Spring 5-9-2020

Testing Kalanchoe laxiflora U6 Promoters for Optimal sgRNA Expression in Tobacco

Marlee E. Jones

Testing *Kalanchoe laxiflora* U6 Promoters for Optimal sgRNA Expression in Tobacco

by

Marlee E. Jones

A Thesis Submitted in Partial Fulfillment
Of the Requirements for the
University Honors Program

Department of Biology
The University of South Dakota
May 2020

The members of the Honors Thesis Committee appointed
to examine the thesis of Marlee E. Jones
find it satisfactory and recommend that it be accepted.

Bernie Wone, Ph. D.
Assistant Professor of Biology
Director of the Committee

Beate Wone, M. S.
Instructor of Biology

Scott Breuninger, Ph. D.
Professor / Director of
Honors Program

ABSTRACT

Testing *Kalanchoe laxiflora* U6 Promoters for
Optimal sgRNA Expression in Tobacco
Marlee E. Jones

Director: Dr. Bernie Wone

Within studies of plant genome modification there is a method of gene alteration which involves using CRISPR/Cas systems in order to target specific gene loci and implement the intended modification. This process is driven by the ability of a single guide RNA (sgRNA) to find this target location within the plant's genome. This sgRNA strand is expressed due to the presence of a proper RNA promoter. In an effort to understand the effect of an endogenous RNA promoter's sequence on the successful expression of a sgRNA strand, an experiment was developed involving twelve U6 promoter sequences from a crassulacean metabolism acid (CAM) plant, *Kalanchoe laxiflora*. This study was designed to test each promoter sequence individually to determine if it would successfully express the guide RNA (gRNA) strand within a construct. This success would have been verified through the transient expression of the sgRNA/Cas-GFP fusion construct using confocal microscopy.

KEYWORDS: promoter, plasmid, endonuclease

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to my project advisor, Dr. Bernie Wone, for his support, guidance, and contributions towards my research over the duration of this project. Over the course of this research, he has been a valuable source of both information and encouragement and has aided me in becoming a more knowledgeable and well-rounded student. This research opportunity was a result of his time and efforts, and I am honored to have had the opportunity to work with him.

I would also like to express my appreciation to Beate Wone and Scott Breuninger for serving on my thesis committee and setting aside their time to assist me throughout this process. Their input and support have been very beneficial to me.

I would like to show my appreciation for Kumudu Rathnayake for her guidance throughout our research and her willingness to assist me in any way possible. Her expertise was an asset throughout this process.

Funding for the research was supported by the South Dakota Board of Regents Competitive Research Grant, #A18-0015-001 to Dr. Wone.

CHAPTER ONE

Introduction

sgRNA/Cas Constructs as a Genome Editing Tool

Gene modification within plant genomes is continuously advancing in the field of plant molecular biology research. There are currently several effective methods in use that can produce the desired effect of manipulating a genome. Within molecular biology research, TALENs (transcription activator-like nucleases), zinc-finger nucleases, and Clustered regularly interspaced short palindromic repeats (CRISPRs) dominate the field of genetic modifications, with CRISPRs being the most favorable method of use for these alterations. CRISPRs were discovered as a component of prokaryotes that help with immunity against external invaders (Mojica et al., 1995; Jinek et al., 2012).

CRISPRs are segments of DNA that contain palindromic sequences and are used as viral protection within bacteria (Mojica et al., 1995). The origin of CRISPR/Cas 9 mediated gene modification is derived from a DNA cleavage complex which is a natural component of certain species of prokaryotes (Mojica et al., 1993; Schaeffer et al., 2015). The Cas 9 protein is used to cause double-stranded breaks within DNA. A CRISPR/Cas system is used to cut the DNA upstream of the promoter region and insert genes via a complementary single guide RNA (sgRNA) strand. It can target a specific region because the guide strand contains a sequence that is complementary to the one of interest.

CRISPR modification methods can alter a genome in two ways. Using a knockout procedure involves disrupting the gene in question in order to delete the function within the system. This is done *via* non-homologous end joining (NHEJ). NHEJ is used as a

default repair system within most organisms because it quickly and efficiently repairs broken DNA, which can be lethal if not fixed. By using a sgRNA/Cas construct to cut within a gene in question, the NHEJ repair system causes insertions or deletions of nucleotides upon repair. This changes the sequence of the target gene and disrupts its normal function. The other method of CRISPR modification utilizes a gene knockin procedure which involves adding a gene and its associated function into a target genome. This is done *via* homologous directed repair (HDR). HDR is not regularly used within most organisms, such as plants and animals, which makes this method harder to implement within complex organisms. HDR in this case utilizes a designed template strand to insert the new gene into the genome. The structure binds to the broken DNA and complementary bases to the template strand are recruited to the DNA strand. This process allows a new function, not previously present, to be added to an organism's genome (Lino et al., 2018).

The background on CRISPR systems is relevant to our experiment because it highlights one of the end goals of the study. A sgRNA/Cas system serves as a sort of prelude to implementing CRISPR knockout within a modification process. Our attempts to construct one of these systems through the study of U6 promoter sequences allow us to specifically observe the effects of the promoter sequence. This study is important because when implementing sgRNA/Cas modification techniques, the RNA must be expressed in order to find the target location. The type of promoter used can affect the efficiency of the delivery of this plasmid (Mefferd et al., 2015).

Concerning sgRNA/Cas systems, the process of modification, which takes place within the plant's genome, is straightforward. The sgRNA is used to find the desired gene

for the Cas9 protein to modify, and a U6 polymerase promoter drives the expression of this gRNA. Once the gRNA binds to the target site, the strands are unwound and cut by Cas 9. Once the cut is made by the Cas 9 protein the strands undergo NHEJ in order to repair the cut. The process of designing this plasmid, however, is where complications can happen. In order for this plasmid to work, not only does the promoter need to express the sgRNA strand, but these strands—the promoter, gRNA, Cas 9—must all be ligated together to an expression vector using proper endonucleases. Within our experiment double digestion was done in order to insert and correctly orient our strands into our expression vector.

Determining an efficient RNA promoter sequence to drive a sgRNA/Cas construct was the overall goal of our experiment. Within a sgRNA/Cas plasmid, there are two promoters; one expresses the Cas endonuclease and one expresses the guide RNA strand. The promoter used to express the Cas is a constitutive promoter so it is always expressing the associated Cas gene (Jiang et al., 2018). The sgRNA, however, requires a promoter that is preferably native to the organism in which it is being introduced. In our experiment, we used endogenous U6 promoters because they are native to our long-termed goal of editing CAM plants and will allow the expression of the sgRNA strand within the plasmid insert. Without the proper expression of the sgRNA strand, the target gene of the plant genome cannot be located. For our study, we set out to test the ability of our proposed sgRNA promoters through transient expression of our sgRNA/Cas-GFP fusion construct. Theoretically, our results would have shown that the promoter activated expression of the gRNA, and the plasmid construct—which contained a GFP reporter—would cause the tobacco leaves to glow green when viewed with confocal microscopy.

The goal of my research was to determine which CAM plant *Kalanchoe laxiflora* U6 promoters express sgRNA. To achieve this goal, three objectives were pursued.

Objective 1: Extract synthesized U6 promoter from the plasmid of *E. coli*.

Objective 2: Ligate the U6 promoter into an expression vector with a sgRNA and the Cas 9 gene fused with a GFP reporter.

Objective 3: Transiently express the U6/sgRNA/Cas 9/GFP fusion construct in tobacco leaves.

Understanding how sgRNA/Cas constructs work and what expresses them is important in being able to implement these modifications on a wider scale. This experiment is relevant within plant molecular biology research because it shows how the implementation of CRISPR methods within plants is still being studied. Before genetic modification methods, plants were modified through selective breeding, which relied on time and chance. Due to decreases in genetic variation, this method has quickly shown to be inefficient. Being able to induce CRISPR methods within plants will allow plants researchers to impact the way many plants function. This could mean better survival rates and adaptation to abiotic stressors in the future (Zhang et al., 2018).

CHAPTER TWO

Materials and Methods

Experimental Procedures

In an attempt to study and determine a sgRNA strand's ability to target genes for crassulacean acid metabolism (CAM) plant modifications using the CRISPR approach, an experiment was done to determine which *Kalanchoe laxiflora* U6 promoter is most effective in the expression of a specific RNA strand. The goal was to create a plasmid that would contain a promoter sequence which would robustly express the designated sgRNA. The expression would allow the RNA strand to be used to locate a specified gene within the target CAM plant genome.

The experiment was done testing 12 different U6 promoter sequences, of which I was responsible for U6-6, U6-7, and U6-8. These promoters were isolated from different *E. coli* cultures that contained the synthesized promoter plasmid and were then added to an LB broth/ampicillin mixture and shaken overnight at 250 rpm and 37°C. The LB broth was prepared using 250 mL nanopure water and 6.25g of broth powder. Each promoter colony was added to its own designated sterilized flask of broth. From here, 5 milliliters (mL) from each individual U6 promoter flask were added to four correlating 50 mL tubes, leaving 12 tubes in total. The steps from the *QIAprep Spin Miniprep Kit* were then followed, with slight modifications made (QIAGEN, Hilden, Germany). The specific modifications implemented while following this protocol were as follows. In Step 1, the time was modified from three minutes to five. During Step 6, 7, and 8, the tubes were centrifuged for one minute, using 500 microliters of buffer PB in Step 7 and 750

microliters in Step 8. The final modification to the protocol occurred in Step 10 when nanopure water was used as the buffer. After following this procedure, the concentrations of each sample were recorded using a NanoPhotometer NP80 (Imegen, Munich, Germany) (Table 1).

Table 1.

Isolated plasmid concentrations as determined by a NanoPhotometer.

Promoter	Concentration (ng/µL)	260/280	260/230
U6-1	32.25	1.838	1.539
U6-2	24.00	1.798	2.133
U6-3	28.64	1.813	2.256
U6-4	15.85	1.668	1.546
U6-7-1	24.50	1.750	2.103
U6-7-2	38.50	1.812	2.452
U6-7-3	34.55	1.868	2.450
U6-7-4	25.90	1.867	2.582
U6-8-1	35.15	1.836	2.735
U6-8-2	30.15	1.822	2.680
U6-8-3	27.50	1.821	2.926
U6-8-4	--	2.435	.812

Due to low concentration values, the four tubes of each individual promoter sequence of the same promoter were combined into a single tube (e g., U6-7-1, U6-7-2,

U6-7-3, and U6-7-4). They were then placed into a speed-vac to dry, and concentrations were re-calculated (Table 2).

Table 2.

Plasmid concentrations recorded after combining samples of the same promoter.

Promoter	Concentration (ng/μL)	260/280	260/230
U6-6	88.85	1.815	1.528
U6-7	225.70	1.857	1.778
U6-8	199.80	1.859	1.767

With these concentration values being more ideal than previously recorded, the next step was to determine which endonucleases could be used to splice the sgRNA into the expression vector. These endonucleases needed to cut the promoter sequence out of its promoter plasmid for it to be placed within the expression vector pGWB402, which would be used downstream for the gene modifications. These endonucleases were determined by finding which restriction enzymes would cut the ends of the promoter insert, but not cut within the promoter itself. They also must cut within the Multiple Cloning Site (MCS) of the receiving plasmid, but not elsewhere in the plasmid (Tóth et al., 2014).

By using two different restriction enzymes, we can properly orient the promoter behind the sgRNA and Cas segments. It was determined that for promoters U6-6 and U6-8, restriction endonucleases *EcoRI* and *XmaI* would be used and for promoter U6-7 *XmaI* and *KpnI* would be used. Using these designated restriction endonucleases, double

digestion was performed in order to splice and insert the segments into the pGWB402 vector (Table 3).

Table 3.

Reagents and the concentrations used to perform the double digestion.

Reagents	U6-6	U6-7	U6-8
EcoRI	1 µL	--	1 µL
XmaI	1 µL	1 µL	1 µL
KpnI	--	1 µL	--
DNA	22.5 µL	8.9 µL	10.0 µL
Cutsmart	5 µL	5 µL	5 µL
Free Water	20.5 µL	34.2 µL	33.0 µL

After this procedure, all components were mixed by pipetting up and down. From here, the tubes were incubated at 37 °C for 2 hrs. The double digestion, in this case, was done in order to both cleave the plasmid and ensure correct orientation upon insertion of the components. A double digestion ensures that the inserts are correctly placed because two different endonucleases used on either end would create mismatches. This way, the inserts cannot be ligated in upside down or backward.

After the digestion of the strands was complete, an agarose gel was prepared in order to check that the plasmid was successfully digested. The agarose gel was prepared using 1.5 grams of agarose. This was mixed with 100mL of a pre-prepared 1X TAE buffer in a 250 mL flask. The buffer was made using the following protocol:

1. Concentrated 50X stock solution of TAE was made by weighing out 242 grams of Tris base and dissolving it in 750 mL deionized water.
2. 57.1 mL of glacial acid and 100 mL of 0.5M EDTA were added
3. To create 1X TAE, 20 mL of 50X TAE was added to 980 mL of water

Gel extraction of the electrophoretic bands was done following the *QIAquick Gel Extraction Kit Protocol* (QIAGEN, Hilden, Germany).

CHAPTER THREE

Results

Experimental Outcomes

At this point in our experiment, no definitive results could be observed because the process was stymied during attempts to amplify the promoters using polymerase chain reaction (PCR). Apparently, methylation of the restriction sites was inhibiting our endonucleases from excising out the promoters. Because of this, the promoters could not be successfully ligated into our expression vectors during the double digestion. In order to resolve this issue, PCR needed to be run to amplify the promoters and create unmethylated strands so we could retry the ligation. The PCR process was needed in order to amplify our promoters, however, dimerization of the primer strand occurred during the process, meaning the primers—which are used to direct the construction of the replica strand—annealed to one another instead of the promoter strand. This meant the promoter could not be amplified for use within our target vector. With appropriate resources and time, I assume we would have been able to identify what was causing these dimers and prevent it from happening again.

Had these obstacles not occurred, we likely would have been able to construct a sgRNA/Cas system and determine if it was successful expression of the gRNA strand by the promoter sequences. The final steps in this experiment would have been to subclone the Cas 9 gene of the pUC57 vector to the pGWB402 and ligate in the U6 promoter. If the promoter had been successful in expressing the sgRNA, the construct would have been transiently expressed within the leaves of our *Arabidopsis* plant.

CHAPTER FOUR

Discussion and Conclusions

Options and Solutions Moving Forward

An experiment was conducted to test twelve different *K. laxiflora* U6 promoter sequences for their capability to successfully and efficiently express a designated sgRNA strand. In this case, the efficiency of the promoters could be determined based on a sgRNA/Cas-GFP fusion construct's ability to express GFP fluorescence when viewed *via* confocal microscopy. This experiment is important because the components of this study can be applied to genetic modification efforts involving the use of CRISPR technology in CAM plants. The CRISPR method is quickly becoming a popular process of modification within plant molecular biology research due to its specificity and ability to be easily manipulated by researchers, and it holds great promise for further applications of gene editing and targeting in almost all living species in the future (Rozov et al., 2019).

The goal of the experiment was to determine which of the U6 promoters can efficiently express sgRNA. The rationale for testing these sequences is because, without proper expression of the sgRNA strand, the construct would be unable to implement any gene modification within the CAM plant. The role of the U6 promoter within a sgRNA/Cas system is to express the guide RNA so it can locate the target gene within the plant genome.

Designing our construct for this experiment involved first isolating our promoter plasmids from bacterial cultures. From there, these promoters were to be cut out of their plasmids via restriction enzymes and inserted into the expression vector along with the

sgRNA and Cas sequences. For this to be done, we had to first determine which endonucleases would work properly for the digestion. These were chosen by matching which endonucleases would cut within the cloning site of the target vector with which ones would cut the promoter out of its original one. Choosing proper endonucleases can be challenging as they cannot cut the target plasmid outside of the cloning site of the expression vector. We need to also ensure that the promoter sequence itself is not cut either.

The endonucleases in this procedure were used to transfer the promoter from its original plasmid to a new one, which would become our sgRNA/Cas construct used for gene modification. During our experiment, methylation of the restriction sites was preventing proper excision and ligation into the expression vector. Methylation within restriction sites can inhibit the function of restriction enzymes. To circumvent this issue, we attempted to run PCR because PCR products are unmethylated (Marinus and Olesen, 2014). This would have resolved the excision issue and we should have been able to ligate the excised promoters into expression vectors.

However, the formation of primer-dimers inhibited our attempts to amplify the promoter sequences. The PCR process should have resulted in many duplicates of our original promoter strands. During PCR, the primers should have annealed to the 5’ ends of the denatured target strands. Without being able to properly amplify our promoters, we were unable to move forward with the experiment.

With more time, we likely would have been able to fix these issues and proceed with our experiment. Preventing primer-dimers from forming during PCR is straightforward, and there are multiple solutions. The design of the primers being used

impact the likeliness of dimers forming. Because our current primers were an issue, we could have designed longer primers and have done a gradient analysis to determine which temperature would be optimal for the target strands to anneal to the new primers. Also, if PCR is run at higher temperatures, the binding of primers to one another is less probable (Brownie et al., 1997). These alterations to our experiment would likely have allowed us to use PCR to amplify our promoters so they could be ligated into expression vectors.

As for the issue of the methylation itself, we also had an option to overcome this obstacle. An alternative to prevent this methylation from occurring would have been to grow the promoter plasmids within *E. coli* cultures that lacked any DNA methyltransferases (DNA MTase). DNA MTase is the enzyme that adds methyl groups to nucleotides. Without this enzyme present, there would have been no methylation of the restriction sites. While this process is ideal, it would have been a last resort due to the fact that it would put the promoters at high risk for mutations. The methylation of DNA strands is how mismatched bases are flagged for repair. Without this enzyme, the chance of mutations occurring within the sequences would have been high (Ding et al., 2016).

Had these obstacles not occurred, and the promoters had been properly ligated into our expression vectors, we would have been able to observe the transient expression of our sgRNA within our tobacco leaves. This would have allowed us to determine which of our U6 promoters were successful in robustly expressing their associated sgRNA. That is, if they showed fluorescence, it meant that the construct found its target DNA, and all of the components were being read and expressed, including the fused GFP reporter. The results of this study would have allowed us to further study sgRNA/Cas systems and attempt to conduct gene modifications using a designed construct. Our long-term goal

was to modify CAM plants, hence the reason CAM plant *Kalanchoe laxiflora* promoters were tested. Once we confirmed which promoter sequences were most successful in expressing our sgRNA strand, we would be able to use those promoters in our attempts to induce gene alterations.

Using sgRNA/Cas constructs to genetically modify organisms is a simplified version of CRISPR methods. The plasmid consists of a sgRNA, which contains a complementary sequence to that of the target gene, a Cas 9 endonuclease, and two promoters to express these two strands. The Cas gene is expressed by a constitutive promoter, which allows it to be continually expressed. The sgRNA, however, must be expressed by an RNA promoter that is native to the organism in question. Our experiment was conducted in order to test twelve U6 promoter sequences for their ability to express a sgRNA sequence. In our study, we chose CAM plant *Kalanchoe laxifora* U6 promoters. This study is important to any gene modification attempts because the success of an endogenous RNA promoter in sgRNA expression affects the overall efficiency of a sgRNA/Cas construct in genetically modifying a target gene.

References

Brownie J, Shawcross S, Theaker J, et al. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Research*. 1997;25(16):3235-3241. doi:10.1093/nar/25.16.3235

Campenhout, CV, Cabochette P, Veillard A-C, et al. Guidelines for Optimized Gene Knockout Using CRISPR/Cas9. *BioTechniques*. 2019;66(6): 295–302. doi:10.2144/btn-2018-0187.

Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. *Journal of Molecular Cell Biology*. 2016;8(3):244-254. doi:10.1093/jmcb/mjv050

Jiang P, Zhang K, Ding Z, et al. Characterization of a strong and constitutive promoter from the *Arabidopsis* serine carboxypeptidase-like gene AtSCPL30 as a potential tool for crop transgenic breeding. *BMC Biotechnology*. 2018;18(1). doi:10.1186/s12896-018-0470-x

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-821. doi:10.1126/science.1225829

Lino CA, Harper JC, Carney JP, Timlin JA. Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery*. 2018;25(1):1234-1257. doi:10.1080/10717544.2018.1474964

Marinus MG, Løbner-Olesen A. DNA Methylation. *EcoSal Plus*. 2014;6(1):10.1128/ecosalplus.ESP-0003-2013. doi:10.1128/ecosalplus.ESP-0003-2013

Mefferd AL, Kornepati AV, Bogerd HP, Kennedy EM, Cullen BR. Expression of CRISPR/Cas single guide RNAs using small tRNA promoters. *RNA*. 2015;21(9):1683-1689. doi:10.1261/rna.051631.115

Mojica F, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Holeferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology.* 1995;17(1):85-93. doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x.

Mojica FJM, Juez G, Rodrıguez-Valera F. Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular Microbiology.* 1993: (3):613–621. doi.org/10.1111/j.1365-2958.1993.tb01721.x.

Rozov SM, Permyakova NV, Deineko EV. The Problem of the Low Rates of CRISPR/Cas9-Mediated Knock-Ins in Plants: Approaches and Solutions. *International Journal of Molecular Sciences*. 2019;20(13):3371 doi:10.3390/ijms20133371.

Schaeffer SM, Nakata PA. CRISPR/Cas9-Mediated Genome Editing and Gene Replacement in Plants: Transitioning from Lab to Field. *Plant Science.* 2015;240:130-142. doi:10.1016/j.plantsci.2015.09.011.

Tóth E, Huszár K, Bencsura P, et al. Restriction enzyme body doubles and PCR cloning: on the general use of type IIs restriction enzymes for cloning. *PLoS One*. 2014;9(3). doi:10.1371/journal.pone.0090896

Zhang Y, Massel K, Godwin ID, Gao C. Correction to: Applications and Potential of Genome Editing in Crop Improvement. *Genome Biology* 2019;20(1). https://doi.org/10.1186/s13059-019-1622-6.