rnl2trunc when cultures were grown to no more than OD_{600}=0.45 and induced with an IPTG concentration no more than 0.1mM. We were also more successful when using LB media for growth rather than 2X YT. Surprisingly, media volume was the most important factor for solubilizing DM Rnl2trunc. We determined transformed Rosetta competent cells containing the recombinant DM Rnl2trunc clone should not be inoculated and induced in more than 200mL growth media. It is possible that increased cellular oxygenation in the smaller media volume accounts for increased DM Rnl2trunc solubility, since changes in cellular metabolism, which consequently affect gene expression, is correlated to oxygen levels (Al-ani et al., 2018). Ethanol addition to 2% and temperature control before induction slightly enhances solubility but does not have a large effect.

With the conditions determined for purifying active DM Rnl2trunc, the next steps would be to test the effectiveness of the enzyme across various RNA-sequencing techniques, including 3' UTR and TSS mapping. DM Rnl2trunc is becoming an essential RNA-seq tool for researchers who wish to target the mapping and analysis of 5' and 3' RNA ends. Unlike WT T4 RNA ligase 2, utilization of DM Rnl2trunc results in a highly controlled ligation environment by eliminating 5'-adenylation and de-adenylation activities. The desired ligation product necessary for analyzing unknown RNA sequences with RNA-seq is achieved with fewer reactant components and greater confidence in success. These benefits make DM Rnl2runc an efficient tool for RNA ligation (Viollet et al., 2011).

RNA-seq can be implemented to gather a multitude of information, including the status of transcription, the occurrence of SNPs, gene expression variances across samples and alternative splicing events. Utilizing DM Rnl2trunc adds to the list of RNA-seq applications by allowing researchers to analyze RNA ends, where many regulatory elements are clustered. Transcription initiation and polyadenylation can change the 5' and 3' ends of mRNAs. Altering either process has been associated with disease occurrences (Li and Capetanaki, 1994; An et al., 2013; Akman et al., 2012). To identify if changes in the ends of mRNA can result in gene mis-expression that lead to disease, RNA sequences for diseased cell lines obtained from RNA-seq can be compared to RNA sequences of WT cell lines. New questions concerning the mechanisms behind disease progression can be raised and sought, allowing researchers to begin piecing together a story for the disease's mechanism. For example, mutations in the 5' end of the gene that prohibit TF binding may deregulate the cell cycle, leading to tumor advancement. DM Rnl2trunc complements this research by acting as an efficient tool for achieving highly predictable, desired ligation products with minimal contaminating products.

Determining which mRNA sequence changes are correlated to disease progression provides opportunities to identify which factors have a critical role in regular cell development and leads to a more comprehensive understanding of cellular mechanisms important for disease progression, which can be used for disease treatment. Researchers can begin to consider the earliest mechanistic step diseases should be targeted to eliminate or reverse effects. This more pinpointed methodology for addressing disease progression may also decrease the amount of side effects experienced with current medical drug options. Alternatively, medical experts can begin to consider gene editing with CRISPR technology to reverse sequence mutations and restore wildtype function.

Dr. David Bentley has suggested a preliminary procedure for identifying 5' transcription start sites with RNA-seq, utilizing DM Rnl2trunc. This procedure consists of pulling down polyA

RNA using oligo-T beads to separate coding RNA. Additionally, Xrn1 is added to digest unstable RNAs that are uncapped and deadenylated to decrease the amount of noise in cDNA library development. A small adaptor is ligated to the AppRNA using DM Rnl2trunc. cDNA is generated using a random, universal 9-mer primer and amplified with PCR using the same universal primer and a primer complementary to the adaptor. Subsequent RNA-seq data provides information about the nucleotide transcript of the 5' region, including transcription start site and RNA regulatory elements in the 5' UTR.

Alternative polyadenylation (APA) is another regulatory system that has been identified as a large contributor to disease. APA leads to the formation of different mRNA isoforms that may change the amount, the activity and/or localization of the protein. For example, the translation of mRNAs can be regulated by miRNAs that bind to complementary regions in 3'-UTRs. When the 3' end of RNAs are truncated through APA, miRNA binding sites are eliminated, and genes are overexpressed. RNAs can also be elongated through APA, possibly exposing more miRNA binding sites. With DM Rnl2trunc, researchers can map 3' RNA ends to identify which mRNA isoform is present in the cell and which miRNA regulation is bypassed or overactive through APA-mediated truncation or elongation.

Mapping 3'RNA ends with DM Rnl2trunc has become of interest in light of a recent study that showed APA is associated with breast cancer progression (Begik et al., 2017). Breast cancer (BC) is the most commonly diagnosed disease in women (American Cancer Society, 2017). According to the American Cancer Society (2017), it is the second leading cause of cancer-associated death. Within a woman's lifetime, there is a 1 in 8 chance she will be diagnosed with BC and a 1 in 37 chance she will pass away from this disease. In BC cells, uncontrolled proliferation is mediated by bypassing miRNA negative regulation through APA-mediated 3' UTR shortening. Alternative polyadenylation in BC cells can be analyzed by utilizing DM Rnl2trunc to ligate a known adaptor to the 3' end of BC RNAs. cDNA can then be developed and amplified for sequencing. In their 2013 paper, *Hoque et al.* describes an improved method for mapping 3' ends that troubleshoots for internal priming and oligo-A tail complications (Hoque et al., 2013). The authors developed this new technique, 3' Region Extraction and Deep Sequencing (3' READS), by excluding oligo(dT) in reverse transcription and sequencing steps and separating RNAs with longer polyA tails. Using 3'READS along with DM Rnl2trunc can optimize RNA sequencing results for BC study in the future. By determining which genes are overexpressed in BC when miRNA regulation is bypassed, researchers can design novel miRNAs to control tumor development. One difficulty is specifying miRNAs to the target gene and eliminating promiscuous binding to alternative genes. Yet, miRNA therapy is a subject of growing popularity, especially as miRNA

association with several diseases becomes more concrete. As miRNA become a more viable treatment option in the future, RNA sequencing with DM Rnl2trunc will become standard for mapping 3'UTRs and identifying miRNA binding site sequences that are critical for cell regulation and maintenance.

TABLES AND FIGURES

Table 1 Glycerol Stocks of *E. coli* competent cells transformed with DM Rnl2trunc clone

Strain	Plasmid	Insert
DH5α	pET28a	T4 RNA Ligase 2, truncated (1-249) between NdeI and BamHI
Rosetta		

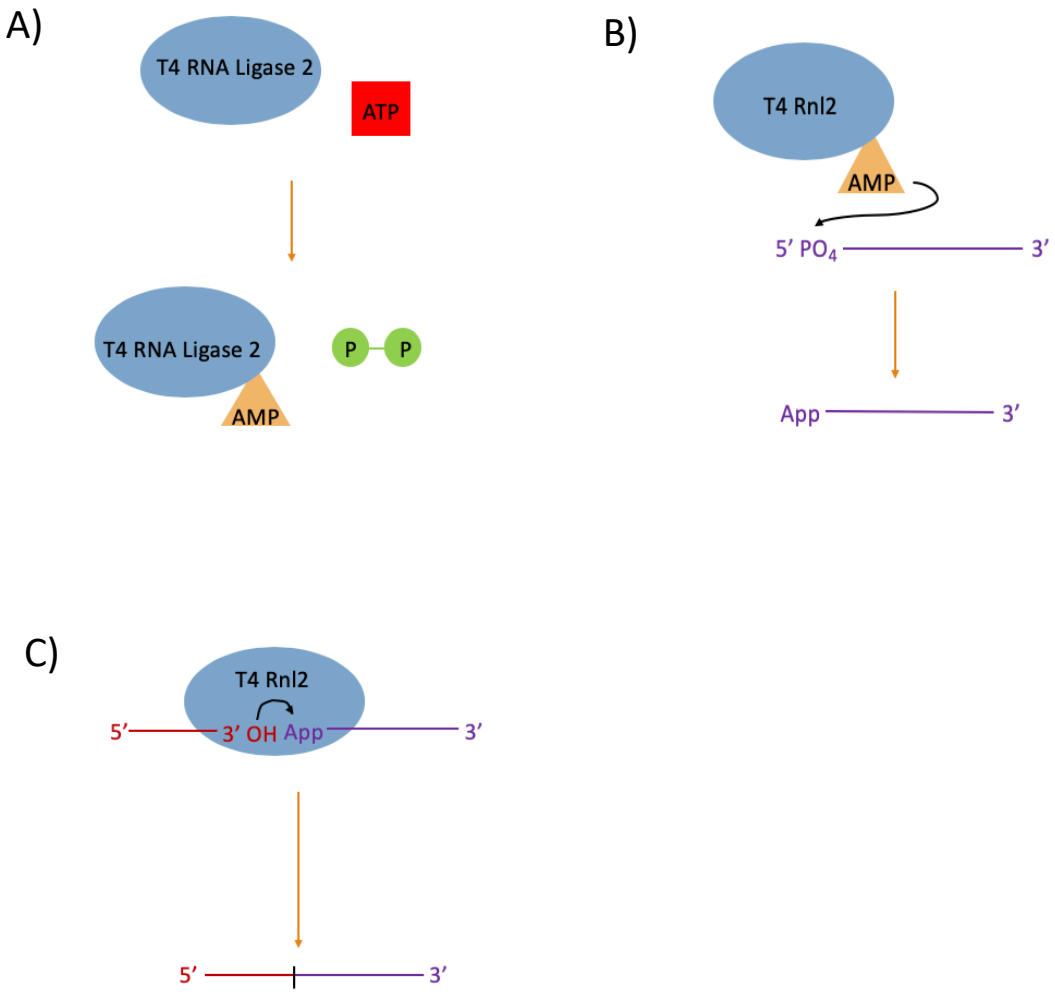

Figure 1 *Rnl2 ligates RNA strands using a 3-step enzymatic process.* A) During the first enzymatic step, Rn2 interacts with ATP to create a ligase-(lysyl-N)-AMP intermediate and release pyrophosphate. B) In the second enzymatic step, Rnl2 transfers AMP to the 5'-end of an RNA to create the second intermediate, a 5'-adenylated RNA. C) During the third enzymatic step, Rnl2 catalyzes the attack of the 3'-OH end of another RNA to the 5'-end of the adenylated RNA, ligating the two strands together.

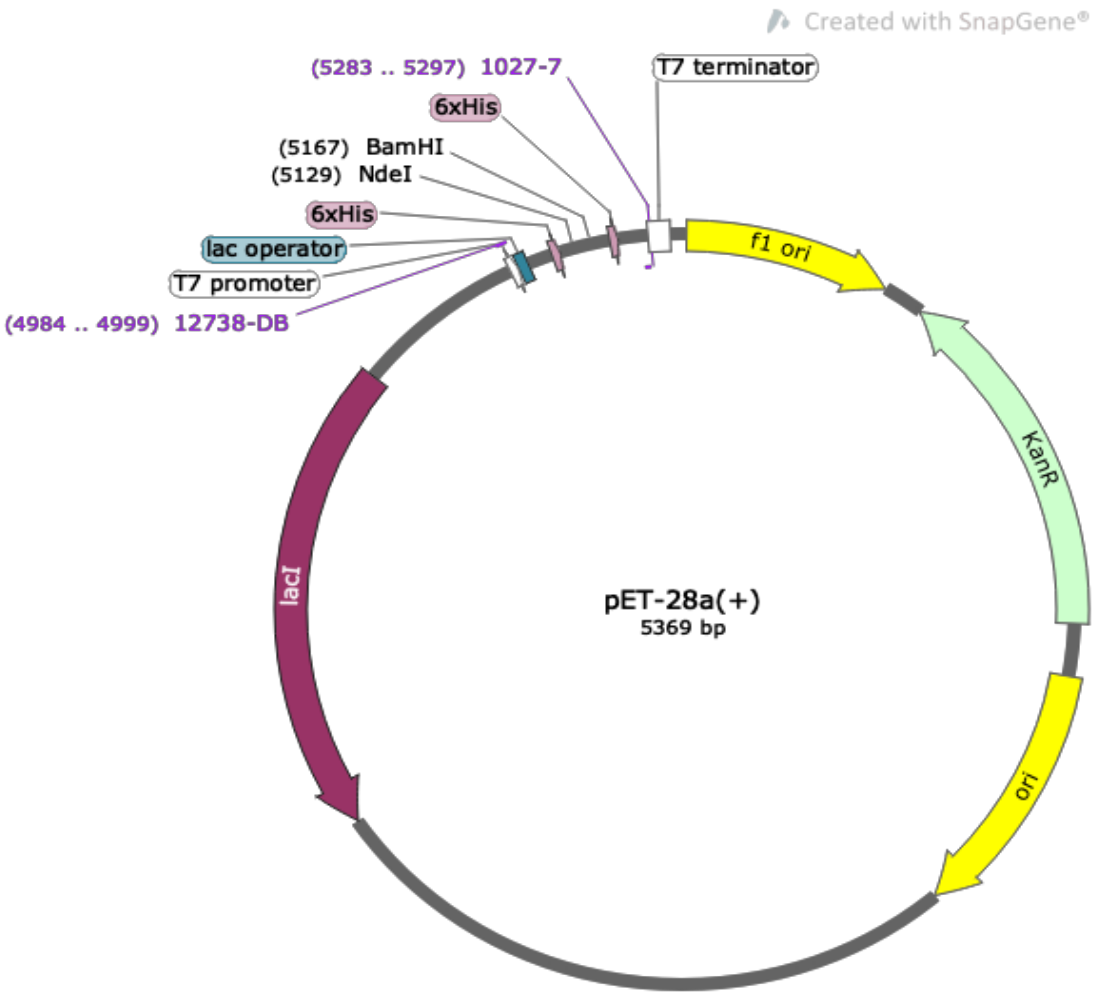

Figure 2 *DM Rnl2trunc inserted between NdeI and BamHI sites maintain N-terminus 6-HIS tag and translational frame.* pET18a was chosen for its kanamycin resistance, 6-HIS tag, and inducible T7 promoter. 12738-DB and 1027-7 are the forward and reverse primers, respectively, utilized to determine if DM Rnl2trunc DNA had been inserted into the vector, as expected.

Rnl2 DNA	+	-
pET28a	+	+
DNA Ligase	+	+

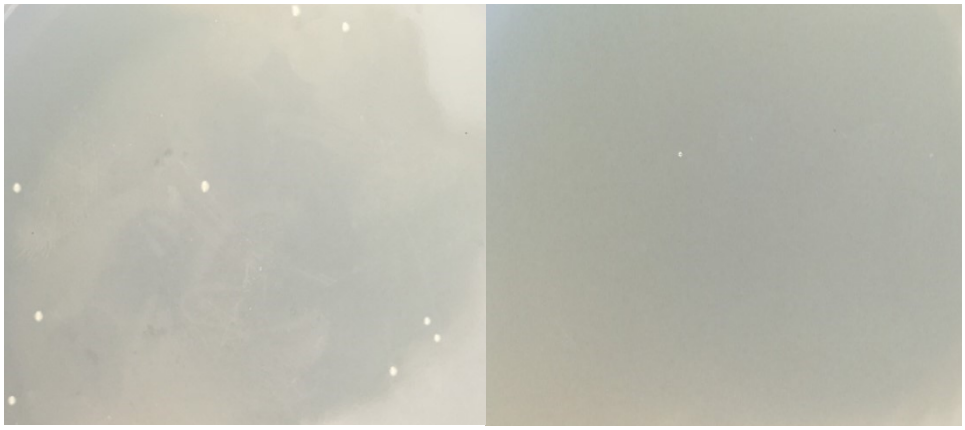

Figure 3 *Colony growth on LB+KAN plates is a primary indicator of successful subcloning.* Two ligation reactions, one with Rnl2 DNA and one without Rnl2 DNA, were completed. pET28a ends were phosphatized and thus not expected to re-circularize without presence of Rnl2 DNA. Un-circularization compromises the plasmid's integrity and kanamycin resistance is lost. Thus, colony growth of +Rnl2 plasmids following transformation into DH5α competent cells indicates that Rnl2 has been successfully inserted into the pET28a vector.

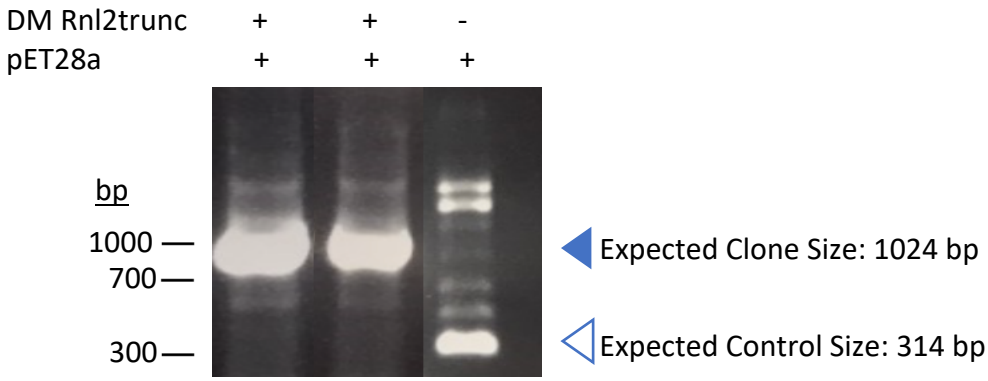

Figure 4 *PCR confirms development of DM Rnl2trunc+pET28a clone.* The expected clone size is 1024 base pairs. T7 promoter and termination primers were utilized to amplify the PCR product. All sixteen colonies that were tested contained the expected clone size. Colonies three and eight, depicted above, were chosen for continued experimentation. pET28a vector (-insert) served as a positive control to ensure experimental conditions did not affect outcomes visualized. The expected control product is 314 bp. Other bands visualized in the control lane may be due to nonspecific binding.

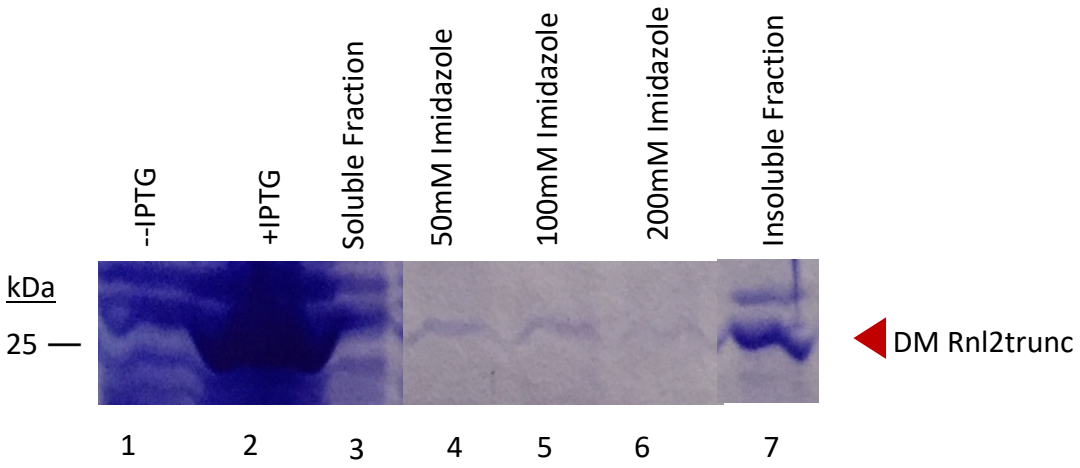

Figure 5 *Initial inoculation and induction conditions are impermissible for solubilizing DM Rnl2trunc.* DM Rnl2trunc (1-249) is indicated by the ~25 kDa band. The recombinant protein was strongly induced but is faintly present in the soluble fraction. Almost no protein was purified. However, there is preliminary evidence that the protein is eluted in 100mM imidazole. Most of the protein remains in inclusion bodies.

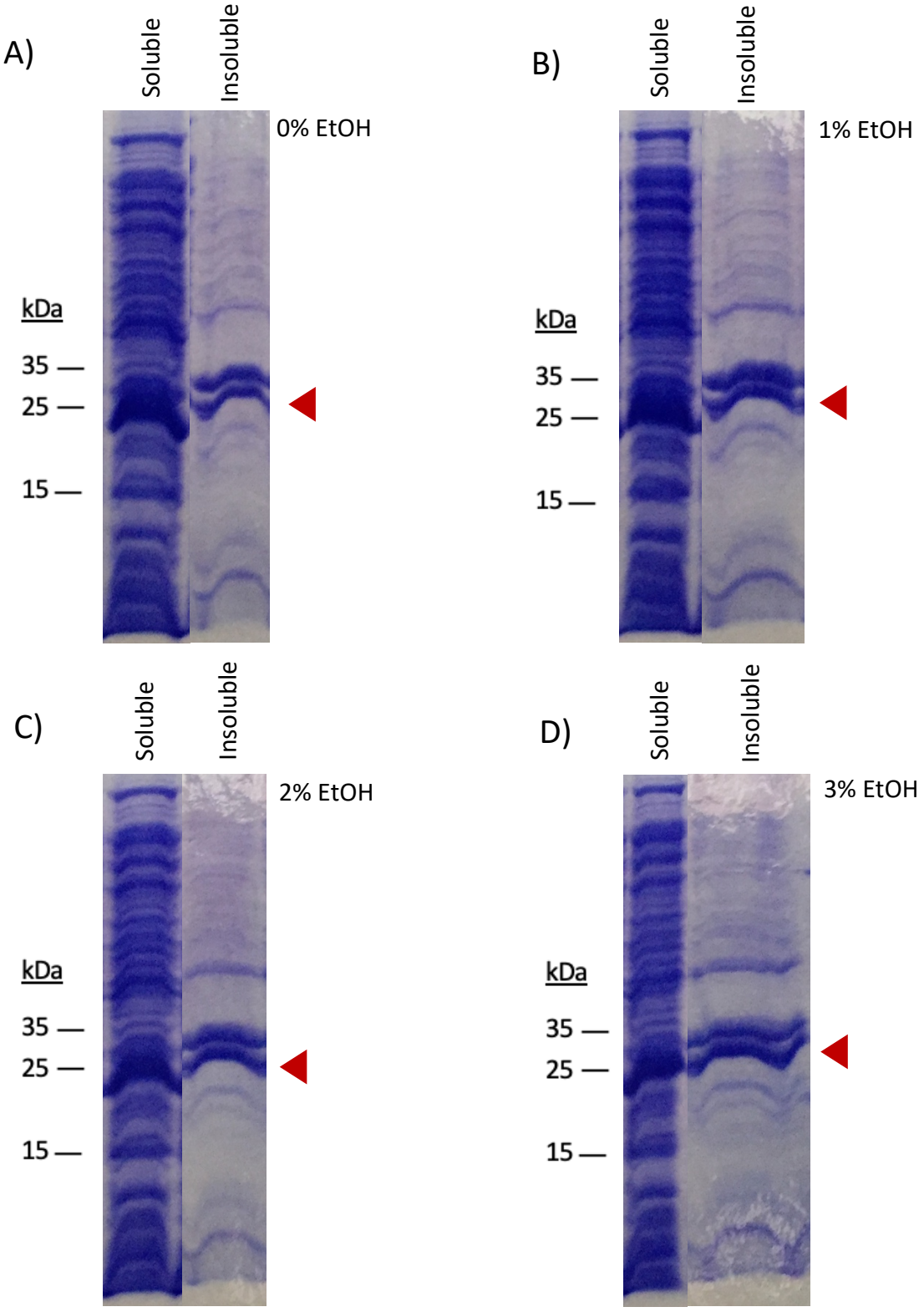

Figure 6 *Ethanol addition and temperature control before induction in smaller media volumes slightly increases DM Rnl2trunc solubility.* A) No ethanol is added before induction. B) Ethanol is added to 1% the total volume of growth media before induction. C) Ethanol is added to 2% the total volume of growth media before induction. D) Ethanol is added to 3% the total volume of growth media before induction.

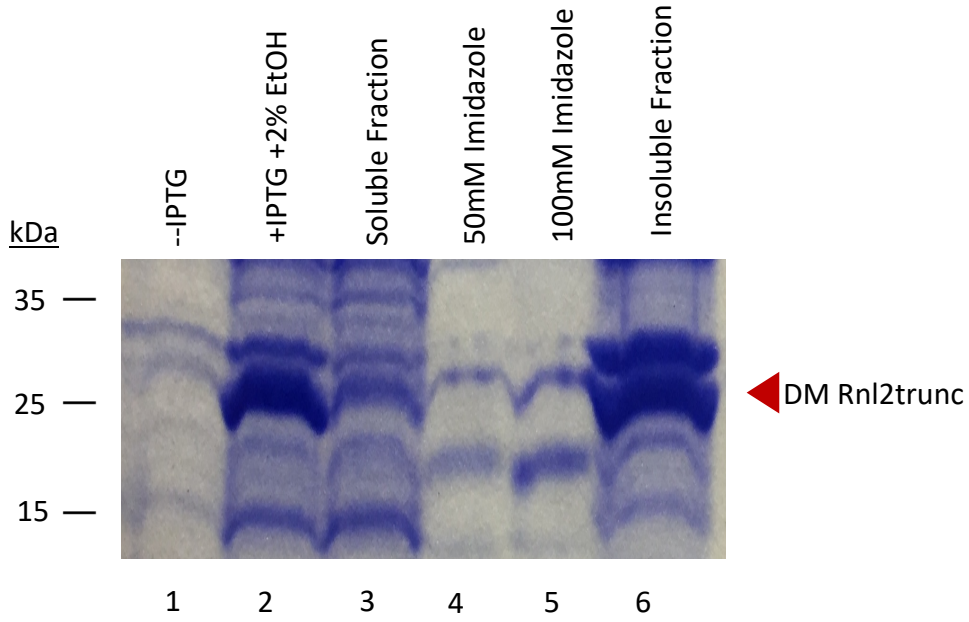

Figure 7 *Protein solubility is greatly reduced when cultures are inoculated and induced in 500 mL growth media, with ethanol addition and temperature control, in comparison to small-scale assays.* Most DM Rnl2trunc is retained in the insoluble fraction even with ethanol addition and temperature control before induction. However, more DM Rnl2trunc is visualized in the soluble fraction than under initial conditions. Further, darker DM Rnl2trunc bands are observed during purification with imidazole when compared to the bands seen under initial conditions. It seems that more recombinant protein is eluted in 100mM imidazole than in 50mM imidazole. Comparison of large-scale experiments under initial conditions and under modified conditions (+EtOH, temperature reduction to 0°C) to small scale assays suggests that media volume may be the largest contributing factor to solubilizing DM Rnl2trunc.

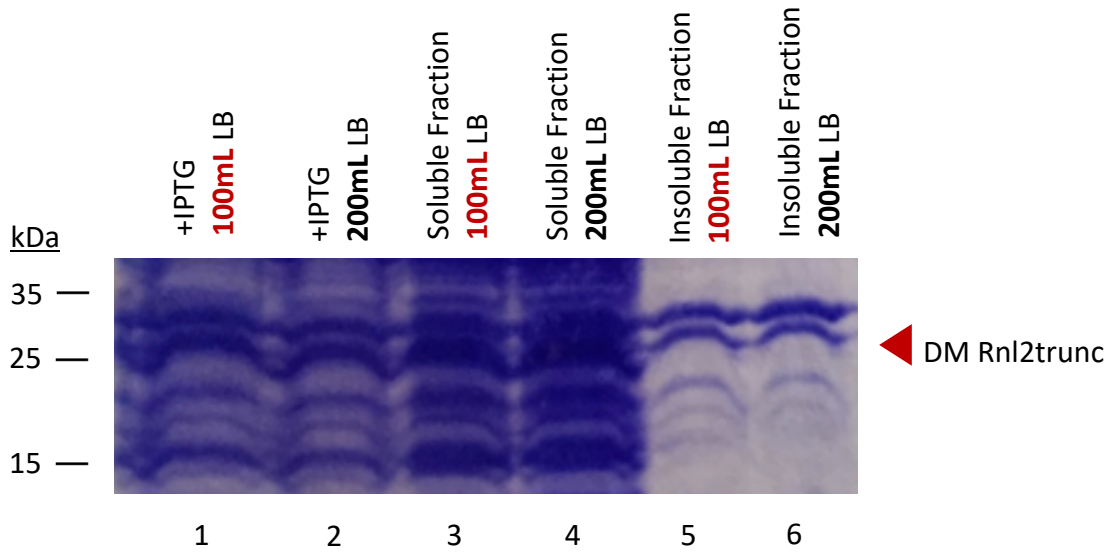

Figure 8 *Inoculating and inducting cultures in 200mL growth media is the optimal volume for solubilizing the most DM Rnl2trunc.* Bands of similar intensity are visualized after inductions in both 100mL and 200mL, suggesting a similar amount of protein is initially present. However, a slightly darker band is observed in the 200mL soluble fraction in conjunction with the 100mL soluble fraction. A significantly lower amount of protein is present in both insoluble fractions when compared to previous experimental results.

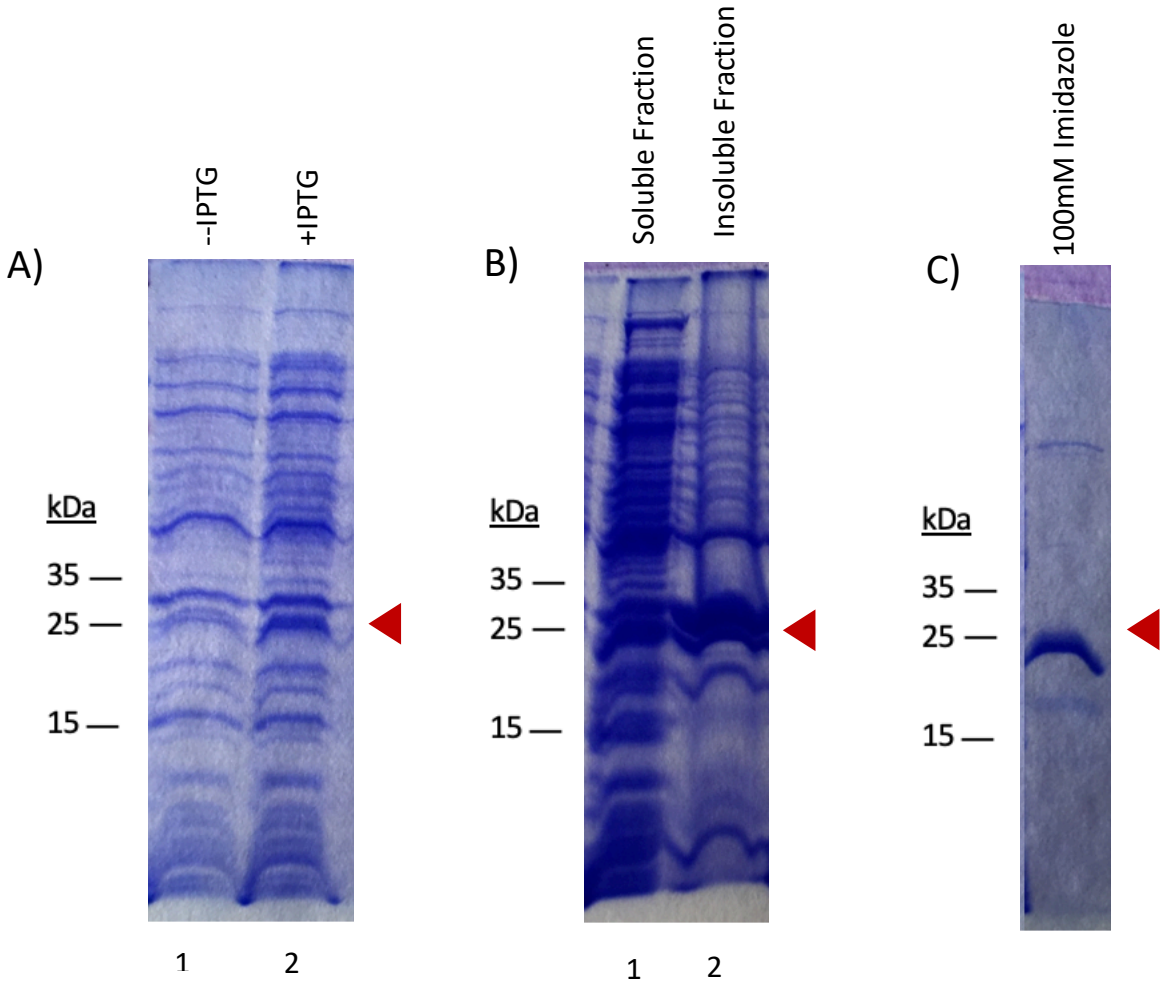

Figure 9 *SDS-PAGE results suggest DM Rnl2trunc is eluted in 100mM imidazole.* A) The expected protein size is indicated by the red arrow. Left lane shows uninduced protein extracts and right lane depicts proteins after induction with IPTG. Stronger ~25 kDa protein band after IPTG addition implies DM Rnl2 expression was induced. B) Left lane contains soluble fraction proteins. Right lane presents insoluble fraction proteins. Most DM Rnl2trunc is present in the soluble fraction. Some protein remains in insoluble bodies. C) DM Rnl2trunc is eluted with 100mM of imidazole in Buffer B. A strong ~25 kDa band remains with very few contaminating bands. This result signifies successful purification of the recombinant protein.

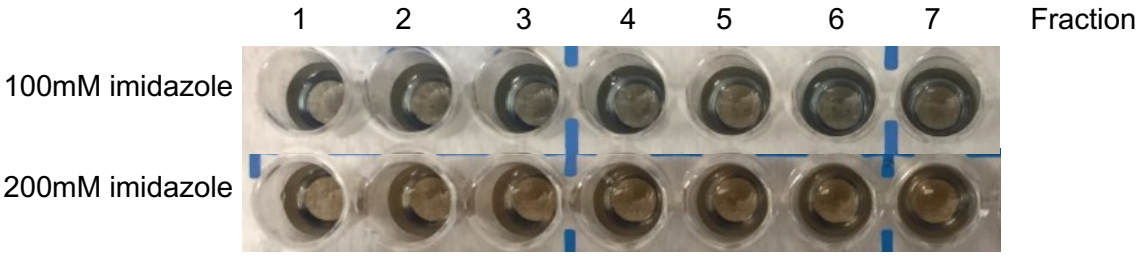

Figure 10 *DM Rnl2 is eluted from Ni-affinity column with 100mM imidazole.* Fractions from the 100mM and 200mM imidazole elations were tested with Bradford Reagent, a color indicator for protein presence. Fractions 1-7 of the 100mM elution turned bluest in the Bradford. None of the fractions from the 200mM elution changed the color of the Bradford.

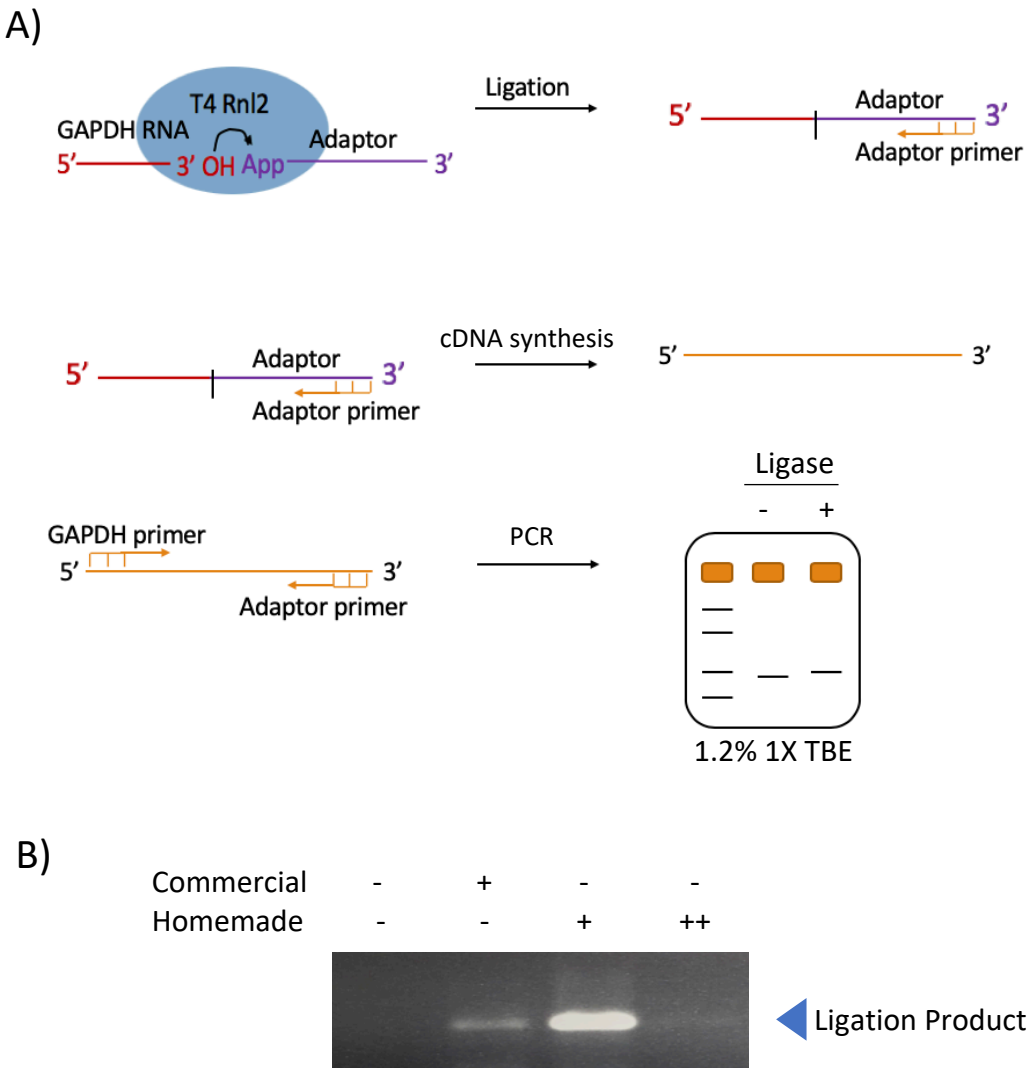

Figure 11 *Enzymatic activity of DM Rnl2trunc is dose-dependent.* A) Diagram of the ligation assay completed to determine enzymatic activity integrity. Synthetic GAPDH RNA was ligated to a pre-adenylated adaptor. B) 0.025 μg/μL of the commercial ligase (recommended concentration) is utilized for ligation assay in comparison to 0.03 μg/μL of the homemade enzyme. A stronger ligation product band is visualized for the homemade enzyme. Larger concentration of homemade enzyme (0.12 μg/μL) depicts almost no ligation product.

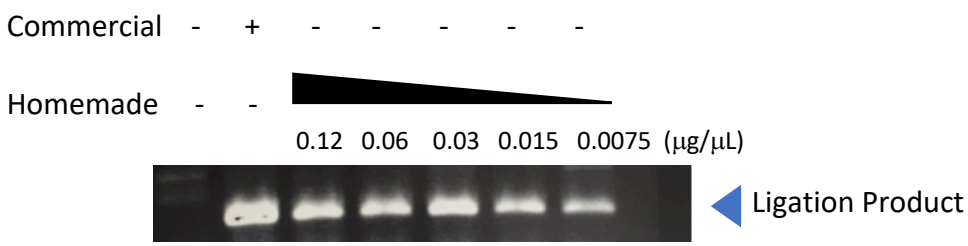

Figure 12 *A protein concentration of 0.3 μg/μL obtains optimal results.* Decreasing concentrations of the homemade ligase is tested in ligation assays to characterize protein activity. A concentration of 0.3 μg/μL (lane 5) yields the strongest ligation product band. Using this concentration of the homemade ligase results in a comparable amount of ligation product as when the recommended concentration of commercial ligase is employed (0.25 μg/μL). Enzymatic activity of the homemade ligase to the commercial ligase is 1:1.

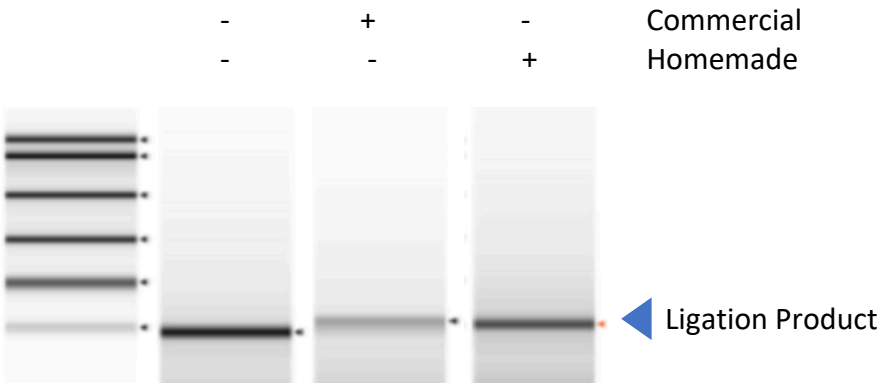

Figure 13 *Bioanalyzer results supports homemade DM Rnl2trunc converts almost 100% of reactants into the ligation product.* Only one band visualized in the + ligase reaction lanes, which are larger in size when compared to the – ligase lane band, suggests almost all reactants were converted to the expected product. According to bioanalyzer results, the unligated product is 179 nucleotides. The ligated products with commerical and homemade enzymes are 226 and 225 nucelotides, respectively.

WORKS CITED

Akman et al (2012). Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res.* 40(21): 10679-88

Al-ani et al (2018). Oxygenation in cell culture: Critical parameters for reproducibility are routinely not reported. *PLOS ONE.* 13(10): e0204269

American Cancer Society. (2017, July 1). *How Common is Breast Cancer?* Retrieved from hhttps://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html.

An et al (2013). A dynamic interplay between alternative polyadenylation and microRNA regulation: Implications for cancer. *International Journal of Oncology.* 43: 995-1001

Arcasoy et al (1997). High levels of human γ-globin gene expression in adult mice carrying a transgene of deletion-type hereditary persistence of fetal hemoglobin. *American Society for Microbiology.* 17(4): 2076-2089

Begik et al (2017). Alternative Polyadenylation patterns for novel gene discover and classification in cancer. *Neoplasia.* 19(7): 574-582

Chhetri et al (2015). An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli. *MethodsX*. 2: 385-91

Eichbaum et al (1994). Restriction of interferon γ responsiveness and basal expression of the myeloid human FcγR1b gene is mediated by a functional PU.1 site and a transcription initiator consensus. *J. Exp. Med.* 179: 1985-1996

Etienne Dako, Anne-Marie Bernier, Adjéhi Thomas Dadie and Christopher K. Jankowski (2012). The Problems Associated with Enzyme Purification, Chemical Biology, Prof. Deniz Ekinci (Ed.), ISBN: 978-953-51-0049-2, InTech, Available from: http://www.intechopen.com/books/chemical-biology/the-problems-associated-with-enzyme-purification

Ho et al (2004). Structure and mechanism of RNA ligase. *Structure.* 12: 327-339

Hoque et al (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature Methods.* 10(2): 133- 139

Jain et al (2018). The role of alternative polyadenylation in cancer progression. *Gene Reports.* 12: 1-8

Li and Capetanaki (1994). An E boc in the desmin promoter cooperates with the E box and MEF-2 sites of a distal enhancer to direct muscle-specific transcription. *The EMBO Journal.* 13(15): 3580-3589

Miles et al (2016). Alternative Polyadenylation in triple-negative breast tumors allows NRAS and c-JUN to bypass PUMILIO posttranscriptional regulation. *Cancer Res.* 76(24): 7231-7241

Nandakumar et al (2004). RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA Ligase 2. *The Journal of Biological Chemistry.* 279(30): 31337-31347

Viollet et al (2011). T4 RNA Ligase 2 truncated active site mutants: Improved tools for RNA analysis. *BMC Biotechnology*. 11: 72

ACKNOWLEGMENTS

I would like to thank Dr. David Bentley for mentoring me and providing this wonderful opportunity. I give thanks to the members of the Bentley Lab, Nova Fong, Dr. Michael Cortazar, Ryan Sheridan, Ben Erickson and Dr. Tassa Saldi, for their willingness to support and contribute feedback during the course of this project. I would like to extend my gratitude especially to Nova Fong, who was a constant resource and secondary mentor. I acknowledge Dr. Jennifer Garcia and Dr. Darrel Killian for their roles in aiding me during the writing and revision of this thesis. This research was supported by University of Colorado, Denver Anschutz Medical School and the UC Denver RNA Bioscience Initiative.