Proceedings of GREAT Day

Volume 2021 Article 13

2022

Identifying Potential RNA Binding Domains in the Thumb Region of R2 Protein

Brooke Demetri
SUNY Geneseo

Jessica Palmeri
SUNY Geneseo

Identifying Potential RNA Binding Domains in the Thumb Region of R2 Protein

Cover Page Footnote

sponsored by Varuni Jamburuthhugoda, PhD

Identifying Potential RNA Binding Domains in the Thumb Region of R2 Protein

Brooke Demetri

Jessica Palmeri

sponsored by Varuni Jamburuthhugoda, PhD

Abstract

Transposable elements are selfish mobile genetic elements able to replicate in the host genome and are classified as either DNA type elements or retrotransposons. In our study, we focus on R2 retrotransposable elements. Retrotransposable elements can reverse transcribe an RNA intermediate into DNA either before or during integration into the target genome. The R2 element exclusively inserts in the 28S rRNA genes via the mechanism of target primed reverse transcription (TPRT). For the TPRT mechanism to occur, the 5' and 3' ends of the RNA intermediate must bind to R2 protein before cleavage and insertion into a new genomic site can occur. Despite its importance in TPRT, RNA binding sequences of the R2 protein are not well understood. The objective of this study was to create single alanine replacements via site-directed mutagenesis in both the RYGLV and KPQQR sequences, which are highly conserved in the thumb domain of the R2 protein, and to isolate this mutated R2 protein for use in future assays. By examining the RNA binding properties of the R2 protein, we can further understand the TPRT mechanism and its overall role in retrotransposon success.

Background

When considering the importance of DNA, it is first regarded for its instructional role in the production of proteins, which is critical to sustaining life. However, only 1.5% of the human genome actually encodes for proteins, meaning that 98.5% of the human genome is composed of non-protein encoding genes (Gregory, 2005). Of these 98.5% of non-coding genes, roughly 45% are composed of transposable elements (TE), sometimes referred to as "mobile genetic elements," "jumping genes," "selfish DNA," and "junk DNA" (Belancio et al., 2009). TE are mobile genetic DNA sequences that are able to insert themselves into different sites throughout

the genome and are present in almost all prokaryotic and eukaryotic organisms, accounting for vast amounts of genetic material.

TE are classified as either DNA type elements or retrotransposons, as these respective subtypes differ in the mechanism of which they transpose themselves. Whereas DNA type elements move via a DNA intermediate, retrotransposons move via an RNA intermediate, through a mechanism known as retrotransposition. Retrotransposons encode their own reverse transcriptase (RT), allowing them to reversely transcribe the RNA intermediate into cDNA when transposing themselves to the target site (Brooker, 2018). Retrotransposons are further classified into two groups: Long Terminal Repeat (LTR) and non-LTR retrotransposons. For our study, we are focusing on non-LTR retrotransposons, specifically R2 retrotransposable elements.

Retrotransposons are highly abundant and found in many eukaryotic genomes. In fact, retrotransposons compose roughly 41.8% of the human genome, with non-LTR retrotransposons representing 17% of the human genome, highlighting the importance and applicability of retrotransposon research in the study of human genetics (Cordaux & Batzer, 2009; Eickbush & Jamburuthugoda, 2008). In fact, understanding the way that retrotransposons move and insert within the human genome has important applications in medicine because many types of cancers and other diseases can be caused by insertion of these TEs within critical genes. For example, a study found that frequent somatic insertion of L1 elements (abundant TE in the human genome) into critical genes may play a role in lung-tumorigenesis (Iskow et al., 2010).

Additionally, there are many applications of TE research in the field of human genetics, such as its promise in gene therapy as a gene delivery tool. R2 retrotransposable elements exclusively insert into a conserved region of the 28S rRNA genes (Eickbush et al., 2013). Since this R2 28S rRNA insertion site is conserved across all eukaryotes, including humans, this integration consistency allows for detailed analysis of its retrotransposition mechanism, thereby helping researchers further understand how the R2 integration mechanism can be used as a potential gene delivery tool (Christensen et al., 2006; Jamburuthugoda & Eickbush, 2014). Therefore, by studying the R2 integration mechanism, it can provide insight into how other retrotransposons with similar integration mechanisms integrate into the genome and how this could be exploited for gene delivery purposes. The goal of our study is to learn more about the integration mechanism of the R2 element, specifically, the domains of the R2 protein that are important for RNA binding and the R2 elements' overall success. In our study, we are using the R2 protein encoded by *Bombyx mori* to study the integration mechanism of the R2 element into the host genome.

R2 elements encode a single open reading frame (ORF) with a central RT domain, C-terminal restriction-like endonuclease (RLE) and cysteine-histidine rich domain, and N-terminal zinc-finger (ZF) and Myb nucleic-acid binding domains (Jamburuthugoda & Eickbush, 2014; Khadgi et al., 2019). When the R2 element gets inserted into a new location, the 28S rRNA gene in the eukaryotic chromosome is first transcribed into mRNA in the nucleus. This mRNA then leaves the nucleus and enters the cytoplasm

where the R2 RNA is translated into protein by the ribosomes. After translation, the newly formed R2 protein binds to the same RNA strand that it was translated from, forming an integration-competent ribonucleoprotein (RNP) complex that goes back into the nucleus and inserts into a new site within the 28S rRNA gene in the eukaryotic chromosome. R2 integrates via the mechanism of target primed reverse transcription (TPRT), where target DNA is first cleaved at the bottom strand by the upstream bound R2 protein subunit and the released 3' OH group is used to prime the reverse transcription of the element RNA onto the target site (Khadgi et al., 2019). Next, the downstream bound R2 protein subunit cleaves the top DNA strand and the released 3' OH group is used to prime second-strand DNA synthesis, displacing the RNA strand as the newly synthesized DNA is extended. This newly synthesized double stranded DNA is then inserted at the target site, and the nicks that were created during the double stranded breaks in DNA are repaired by a host repair mechanism (Yamaguchi et al., 2015).

Important to the TPRT mechanism is the activity of the R2 protein, which directs RNA binding of the protein upstream and downstream of the target site upon association with the 3' and 5' end of R2 RNA, respectively, as well top and bottom DNA strand cleavage to prime reverse transcription (Jamburuthugoda & Eickbush, 2014). In our study, our goal was to mutate two highly conserved regions within the thumb domain of the R2 protein to identify whether these specific regions are important in R2 RNA binding of the R2 protein, and therefore can be classified as an RNA binding domain. If the mutant R2 protein has decreased TPRT activity compared to that of the wild type R2 protein, then these conserved regions can be identified as an RNA binding domain.

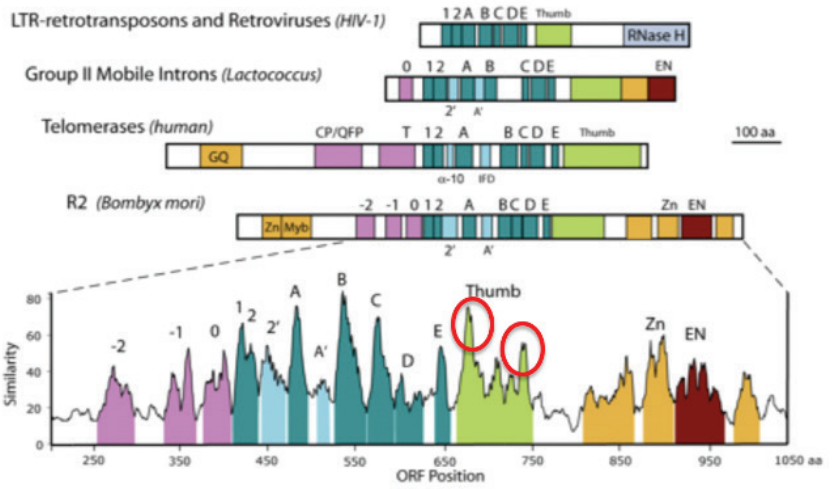


Figure 1. Comparison of R2 protein reverse transcriptase domain with retroviruses, group II introns and telomerases (Jamburuthugoda unpublished figure)

As displayed in *Figure 1*, phylogenetic studies have shown highly conserved protein sequences of RT domains in the fingers, palm and thumb domain encoded by TEs, telomerases and group II introns (Jamburuthugoda & Eickbush, 2014). Previous research has characterized two highly conserved protein motifs N-terminal to the RT domain in

R2, the 0 region and -1 region, as an RNA binding region (Jamburuthugoda & Eickbush, 2014). By analogy to group II introns, the putative thumb domain of R2 protein is a candidate for RNA binding (Eickbush & Jamburuthugoda, 2008). In this study, we are focusing on the thumb domain of the R2 protein, specifically both the RYLGV and KPQQR sequence, as potential RNA binding regions of R2 protein to the 3' and 5' end of R2 RNA.

Our goal was to perform site-directed mutagenesis of the RYLGV sequence and KPQQR sequence. Specifically, we sought to create single alanine replacements of arginine to alanine in the RYGLV sequence, and lysine to alanine in the KPQQR sequence. After confirming that the single mutation was successful, we ultimately want to use this single mutant as a template to create a second mutation within the RYGLV sequence (tyrosine to alanine) and KPQQR sequence (glutamine to alanine). If site-directed mutagenesis of the RYLGV and KPQQR sequences is successful, we can then express the mutant protein to study its RNA binding ability compared to that of the wild type R2 protein. We hypothesize that the mutant R2 protein will have significantly reduced ability to carry out activities that require specific RNA binding, such as the TPRT integration mechanism. Additionally, the functionality of R2 protein, such as its ability to carry out its reverse transcriptase activity and endonuclease activity, should be retained after mutagenesis.

Experimental Design

pR2cdn-B plasmid construct. Our goal is to identify additional RNA binding protein motifs important in the R2 RNA binding ability of the R2 protein and its subsequent integration into the host genome. In order to mutate the R2 protein and to produce this mutant protein for use in future biochemical assays, we used a pR2cdn-B plasmid construct obtained from our collaborator, Dr. Shawn Christensen, from the University of Texas at Arlington (*Figure 2*). The pR2cdn-B plasmid is a genetically modified plasmid that can be transformed into different *Escherichia coli (E. coli)* cell lines, such as JM109 and BL21, to produce the R2 protein within *E. coli* cells. The pR2cdn-B plasmid is 7.3 kilobases and contains the open reading frame of the R2 protein, which has the same exact amino acid sequence as that found in the R2 protein of *Bombyx mori.* However, since the pR2cdn-B plasmid construct is codon optimized for *E. coli,* meaning that the DNA sequence was altered to match the codons most frequently used in *E. coli* when synthesizing R2 protein, the pR2cdn-B plasmid allows for effective production of the R2 protein within *E. coli* cells.

Important features of pR2cdn-B include the T5-lac promoter, *lacI* gene, and Kanamycin resistance gene. The hybrid T5-lac promoter is recognizable by the *E. coli* RNA polymerase and is important for the transcription of the R2 protein. This hybrid promoter contains three *lacI* binding sites, where the *lacI* repressor will bind and prevent transcription of the R2 protein. In normal circumstances, when the lac repressor is bound to the operator site, transcription and subsequent translation of the plasmid genes is prevented. Therefore, in order to induce R2 protein production in *E. coli* cells, IPTG

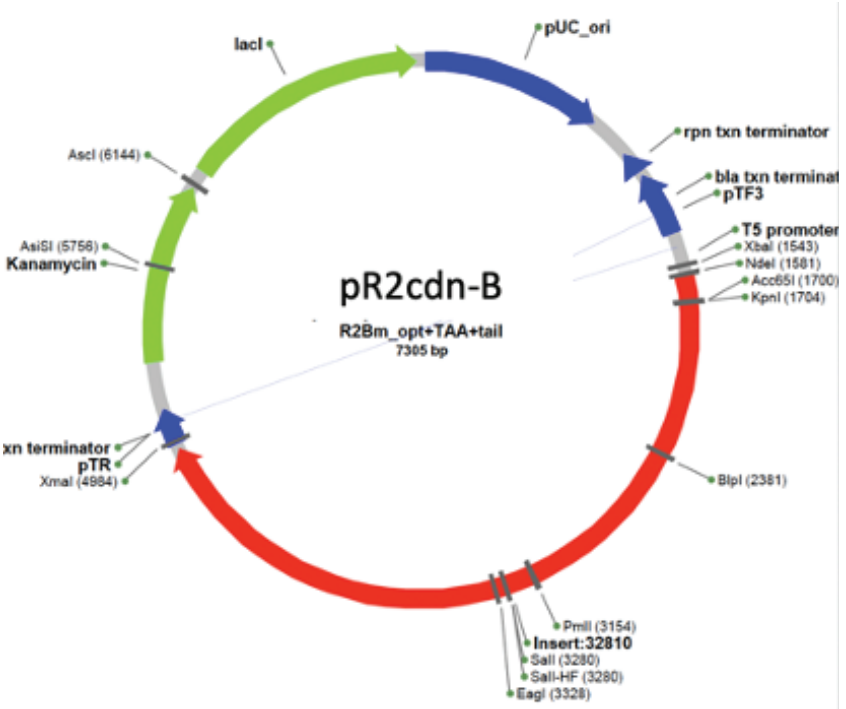


Figure 2. Plasmid/template contain R2 protein sequence, codon optimized for E. coli, with open reading frame (ORF) of R2 protein (red)

(a lactose analog) is added, which will bind to the lac repressor and thereby remove it from the *lacI* binding sites, ultimately inducing subsequent R2 protein transcription and translation. Additionally, the Kanamycin resistance gene allows for selection of successful transformants of the pR2cdn-B plasmid grown on LB/Kanamycin plates. *E. coli* cells that did not successfully uptake the plasmid will not have the Kanamycin resistance gene and therefore would not survive on these plates.

Primer design. The forward and reverse mutagenic primers for the RYGLV and the KPQQR sequence of the thumb region of the R2 protein of the pR2cdn-B plasmid construct were designed (Invitrogen) (*Figure 3A*). Primers were designed such that within the RYLGV sequence, arginine (R) was mutated to an alanine (A), and within the KPQQR sequence, lysine (K) was mutated to an alanine (A) (*Figure 3B*). After confirming that the single mutation was successful, we ultimately want to use this as a template to create the second mutation, tyrosine to an alanine, within the RYGLV sequence, and glutamine to alanine, within the KPQQR sequence (*Figure 3C*).

Creating more template DNA. A streak plate was made on LB/ Kanamycin plates using the WT R2 glycerol stocks obtained from our collaborator, Dr. Shawn Christensen. After incubating overnight at 37°C overnight, a single colony was selected and suspended in a 50µg/mL LB/Kanamycin stock solution to allow the plasmid to amplify overnight. 3mL of bacterial overnight culture was used in the QIAprep Spin Miniprep Kit to extract

the WT R2 plasmid DNA. DNA was quantified using the NanoDrop Spectrophotometer and an aliquot of the plasmid DNA sample was also ran on a 0.75% agarose gel.

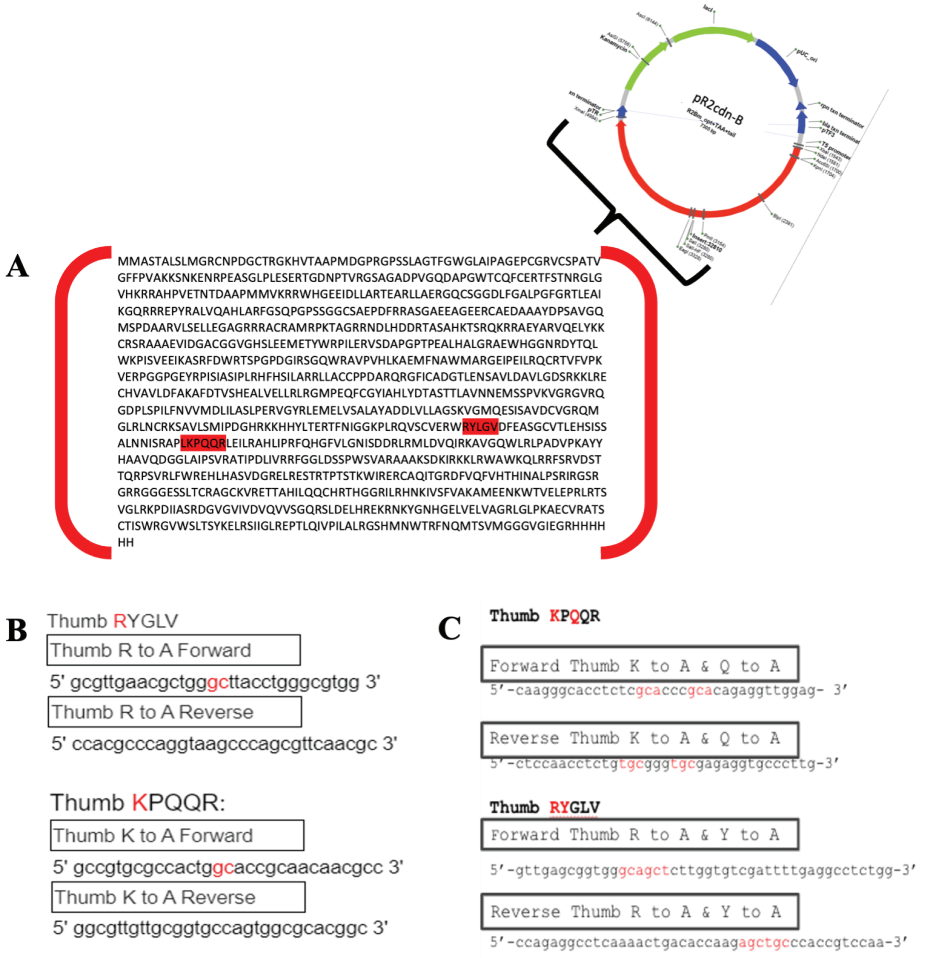


Figure 3. Mutagenic primer design for site-directed mutagenesis of conserved sequences within the thumb domain of R2 protein. (A) Open reading frame of R2 protein. (B) Single mutant primer design. RYGLV sequence: arginine (R) to alanine (A); KPQQR sequence: lysine (K) to A. (C) Double mutant primer design. RYGLV sequence: R and tyrosine (Y) amino acids mutated to A; KPQQR sequence: K and glutamine (Q) amino acids mutated to A.

Site-directed mutagenesis

For the site-directed mutagenesis reaction, the pR2cdn-B plasmid construct was used as a template that the mutagenic primers annealed to. After preparation of the plasmid, the double stranded DNA of the plasmid was first denatured, followed by annealing of the mutagenic primers. At 68°C, the mutagenic primers were extended by *PfuTurbo* DNA polymerase to synthesize the mutant DNA strand. *PfuTurbo* DNA polymerase was used

because it does not have strand displacement properties and extends at a higher fidelity than the commonly used *Taq* DNA polymerase for PCR. It is important to note that once the primer starts to get extended, it will extend throughout the entire plasmid until it reaches the primer, which generates the full-length linear product of 7.3 kilobases. Only when transformed do the nicks in the plasmid become repaired by the ultracompetent cells, thereby reforming the circular plasmid. After temperature cycling, the PCR reaction was treated with *Dpn1* endonuclease to digest the parental DNA strand. Unlike the newly synthesized strand, the parental strand is methylated and thus cleaved by *Dpn1*, allowing for selection of the newly synthesized template with the desired mutation. After the *Dpn1* digest, the reaction was transformed into XL10-Gold ultracompetent cells.

QuikChange XL site-directed mutagenesis kit. For the initial set of site-directed mutagenesis attempts, we used the PCR cycling parameters outlined by the QuikChange XL method with the annealing phase at 68°C (segment 2) for 7 minutes for the first trial and 10 minutes for the second trial. However, when visualizing the PCR on 0.75% agarose gel, no bands were present for both the PCR control product and the PCR reactions and the transformation reaction we performed to confirm whether or not the cells had the plasmid was also unsuccessful. Since the pR2Cdn-B plasmid has the Kanamycin resistance gene, successful transformants should be able to grow on the Kanamycin plates.

Therefore, for the second site-directed mutagenesis trial using the QuikChange XL method, we increased the amount of template DNA from 3ng to 9ng and used a 10-minute extension phase at 68°C rather than the 7-minute extension phase used previously. However, as we saw previously, no band around 7.3 kilobases was present for the PCR product for the mutagenesis reactions, indicating that our site-directed mutagenesis did not work. *DpnI* endonuclease was added to the PCR reaction to digest the parental DNA template and thereby select for the mutation-containing template only. After the *DpnI* digest was incubated for 1 hour, the reaction was transformed into XL10-Gold ultracompetent cells. For controls, the R2 template DNA was directly transformed into the competent cells, as well as the pET8a plasmid. However, both transformations were unsuccessful.

Two-step PCR. Since the transformations from the previous site-directed mutagenesis reactions performed using the kit parameters were unsuccessful, for the next site-directed mutagenesis reaction a two-step PCR was used. The rationale for using a two-step PCR was that possibly lowering the initial annealing temperature would help the primers adhere better to the template. Additionally, more cycles were utilized than the kit recommended. However, like the previous site-directed mutagenesis reactions, there was no band around 7.3 kilobases for the PCR reaction and the transformation reactions were unsuccessful.

Gradient PCR. Since the site-directed mutagenesis and transformation reaction using the two-step PCR parameters was also unsuccessful, for the next set of reactions instead of using *PfuTurbo* DNA polymerase, we used a new polymerase, Platinum SuperFi DNA polymerase. Platinum SuperFi DNA polymerase is better at copying larger templates than *PfuTurbo* DNA polymerase, which might be helpful with the R2 template that we

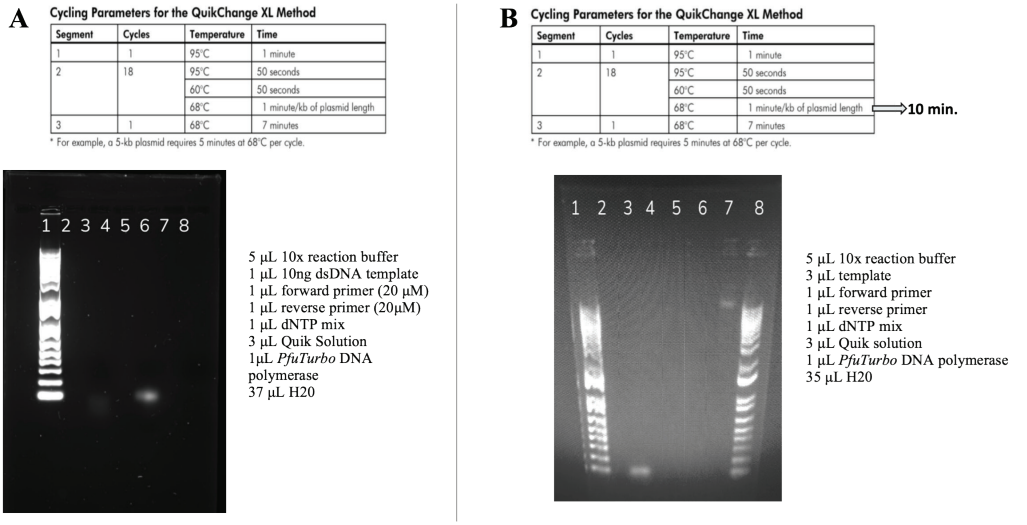
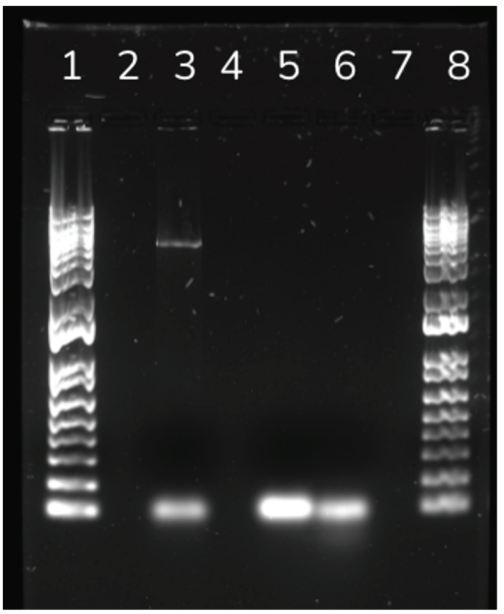


Figure 4. Site-directed mutagenesis using QuikChange XL kit. 0.75% agarose gel stained with Gel Red. (A) PCR cycling parameters and gel analysis for first trial of site-directed mutagenesis. Lane 2, DNA ladder; Lane 4, PCR reaction (RYGLV); Lane 6, PCR reaction (KPQQR). (B) PCR cycling parameters and gel analysis for second trial of site-directed mutagenesis, with extension phase at 68°C for 10 minutes. Lane 1 and Lane 8, DNA ladder; Lane 3, PCR control; Lane 4, PCR reaction (KPQQR); Lane 5, PCR reaction (RYGLV); Lane 7, template DNA.

are using since it can form concatemers. Additionally, Platinum SuperFi DNA polymerase has higher fidelity than *PfuTurbo* DNA polymerase and is more cost effective for the amount of reactions that can be performed. In addition to changing the polymerase, a gradient PCR was performed. Gradient PCR involves incremental increases in temperature for the annealing step, which is when the mutagenic primers bind to the template. As such, gradient PCR allows for multiple reactions to take place at a time, but these separate reactions are subjected to different annealing temperatures. For our gradient PCR, eight PCR reactions for each conserved thumb region were performed, with incremental increases in annealing temperature from 53°C to 63°C (*Figure 4*). The PCR reaction was visualized on a 0.75% agarose gel stained with Gel Red at long exposure. As shown in *Figure 4*, faint bands and smearing were present in lanes 5-12 and in lane 20, indicating that there was something amplifying in the PCR reactions.

Since a faint band located around the expected 7.3 kilobases in lanes 5-12 and in lane 20 was present, we concentrated the PCR product by combining some of the reactions that were run separately. Specifically, for the mutagenic reactions of the KPQQR sequence, the PCR reactions in lanes 9 and 10 were combined, as well as lanes 11 and lanes 12. For the mutagenetic reactions of the RYGLV sequence, only the PCR reaction in lane 20 was used since it was the only lane with a visible band present. The purpose of concentrating the PCR product was to yield a stronger band when running these reactions again on a different 0.75% agarose gel stained with Gel Red, which we could then extract from the gel and purify (*Figure 5*). Each combined sample was subjected to *DpnI* digest to



Segment	Cycles	Temp. (° C)	Time
1	1	95	1 min
2	10	95	50 sec
		50	50 sec
		68	7 min
3	25	95	50 sec
		60	50 sec
		68	7 min
3	1	68	7 min

5 µL 10x reaction buffer
1 µL template
1 µL forward primer
1 µL reverse primer
1 µL dNTP mix
3 µL Quick Solution
1 µL PFU Turbo
37 µL ddH20

Figure 5. Site-directed mutagenesis using two-step PCR.0.75% agarose gel stained with Gel Red. Lane 1 and Lane 8, DNA ladder; Lane 3, PCR control; Lane 5, PCR reaction (RYGLV); Lane 6, PCR reaction (KPQQR)

eliminate any template DNA, and then the concentrated samples were run on a 0.75% agarose gel. The gel was visualized at quick exposure to avoid prolonged exposure to ultraviolet light, which could damage the DNA. Stronger bands around the 7.3 kilobases were present for the combined gradient PCR for the KPQQR sequence in lanes 4-6, as well as for the gradient PCR reaction at 62.2°C for the RYGLV sequence in lane 11 (*Figure 5*). These PCR products were then cut out from the gel, and DNA was purified using the QIAquick Gel Extraction (QIAGEN) kit. Purified plasmid DNA was then transformed into JM109 cells.

Transformation and plating of gradient PCR. As shown in *Figure 6*, two different controls were used, the pET28a control (*Figure 6A*) and the WT R2 plasmid DNA itself (*Figure 6B*). Since both control plasmids have a Kanamycin resistance gene, the JM109 cells that take up either of these plasmids should survive on the LB/ Kanamycin media, growing and forming colonies, which did occur. When transforming into the JM109 cells, serial dilutions of both the pET28a control plasmid and the R2 control were created and 100 µL was directly plated into each of the petri dishes. Next, for each control a 1/10 dilution, a 1/100 dilution, and a 1/1000 dilution were created starting with 100 µL. The transformation efficiency of the JM109 cells was calculated by multiplying the number of colonies grown on each plate by their dilution factor and plating dilution, which was determined to be 2 x 10^7 colony forming units per microgram (cfu/µg) when using the 1/10 dilution plate for the wild-type template. Although this transformation efficiency was slightly lower than the expected 2 x 10^8 value, overall, the transformation of the PCR reactions was successful (*Figure 7*), indicating that the PCR products minimally have the Kanamycin gene, which allowed the transformants to survive on the LB/ Kanamycin media.

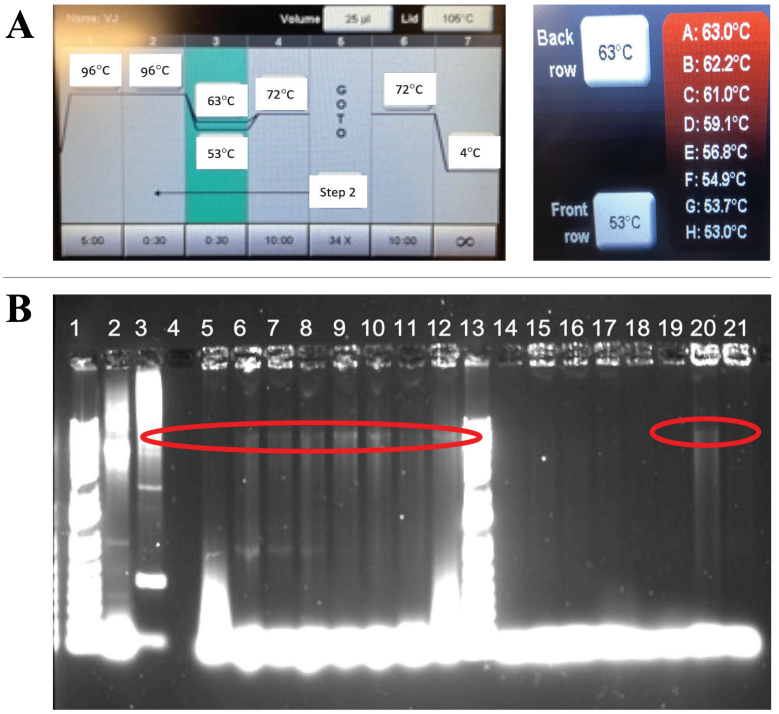


Figure 6. Gradient PCR. (A) Cycling parameters for gradient PCR. Each PCR reaction underwent 35 cycles in the thermocycler with a 10-minute extension phase. The annealing temperature spanned 53°C to 63°C. (B) Long exposure image of gradient PCR. The PCR reaction was visualized a 0.75% agarose gel stained with Gel Red to try and find the ideal annealing temperature for the mutagenic primers for each conserved region. Red circle indicates faint bands present around 7.3 kilobase mark. Lane 1 and 13, DNA ladders; Lane 2 and 3, positive controls. Lanes 5-12, PCR reactions for the KPQQR sequence with increasing annealing temperature (53°C→ 63°C); Lanes 14-21, PCR reactions for the RYGLV sequence with increasing annealing temperature (53°C→ 63°C). For lanes 5-12, 5 µL of the PCR sample and 2 µL loading dye was loaded into each well. For lanes 14-21, 5 µL PCR sample and 4 µL loading dye was loaded into each well.

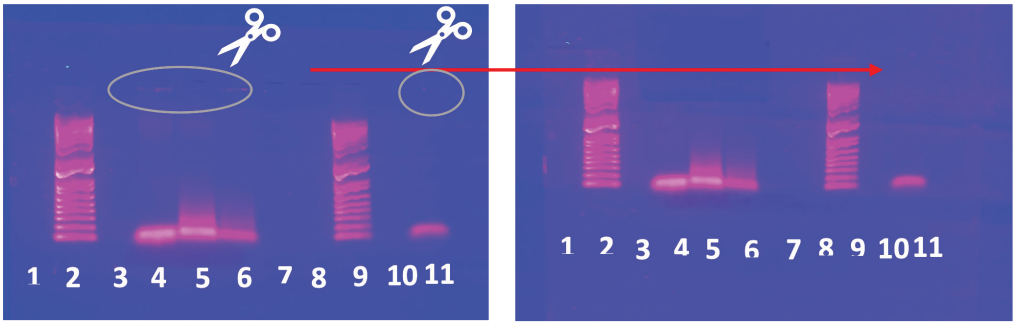


Figure 7. PCR product concentration (left) and gel extraction (right). PCR ran on 0.75% agarose gel. Lane 2 and 9, DNA ladder; Lane 4-6, combined gradient PCR for KPQQR sequence (30 µL sample and 6 µL of loading dye in each lane); Lane 11, gradient PCR reaction at 62.2°C for RYGLV sequence (30 µL sample and 6 µL of loading dye).

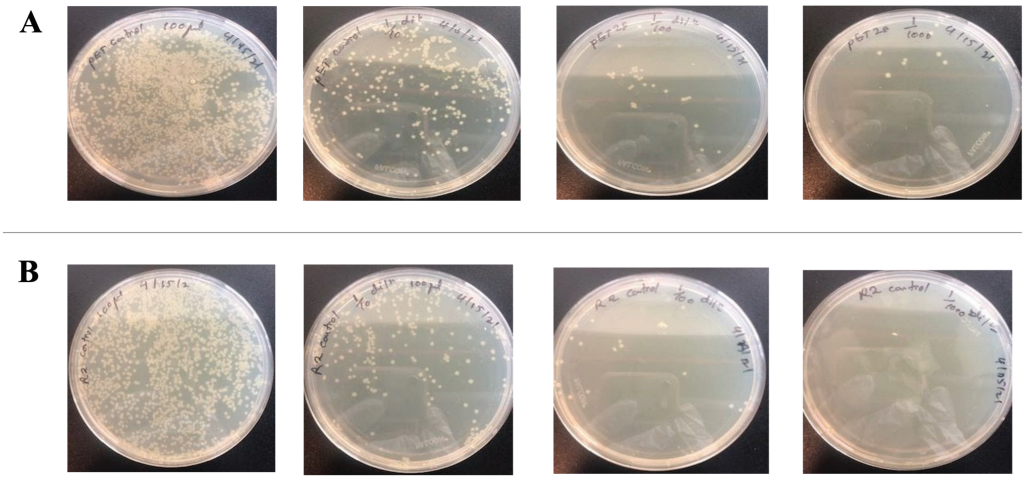


Figure 8. Serial dilution of gradient PCR concentration transformation for controls (from left to right: 100 µL → 1/10 µL → 1/100 µL → 1/1000 µL. (A) serial dilution of pET28 control (B) serial dilution of wildtype template control. Transformation efficiency: 2 x 107 cfu/µg.

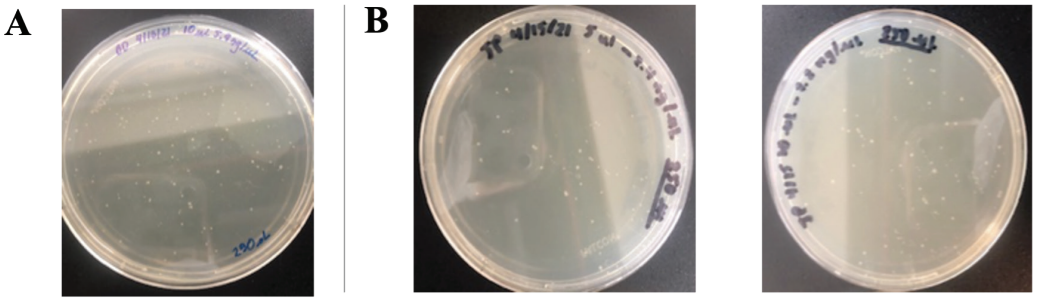


Figure 9. Gradient PCR concentration transformation of PCR products for (A) RYGLV sequence and (B) KPQQR sequence. Transformants plated on LB/Kanamycin.

Since the transformations were successful, a few colonies from these successful transformants of the mutant plasmid were selected and grown in an overnight LB/Kanamycin broth, and then miniprep protocol to extract plasmid DNA was performed. After, the DNA was quantified using the nanodrop spectrophotometer before sending the mutants out for sequencing to confirm that the desired mutation was achieved. We will first sequence the thumb region and once the mutation is confirmed in that region, the entire plasmid will be sequenced to confirm that there are no other mutations. Once we confirm the mutation, we will then use this as a template to generate the second mutation to make the double mutants in the thumb region. Then we will transform those cells for protein expression and purification.

Conclusions and Future Directions

Future research will focus on expression and purification of the mutant protein. Once expression and purification of the R2 protein is achieved, biochemical assays can be performed to compare mutant R2 protein's ability to bind RNA to that of the WT R2

protein. First, strand DNA cleavage and primer extension assay will assess if the endonuclease activity and reverse transcriptase activity of the R2 element has not been affected by the site-directed mutagenesis of the thumb region of the R2 protein. This would mean that the functionality and structure of the R2 protein was retained. After confirming that the thumb domain does not affect these other activities, a TPRT assay will be conducted to examine the effect of mutating a specific amino acid within each of the conserved regions of the thumb domain on RNA binding. Additionally, we will also perform an electrophoretic mobility shift assay (EMSA) to study R2 protein binding ability to 3' UTR RNA.

Performing these assays will allow us to test our hypothesis which is whether the thumb region is a potential RNA binding domain, as well as our understanding about the TPRT mechanism, which is how the R2 element integrates into the host genome. By understanding this mechanism, many new innovations can arise in the prevention and treatment of human disease.

References

Belancio, V. P., Deininger, P. L., & Roy-Engel, A. M. (2009). LINE dancing in the human genome: Transposable elements and disease. *Genome Medicine, 1*(10), 97.1–97.8. https://doi.org/10.1186/gm97

Brooker, R. J. (2018). *Genetics: Analysis & Principles* (6th ed.) McGraw-Hill Education.

Christensen, S. M., Ye, J., & Eickbush, T. H. (2006). RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proceedings of the National Academy of Sciences, 103*(47), 17602–17607. https://doi.org/10.1073/pnas.0605476103

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature reviews: Genetics, 10*(10), 691–703. https://doi.org/10.1038/nrg2640

Eickbush, D. G., Burke, W. D., & Eickbush, T. H. (2013). Evolution of the R2 retrotransposon ribozyme and Its self-cleavage site. *PLoS ONE, 8*(9), 1–16. https://doi.org/10.1371/journal.pone.0066441

Eickbush, T. H., & Jamburuthugoda, V. K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research, 134*(1-2), 221–234. https://doi.org/10.1016/j.virusres.2007.12.010

Gregory, T. R. (2005). Synergy between sequence and size in Large-scale genomics. *Nature Reviews: Genetics, 6*(9), 699–708. https://doi.org/10.1038/nrg1674

Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., Van Meir, E. G., Vertino, P. M., Devine, S. E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell, 141*(7), 1253–1261. https://doi.org/10.1016/j.cell.2010.05.020

Jamburuthugoda, V. K., & Eickbush, T. H. (2014). Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucleic Acids Research, 42*(13), 8405–8415. https://doi.org/10.1093/nar/gku514

Khadgi, B. B., Govindaraju, A., & Christensen, S. M. (2019). Completion of LINE integration involves an open "4-way" branched DNA intermediate. *Nucleic Acids Research, 47*(16), 8708–8719. https://doi.org/10.1093/nar/gkz673

Yamaguchi, K., Kajikawa, M., & Okada, N. (2015). LINE retrotransposition and host DNA repair machinery. *Mobile Genetic Elements, 5*(6), 92–97. https://doi.org/10.1080/2159256X.2015.1096998