Toward the Identification of Anionic Guanines *In Vivo*

Glenn Duncan

Undergraduate thesis submitted in fulfillment of requirements for the

Millennium Scholars Program.

Eberly College of Science

The Pennsylvania State University

Submitted on _____

ABSTRACT

The charge state of RNA nucleobases can impact RNA folding, which in turn may affect gene regulation. 1-Ethyl-3-(3- dimethylaminopropyl) carbodiimide (EDC) and glyoxal have been developed as chemical reagents that react with and covalently modify guanine nucleobases *in vivo*. EDC tends to react with uncharged guanines found in the neutral intracellular environment, whereas glyoxal reacts with anionic guanines that are not typically found in the intracellular environment. Previous *in vivo* treatment of rice (*Oryza sativa*) seedlings with glyoxal and EDC have identified potential anionic guanines. The experiments described herein were designed to use a cost-efficient approach of finding more anionic guanine candidates *in vivo*. The first step was to determine the efficacy of EDC reactivity *in vitro* using RNA extracted from rice seedlings. Upon finding that EDC selectively modified uridines and guanines *in vitro*, the next steps were to treat rice seedlings with EDC, determine whether EDC modified RNA within intact cells, and pull down RNA targets. Chapter 1 of this thesis introduces RNA structure and the use of probing reagents to elucidate structure. Next, chapter 2 covers crystal structure analysis conducted in this research to examine the environment surrounding these guanines, and several interactions that were identified that could promote an anionic charge. Chapter 3 discusses the observation of EDC modification with uridines and guanines *in vitro*. This result allowed further *in vivo* treatment of rice seedlings with EDC, the extracted RNA from which would be used for all subsequent experiments. Lastly, Chapter 4 mentions experiments that have yet to be performed and future directions. Specifically, probe design methodology to pull down selected target RNAs, pulling down the RNAs of interest, isolating RNAs, and preparing them into cDNA libraries for sequencing. The final goal would be to analyze sequencing data to indicate the presence and location of anionic guanines.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I would like to express my gratitude for the knowledge, mentorship, and research exposure I received from the Bevilacqua lab during my undergraduate years. Dr. Philip Bevilacqua has provided me with ample resources to excel as a researcher and future professional. Additionally, I would like to specifically thank Andrew Veenis and Dr. David Mitchell III for their guidance throughout my undergraduate research experience and creation of this undergraduate thesis. Andrew Veenis has helped me acquire computational skills, knowledge of chemical reactions, and graduate school advice. Dr. David Mitchell III helped me design experiments, learn wet lab techniques, and gain knowledge on chemicals and reactions. All of these mentors helped me to gain confidence as a scientist and I am very thankful for that. They enabled improvement of my presentation and writing skills. I would like to thank the other postdoctoral, graduate, and undergraduate members of the lab. Everyone in the lab has provided me with assistance for some aspect of my project and supported me on my research journey in some capacity. Lastly, I would like to thank Dr. David Boehr for reviewing my thesis and assisting with edits.

Chapter 1
Introduction To Thesis

1.1 RNA performs various functions

RNA is well-known for its intermediary role in the Central Dogma of Molecular Biology. In this model, RNA is transcribed from DNA and used to synthesize proteins (Figure 1.1). These proteins carry out specific cellular functions such as serving as receptors for cell signaling, ion channels, and ribosomal proteins. However, RNA has many other cellular functions that are imperative to cell fitness and survival.

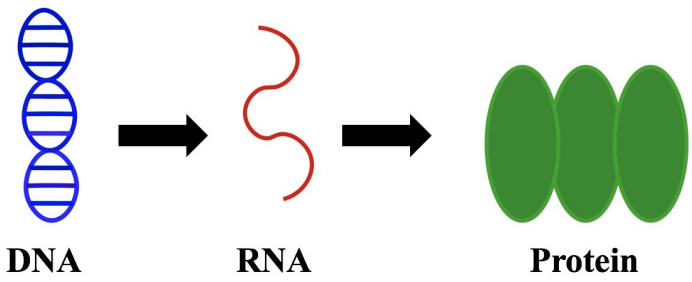

Figure 1.1 Scheme of Central Dogma. Diagram of RNA's intermediary role between DNA and proteins. In this process, RNA is transcribed from a DNA template and translated into a protein.

RNA takes on many forms and carries out various functions. Types of RNA include riboswitches, messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), ribosomal RNA (rRNA), and small interference RNAs (siRNAs). Riboswitches play a role in gene expression by binding to metabolites resulting in a change in the riboswitch conformation.[1] mRNAs carry the genetic code that is translated into proteins. lncRNAs are RNAs that contain more than 200

nucleotides and sometimes aid in developmental processes.[2] rRNAs along with ribosomal proteins make up ribosomes which synthesize peptides. Lastly, siRNAs silence genes and serve as antiviral defense.[3] For example, some siRNAs join an RNase complex that degrades viral RNAs and prevents their expression.[3] Thus, learning about RNA structure can shed insight on the relationship between structure and function.

1.2 Details of RNA structure

RNA folds into complex structures that promote stability and proper function. Folding results from different intramolecular and intermolecular interactions. The RNA structural hierarchy consists of primary, secondary, and tertiary structure (Figure 1.2). Primary structure is the RNA sequence alone that is formed by the sugar-phosphate backbone, a series of covalent bonds. Secondary structure consists of hairpin loops, bulges, and stems that arise from hydrogen bonding between residues within the sequence. Lastly, tertiary structure is often visualized in 3-dimensional space and consists of interactions between secondary structure components. RNAs are able to function within cells because of their structure. For example, the ribosome structure allows for a catalytic component to facilitate the formation of amino acids by matching RNA codons with tRNA anticodons and forming peptide bonds between each amino acid.

Studies have shown that mRNAs undergo an enthalpy-driven process to unfold within cells.[4] RNA folding usually increases its stability, but mRNAs must deviate from this energetic norm. This unfolding may increase efficiency of transcription, such that the sequence can be easier for the ribosome to access and pair with tRNA.

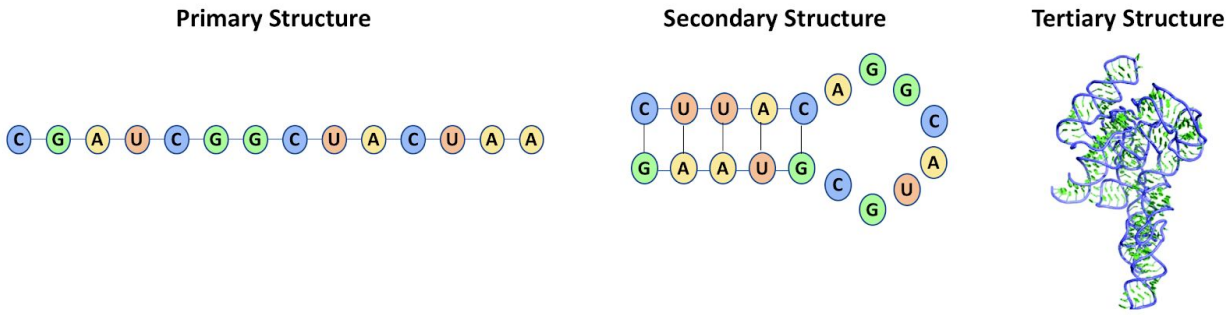

Figure 1.2 RNA structural hierarchy. Primary, secondary, and tertiary structures of RNA that show base pairing interactions that make up RNA folding. The rightmost panel was adapted from reference #5.

1.3 Probing reagents elucidate RNA Structure

Studying RNA structure provides insight into RNA function. Probing reagents have been developed to examine RNA structures by selectively attaching themselves to certain components of the RNA. These probing reagents can either react with the nucleobase itself or the ribose sugar of the nucleotide (Figure 1.3). Selective 2′-hydroxyl acylation and primer extension (SHAPE) reagents add an acyl group to the 2′-oxygen in the ribose ring (Figure 1.3).[6] Early RNA probing utilized dimethyl sulfate (DMS) *in vivo* and other membrane impermeable reagents *in vitro*.[6] DMS modifies the Watson-crick face of cytosine and adenine as well as the Hoogsteen face of guanine in DNA and RNA (Figure 1.3). When a residue is modified by a probing reagent, the chemical probe becomes covalently attached to the nucleobase or substituent. This creates steric hindrance that makes the nucleotides after the modified nucleotide inaccessible for the reverse transcription polymerase. In turn, this halts cDNA elongation during reverse transcription resulting in cDNAs of different lengths.

Other probing reagents like kethoxal react strictly *in vitro* because they are unable to traverse the cell membrane or cell wall. CMCT is a lipid soluble carbodiimide reagent that has been used to probe uridine and guanine bases.[7,8] CMCT has primarily been used for *in vitro* probing of RNA structure and has seen limited use *in vivo* due to the requirement to permeabilize the cell membrane prior to reagent application.[9,10] Permeabilizing the cell membrane is not ideal because it can induce cellular damage response mechanisms that may significantly alter data obtained. Extremely high concentrations of CMCT are required to penetrate the cell membrane but using high concentrations can damage cells prior to reacting with the RNA and may also present hazards to the experimenter.[11,13] Reagents that probe RNA *in vivo* tend to be less water soluble and able to traverse the cell membrane. This is because the lipid bilayer is hydrophobic and has no charge in the interior. A sufficient proportion of EDC is neutral at pH > 7 which means it has a greater chance of crossing the cell membrane than CMCT, which carries a perpetual positive charge.[12] The lack of *in vivo* reagents impeded the study of RNA in a biological context because the effect of cellular protein and cation interactions on structure could not be taken into account. It was necessary to create a collection of *in vivo* probing reagents that could modify different areas of the four nucleobases to receive more comprehensive probing data, which would lead to more accurate structural predictions. As a resolution, the Bevilacqua and Simon labs demonstrated that 1-ethyl-3-(3- dimethylaminopropyl) carbodiimide (EDC) and glyoxal could be used to probe the Watson-Crick face of guanine (N1) and uracil (N3) *in vivo* (Figure 1.3).[6,13,14]

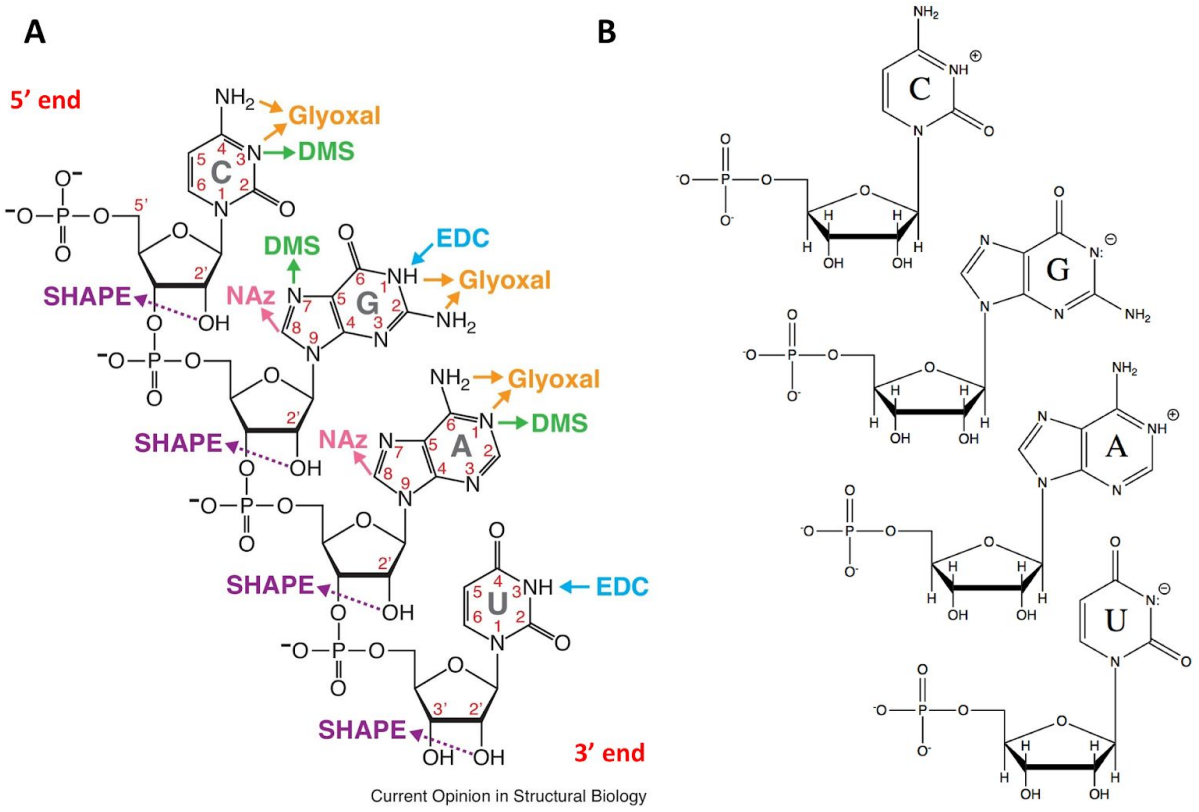

Figure 1.3 Reactivity sites of chemical reagents. This figure was adapted from reference # 6. (A) Arrows in the panel represent the direction of nucleophilic attack with the probing reagents, EDC, glyoxal, SHAPE, and DMS.(B) This panel shows the charges that these bases are able to adapt.[15]

Previous studies have shown that there are charged nucleobases of RNA and that they may assist in RNA catalysis.[15] Adenine and cytosine nucleobases are ale to gain a proton at pHs near 4 and possess a positive charge.[15] In contrast, guanines and uridines can lose a proton at pHs near 9 and bear a negative charge. Although charged nucleobases can exist at varying pHs, they also may exist at neutral pHs to play catalytic roles in cells. This prompts the question as to if they exist in living cells around the pH of 7. Since previous probing data from the Bevilacqua lab uses glyoxal modification to identify potential anionic guanines, this study uses PyMOL to

further examine these candidates. Additionally, this study uses EDC and glyoxal in an attempt to find more anionic guanines in vivo.

As mentioned previously, glyoxal and EDC are both molecules that can be neutral at a pH around 7. Although these two reagents are able to traverse the cell membrane without prior permeabilization, the structures and sizes of the two molecules are very different. Glyoxal is a symmetric molecule that contains two carbonyls making it a strong electrophile. Because glyoxal behaves as an electrophile, it is able to accept the lone pair of electrons from an anionic guanine (Figure 1.3). Glyoxal has many derivatives that are characterized by the difference in R groups attached to either of the carbonyls. For example, phenylglyoxal has a phenyl group attached to one carbonyl (Figure 1.4). The experiments in this paper prepared for the use of methylglyoxal, a glyoxal derivative that contains a methyl group attached to one of the carbonyls. In contrast to the structure of glyoxal derivatives, EDC is a carbodiimide with three nitrogens and more alkane branching. This makes EDC a better nucleophile because nitrogens are able to gain a proton more easily, which is why EDC reacts with protonated neutral guanines (Figure 1.3). Even though these structures have different physical properties, they are able to carry out the same function of modifying uridines and guanines *in vivo.* The work detailed herein entails using EDC and glyoxal reagents to probe rice RNA.

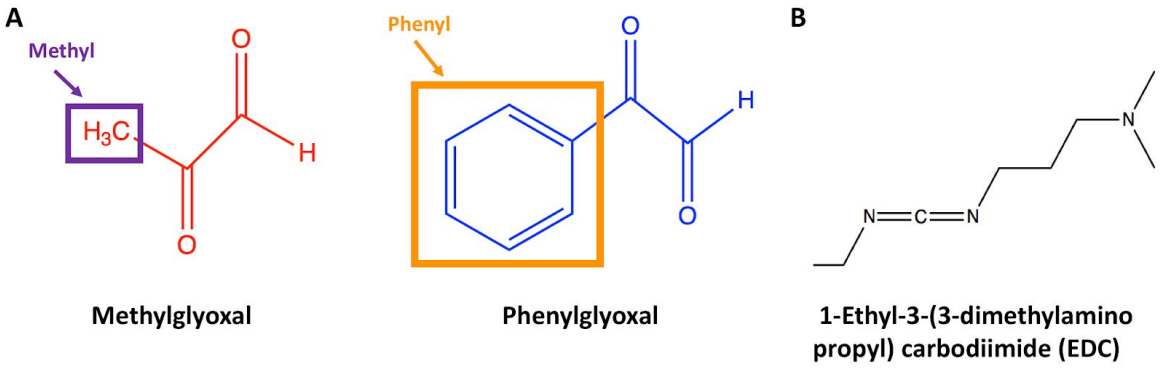

Figure 1.4 Derivatives of glyoxal chemical probes and EDC chemical probe. (A) Variations of glyoxal probes, methylglyoxal and phenylglyoxal, which were used to modify RNA *in vivo* in previous experiments.[13] (B) Structure of EDC chemical probe.

Chapter 2
Using Crystal Structure Analysis to Investigate Anionic Guanines

2.1 Abstract

In the neutral intracellular environment, guanine nucleobases generally have no charge. Previous probing data produced by the Bevilacqua lab has shown that some guanines may be appreciably anionic at neutral pH.[14] Anionic guanines are characterized by reacting with glyoxal. Although anionic guanine candidates have been identified, how prevalent they are within the rice transcriptome is not well understood. *In vitro* studies have shown that anionic guanines are favored when the RNA is in a basic environment. In order for anionic guanines to exist in a neutral cellular environment, there has to be the local environment that promotes this anionic state. In this chapter, crystal structure analysis of previously identified glyoxal-reactive anionic guanine candidates is provided revealing H-bonding interactions that may promote the deprotonated state.

2.2 Introduction

Anionic guanines are characterized by the loss of a proton at the N1 of guanine (N1G) which leaves the guanine with a negative charge. EDC and glyoxal are both chemicals that modify N1G *in vivo*.[6] N1G has a pK_a around 9, so it is expected to be primarily protonated at an intracellular pH around 7.[14,15] EDC reacts with a protonated N1 in neutral conditions while glyoxal does not. Prior to more recent findings, glyoxal was only known to modify N1G at a more basic pH when G is deprotonated and anionic (Figure 2.1).[14] However, probing data in *Escherichia coli (E. coli)* 5.8S ribosomal RNA shows that three guanines (G13, G41, and G81)

react with glyoxal *in vivo* at neutral pH.[14] These findings support a model in which these guanines are anionic at intracellular pH near 7, prompting questions of how this is possible.

The environment surrounding these guanines may cause the N1 to deprotonate. This could happen when various atoms within the purine interact with proteins, ions, or other nucleotides. PyMOL, a structural modeling and visualization software, was used to examine if there are any interactions in an *Escherichia coli* ribosome that could favor this anionic character.

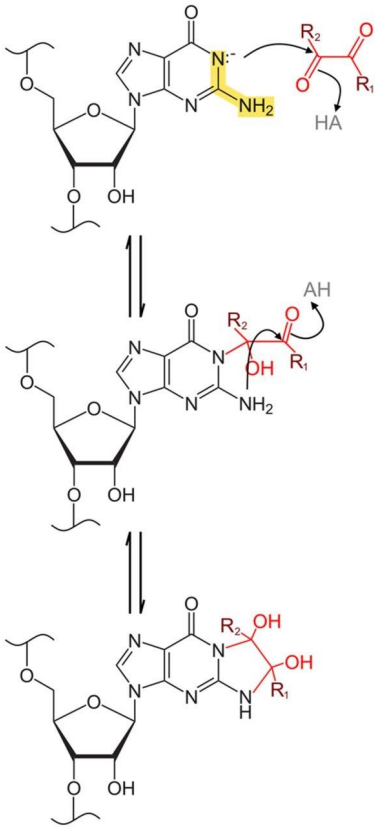

Figure 2.1 Reaction mechanism of glyoxal with an anionic guanine. An illustration of the nucleophilic attack of the anionic guanine on the glyoxal chemical probe. This figure was adapted from #14.

2.3 Materials and Methodology

PyMOL was used to analyze the crystal structure of the *Escherichia coli (E. coli)* ribosome LSU, retrieved from RCSB PDB (PDB ID 4ybb). G13, G41, and G81 were located in the 5S ribosomal RNA and interactions around them were identified. The Wizard tool in PyMOL was used to measure lengths and dihedrals to assess the likelihood of H-bond formation.

2.4 Results and Discussion

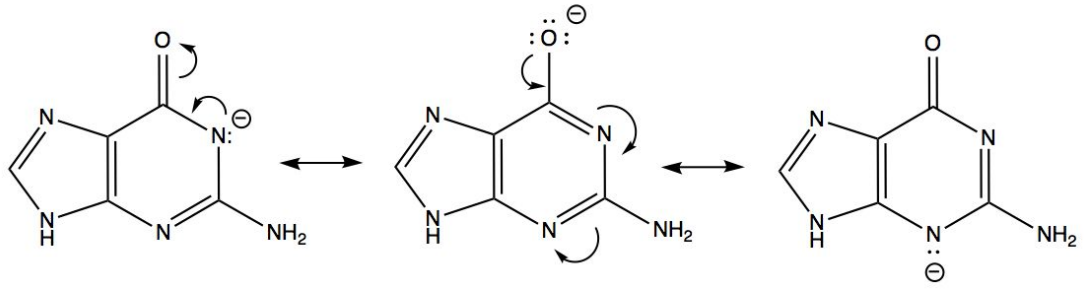

Figure 2.2 Resonance forms of an anionic guanine. Resonance forms of anionic guanines and arrow pushing that depicts delocalization of the electrons from N1G.

When guanine is deprotonated, the negative charge is distributed amongst the N1, N3, and O6 of the purine ring (Figure 2.2). Thus, interactions with these atoms may stabilize the anionic state. I observed that G81 did not display interactions that would promote the anionic state during crystal structure analysis (Figure 2.3A). The G81(N1) donates a H-bond to the U95 (O2) , meaning the N1 is protonated. This N1 has a trigonal planar geometry and is in the same plane as U95.

Despite crystal structure analysis disproving an anionic state of G81(N1), interactions were observed surrounding G13(N1) and G41(N1) that promoted the anionic state. The next

glyoxal reactive guanine studied, G13, interacts with another guanine, G69 (Figure 2.3B). G69(N2) donates a H-bond to G13(N3), with a bond length of 3.0 Å. This H-bond could promote deprotonation of G13(N1) by stabilizing the resultant negative charge on the N3. G13(N1) resides 3.0 Å from the G69(2′OH). Because the 2′OH has a proton and two electron pairs, the moiety can donate or accept a hydrogen bond to or from G13(N1). If G69(2′OH) donates a H-bond to G13(N1), the N1 must be deprotonated to accept this H-bond, and this H-bond would stabilize the negative charge. Based on the image, the 2′OH and G13(N1) are positioned in a way that would not favor G13(N1) as the H-bond donor. G13(N1) has a trigonal planar geometry and is not free to rotate, whereas the hydroxyl bond is tetrahedral and free to rotate. G13(N1) and 2′OH are not in the same plane, therefore the hydroxyl would have to rotate to form this H-bond because the G13(N1) could not. If they were all in the same plane, the dihedral angle would be near 180 ◦ , however, it was measured to be 135 ◦ in PyMOL (Figure 2.3D). More evidence that supports G13 being anionic is the G13(N2) interaction with G69(N3). If N2 takes part in the resonance of the guanine, it would be unable to interact with G69(N3). This is because the partial double bond character between the C2 and N2 would make the bond unable to rotate. Thus, it could not interact with G69(N3) that is out of its plane. However, there is a bond observed between the two entities, meaning that G13(N2) has a tetrahedral geometry and is able to rotate. This would mean that the lone pair of electrons on G13(N2) is not delocalized throughout the ring, which lowers electron density of the purine. This favors the deprotonation of G13(N1) because if the electrons were delocalized that would increase electrostatic forces.

The second guanine studied, G41, interacts with K69 from the S2 protein of the 30S ribosome, a positively charged amino acid (Figure 2.3C). The amino of K69 resides 3.1 Å from G41(N7) and 3.2 Å from G41(O6). As mentioned previously, the H-bond donation from the

amino to G41(O6) can help to stabilize the anionic guanine. Additionally, the positive charge on K69 would stabilize the negative charge purely through electrostatic means.

The ability to target other RNAs with well designed probes would enable the identification of other putative anionic guanines for future crystal structure analysis. This could reveal other interactions that favor the anionic state of guanine.

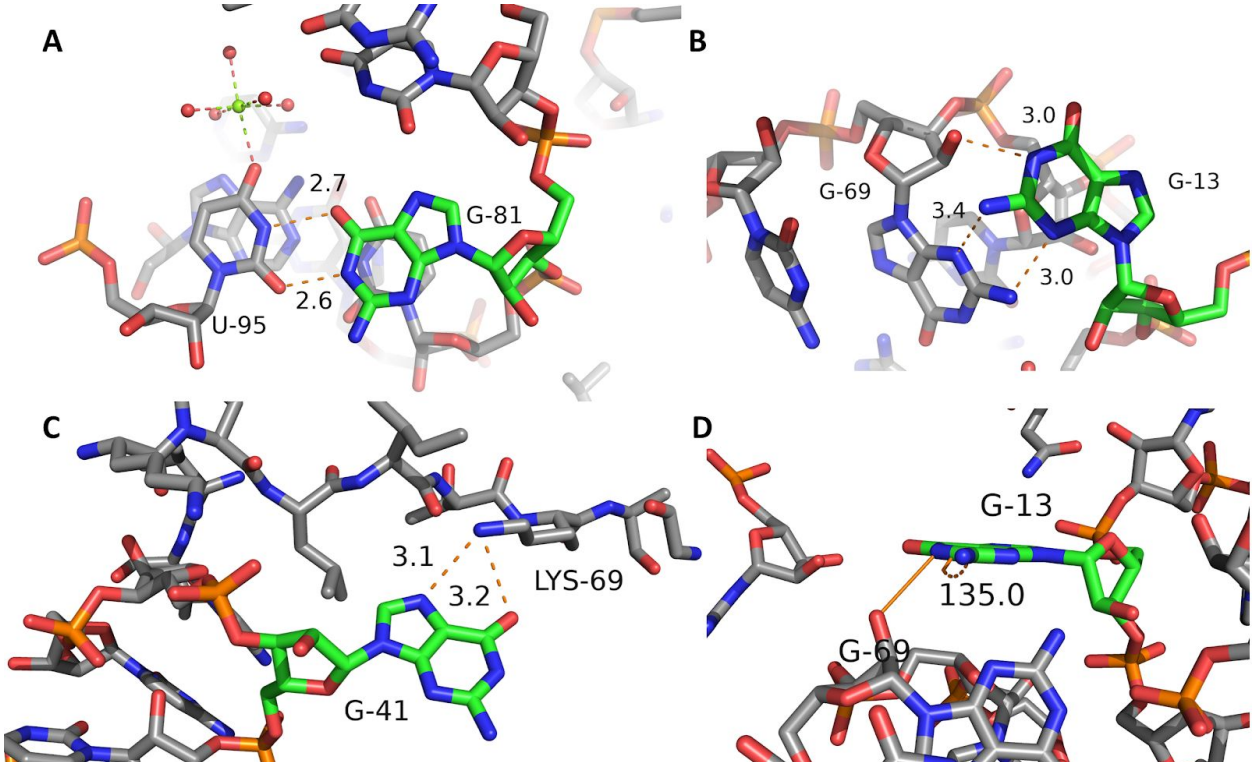

Figure 2.3. PyMOL images illustrate interactions around potential anionic guanines. Image displaying interactions with G13, G41, and G81 that can give rise to its potential anionic character in the *E. coli* ribosome (PDB ID 4ybb). (A) G81(N1) interacting with U95(O2). (B) G13(N1) interacting with G69(2′OH). (C) G41(N1) interacting with an amino acid, lysine (K69). (D) Dihedral angles between G13 and G69.

Chapter 3
Treating Rice RNA with Chemical Probe EDC

3.1 Abstract

Chemical probing reagents allow scientists to elucidate RNA structure. Their ability to react with and chemically attach to different constituents of RNA nucleobases enables identification of chemical properties and interactions. The probing reagents EDC and glyoxal have previously been identified as chemical probes for solvent accessible guanine nucleobases. Glyoxal reacts exclusively with anionic guanines, whereas EDC reacts with neutral guanines. This project is designed to identify and locate additional anionic guanines within a genome, thus the effective usage of glyoxal along with EDC is paramount for conducting this research. However, in order to obtain usable data from *in vivo* experiments, the RNA must react with an appropriate quantity of reagent and for a proper exposure time. This section highlights the methodology I used to modify uridines and guanines of *Oryza Sativa* (rice) RNA with EDC both *in vitro* and *in vivo*. Polyacrylamide gels were run to verify modification of these bases, and modification sites were mapped on a secondary structure model from previous literature.

3.2 Introduction

1-Ethyl-3-(3- dimethylaminopropyl)carbodiimide (EDC) and glyoxal are chemical reagents that modify anionic guanines. In order to be characterized as an anionic guanine, the guanine needs to react with glyoxal. EDC can modify protonated N1s and deprotonated N1s of guanine so its reactivity alone cannot dictate whether or not it is an anionic guanine. Additionally, glyoxal is a much smaller molecule than EDC. Consequently, a guanine that reacts solely with glyoxal could still be anionic but sequestered from EDC due to steric hindrance.

In addition to being able to modify anionic guanines, these chemicals were selected because of their applicability *in vivo*. Chemical reagents that are membrane permeable are very important because they enable RNA study within a living cell without making drastic changes to the cellular environment.

EDC was used in these experiments as a chemical probe of RNA to further identification of anionic guanines, specifically by comparing the reaction profiles of the two reagents. The techniques developed in these experiments were expected to be used for methylglyoxal treatments that have not yet been performed. This chapter details experiments designed to chemically modify RNA of *Oryza sativa* (rice) seedlings with EDC. Rice seedlings were selected as the model organism because my postdoctoral mentor, Dr. David Mitchell III, had previously been working with rice. Additionally, rice is relatively inexpensive and large amounts of it can be easily obtained. RNA extracted from rice seedlings was modified with EDC and *in vitro* to test reactivity of the reagent. EDC was then applied to intact rice seedlings for *in vivo* treatments, followed by RNA extraction and analysis of the modified RNA by gene-specific reverse transcription and polyacrylamide gel electrophoresis (Figure 3.2). As mentioned before, EDC is able to traverse the cell membrane without permeabilization and modify bases without complete denaturation of the RNA.

3.3 Materials and methods

Rice Plant Cultivation

Thirty to forty rice seeds were placed into a standard 100 mm x 15 mm petri dish with filter paper. The dishes were flipped upside down and the paper was placed in the lid which became the bottom side. They were transported to a greenhouse, which is around 30 °C, and positioned directly under a light with intensity of ~500 μmol photons m^{-2} s^{-1}. About 100 mL of water was added to each dish until the seeds were submerged. These plants were monitored for a week and watered every other day until seeds were submerged again. Plants were then moved to cartons of soil and placed in a bed over water for another week.

EDC and Glyoxal Reaction Set-ups and Total RNA Extraction

In vitro reactions

All reactions with EDC were performed in a chemical fume hood. The plants were cut about 1-2 cm above the soil and placed in a tube in the freezer. Total RNA extraction was then performed on the plant tissue. Various concentrations of EDC, 0.5% (v/v), 1%, 2%, and 3%, were added to the RNA in separate reactions. After addition of EDC, the tubes were shaken to ensure the chemical mixed with the RNA. The tubes were then moved to powdered dry ice for about an hour for the following ethanol precipitation step. They were then centrifuged for 20 minutes at 16000g at 4°C. The supernatant was removed, and ice cold 70% ethanol was used to wash the RNA pellet. The goal of this step was to purify the RNA and rid of the EDC.

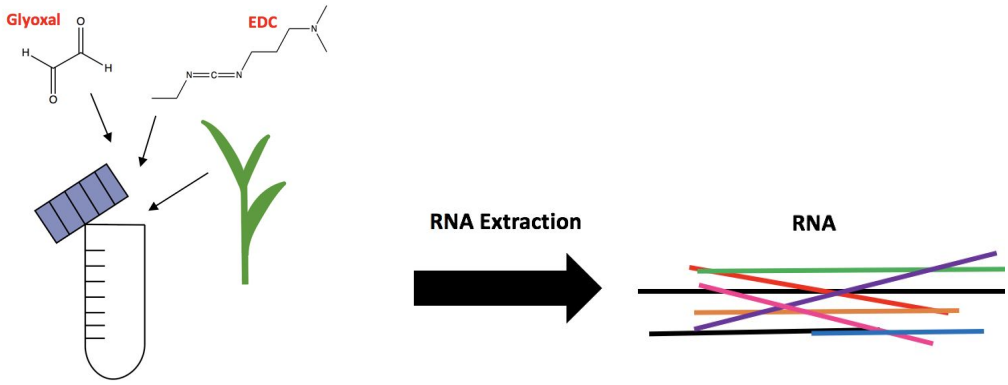

Figure 3.2 *In vivo* EDC and glyoxal reaction steps. The *in vivo* process of treating RNA of rice plants with EDC or glyoxal *in vivo*, involving the addition of probing reagents to rice tissue. Total RNA extraction was then performed to isolate the RNA from the plant tissue.

In vivo reactions

All reactions with EDC were performed in a chemical fume hood. The plants were cut about 1-2 cm above the soil and distributed amongst three 50 mL conical tubes, each designated for the three treatments with the following final concentrations: 1) 25 mM methylglyoxal, 2) EDC (2% or 2.5%), and 3) water. These tubes consisted of 9.8 mL of reaction buffer (HEPES, ph 8.0, potassium chloride, magnesium chloride, water) and 0.2 mL of the stock reagent (Figure 3.2). Each tube was sealed with parafilm to prevent buffer leakage and shaken vigorously for 10 minutes in the fume hood. Afterwards the plants were washed 6 times with water and dried with a paper towel. They were then placed on dry ice for about 30 minutes. Lastly, the plant material was moved to a new 50 mL conical tube and stored at -70 °C.

Total RNA extraction

Untreated (for *in vitro* reactions) or reagent- or control-treated (for *in vivo* reactions) rice seedlings that were quickly frozen in liquid nitrogen were ground up using a mortar and pestle.

Afterward, the protocol contained within the Macherey-Nagel NucleoSpin RNA Plant kit was followed for all steps of extracting total RNA (Figure 3.2).

Gene-specific Reverse Transcription

The procedures for gene-specific reverse transcription are identical to that published in reference #14. The extracted *in vitro* or *in vivo* rice total RNA was added to RNase-free water to give 0.5–1 μg RNA in 5.5 μL total volume. Next, 1 μL of ~500,000 cpm/μL ^{32}P-radiolabeled 5.8S rRNA primer (5′-GCGTGACGCCCAGGCA-3′) was added to the mixture. The solution was placed in the thermocycler at 75 °C for 3 min then cooled to 35°C. Then 3 μL of reverse transcription reaction buffer was added to a final concentration of 20 mM Tris–HCl (pH 7.5), 1 mM DTT, 100 mM KCl, 8 mM $MgCl_2$, and 1 mM dNTPs. Annealing proceeded at 35°C for 5 min, then the solution was heated to 55 °C for 1 min, and 0.5 μL of Superscript III reverse transcriptase (Invitrogen; 100 U total) was added to the reaction. The reverse transcription reaction continued at 55°C for 15 min. Next, 1 μL of 5M NaOH was added to the solution, which was heated to 95°C for 7 min and then cooled to 4°C for ≥3 min to hydrolyze all RNAs and inactivate reverse transcriptase. Lastly, 1 μL of 5M HCl was added to neutralize the solution, and an equal volume of 2× stock solution with 100% deionized formamide was added along with 20 mM Tris–HCl (pH 7.5), 40 mM EDTA, 0.1% xylene cyanol and 0.025% bromophenol blue. The mixture was then run at 80 W for ~2 h on a 10% denaturing polyacrylamide gel (8.3 M urea).

3.4 Results and Discussion

EDC was used in these experiments to modify guanines and uridines of rice RNA both *in vitro* and *in vivo*. These experiments were conducted to test the ability of the reagent to actually modify RNA and provide experience with handling chemical reagents. The goal was to develop this method and adapt its use in future methylglyoxal treatments with hopes of finding more anionic guanines. Methylglyoxal treatments have yet to be performed but the modification technique for EDC proved to be effective so it could be used for future methylglyoxal treatments.

For *in vitro* studies, RNA was extracted from plants, then treated with reagents and reverse transcribed. The reverse transcription stops are revealed as bands on the gel that indicate modification (Figure 3.3A). Darker bands represent a larger frequency of stops at a particular modified nucleobase and thus the number of cDNAs of a given length. The lanes labeled with U, G, C, and A represent the location of the uridine, guanine, cytosine, and adenine nucleotides, respectively. There are more dark bands in the 2% EDC lane than the 0% lane that represents uridines and guanines, indicating that these nucleotides are modified successfully (Figure 3.3A). There are also reverse transcription stops that occur without EDC in the control lanes but the reason for this isn't well understood.

The rice RNA was modified by various concentrations of EDC *in vitro*, followed by examination of 5.8S ribosomal RNA on a polyacrylamide (rRNA). The concentrations tested in these experiments were 0.5%, 1%, 2%, and 3%. No RNA modification was observed when using 0.5% EDC in lane 2 (Figure 3.3A). RNA modifications were observed at EDC concentrations of 1% to 3% (lanes 3-5), with modifications occurring exclusively at guanine and uracil nucleobases. The modifications in these lanes provide evidence that the RNA was successfully modified by the 1%, 2%, and 3% EDC. The *in vitro* experiments were used to test EDC

reactivity with the RNA. If these reactions were not successful, they would have to be repeated with different values of EDC concentration and could not be used *in vivo* until they are successful. These *in vitro* experiments may be used to inform some parameters for EDC concentration used in subsequent *in vivo* rice treatments and potential protein binding interactions.

Ideally, *in vivo* treatments would use an EDC concentration that gives single-hit levels of RNA modification. Single hit kinetics is defined as the concentration of reagent which modifies 1 nucleotide per 100 to 200 nucleotides along a stretch of RNA. This is to make sure just enough EDC was used to modify the guanines. Higher concentrations of reagents could result in the misfolding of the RNA because they disrupt RNA-RNA and RNA-protein interactions that existed before treatment. Because the goal is to later use these techniques to learn about native RNA folding that is usually found in cells, it is important to avoid disruption of natural folding in future *in vivo* experiments. Under modifying RNA also isn't ideal because nucleobases that should be modified may be missed. If the gel did not achieve this goal, it would be visible by either excess dark bands (too much EDC) or no bands/very faint bands (too little EDC). The extent to which EDC modifies RNA can be calculated by quantifying the intensity of the bands on the gel using the following formula:

$$M_{\%} = I_{mod} / I_{total}$$

Here, $M_{\%}$ represents percent modification, or the percent of the RNA which is modified by EDC. $M_{\%}$ is calculated by taking the sum of band intensity for each band except the topmost band (I_{mod}) and dividing it by the total intensity of all bands including the topmost band (I_{total}). Ideally, this should result in a value between 10% and 20% modification. While percent

modification was not quantified, comparing the gel to similar gels gives a qualitative percent modification that may be within the acceptable range.

An important takeaway is that a small amount of modification *in vitro* could result in almost no modification *in vivo*, whereas excessive modification *in vitro* may also lead to excessive modification *in vivo*. Based on this gel, somewhere between 2-3% EDC may be appropriate for *in vivo* studies. This is because they are slightly over modified so *in vivo* it may be an appropriate amount of modification (Figure 3.3A). However, since *in vivo* and *in vitro* systems greatly vary the results from *in vitro* experiments cannot be directly translated to *in vivo*. *In vitro* studies can provide a ballpark for *in vivo* studies and eliminate some options. As mentioned before, if 0.5 % EDC does not modify RNA *in vitro*, it will certainly not work *in vivo* because the RNA is a lot less accessible.

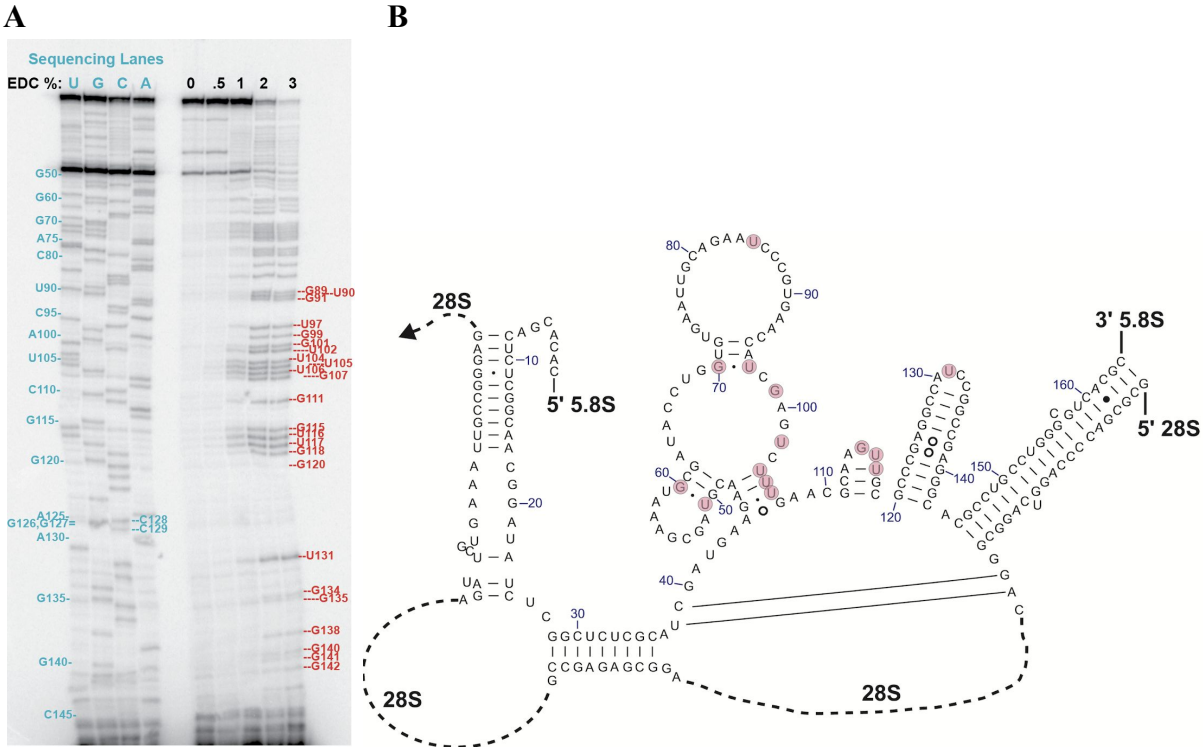

Figure 3.3. *In vitro* reactions of EDC with rice 5.8S rRNA. (A) Polyacrylamide gel showing modifications of 5.8S rRNA at EDC concentrations from 0.5% to 3%. The leftmost four lanes

show the sequence of 5.8S rRNA. (B) Secondary structure of 5.8S rRNA with red circles indicating bases modified by EDC. Modified from reference #16.

Mapping the EDC reactive nucleotides onto the secondary structure of rice 5.8S rRNA shows that EDC predominantly targets single-stranded nucleotides (Figure 3.3B), which is expected. This result is similar to results obtained for *in vitro* and *in vivo* reactions of rice, mouse, and bacterial RNA with EDC.[11,13] Interestingly, a double-stranded helical region of RNA appears to react with EDC. However, this phenomenon was also observed in previous work with glyoxal as well as with EDC, and indicates that this region of 5.8S rRNA may not actually form duplex structure *in vitro*.[13,14]

For *in vivo* studies, rice shoots were treated with EDC, then the RNA was extracted and reverse transcribed. EDC appeared to modify 5.8S rRNA *in vivo* based on the band pattern (Figure 3.4). Due to the extensive banding of the control (0% EDC) lane, it is difficult to determine quantitatively the extent of EDC modification. However, the band pattern for the *in vivo* reaction appears similar when compared to the pattern shown in the *in vitro* reaction (Figure 3.3A). Furthermore, for the bases where measurement is possible, it appears that EDC modified only guanine and uracil nucleobases *in vivo,* as anticipated.

Since the gel showed that the experimental method for treating the rice plants was effective, the next step would be to treat more rice plants with 2.5% EDC. Additionally, similar conditions developed here could also be used to treat other plants with 2.5% glyoxal. Now that the chemical reagent has to traverse the cell membrane there is a greater risk of it not modifying the RNA. To ensure modification and create enough chemically probed RNA to pull down and sequence, 2.5% EDC was used instead of 2%. After successfully modifying the RNA, the next

step is to isolate RNAs of interest. The methodology designed to achieve this goal is described in the next chapter.

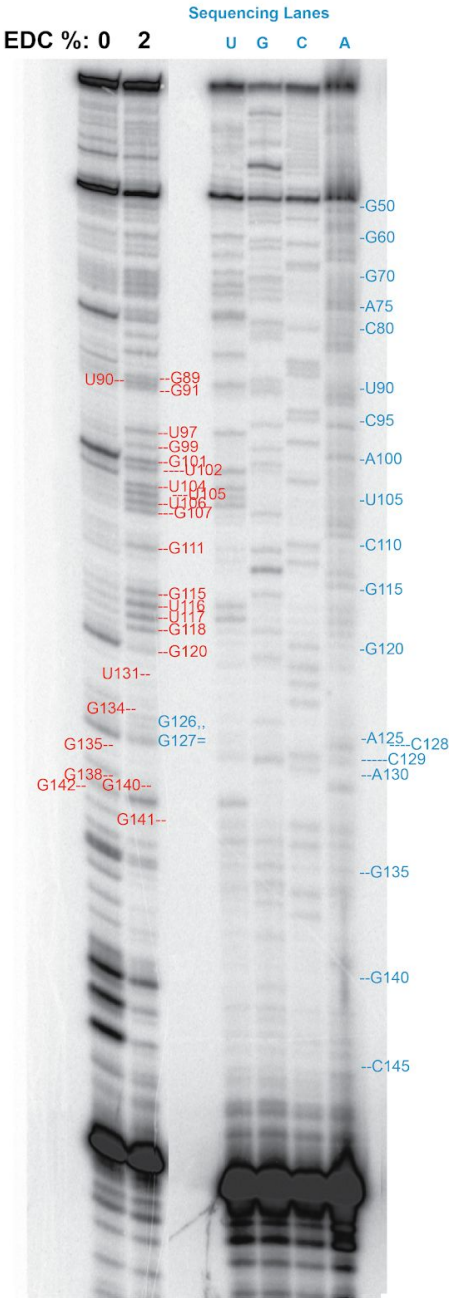

Figure 3.4. *In vivo* reaction of EDC with rice 5.8s rRNA. Polyacrylamide gel in *vivo* that exemplifies single hit kinetics. The reverse transcription stops (dark bands) show that the RNA bases were effectively modified with EDC. The rightmost four lanes indicate the sequence of 5.8S rRNA.

Chapter 4
RNA Pull-Down and Future Directions

4.1 Abstract

Probes need to be designed to target specific RNAs for future analysis. As mentioned previously, designing probes can be expensive and take a long time to develop. In this project, the methods performed for probe generation were lengthy and susceptible to error. The methodology carried out in this experiment along with troubleshooting mechanisms is described in this section. Additionally, a more efficient and cost effective approach to generate probes to target RNAs of interest is proposed as a future direction. These methods eliminate steps from the original design without significantly adding to the cost. The next steps and future directions after successfully creating probes are also proposed in this section. These steps involve isolating the RNA and preparing it into cDNA.

4.2 Introduction

The goal with these experiments was to find a cost-effective way to design cDNA probes to target RNAs of interest. Originally, the IC4R database was used to select genes with high transcript levels in the leaves and shoots. After inability to form substantial PCR products, mRNAs that code for housekeeping genes were selected because of their high abundance in cells that increases the likelihood of their capture. The Khurana lab provided a list of housekeeping genes for *Genus species*, of which five were chosen for study here.[17] The five genes I selected to target encode; actin 11, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), citrate synthase, polyubiquitin (UBQ), and ubiquitin conjugating enzyme. Actin 11 forms muscle filaments while GAPDH and citrate synthase are involved in the glycolysis and citric acid cycles. UBQ and

ubiquitin conjugating enzymes assist in the degradation of proteins and other cellular components.

After selecting target genes, I began the probe design process. First, I used NCBI Primer-BLAST to create primers to amplify DNA material.[18] I designed multiple primer sets for different regions on each selected gene. This helps to increase the likelihood of amplification and generate more probe material, which increases the chance of capturing target RNAs. The primer sets were designed to be at least 100 nucleotides apart and have a melting temperature (T_m) around 60-63℃. The self-complementary and 3′ self-complementary numbers were chosen at the lowest values available because this means that they are not likely to dimerize or form interactions within themselves.

Name	Region	Length	Primer Name	Tm (Celcius)	%GC	Self Comp.	3' Self Comp.	Sequence
			60SARP_Fwd_1	69.01	64	4	0	TGTATCTAATACGACTCACTATAGGGGCGATCAAGAGGACCAAGGCGGAG
Probe 1	3-226	224	60SARP_Rev_1	68.63	63.64	4	0	TGTTGCCGGTGTTGTCGGCGTG
			60SARP_Fwd_2	67.09	70	6	2	TGTATCTAATACGACTCACTATAGGGGTGCACGCCGACAACACCG
Probe 2	201-430	230	60SARP_Rev_2	66.01	56	4	1	GCACCTGGAAGAAAGAGGTCTGGGA
			60SARP_Fwd_3	65.55	56	4	0	TGTATCTAATACGACTCACTATAGGCCAGGTGCTTAACATCCCCACCAAG
Probe 3	423-725	303	60SARP_Rev_3	65.93	54.17	7	0	TGTGGTGCAGCAGCAATAGTTGGG
			60SARP_Fwd_4	66.49	63.64	4	3	TGTATCTAATACGACTCACTATAGGGGTTGCCTCAGTCTCCCTGGCA
Probe 4	672-932	261	60SARP_Rev_4	67.51	68.18	5	0	CCATCGGACTCCTCCTCGGGCT
			60SARP_Fwd_5	67.6	70	5	3	TGTATCTAATACGACTCACTATAGGCCCCAGTTGCCGCCGACAGT
Probe 5	845-958	114	60SARP_Rev_5	67.62	58.33	5	1	AGTCGAAGAGGCTCATGCCCAGGT

Table 4.1. Primer sets for 60S Acidic Ribosomal Protein P0 (IC4R-OSJ08G025800). Original gene selected from IC4R based on its high transcript level in rice leaves.

Name	Region	Length	Primer Name	Tm (Celcius)	%GC	Self Comp.	3' Self Comp.	Sequence
Probe 1	450-550	80	GMD_CS_P1fwd	62.35	48	4	2	TTCGTTTTAGGGGTCTCTCGATTCC
			GMD_CS_P1frev	62.22	60	3	1	CTCAGGCAAAGGCTCCCCAT
Probe 2	850-950	90	GMD_CS_P2fwd	60.57	44	8	3	CTATAGCAGCTGATAATGCACTGGA
			GMD_CS_P2rev	61.66	45.83	3	0	GTCGCATCAACTCAAGCATTTTGG
Probe 3	1500-1600	80	GMD_CS_P3fwd	60.8	60	4	0	CCGCTCGAAAGACCGAAGAG
			GMD_CS_P3rev	62.87	57.14	4	2	CCGTGTAGGCGTGTGTAGCTT
Probe 4	1700-1820	82	GMD_CS_P4fwd	62.09	45.83	6	0	TGCTCTCAGCTGGTAGGATTTTGT
			GMD_CS_P4rev	60.82	45.45	6	3	AATCACATTGTTGCCCAATGGC
Probe 5	1200-1300	83	GMD_CS_P5fwd	62.16	50	4	2	TGTTCCTGGCTTTGGTCATGGA
			GMD_CS_P5rev	61.95	44	8	0	GCAAGTACTTCAAAGCAAACTCCCT

Table 4.2. Primer sets for five housekeeping genes. Information pertinent to the genes that encode a) actin 11, b) GAPDH, c) citrate synthase, d) UBQ, and e) ubiquitin conjugating enzyme are shown (not yet used in pull down).

4.3 Methods

Primer Design

On NCBI Primer-BLAST, I entered the sequences of RNA after downloading them from the IC4R database. After entering the sequence, the Primer-BLAST tool identified and verified that the sequence was correct and coded for the selected gene. The goal was to have similar Tms for the primers and within the range of 60-62 ℃.

Probe Generation

After designing the primers to target the probes, genomic DNA was extracted in rice plants following the protocol listed within the Macherey-Nagel NucleoSpin RNA Plant kit. The primers were then added to the solution and PCR was used to amplify the targeted material. PCR

was performed many times but DNA extracted from preparative gels produced low yields. To increase PCR product, the final PCR used 92 µL of Master Mix (water, genomic DNA template, Q5 reaction buffer, dNTPs, Q5 DNA polymerase) and 8 µL of primers. There were 30 cycles, a 98 ℃ denaturing step, 62 ℃ annealing temperature, and 72 ℃ for elongation. Next, a 2% agarose gel was run to ensure the amplification of DNA. The bands were compared with those of a low MW ladder and a 100 base pair ladder. For this step, only two primer sets were successful in amplifying the DNA. I attempted this more than once and later attempts were not more successful. A 2% preparative gel was then run to purify the DNA that was successfully amplified and it was extracted from the gel using a kit. Only 20 ng/ µL DNA was effectively isolated.

Transcription was then performed by adding T7 polymerase to the DNA solution and leaving it at 37 ℃ for 4 hours. This allowed generation of the single stranded RNA. The placement of the T7 promoter specifies transcription of the antisense strand, resulting in sense RNA.

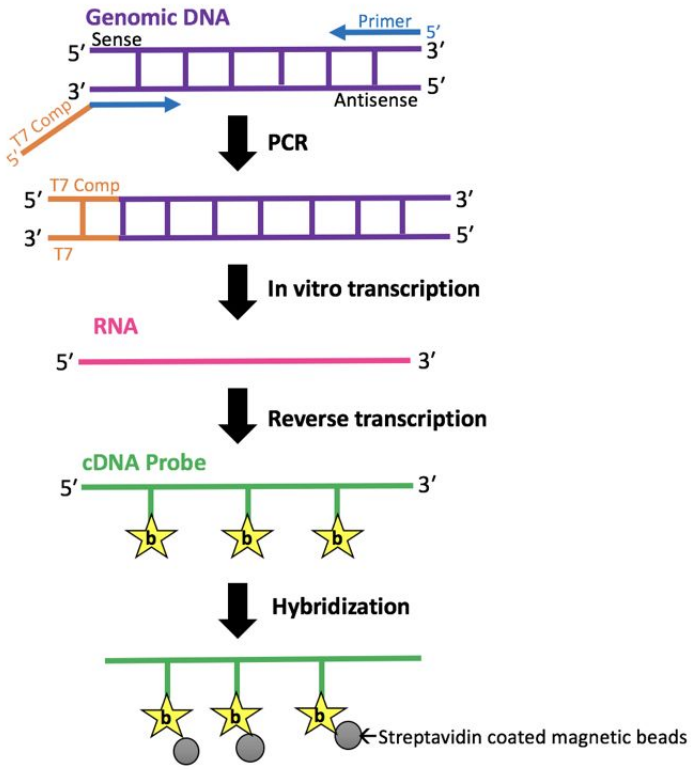

Figure 4.1. Probe design flowchart. The DNA material is amplified in PCR then transcribed into RNA then biotinylated in reverse transcription to make cDNA probes.

4.4 Results and Discussion

Some headway was made with probe generation for two of the probes. The DNA was successfully amplified for probes 1 and 3 (Figure 4.2). In vitro transcription was performed on the two primer sets that had amplification but only the transcription for probe 3 worked (gel image not recorded). This initial method was very cost-effective but not time-efficient. The process was very difficult and required a lot of PCR and transcription adjustments. PCR and in vitro transcription parameters were adjusted on numerous occasions to maximize yield. For PCR, the number of cycles and PCR solution volume were increased and the annealing temperature

was lowered (Table 4.3). Lastly, the reaction time for transcription and amount of T7 were increased.

PCR Parameters	Original Trial	Revised Trial
Annealing Temperature (℃)	65	62
Number of Cycles	25	30
PCR Volume (µL)	25	200

Table 4.3 Adjusted PCR Parameters. The PCR parameters that were adjusted to maximize yield of product.

Due to troubles with the transcription, the next steps were not attempted. Reverse transcription with biotinylated dUTPs could then be used to biotinylate the probes and aid in future pull down with streptavidin coated magnetic beads. After RT, the cDNA could be antisense and able to hybridize to the RNA targets. Streptavidin coated magnetic beads could then be added to isolate the target RNA-probe complexes by binding the biotinylated probes and removing them from solution. A magnet would then be used to isolate these complexes. To isolate the RNA from the complex, excess base would be added to disrupt the hydrogen bonding between the RNA and DNA.

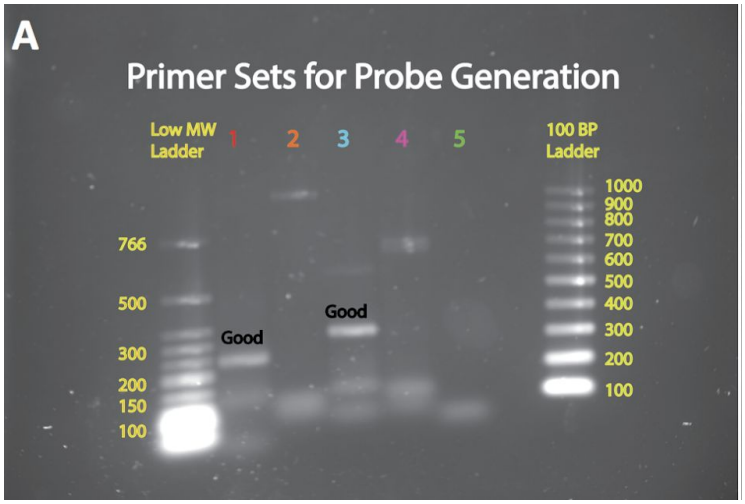

Figure 4.2. Verification of PCR amplification from genomic DNA template. (A) Depiction of an agarose gel to detect the presence of PCR amplicons for probe generation (primers used from **Table 4.1)** . Regions 1 and 3 of the DNA were amplified effectively because they match the predicted number of base pairs (~300 bp).

4.5 Future Directions

Revised Pull Down Methodology

The original probe design process presented challenges that were difficult to overcome. A way to troubleshoot this problem was to eliminate the number of steps to decrease possibilities of error. A shorter, more efficient alternative method was devised to generate the probes (Figure 7). In the revised method design, the in vitro transcription and reverse transcription steps would be removed. This process involved using biotinylated primers during PCR to avoid the need to perform in vitro transcription and reverse transcription. It is important to note that only the antisense strand is biotinylated so that it can bind to the sense RNA targets. After PCR, the biotinylated probes would be hybridized with the streptavidin coated magnetic beads. The DNA would then be treated with excess base to remove the sense strand, freeing the antisense strand to ultimately bind the target RNA. After receiving the target RNA, cDNA libraries would be prepared and sent off for sequencing. This would enable further analysis and identification of anionic guanines.

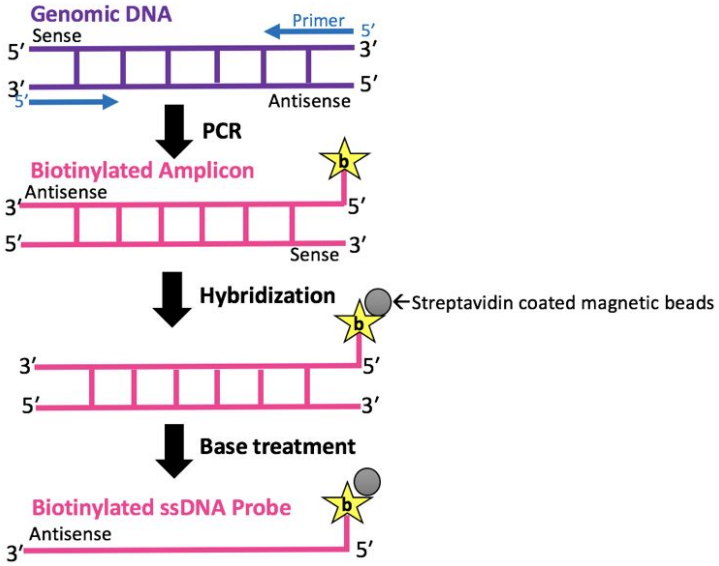

Figure 4.3. Revised probe design flowchart. Revised methodology to improve efficiency of RNA pull down. The *in vitro* transcription and reverse transcription steps are eliminated.

Anionic Guanine Identification

Once the probe design is mastered, multiple types of RNAs can be captured for the anionic guanine analysis. Once the target RNA is isolated, reverse transcription (RT) will be performed to reveal chemical modification. Guanines that reacted with glyoxal or EDC will prevent the RT polymerase from completing the RT. This happens because the modification of the nucleobase causes steric hindrance. Thus, the cDNA strands from RT are expected to be of varying lengths, displaying which guanines reacted with glyoxal or EDC. The cDNAs will then be sent off for sequencing which will confirm whether the selected RNAs were successfully targeted. The objective is to receive sequences of varying lengths, this means that certain nucleotides were modified. As a result of modification, the reverse transcription enzyme was halted by steric hindrance at the modified site and unable to transcribe the entire length of RNA into cDNA. In the long term, this method can be extended to other RNAs and even RNAs in different organisms. Pulling down multiple types of RNAs from different organisms can provide insight into how prevalent anionic guanines are within each organism's transcriptome.

REFERENCES

1. Serganov, A.; Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet.* **2007**, 8(10), 776-790. 10.1038/nrg2172
2. Kung, J. T.;Colognori, D.;Lee, J. T. Long noncoding RNAs: past, present, and future. *Genetics*. **2013**, 193(3), 651–669. https://doi.org/10.1534/genetics.112.146704
3. Agrawal, N.; Dasaradhi, P. V.; Mohmmed, A.; Malhotra, P.; Bhatnagar, R. K.; Mukherjee, S. K. RNA interference: biology, mechanism, and applications. *Microbiology and molecular biology reviews: MMBR*. **2003**, 67(4), 657–685. https://doi.org/10.1128/mmbr.67.4.657-685.2003
4. Rouskin, S.; Zubradt, M.;Washietl, S. *et al.* Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature*. **2014**, 505, 701–705. https://doi.org/10.1038/nature12894
5. Butcher, S. E.; Pyle, A. M. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res*. **2011**, 44, 1302−1311.
6. Mitchell III, D.; Assmann, S. M.; Bevilacqua, P. C. Probing RNA Structure In Vivo. *Curr. Opin. Struct. Biol.* **2019**, 59, 151-158. https://doi.org/10.1016/j.sbi.2019.07.008
7. K.A. Harris Jr.; D.M. Crothers, E. Ullu. In vivo structural analysis of spliced leader RNAs in *Trypanosoma brucei* and *Leptomonas collosoma*: a flexible structure that is independent of cap4 methylations. *RNA*. **1995**, 1, 351-362.
8. Ziehler, W.A.; Engelke, D.R. Probing RNA structure with chemical reagents and enzymes. *Curr Protoc Nucleic Acid Chem*. **2001**, 1-21.

9. Antal, M.; Boros, E.; Solymosy, F.; Kiss, T.Analysis of the structure of human telomerase RNA in vivo. *Nucleic acids research*. **2002**, *30*(4), 912–920. https://doi.org/10.1093/nar/30.4.912
10. Incarnato, D.; Neri, F.; Anselmi, F.; Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol*. **2014**,15(10):491. doi: 10.1186/s13059-014-0491-2. PMID: 25323333; PMCID: PMC4220049.
11. Wang, P.Y.; Sexton, A.N.; Culligan, W.J.; Simon, M.D. Carbodiimide reagents for the chemical probing of RNA structure in cells. *RNA*. **2019**, 25, 135-146.
12. Szeto. 1-Cyclohexyl-3-(2-morpholinoethyl)-carbodiimide Metho-p-toluenesulfonate. **2001**, e-EROS.
13. Mitchell, D.; Renda, A. J.; Douds, C.A.; Babitzke, P.; Assmann, S.M.; Bevilacqua, P.C. In vivo RNA structural probing of uracil and guanine base-pairing by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). *RNA*. **2019**, 25:147-157.
14. Mitchell, D.; Ritchey, L. E.;Park, H.; Babitzke, P.; Assmann, S. M.; Bevilacqua, P. C. Glyoxals as in Vivo RNA Structural Probes of Guanine Base-Pairing. *Rna*. **2018**, 24(1), 114–124. https://doi.org/10.1261/rna.064014.117.
15. Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P.C. Charged Nucleobases and Their Potential for RNA Catalysis. *Acc. Chem. Res*. **2011**, 44(12), 1270-1279.
16. Cannone, J. J.; Subramanian, S.; Schnare, M. N.; Collett, J. R.; D'Souza, L. M.; Du, Y.; Feng, B.; Lin, N.; Madabusi, L. V.; Müller, K. M.; Pande, N.; Shang, Z.; Yu, N.; Gutell, R. R. The comparative RNA web (CRW) site: an online database of comparative

sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics*. **2002**, *3*, 2. https://doi.org/10.1186/1471-2105-3-2

17. Jain, M.; Nijhawan, A.; Tyagi, A. K.; Khurana, J. P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR, *Biochemical and Biophysical Research Communications*. **2006**, 345, 2, 646-651, ISSN 0006-291X, https://doi.org/10.1016/j.bbrc.2006.04.140.
18. Ye J.; Coulouris G.; Zaretskaya I.; Cutcutache I.; Rozen S.; Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. **2012**, 13, 134. doi: 10.1186/1471-2105-13-134. PMID: 22708584; PMCID: PMC3412702.

Glenn Duncan
13815 Notley Rd Silver Spring, MD 20904 | gfd5078@psu.edu | (301)804-7708

EDUCATION

Fall 2017-Fall 2020 — **The Pennsylvania State University, University Park, PA**
Bachelor of Science, Biology: Neuroscience

RESEARCH EXPERIENCE

Fall 2017-Present — **Department of Chemistry, Penn State University,**
Undergraduate Researcher, Dr. Philip Bevilacqua

- Conducted independent research to identify anionic guanines in RNA. Performed PCR, gel electrophoresis, *in vitro* transcription, chemical reactions, and reverse transcription.
- Created presentations for lab meetings and completed a report to earn credit at the end of each semester.
- Drafted a Seniors Thesis to submit by Nov 30, 2020.

Summer 2020 — **Perelman School of Medicine, University of Pennsylvania,**
Summer Intern, Dr. Kelly Jordan-Sciutto

- Designed an experiment to identify RNAs that are responsible for the excitotoxic response neurons produce in neurological disease, specifically HIV Associated Neurocognitive Disorders (HANDs).

Summer 2019 — **Department of Chemistry, Penn State University,**
Summer Intern, Dr. Philip Bevilacqua SROP

- Worked with a writing coach and completed writing and professional development workshops while conducting full time research.

PROFESSIONAL/TEACHING EXPERIENCE

Fall 2019-Present — **Millennium Scholars Tutor**

- Tutored Millennium Scholars in chemistry and physics.

Fall 2019-Present — **Millennium Scholars Mentor**

- Mentored underclass Millennium Scholars and provided assistance with college transition and coursework.

PRESENTATIONS AND CONFERENCES

SUMMER 2020 — **SUIP Final Presentation**

- Completed a final abstract video to present my work over the summer as an SUIP intern.

FALL 2019 **American Chemical Society (ACS) Poster Presentation**

- Presented posters at this conference with students from other colleges, hosted at Penn State.

FALL 2019 **Rustbelt RNA Conference**

- Presented my poster at this conference hosted at Case Western Reserve University.

FALL 2019 **National Organization for the Professional Advancement of Black Chemists and Chemical Engineers**

- Received travel grant that covered lodging expenses to attend this conference in St. Louis, Missouri.

FALL 2019 **Eberly College of Science Undergraduate Poster Exhibition**

- Hosted at Penn State, students are invited to present when they receive grant money from the Eberly College of Science.

SUMMER 2019 **SROP Final Presentation**

- Presented research that I conducted over the summer to Penn State faculty members at the final conference.

SCHOLARSHIPS

Fall 2017-Fall 2020 **Millennium Scholars Program**

- Academic scholarship awarded to high achieving STEM students that are interested in pursuing PhDs and requires completion of 6 week summer bridge program.

Fall 2017-Fall 2020 **4YR Provost's Award**

- Award given to selected students when applying to Penn State.

LEADERSHIP POSITIONS

Fall 2020 **Delta Sigma Theta Sorority, Inc. (DST) Vice President**

- Simultaneously served as Membership Committee Chair and planned events for chapter members to build a stronger bond.

Fall 2020 **LOCKs Natural Hair Club Treasurer**

- Organized fundraisers to raise money for the organization to host events.

Fall 2019-Spring 2020 **DST Programming Chair**

- Planned events and organized collaborations to educate the Penn State community on various topics like COVID-19 and impact of climate change on people of color.

Fall 2019-Spring 2020 **NOBCChE Secretary**

- Maintained and kept detailed notes for executive board meetings where we all planned events to educate the Penn State community on science topics.

Fall 2019-Spring 2020 **LOCKs Social Media Chair**

- Created posts for the social media page and engaged with general body members.