5-2016

Spider Aggregate Glue Sequence Characterization and Expression

Kyle Berg
Utah State University

SPIDER AGGREGATE GLUE SEQUENCE CHARACTERIZATION AND EXPRESSION

by

Kyle R Berg

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTERS OF SCIENCE

in

Molecular Biology

Approved:

Randolph V. Lewis Major Professor	Jon Takemoto Committee Member
H. Scott Hinton Committee Member	Mark McLellan Vice President for Research & Dean of the School for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2016

ABSTRACT

Spider Aggregate Glue Sequence Characterization

and Expression

by

Kyle R. Berg, Master of Science

Utah State University, 2016

Major Professor: Dr. Randolph V. Lewis
Department: Biology

Spider aggregate glue is secreted on to the webs of many spiders in the superfamily Araneoidea. Aggregate glue is sticky when wet and has a unique stress strain curve that varies depending on the relative humidity and speed of probe retraction. These features make it an attractive target for applications such as underwater adhesives or surgical glues. However, little is known about the genetic sequence of the glue protein itself. In this work, a method is identified to isolate the remainder of the aggregate gene sequence, and genetic constructs glue are created using known aggregate gene sequence. One construct is expressed in *E. coli* BL21 cells and the protein is tested for its adhesive properties.

(68 pages)

PUBLIC ABSTRACT

Spider Aggregate Glue Sequence Characterization

and Expression

Kyle Berg

Orb-weaving spiders create a glue which is secreted onto the web to capture and retain insects. This glue is made from aggregate protein. Aggregate glue is sticky when wet and can stretch far when pulled at slow speeds, and pull back with a greater force if pulled at high speeds which helps it achieve its purpose as the adhesive that keeps insects in the web. These features also make it an attractive target for applications such as a surgical glue or underwater adhesive. Unfortunately, very little is known about the aggregate glue gene, and knowledge of the genetic sequence is necessary to create a synthetic aggregate glue. In this work, a method is identified that will likely be able to find the rest of the glue sequence. In addition, using what is known about the gene, a synthetic glue sequence was extrapolated and used to create a synthetic glue using bacteria. Finally, this glue has begun preliminary testing as an adhesive.

ACKNOWLEDGMENTS

I am grateful to the Office of Naval Research and their funding for this project. In addition, much thanks goes to Dr. Michael Hinman for his willingness to share his genetic work expertise with me. I would also like to thank Dr. Paula Oliveira who helped me with expression and purification of the aggregate glue constructs. I am also thankful for Dr. Randolph Lewis, who oversaw the work. Finally, I am grateful for my wife, who took care of my family and me while I worked to complete this project. I could not have done this without her.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Spiders in the superfamily Araneoidea are known for constructing large orb webs to capture and retain their prey. Orb weaving spiders have the ability to synthesize six silks and one glue that are all evolutionarily derived from the same family of proteins.[1,2] The proteins and glue are all extruded from spinnerets on the posterior end of the spider, and these spinnerets have been used for decades as evolutionary evidence for the classification of spiders.[3] The spiders use these proteins to construct different elements of their webs, and each type of silk has distinct mechanical properties. For example, major ampullate silk is used to create the scaffolding of the web, piriform silk is used to lash the web to external objects, and the aciniform and tubuliform silks are used in creating the egg sac. Among orb weavers,

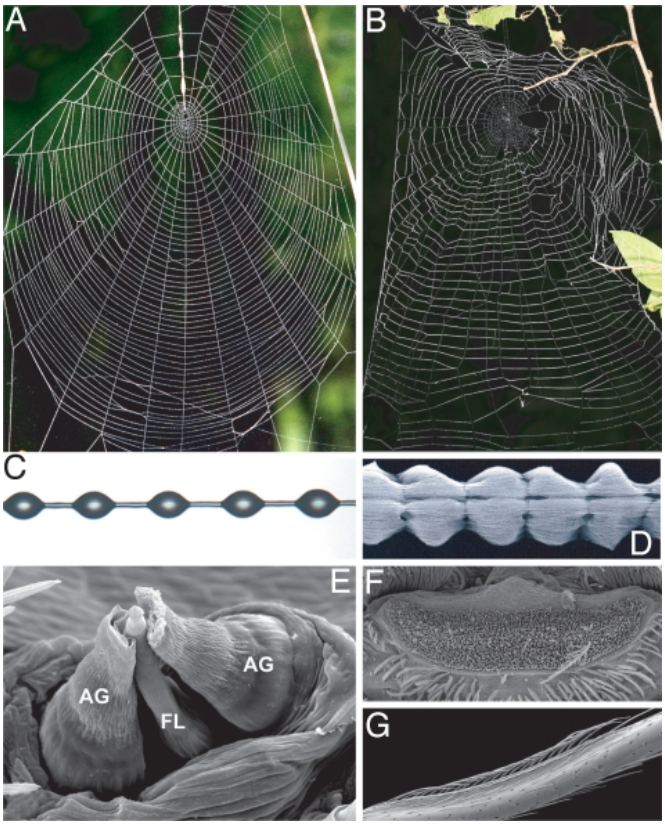

Figure 1.1 A comparison of aggregate glue webs vs cribellate silk webs A) A web utilizing aggregate glue as an adhesive B)A web utilizing cribellate silk as an adhesive C) Aggregate glue droplets on flagelliform silk D) Cribellate silk E) Aggregate gland spigots (AG) flanking a flagelliform silk gland spigot (FL) on the spinneret. F) The cribellum of a cribellate spinning spider G) The pseudoflagelliform core fiber of the cribellate capture spiral.[7]

two different methods are used to add stickiness to the web. The first is the extrusion of cribellate silk onto the capture spiral, and the second is to secrete a solution containing aggregate protein onto the capture spiral. (Figure 1.1)

Cribellate silk consists of thousands of threads extruded from the cribellum[4], an oval like plate on the spinneret of the spider. The cribellate silk surrounds a core axial fiber of pseudoflagelliform silk. Although the core fiber will rupture at moderate extension, the surrounding cribellate silk continues to extend and maintain fiber integrity to extensions up to 500% of its original length.[5] The adhesive ability of the thread is proportional to the number of fibers extruded by the criblellum.[4] The primary adhesive force of these fibers are van der Waals and hygroscopic forces which increase with increasing humidity.[6] However, since the evolution of orb weaving spiders in the late Cretaceous, a general shift has been seen from spiders spinning cribellate capture threads to adhesive capture threads containing aggregate protein.[7] This shift is explained because, although a similar amount of volume is used in creating both types of fibers, aggregate fibers create a stickier web with greater prey capture potential.[4,6,8] Today only 5% of known orb weaving species use cribellate capture threads.[9]

In contrast to cribellate silk which creates weaker adhesion at many points of contact, aggregate adhesive fibers create larger adhesive forces at fewer points of contact. The two adhesives vary substantially in form and function and it is still debated whether the evolution of orb webs happened once, and that the two types of adhesive webs were not the result of convergent evolution,[7] or that orb webs evolved multiple times, with cribellar spiders being much more closely related to mostly webless spiders than adhesive capture thread spiders.[10]

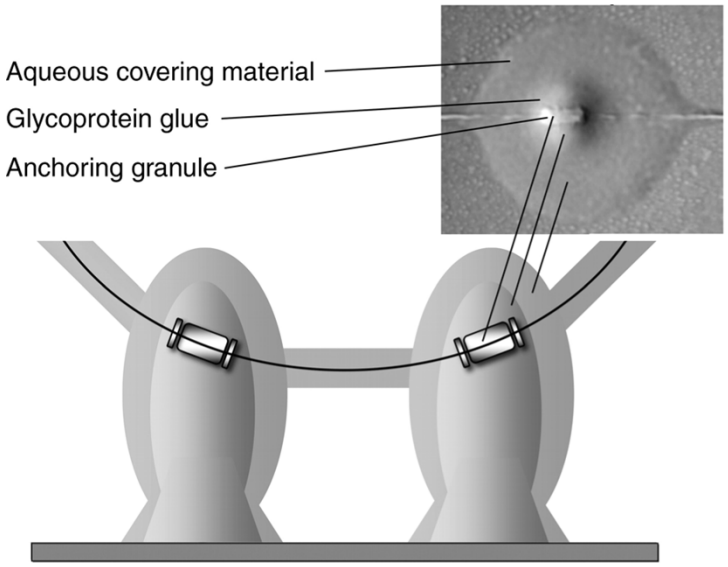

Figure 1.2 An illustration of the three components of aggregate glue.[12]

The capture spiral in adhesive webs is created by the extrusion of a central flagelliform fiber which is simultaneously covered in a viscid aqueous coating by the flanking aggregate gland spigots (Figure 1 E). Droplets may be large and relatively further spaced, or smaller and more closely spaced depending on the species of spider.[11] At the center of each droplet is an opaque granule. This granule was originally thought to harbor the glycoprotein glue responsible for the adhesive forces of the fibers. It was later shown that granule size had no correlation with adhesive force of the fibers, and it was hypothesized that there were, in fact, three layers to this glue.[12] The first granule layer serves as an anchor for the clear glycoprotein glue second layer and on top of this is the aqueous coating. The aqueous coating is reported to contain water soluble compounds related to neurotransmitters, salts, small peptides, and glycoproteins.[13,14] The salts are reported to assist in the solvation of the glycoprotein glue in these droplets and increase their adhesive properties.[15,16] All of these layers together help plasticize the flagelliform core fiber and allow it to extend without breaking when impacted by prey.[12] (Figure 1.2)

One of the most interesting aspects of aggregate glue droplets are their ability to vary the Young's Modulus depending on both the relative humidity and the speed with which a probe is retracted. When pulled with a higher retractive velocity, the stress vs strain curve shows increased stress and decreased strain. In addition, when tested under three different relative humidities (RH), the glue shows highest stress and strain at 40% RH and lower values at both 15% RH and 90% RH[17] (Figure 1.3). These properties make aggregate glue very useful for its intended purpose of keeping insects in the web and it also makes aggregate glue an attractive target for production of a synthetic aggregate glue for use as an underwater adhesive or surgical glue.

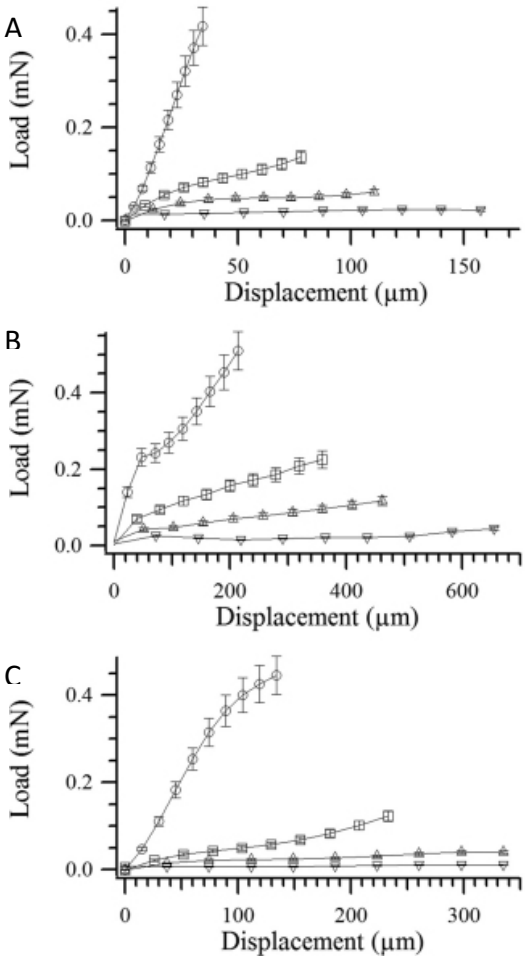

Figure 1.3 Stress vs Strain curves for aggregate glue under a)15% RH, b)40%RH, and c)90%RH. Pull speeds were 100um/s (circles), 50um/s (squares), 10um/s (triangles), 1um/s (inverted triangles).[17]

Much is not known about the glue. Most of the genetic and protein sequence is unknown. The secondary and tertiary structures of aggregate glue protein are unknown as well as how the protein interacts with itself and other materials that come in contact with it at the molecular level.

The work in this thesis does not address the structural data mentioned, rather it focuses on the genetic aspect of aggregate glue. As will be addressed, the genetic work done on the aggregate gene is lacking in both quality and quantity. In this thesis the aggregate glue gene will be partially elucidated, and using that knowledge, a synthetic aggregate construct will be created and expressed in *E. coli.*

References

(1) Gatesy, J.; Hayashi, C.; Motriuk, D.; Woods, J.; Lewis, R. *Science* **2001**, *291* (5513), 2603–2605.
(2) Clarke, T. H.; Garb, J. E.; Hayashi, C. Y.; Arensburger, P.; Ayoub, N. A. *Genome Biol. Evol.* **2015**, *7* (7), 1856–1870.
(3) Coddington, J. A. *J. Arachnol.* **1989**, *17* (1), 71–95.
(4) Opell, B. D. *J. Morphol.* **1994**, *221* (1), 111–119.
(5) Blackledge, T. A.; Hayashi, C. Y. *J. Exp. Biol.* **2006**, *209* (16), 3131–3140.
(6) Hawthorn, A. C.; Opell, B. D. *Biol. J. Linn. Soc.* **2002**, *77* (1), 1–8.
(7) Blackledge, T. A.; Scharff, N.; Coddington, J. A.; Szüts, T.; Wenzel, J. W.; Hayashi, C. Y.; Agnarsson, I. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (13), 5229–5234.
(8) Hawthorn, A. C.; Opell, B. D. *J. Exp. Biol.* **2003**, *206* (22), 3905–3911.
(9) Opell, B. D. *Biol. J. Linn. Soc.* **1997**, *62* (3), 443–458.
(10) Bond, J. E.; Garrison, N. L.; Hamilton, C. A.; Godwin, R. L.; Hedin, M.; Agnarsson, I. *Curr. Biol. CB* **2014**, *24* (15), 1765–1771.
(11) Opell, B. D.; Hendricks, M. L. *J. Exp. Biol.* **2007**, *210* (4), 553–560.
(12) Opell, B. D.; Hendricks, M. L. *J. Exp. Biol.* **2010**, *213* (2), 339–346.
(13) Vollrath, F.; Tillinghast, E. K. *Naturwissenschaften* **1991**, *78* (12), 557–559.
(14) Vollrath, F.; Fairbrother, W. J.; Williams, R. J. P.; Tillinghast, E. K.; Bernstein, D. T.; Gallagher, K. S.; Townley, M. A. *Nature* **1990**, *345* (6275), 526–528.
(15) Sahni, V.; Miyoshi, T.; Chen, K.; Jain, D.; Blamires, S. J.; Blackledge, T. A.; Dhinojwala, A. *Biomacromolecules* **2014**, *15* (4), 1225–1232.
(16) Amarpuri, G.; Chaurasia, V.; Jain, D.; Blackledge, T. A.; Dhinojwala, A. *Sci. Rep.* **2015**, *5*, 9030.
(17) Sahni, V.; Blackledge, T. A.; Dhinojwala, A. *Sci. Rep.* **2011**, *1*.

CHAPTER 2

AGGREGATE SILK WORK AT THE GENETIC LEVEL

Introduction

In relation to its counterparts in the web, the aggregate gland is very much understudied at the genetic level. In order to create a synthetic aggregate glue, genetic sequence data must be obtained to create a construct which can be transformed into an expression organism for production and purification.

The first work to obtain aggregate sequence data was published in 2009[1]. cDNAs were isolated from *N. clavipes* and sequence data obtained for two proteins named Asg1 and Asg2. Short protein sequence information was obtained by collecting spider web raw material and dissolving the glue in water. Since all other web components are insoluble in water it was presumed that any protein found was aggregate protein. The solution was deglycosylated and mass spectrometry was used to determine three common sequences in the protein: Gly-Ser-Ser-Val-Ser, Gly-Leu-Gly-Val, and Ala-Gly-Pro-Gly-Thr. From these sequences, a degenerative probe was created to screen a cDNA library created from the Aggregate gland mRNA. In this manner, two proteins were identified and named Asg1 and Asg2 which both contain N and O-linked glycosylation sites. The authors also reported that the cDNAs isolated contain full length proteins. Using Northern blot analysis, the authors claim that the mRNA transcripts for these two proteins are present only in aggregate gland mRNAs, and not in other silk gland mRNA supporting the hypothesis that these two proteins are components of the glycoprotein glue found on the capture spiral of orb webs. A repetitive region was identified and it was found that Asg1

and Asg2 shared this repetitive region but each protein was translated from a different frame on opposite strands of this region of DNA.

There are several problems with this study. All of the peptide sequences found in the study were from Asg 2. The similarity in DNA sequence to Asg1is only a tenuous link at best to claim that they are expressed on opposite strands of the same DNA sequence. This is especially true considering the number of duplication events that have occurred in the superfamily Aranaeoida only in silk sequences. The statement in the paper that other silk glands were tested by Northern blot is not substantiated by data in the paper or supplemental data. In the analysis of deglycosylated proteins, no control was run of the aggregate glue before deglycosylation, and the claim that other bands in the sample are the deglycosylation enzymes is not substantiated by data.

In addition to these discrepancies, in house work (Supplemental Data) screening an aggregate library showed that a segment of the putative repetitive region is not actually present in the DNA, which, if true, eliminates that section as the "repetitive region" and brings into question whether or not aggregate glue has a repetitive region. In addition, this work shows an open reading frame upstream of the published start codon of Asg2 indicating that the published start codon is not the true start codon, and the protein is likely larger than previously believed.

Additional work with aggregate was needed to determine what data from the 2009 was reliable and to elucidate additional aggregate protein sequence.

Isolating an Aggregate Clone

To isolate a clone containing genetic material of the aggregate gland, spiders of the species *N. clavipes* were anesthetized using CO2 and dissected to isolate the aggregate gland. It was important to remove any flagelliform gland present as the flagelliform gland is found encased by the aggregate gland. The aggregate glands were flash frozen in liquid nitrogen and stored for later RNA isolations.

All surfaces and tools used while working with RNA were treated with RNase Zap and rinsed with Diethylpyrocarbonate (DEPC) treated water. Aggregate glandular material was then digested using Trizol Reagent and total RNA was isolated using the manufacturer's protocol. mRNA was further isolated using the Fast Trac Magnetic bead Kit (Invitrogen). RNA integrity was assessed using a denaturing formaldehyde agarose gel. RNA was visualized using Sybr Green II stain. First strand cDNA synthesis was accomplished by using the reverse transcriptase MMLV. Second strand synthesis was accomplished using the NEBNext mRNA Second Strand Synthesis Module, followed by the NEBNext End Repair Module to produce blunt ended cDNA. cDNA was found to have an average molecular weight of 6Kb by Gel electrophoresis (Figure 2.1).

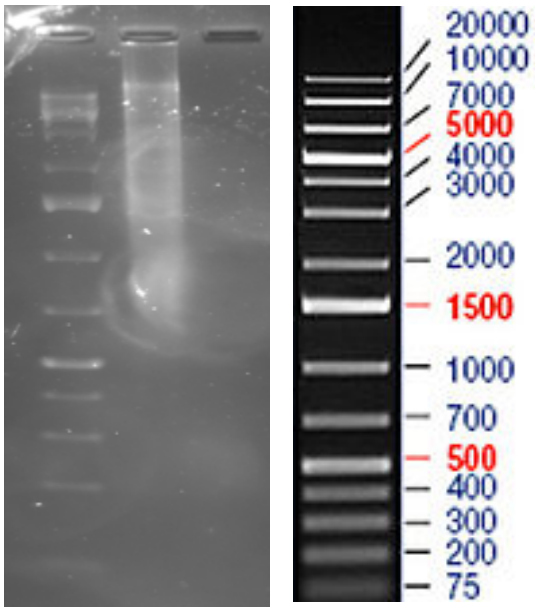

Figure 2.1 Insert cDNA used in library construction.

Vector DNA was prepared by transforming pBluescript II SK + into BL21 *E. coli* cells and plating on ampicillin

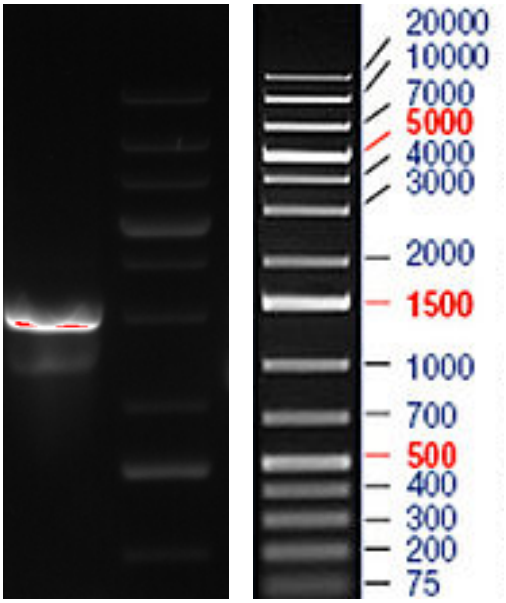

Figure 2.2 Digest of pBluescriptII sk+ with EcoRV.

containing LB agar plates (100ug/mL). An individual colony was selected for growth in ampicillin containing LB broth (100ug/mL). Vector DNA was isolated using an IBI mini plasmid kit and cut using the restriction enzyme EcoRV for blunt ended cuts. It was simultaneously treated with recombinant shrimp alkaline phosphatase (rSAP) to dephosphorylate the plasmid ends to prevent self-ligation. (Figure 2.2)

Vector DNA and cDNA were ligated using T4 DNA ligase(NEB) at a 2:1 molar ratio at 16°C for 8 hours followed by incubation at 4°C overnight. The reaction was heat inactivated at 65°C for 15 minutes and chilled on ice. A series of dilutions of ligated DNA were created and 1uL of each dilution was transformed into ElectroMAX DH10B *E. coli* (Invitrogen) at 1800V. It was found that a 1:250 dilution of the ligation reaction mixture gave approximately 1000 colony forming units (cfu) per plate when electroporated with ElectroMAX DH10B cells (Invitrogen) at 1800V.

Colonies were lifted onto nitrocellulose, and later, nylon membranes by laying 4 strips of plastic wrap on the bench that was then overlaid with Whatman 5mm paper. Each strip of Whatman paper was saturated with a different solution according to the protocol. The first was with a 5mM EDTA, 1%SDS solution, the second was a denaturation solution containing a 0.5M NaOH, 1.5M NaCl solution, the third was a neutralization solution containing 1 M Tris, 1.5 M NaCl, pH 7, and the last was a solution

containing 2x Saline Sodium Citrate (SSC). Membranes were placed on the plates, and then removed and inverted to place the bacteria side up on each Whatman paper saturated with solution for 5 minutes each. Membranes were allowed to dry for 30-45 minutes and DNA was crosslinked to the membrane by exposure to UV light using a VWR UV Crosslinker at 120,000 microjoules per cm^2 for two minutes. Membranes were then ready for hybridization and visualization.

Fluorescent Visualization Using AlexaFluor 647

Probes were designed using a combination of in house work and the published 2009 paper. An approximate 100 base pair region was selected for its high G-C content (>60%) in a region of the gene that was identical in all sources. This was important because the fluorescent dye chemically links to the G and C residues. The probe was then fluorescently labeled by incubation with Alexafluor 647 for 10 minutes at 90°C following the manufacturer's protocol. Probe labeling efficiency was calculated using a NanoDrop 2000 spectrophotometer and found to be acceptable, as per the recommended ratio of acceptable probe labeling of 30-40 bases per dye molecule.

Membranes were heat sealed in hybridization bags (KPL) and prehybridized for 4-6 hours at 37°C in a prehybridization solution (6xSSC, 0.05M sodium phosphate pH6.8, 1mM EDTA, 5x Denhardt's reagent, and 100ug/mL denatured fragmented salmon sperm DNA). Prehybridization solution was removed and replaced with hybridization solution (prehybridization solution + 100mg/mL dextran sulfate and labeled probe at 180pM) and incubated overnight at 37°C in a shaking water bath. Membranes were then rinsed three times in ice cold 6x SSC followed by two 30 minute incubations in ice cold

6xSSC. Filters were then rinsed twice in tetraethylammonium chloride (TEACl)wash solution (2.4M TEACl, 50mM Tris-Cl pH 8, 0.2mM EDTA pH 7.6, 1mg/mL SDS) at 37°C followed by two incubations in TEACl wash solution at 52°C. Membranes were then removed and secured to a strip of exposed x-ray film for support, and covered with plastic wrap to prevent complete drying of the membranes. They were placed in a cassette with fresh x-ray film in a dark room and allowed to develop. 15 minute, 1 hour, 3 hours, overnight, and over the weekend exposures were tested all with negative results.

In order to determine whether there were truly no positive plasmid colonies or if some other step in the process was responsible for the lack of results, a positive control plasmid was synthesized using Geneart (Thermofisher), which contained a 4x repeat of the probe sequence. The plasmid was transformed into ElectroMAX DH10B cells and the hybridization protocol was repeated. Exposures again showed no positive colonies. In a final attempt to determine what was not working, a dot blot using the labeled probe DNA itself was prepared at various concentrations and crosslinked to the membrane using UV light. After attempting the visualization protocol, these once again came up negative; therefore, an alternative visualization method was required.

Chromogenic Detection using a Biotinylated Probe and an Alkaline Phosphatase-Streptavidin Conjugate

The Biotin Chromogenic Detection Kit (Thermofisher) was selected as an alternative method of screening colonies. Probes were ordered from IDT (IDTdna.com) including a biotinylation modification on the 5'end of the probe. A new hybridization method was selected according to the manufacturer's supplemental protocol. Membranes

were heat sealed in hybridization bags containing a new prehybridization solution (6xSSC, 5x Denhardt's Reagent, 0.5% SDS, 50% (v/v) deionized formamide, 50ug/mL sheared denatured salmon sperm DNA) for 2-4 hours at 42°C in a shaking water bath (0.2mL/cm^2). Prehybridization solution was removed and replaced with a new hybridization solution (new prehybridization solution + probe at ~100ng/mL) and incubated overnight at 42°C (0.1mL/cm^2). Membranes were then washed twice at room temperature in 2x SSC 0.1%SDS for 10 minutes. Membranes were then washed twice in 0.1xSSC 0.1%SDS at 65°C for 20 minutes. Membranes were briefly placed on Whatman 5mm paper to remove excess liquid.

Detection was accomplished by first washing each membrane in 15mL blocking/wash buffer for 5 minutes on a shaker plate with moderate shaking. The membranes were then incubated in 15mL blocking solution for 30 minutes with moderate shaking, followed by incubation with 10mL diluted streptavidin-AP conjugate in blocking solution for 30 minutes. Membranes were then washed twice with 30 mL blocking/wash buffer for 15 minutes. They were then washed with 10 mL detection buffer. Membranes were then removed and placed in the hood for detection. Each membrane was developed in a separate petri dish with 5mL substrate solution, covered with aluminum foil, and allowed to develop overnight. Membranes were then rinsed with 18.2 MOhm water and evaluated.

The positive control when subjected to this hybridization and detection protocol showed all positive colonies. The procedure was then done on a series of library transformations. All colonies once again showed up positive. It was then necessary to prepare a negative control to visualize the difference between positive and negative

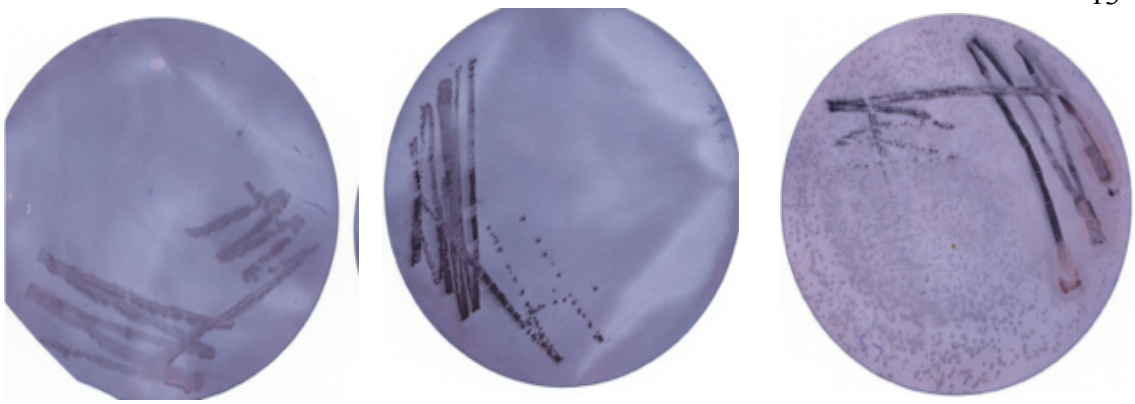

Figure 2.3 Chromogenic detection of empty vector colonies(A), positive colonies(B), and positive colonies overlaid on top of a library transformation(C).

colonies if there was one. Empty vector DH10B bacteria were created and the procedure was repeated on both positive and negative colonies. It was found that all bacterial colonies develop a light purple, whereas truly positive colonies developed a dark, nearly black color. (Figure 2.3) Screening commenced on the aggregate library once again. Due to a dilution error, plates were transformed at 10 times the expected cfu density. (Figure 2.4) As a result, colonies were selected, but needed further separation. Selected colonies were grown up overnight in 5 mL LB-Amp media. They were then streaked out on fresh ampicillin containing plates. These plates were subjected to colony lifts, followed by the chromogenic hybridization and detection protocols above and colonies were further screened. Several of the plates showed positive colonies (Figure 2.5), but most were in the thick band of the streak and no colonies were found that were completely isolated. This procedure was repeated to isolate a clone with aggregate genetic material. False positives were occasionally seen

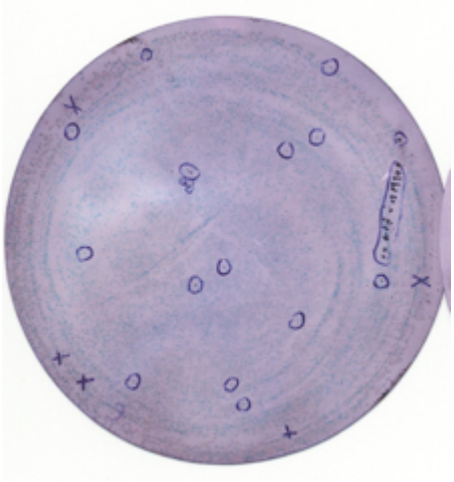

Figure 2.4 Example library post detection at 10x the normal cfu density.

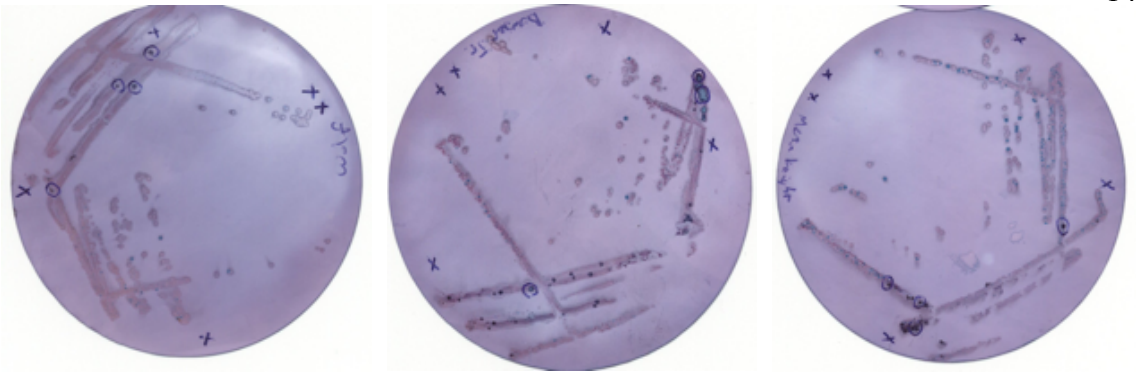

Figure 2.5 Plates resulting from primary colony picks that appear to contain positive colonies in the streaks.

when a colony lift pulled up a chunk of bacteria in the heavy portion of the streak. Although these were followed for subsequent overnight cultures and further screening, the area appearing like a positive result could be predicted before the screening as clumps of bacteria will show up positive in AP-streptavidin chromogenic detection whether or not they contain genetic material recognized by the probe.

Unfortunately, a period of time passed when the lab was unable to obtain a critical component of the screening process (Deionized Formamide) and in that time, many plates went bad and further screenings showed up negative. It is also possible, that on occasion, when attempting to streak out the bacteria to isolate a clone, the bacteria containing the construct of interest were selected against in the overnight culture.

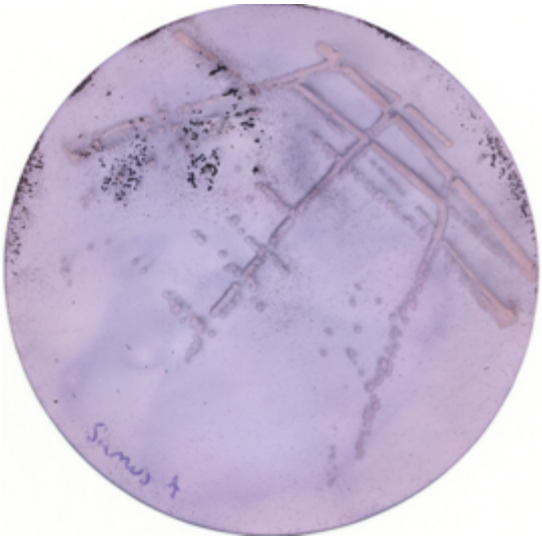

Figure 2.6 Example membrane with false positive results due to absence of shaking in the water bath.

Library screening commenced

again, positive colonies were seen and followed. Because what appeared to be positive colonies in areas of the plate were being visualized where no bacteria existed (Figure 2.6), a negative control consisting of a membrane that was never in contact with a bacterial plate was screened and showed background that appeared positive. During this time, the shaker component of the shaking water bath had malfunctioned and was not shaking the membranes properly in the bath. The bath was repaired and the background positive results disappeared. Screening commenced once again, but was discontinued shortly thereafter, as a different technique was showing more promise for isolating aggregate genetic material.

5' RACE

5'RACE seemed a promising alternative method to detect the unknown 5' end of the mRNA transcript. 5'RACE works by creating several primers based off the reverse compliment of the known 3' end, SP1 (manufacturer prescribed name, not an acronym) is the furthest downstream, followed by SP2, and SP3. Using a reverse transcriptase, reportedly up to 14,000 bases can be reverse transcribed into DNA from a primer SP1. The RNA is degraded and a recombinant terminal transferase enzyme then selectively adds a poly alanine (poly-A) tail onto the reverse complement 3' end of the single stranded DNA which is located at the 5'end of the coding strand. A polymerase chain reaction is used with an anchor primer that binds to the newly formed poly-A tail and the second design primer SP2. From the resulting reaction mixture, a nested PCR reaction

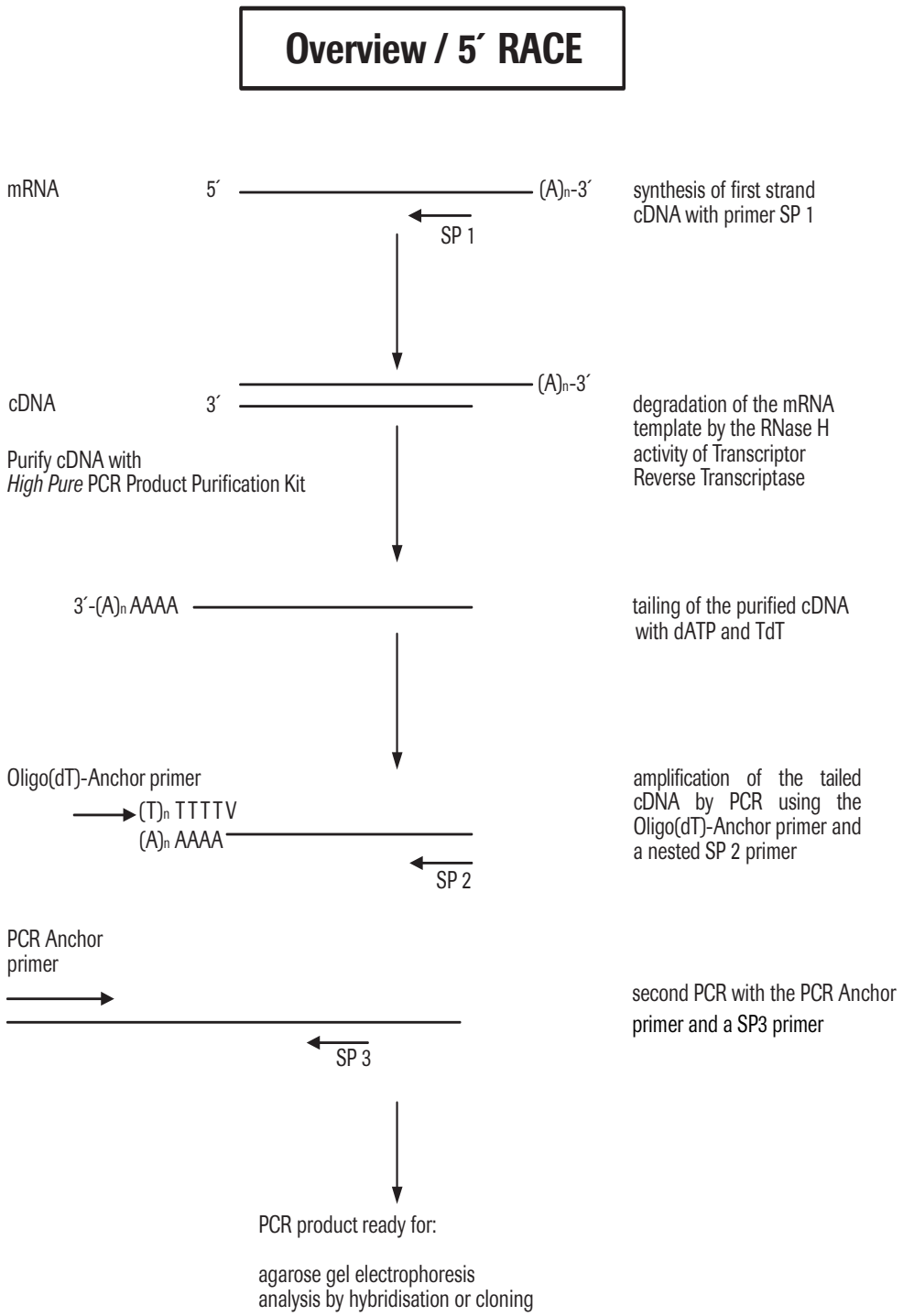

Figure 2.7 Overview of 5'RACE. Reverse transcription using primer SP1, addition of the 5' poly A tail, amplification using primers SP2 and the Oligo(dT)-Anchor primer, and the secondary amplification using the Oligo(dT)-Anchor primer and primer SP3. Figure from Roche 5'/3' RACE Kit instruction manual.

can be accomplished using the anchor primer and design primer SP3 to further amplify the DNA product. (Figure 2.7)

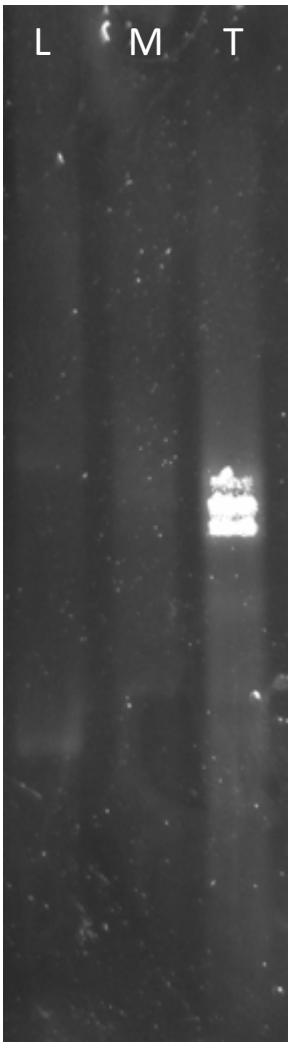

Figure 2.8 RNA used in RACE experiment.
L=ladder (degraded)
M=messenger RNA
T=Total RNA

The 5'/3'RACE Kit, 2nd Generation (Roche) was obtained, as well as designed primers SP1, SP2, and SP3 (Supplemental Data Race Design). Both total RNA and mRNA were isolated from *N. clavipes* aggregate glandular material as previously described. (Figure 2.8) The first round and second round PCR were done with both the High Fidelity Polymerase Mix(Roche) and the Long Template Polymerase Mix (Roche). Manufacturer's protocol was followed. Results of the PCR were run on a 1% agarose gel, and visualized with ethidium bromide. It was found that first round PCR gave smears of DNA in every reaction condition, although a brighter and larger smear was seen when mRNA template was used, rather than total RNA. The Long Template Polymerase Mix resulted in a higher molecular weight product than the High Fidelity Polymerase Mix in every case. This is exemplified in reactions 4 vs 3 of the 1st round of PCR, but is also seen in other lanes. Nested PCR resulted in even brighter bands with molecular weights ranging from 20+ KB to about .5 KB. (Figure 2.9)

Two reaction mixtures, 4L reaction (mRNA LT/LT) and 4H reaction (mRNA LT/HF), were selected for their high molecular weight products, and a portion of the reaction mixture was run on a LMP agarose gel. DNA greater than 3KB was cut out

using a clean razor blade and reisolated by melting the LMP agarose at 65°C then the DNA was isolated on a Qiagen PCR clean-up kit following manufacturer's protocol,

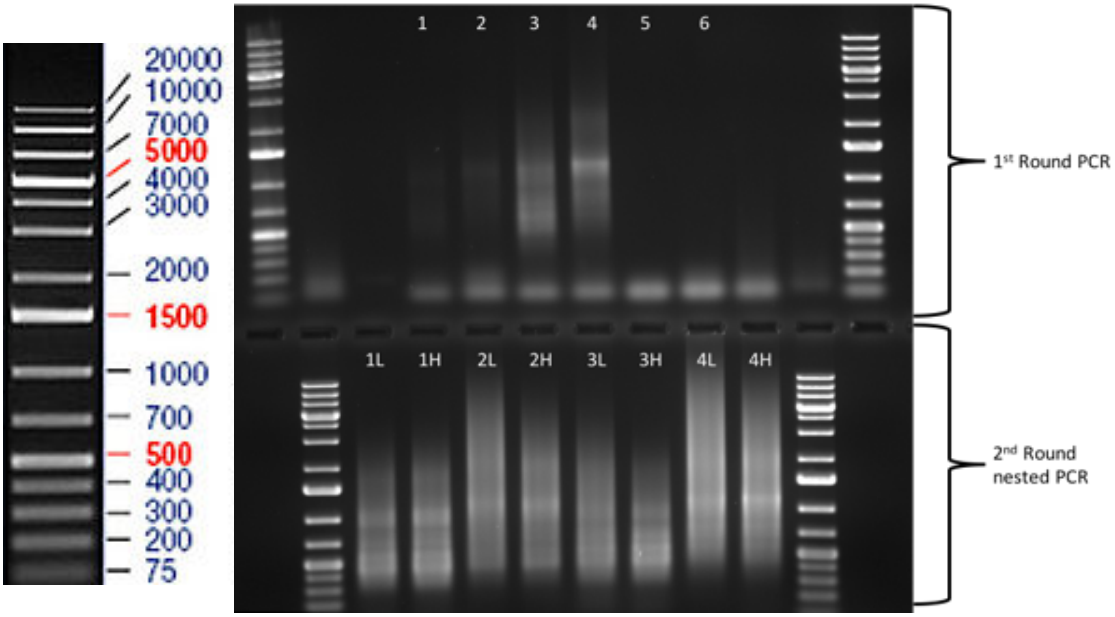

Figure 2.9 Results from 5'RACE experiment. In the first round of PCR, lane 1 used total RNA as a template and the High Fidelity enzyme mix. Lane 2 used total RNA as a template and the Long Template enzyme mix. Lane 3 used mRNA as a template and the High Fidelity Enzyme mix. Lane 4 used mRNA as a template and the Long Template enzyme mix. Lanes 5 and 6 were negative controls. In the second round of nested PCR, reactions corresponding to lanes 1-4 were each run again each with either the Long template enzyme mix (L) or the High Fidelity enzyme mix (H).

except the binding and wash steps were done by heating the solutions to 65° C before centrifuging the column to ensure that the agarose was melted and not able to clog the column. A portion was run on a 1% agarose gel to verify the DNA was still present before ligating into pBluescript II SK+ and transforming into ElectroMAX DH10B e-coli as previously described for cDNA library creation. (Figure 2.10) Blue white screening was used to select colonies containing insert DNA and approximately 100 colonies were

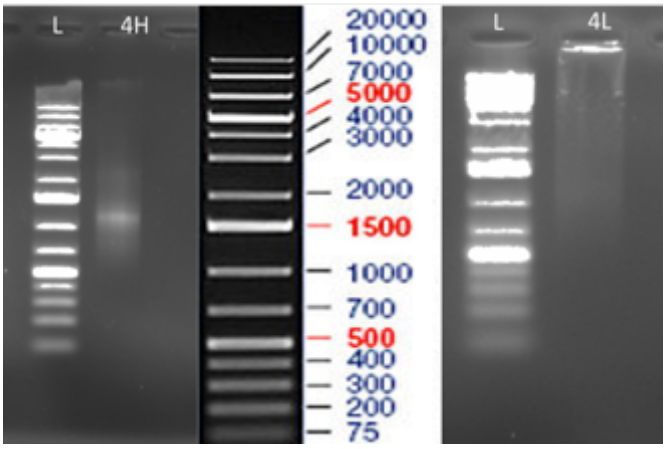

Figure 2.10 Gels of the 4H reaction and 4L reaction after selecting for fragments larger than 3kb.

selected over a series of two weeks. The colonies were grown in an overnight 5 mL LB culture will ampicillin at 100ug/mL, and then DNA was isolated via centrifugal miniprep. Plasmid DNA was cut with EcoRI to get a single cut. Unfortunately, it was found that many of the plasmids contained small fragments of DNA, and only a couple had insert sizes of greater than 1KB.(Figure 2.11) No colonies were selected to send in for sequencing, since 3kB+ was required to find any useful data. During this time a manuscript was sent to our lab from the Hayashi lab at UC Riverside. The paper was published in 2016 and addressed many of the discrepancies of the paper published in 2009.[2] It was discovered that the 353 BP region purportedly shared between Asg1 and Asg2 was not found in the *N. clavipes* genomic contigs of Asg2 and since this was a critical part of the previously reported repetitive region of Asg2, it was reported that it would be impossible for that repetitive region to actually exist. This data corresponds precisely with the in house data performed in the Lewis Laboratory prior to undertaking this work (Supplemental Data). In addition, the researchers found additional sequence data that included the known C-terminus of the protein, a linker region, and about two and a half repeats of what appeared to be a large repetitive region, each repeat consisting of about 100 amino acids. With evolutionary

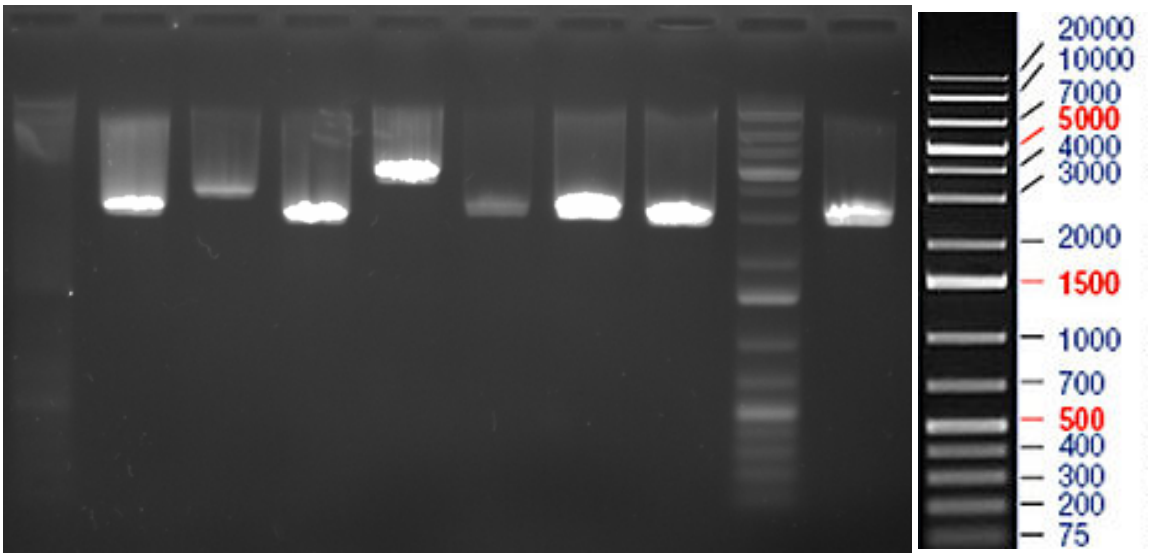

Figure 2.11 Minipreps from multiple colonies transformed with 4H and 4L selected for greater than 3kb fragment sizes ligated into pBluescript II (sk+). Vector size is 3kb, therefore the largest inserts approach 2kb for a total of 5kb.

data comparing the C-termius to other members of the silk protein family, it was proposed to rename Asg2 to Agsp1 (aggregate spidroin protein 1).

The authors also found that, with respect to Asg1, it was found that Asg1 is not specific to aggregate glands, in fact, homologs to the Asg1 protein are found in the *Stegodyphus mimosarum* genome, a spider that does not produce aggregate glue or contain aggregate glands.[2] It was proposed that Asg1 is actually part of a conserved chitin binding domain type 2. It was proposed that this protein is expressed in silk glands to bind chitin in the silk gland duct and walls, which explains the presence of Asg1 mRNA in the aggregate gland cells. It is important to note that the Asg2 genetic material elucidated by Collin et al. is not a full aggregate sequence. (Figure 2.12) The N-terminus of the aggregate gene has yet to be discovered. It is proposed that 5'RACE is a likely way in which this genetic sequence data could be revealed. If time allows, N-terminal

data will be sought in this thesis, however pursuit of this data ceased to allow for the creation of a synthetic construct utilizing the data from Collin et al.

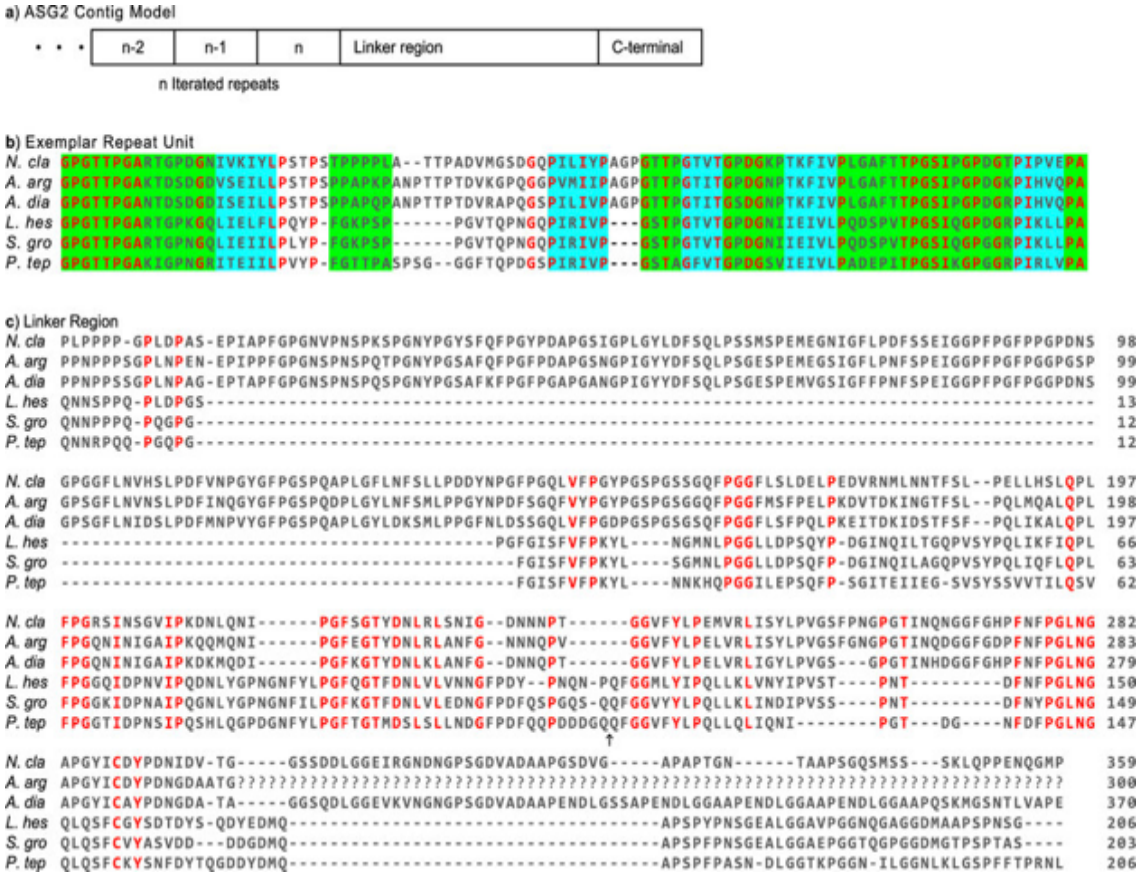

Figure 2.12 Data from Collin et al[2]. (**a**) Diagram of the inferred organization of ASG2. The repetitive region contains an unknown (n) number of repeats followed by the linker and carboxyl (C)-terminal regions. Ellipsis signifies unknown sequence. (**b**) Alignment of exemplar ASG2 repeat units indicating sequences predicted to form β-sheet (blue shading) or β-turn/random coil structure (green shading). (**c**) Alignment of linker regions. Amino acids depicted by single letter IUPAC abbreviations. Red indicates amino acids conserved across species and black arrow indicates location where additional sequence from *P. tep* whole genome assembly contig 63868 was added to the *P. tep* capture contig from Collin et al.

In summary, two methods have been identified able to elucidate additional genetic data on the aggregate gene. DNA sequences of greater than 20 kilobases have been

isolated but not yet cloned. This is much greater than the 3 kilobases identified by Collin et al.

While obtaining large DNA sequences is viewed as progress towards obtaining the genetic sequence data for spider aggregate glue, this also served as a hinderance with large fragments of DNA being more difficult to clone and sequence. Given sufficient time, cloning large portions of the aggregate glue gene should be possible using the techniques described in this chapter. In this thesis the data was not pursued to allow for creation of the synthetic glue.

References

(1) Choresh, O.; Bayarmagnai, B.; Lewis, R. V. *Biomacromolecules* **2009**, *10* (10), 2852ecules
(2) Collin, M. A.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y. *Sci. Rep.* **2016**, *6*, 21589.

CHAPTER 3

CREATION OF A SYNTHETIC AGGREGATE CONSTRUCT

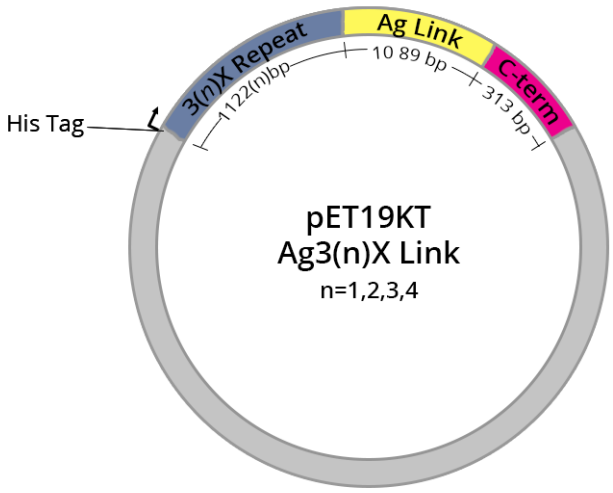

Figure 3.1 A diagram of a finished aggregate gene construct including the repetitive region (blue), linker region(yellow), and c-terminal region(pink). The remainder of the plasmid is represented by the grey region.

Utilizing the data prepared by Collin et al[1] a synthetic analog of aggregate glue gene was created. (Figure 3.1) Assuming the upstream data follows the pattern established by many silk proteins, aggregate should have a C-terminal domain, a repetitive region, a linker region, and an N-terminal domain. Each part would need to be synthesized individually and then sub-cloned together.

Design

The C-terminal domain was determined to be similar enough in structure to use the common C-terminal domain found in the other silk constructs created in the lab.[1,2] Therefore, the C-terminal domain was designed for the purpose of adding it to the constructs if needed, but not ordered. The C-terminal domain used was already cloned into the plasmid pET19K (containing a kanamycin resistance gene) and named pET19KT. The plasmid is also His tagged at the N- terminus of the open reading frame. (Figure 3.2)

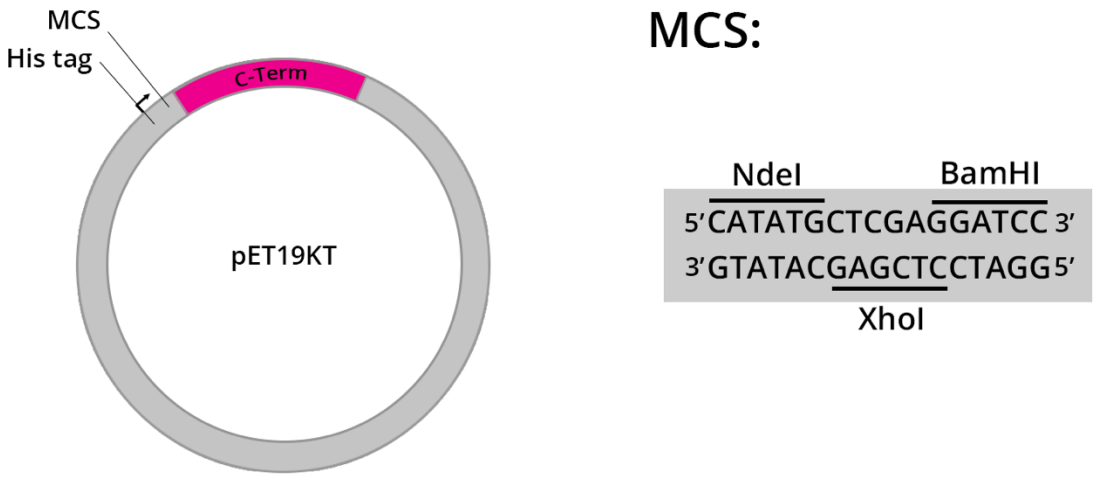

Figure 3.2 Diagram of the expression vector pET19KT along with the multiple cloning site(MCS) used in this experiment. The C terminal region is shaded pink, and the remainder of the vector is shaded grey.

The linker region was designed and ordered from GeneArt with the restriction site BamHI on the outer edges of the sequence for cloning into the pET19KT vector. The insert had a total sequence length of 1089 base pairs.

The repetitive region was designed and ordered from GeneArt in iterations of 3 repeats as follows: NdeI, AgeI, Repeat 1, Repeat 2, Repeat 3, BspEI, and finally XhoI for a total of 1122 base pairs. The outer flanking restriction sites would be used for cloning into the pET19KT vector. The inner two restriction enzymes would be used to create a non-regenerable restriction site for splicing together repetitive regions of the protein DNA sequence.[3]

Methods

pET19KT was transformed by electroporation at 1800V into BL21 cells and plated on kanamycin containing LB plates at 50ug/mL. Islolated colonies were then selected and grown in a 5 mL LB overnight culture with kanamycin at 50ug/mL. Plasmids were then isolated from bacteria, and cut with BamHI and verified to be the correct size of 6087 base pairs (Figure 3.3) Both digested and undigested vector were then stored at -20°C until further cloning steps.

The ordered aggregate linker was delivered in the vector pMA-T with ampicillin resistance. It was transformed at 1800 V into electrocompetent BL21 cells and plated on ampicillin containing LB plates at 100ug/mL. Isolated colonies were then selected and grown in an overnight 5mL LB culture with ampicillin at 100ug/mL. Plasmids were then isolated by centrifugal miniprep and vector was cut with BamHI. pET19KT was also cut with BamHI and dephosphorylated using rSAP to prevent self-ligation. Both were run on

a 0.7% LMP agarose gel. The gel was cut down the edge of the lane containing the linker and visualized using ethidium bromide next to a ruler. The section that was not stained with ethidium was cut at the appropriate area to remove the band containing the linker and the DNA was isolated using a Wizard SC Gel/PCR purification kit. The DNA was then ligated with T4 DNA ligase into the the pET19KT vector. Transformation into BL21 cells resulted in very few colonies, so the transformation was reattempted with ElectroMAX DH10B cells. A 1:500 dilution of ligation reaction product was used to create sufficiently isolated bacteria to select for growth in an overnight culture. Plasmids

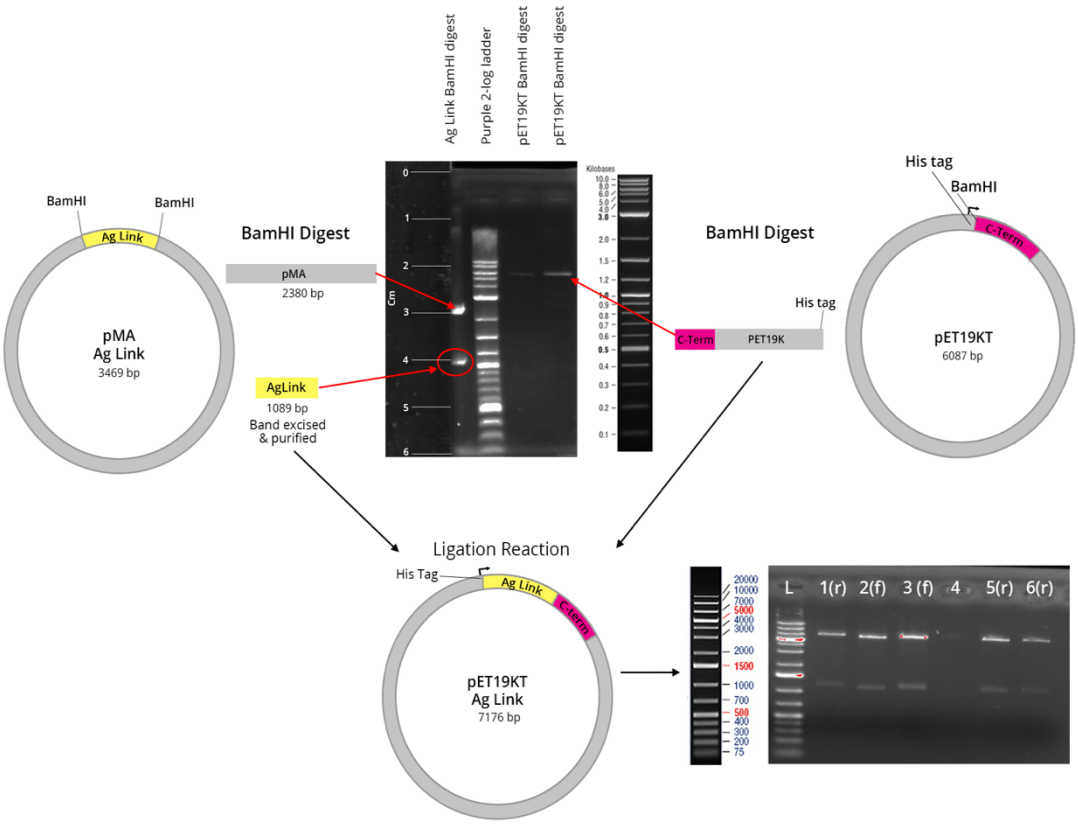

Figure 3.3 Ligation of the linker region(yellow) into the expression vector pET19KT. Reactions 1, 2, 3, 5 and 6 were sequenced. Plasmids with the insert in the forward orientation are denoted with (f). Plasmids with the insert in the reverse orientation are denoted with (r).

were mini prepped and then cut with BamHI to verify the presence of the insert. Of the 6 colonies selected, 5 showed the proper banding pattern and were sent in for sequencing. Of those five colonies, two contained the insert in the proper orientation. These were labeled pET19KTL and used for subsequent cloning steps.

To test the influence of the repetitive region on the protein on adhesive properties, constructs were prepared containing 3x, 6x, 9x and 12x repeats. The ordered 3x repetitive region was delivered in the pMK-RQ-Bb vector which contains a gene that confers kanamycin resistance. (Figure 3.4) The kanamycin resistance gene also contains the restriction site NcoI which was used in subsequent steps. XL1 Blue cells were transformed with this vector. Isolated colonies were selected for an overnight culture in LB broth, and the plasmid DNA was then isolated using the IBR High Speed Mini Kit.

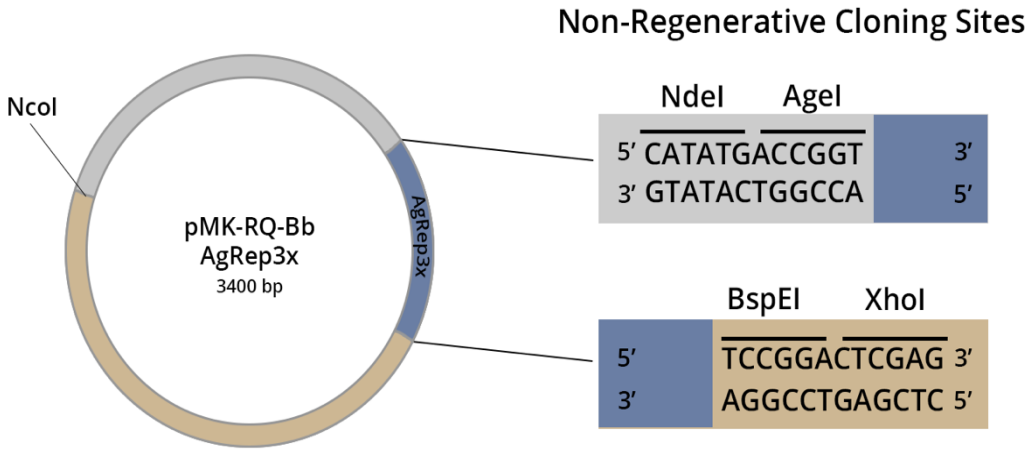

Figure 3.4 Design of the repetitive region vector for sub-cloning. 3x repetitive region is denoted in blue. The grey and tan regions are shaded for convenience in differentiating locations of the plasmid with respect to the NcoI restriction site.

The resulting plasmid was first digested using the XhoI restriction enzyme to determine that the plasmid was the correct length. An aliquot of isolated plasmid was digested with NcoI and AgeI, and a second aliquot was digested with NcoI and BspEI. The digests were then separated by size on 0.7% low melting point agarose using a voltage of 50 V and the desired bands were cut out of the gel with a razor blade and isolated as reported above for the linker. It is important to note that an unexpected cut was observed when digesting with AgeI, but it was determined that the additional cut appeared in the parent plasmid, rather than the desired fragment. AgeI and BspEI both leave identical sticky ends that will ligate together. Therefore, when ligated with T4 Ligase the resulting plasmid contained two repeats of the sticky ends and the internal AgeI and BspEI sites were destroyed, leaving only the flanking AgeI and BspEI sites. The NcoI sites were ligated together as well, reproducing the parent plasmid pMK-RQ-Bb with a 6x repeat of the aggregate repetitive region. (Figure 3.5) The procedure utilizing the non-regenerative restriction sites was repeated with a 6x and a 3x plasmid to produce a 9x repetitive region (Figure 3.6), as well as a 6x and a 6x plasmid to produce a 12x repetitive region (Figure 3.7), to produce the four desired constructs containing 3, 6, 9, and 12 repeats of the aggregate repetitive region. Each was then digested with NdeI and XhoI to remove the entire repetitive region for cloning into pET19KTL.(Figure 3.8) It is important to note that the 6x repetitive region was very close in size to the parent plasmid, the plasmid was cut an additional time using NcoI to ensure sufficient separation for subsequent LMP agarose extraction and purification as previously described.

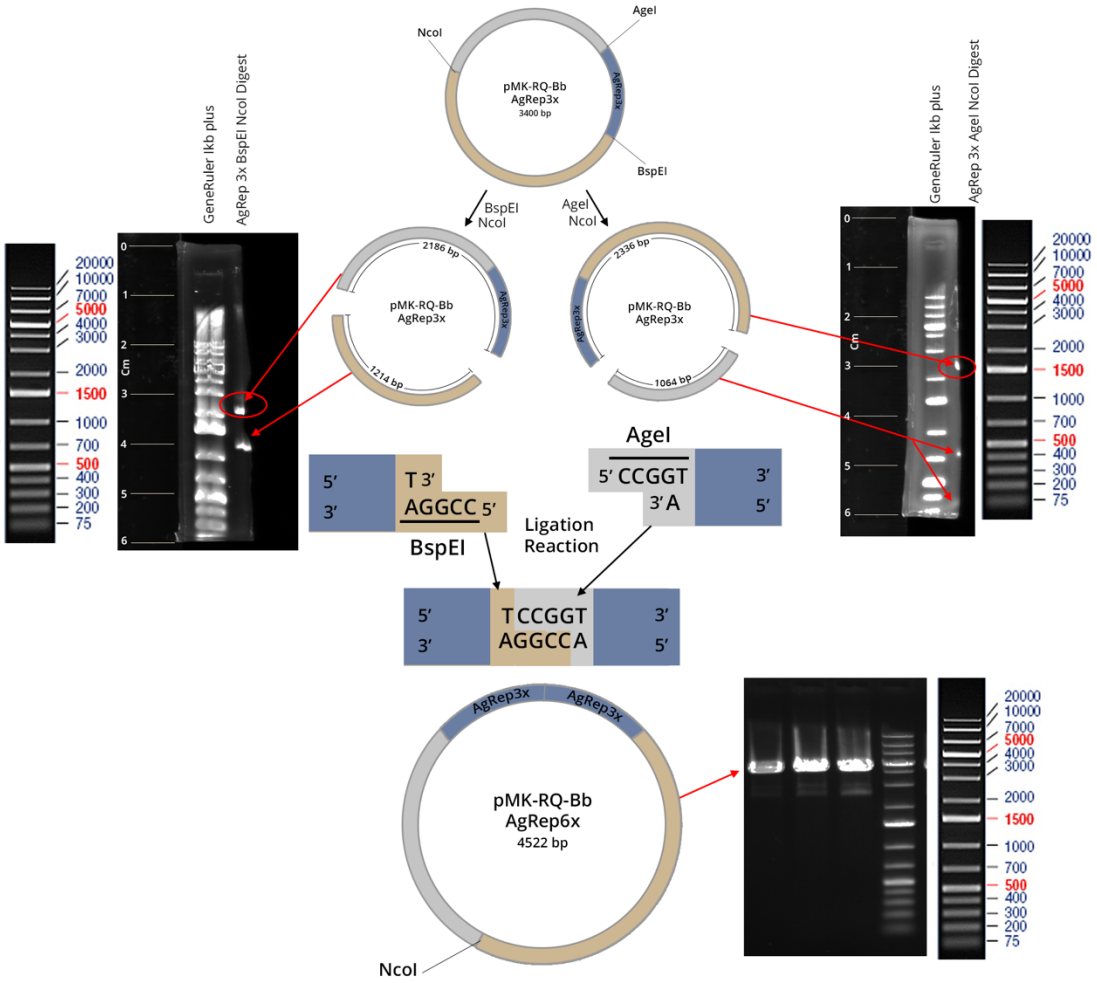

Figure 3.5 Formation of the 6x repetitive region. Repetitive region is shaded in blue. The figure shows the destruction of the AgeI and BspEI sites during ligation by creating a non-palindrome at the site. Sizes of all plasmids were checked by gel electrophoresis and indicated with red arrows. Segments excised and purified are circled in red.

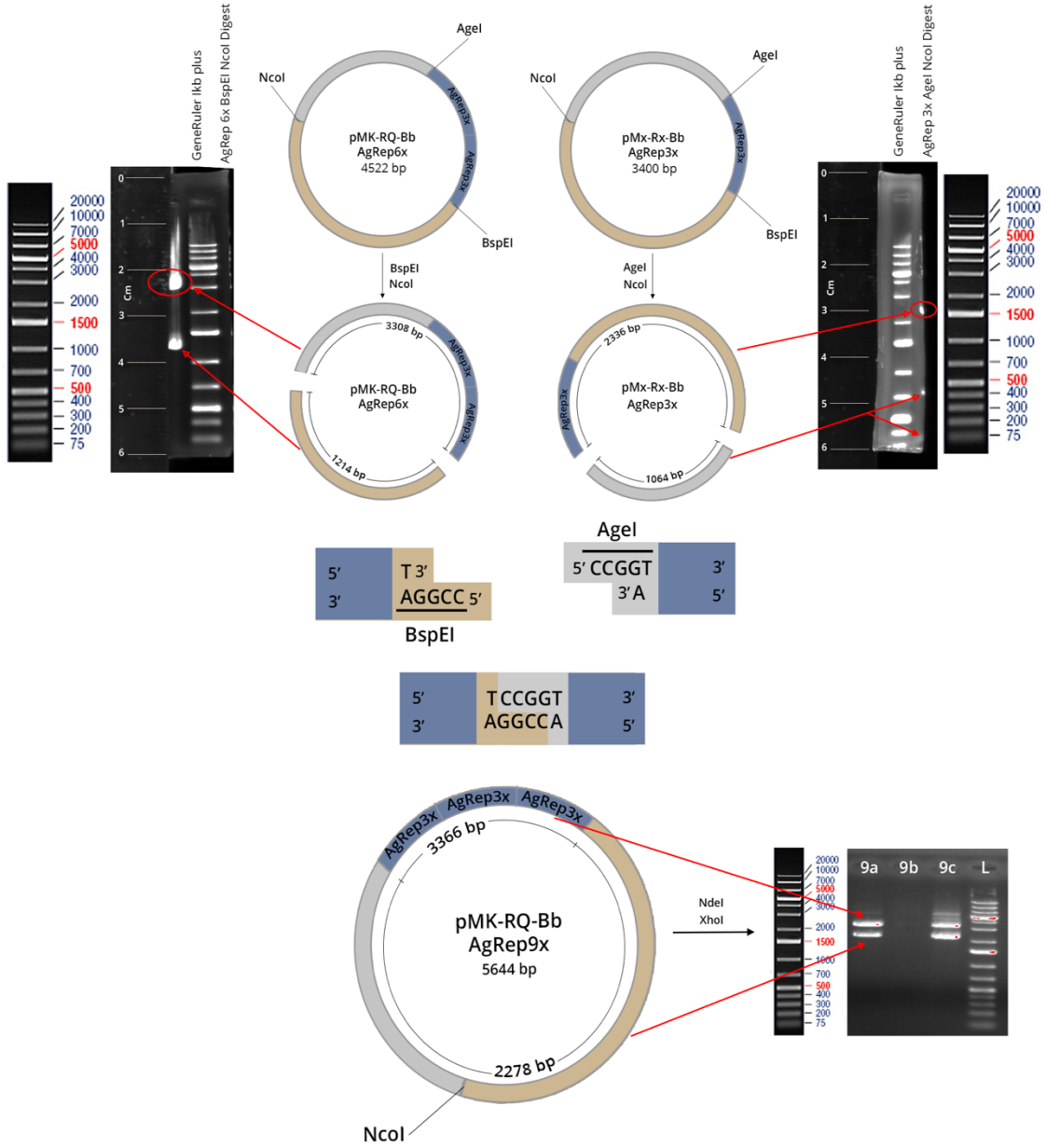

Figure 3.6: Formation of the 9x repetitive region. Repetitive region is shaded in blue. The figure shows the destruction of the AgeI and BspEI sites during ligation by creating a non-palindrome at the site. Sizes of all plasmids were checked by gel electrophoresis and indicated with red arrows. Segments excised and purified are circled in red.

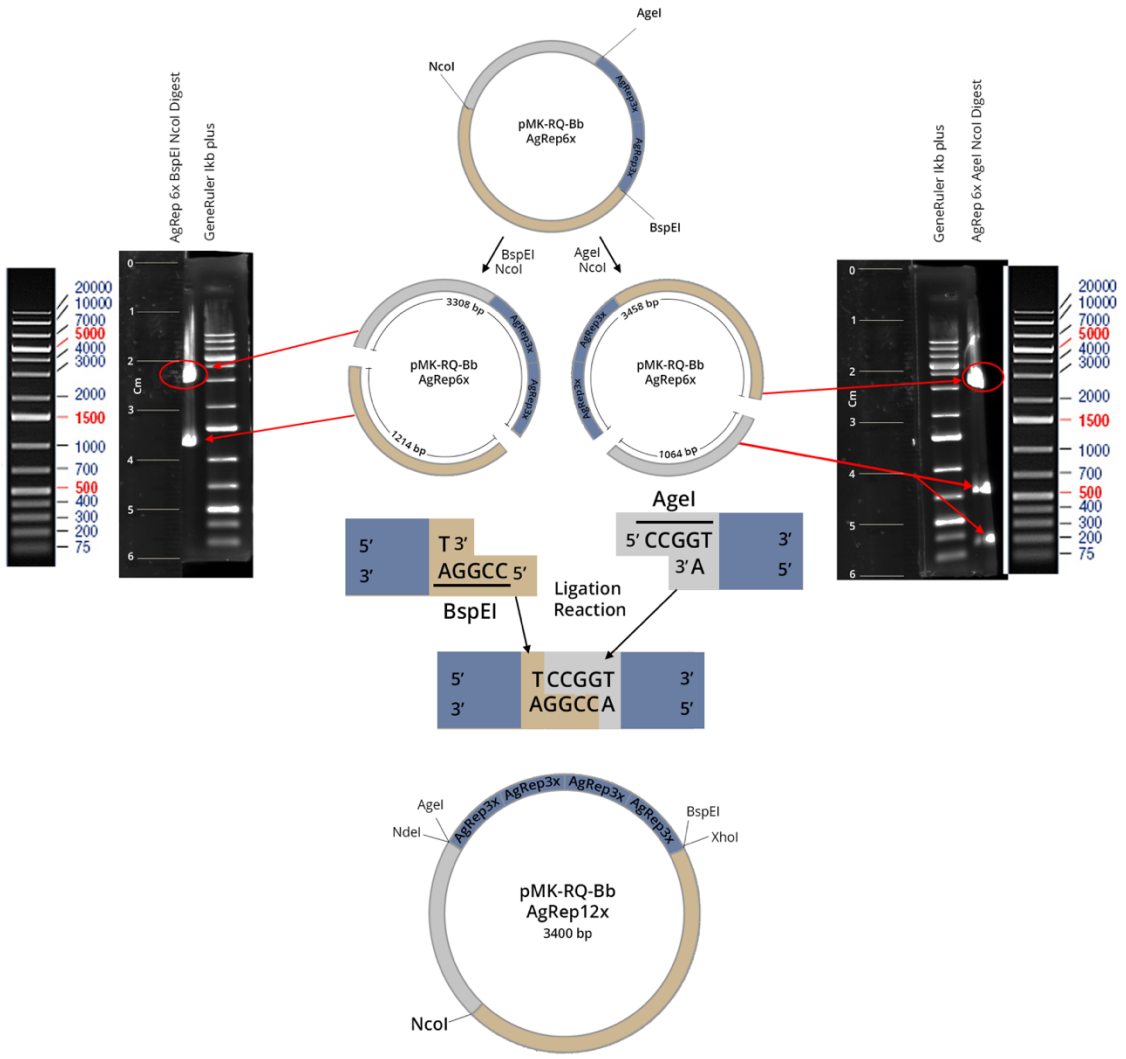

Figure 3.7 Formation of the 12x repetitive region. Repetitive region is shaded in blue. The figure shows the destruction of the AgeI and BspEI sites during ligation by creating a non-palindrome at the site. Sizes of all plasmids were checked by gel electrophoresis and indicated with red arrows. Segments excised and purified are circled in red.

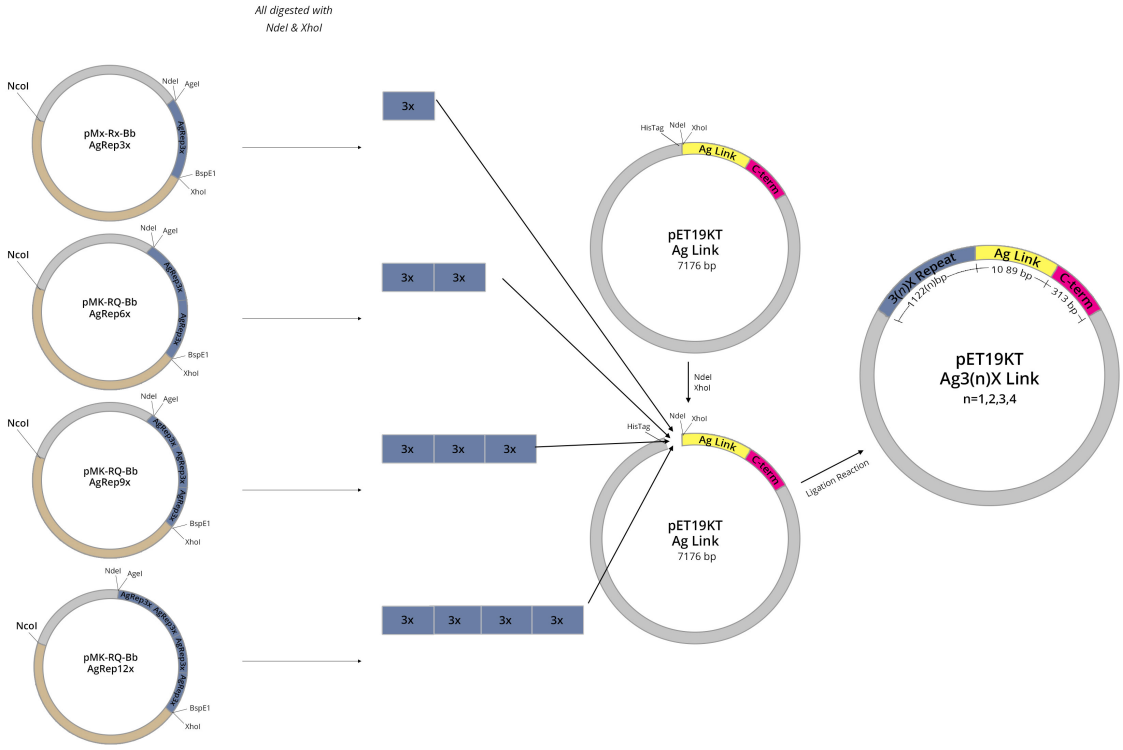

Figure 3.8 Insertion of the repetitive regions into the pET19KT Ag Link vector. Digestion with NdeI and XhoI removed the repetitive region from the pMK-RQ-Bb plasmid after which it was ligated into the pET19KT Ag Link vector cut with the same restriction enzymes.

The vector pET19KTL was prepared by digesting first with NdeI for 1 hour and then adding XhoI and allowing it to digest for an additional two hours. This was done because NdeI requires a larger amount of non-restriction site DNA to function properly and the two sites were found directly adjacent to one another. It was then ligated to each of the repetitive regions in four separate ligation reactions. (Figure 3.8) These were transformed into ElectroMAX DH10B cells and grown on plates with kanamycin at 50ug/mL. A 1:100 ligation reaction dilution usually resulted in sufficiently isolated

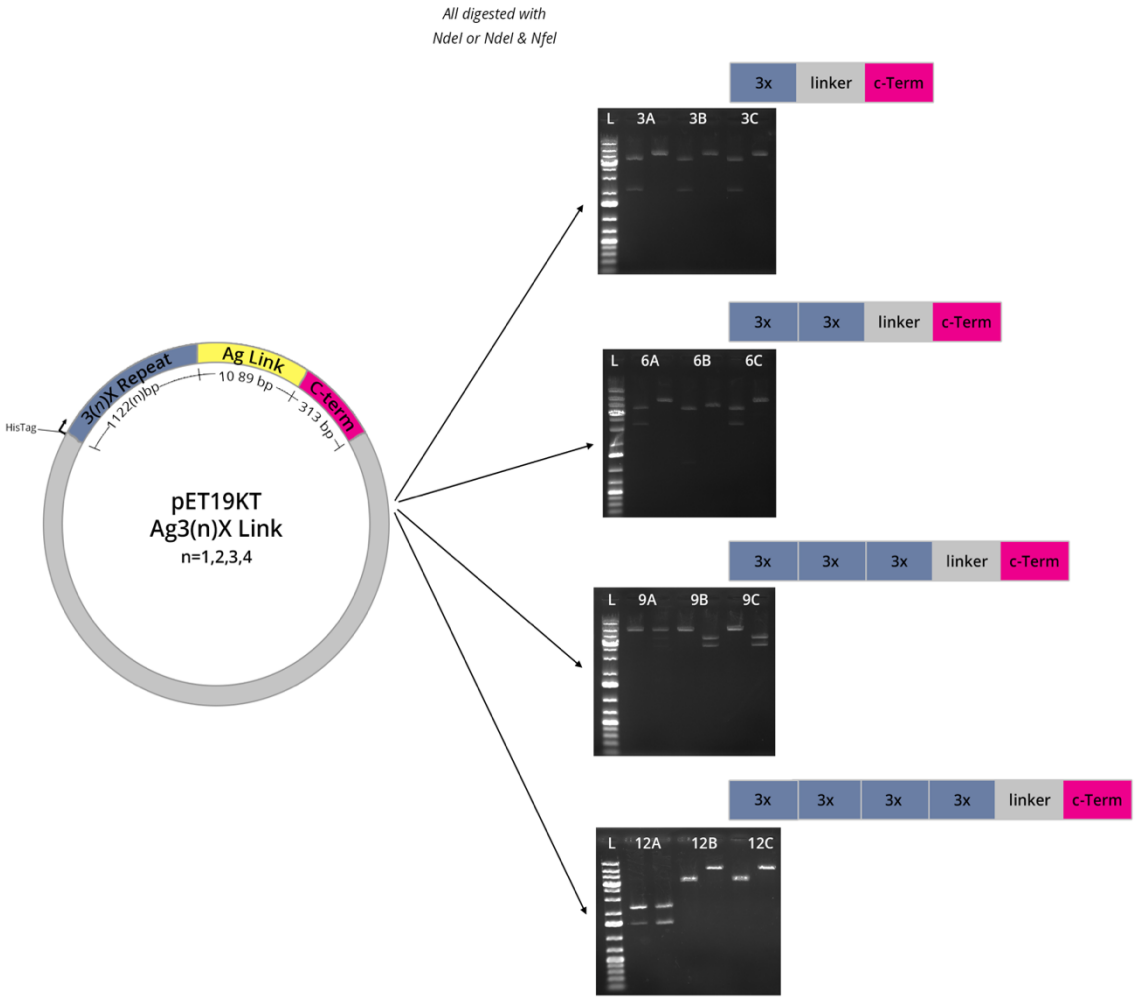

Figure 3.9 Verification of fragment sizes of 3x, 6x, 9x, and 12x repetitive regions inserted into the vector pET19KT Ag Link. Two digests were completed for each reaction. The left reaction was a single cut using NdeI. The right reaction was a double cut on both ends of the construct using NdeI and NfeI.

colonies for subsequent overnight cultures in kanamycin LB broth again at 50ug/mL. The vectors were isolated by centrifugal miniprep and two digestions were done for each selected colony. The first was a single cut using NdeI to verify size. The second was cut with both NdeI and NfeI. The NfeI site is found a few base pairs further downstream

from the BamHI site where the linker was inserted. This removed the insert in its entirety and allowed for the entire construct present in the plasmid to be visualized. Results indicated that construction and ligation of 3x, 6x, and 9x repeats were successful on the first attempt. The 12x repeat reaction was initially unsuccessful but was reattempted and successfully cloned. (Figure 3.9)

In summary, the information from Collin et al. was successfully taken and used to create four synthetic aggregate constructs with repetitive regions ranging from 3-12 repeats of the repetitive region. These constructs can now be used for cloning into bacterial or other expression systems for production of the synthetic glue.

References

(1) Collin, M. A.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y. *Sci. Rep.* **2016**, *6*, 21589.
(2) Ittah, S.; Michaeli, A.; Goldblum, A.; Gat, U. *Biomacromolecules* **2007**, *8* (9), 2768–2773.
(3) Lewis, R. V.; Hinman, M.; Kothakota, S.; Fournier, M. J. *Protein Expr. Purif.* **1996**, *7* (4), 400–406.

CHAPTER 4

PRODUCING A SYNTHETIC AGGREGATE PROTEIN (AGS1)

Few attempts have been made to produce a synthetic aggregate gland protein. This chapter is intended more as a starting point for expression of an aggregate gland protein construct, rather than a comprehensive expression study.

An expression study was done in a 200 mL LB broth flask study. 3x, 6x, and 9x constructs were transformed into *E. coli* BL21 bacteria with 4GPP, a helper expression vector containing the genes for the tRNAs of glycine and proline. Since the protein has a relatively high proportion of proline and glycine it was predicted to support expression. A starter 5mL culture was done overnight and used to seed the flask study. Kanamycin and chloramphenicol were used as the 4GPP vector encodes chloramphenicol resistance and the construct plasmid encodes kanamycin resistance. Flasks were induced with IPTG at an OD of approximately 0.51 and samples were taken every hour for 4 hours. Western Blot analysis with a mouse anti-His tag antibody and a rabbit anti-mouse IgG antibody

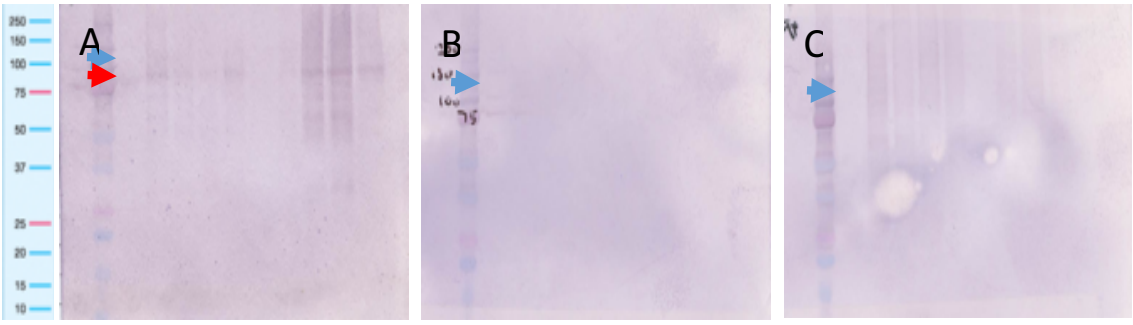

Figure 4.1 Western blots from LB flask study for 3x (A), 6x (B), and 9x (C) repeats. The expected molecular weight of 130 kDa is marked with a blue arrow. The observed band at approximately 80 kDa is marked with a red arrow. Lane 2=Ladder. Lanes 3-7 are the soluble fractions at time t=0, 1, 2, 3, and 4 hours post induction respectively. Lanes 8-12 are the insoluble fractions at time t=0, 1, 2, 3, and 4 hours post induction respectively.

revealed that the 3x construct appeared to express at a molecular weight of approximately 80kDa. (Figure 4.1) The expected weight of the 3x construct protein is 131kDa. The bands did appear to smear both up and down, indicating that a larger protein may be expressed, but the ribosome is terminating translation prior to the stop codon. No expression was seen in the 6x construct. A high molecular weight smear was seen in the 9x Construct.

A second flask study using 500mL Terrific Broth was attempted. In this experiment, only the 3x and 6x constructs were used, but both were done with and without the helper vector 4GPP. Visualization of the western blot used the same method as described above. However, no results were visualized on the positive control nor in the samples of the western blots. It was determined to be ineffective and other routes of production were pursued.

Table 4.1 2L fermentation run data comparing the glucose levels and optical densities (OD) of runs 3xa and 3xc.

Time (hr.)	OD 3xa	OD 3xc	Glucose 3xa (g/L)	Glucose 3xc (g/L)
0	0.186	0.233	18.3	17.4
13.5	21.4	39.4	48.4	30.7
13.75	22.2	38.4	46.9	26.7
14.75	35	68	32.9	5
15.25	44.6	77.4	25.6	<2.3
15.75	56	88.2	17.1	<2.3
16.25	68.8	103.6	8.8	9.8
16.75	81.6	104	8.8	19.7
17.25	98	100.8	4.3	23.8
17.75	98.4	104.4	5.6	18.9
18.25	96.4	101.6	13.8	8.4
18.75	86.4	97.2	15.5	9.9
19.25	81.6	93.6	33.7	13.3
20.25	82	92.4	33.4	17.6
21.25	81.6	NA	22.5	NA

A 2L fermentation run was also attempted. To test if the lower than expected molecular weight of the 3x construct was the result of a premature stop codon introduced by mutation, a different clone containing the 3x construct DNA was also tested. The original was labeled 3xa and the second 3xc. Unfortunately, glucose levels rose to high levels in the fermentation run 3xc and to unacceptable levels in 3xa. (Table 4.1) They also dropped to very low levels in 3xc. Abnormal sugar levels can cause expression anomalies by stressing the bacteria[1]. In the insoluble fraction of 3xa, (Figure 4.2 B) a faint band is seen at approximately 30 kDa, two heavier bands at 50 and 55 kDa were

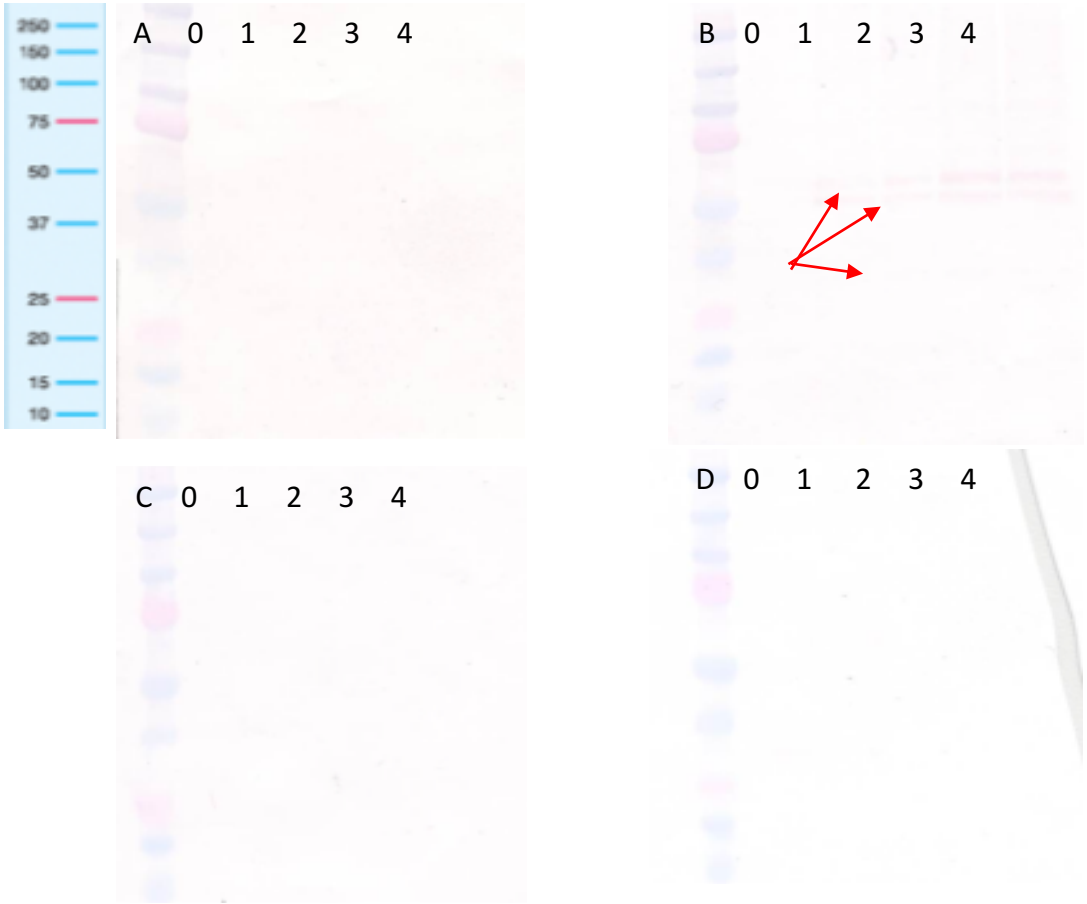

Figure 4.2 Western blots of 3xa soluble fraction using the anti-His tag primary antibody (A), 3xa insoluble fraction (B), 3xc soluble fraction (C), and 3xc insoluble fraction (D) at times t=0, 1, 2, 3, and 4 hours post induction. Arrows indicate the bands at 30, 50, and 55 kDa.

seen and the band appears to smear upward above those bands. No bands were visualized in the soluble fraction, or in the 3xc soluble and insoluble western blots. (Figure 4.2 A, C, and D) The fermentation runs were repeated and care was taken to avoid any large fluctuations in glucose levels. (Table 4.2) Western blots were accomplished by suspending the bacteria from 4 hours post induction in Lysis Buffer (20mM Tris, 0.5M NaCl, 0.5% sarkosyl). Western blot analysis using an anti-His tag primary antibody

in the soluble fraction, or in the 3xc soluble and insoluble western blots. (Figure 4.2 A, C, and D) The fermentation runs were repeated and care was taken to avoid any large fluctuations in glucose levels. (Table 4.2) Western blots were accomplished by

Table 4.2 Fermentation run data comparing the glucose levels and optical densities (OD) of the second set of runs for 3xa and 3xc. Induction time noted with an asterisk (*).

Time(hr.)	OD 3xa	OD 3xc	Glucose 3xa (g/L)	Glucose 3xc (g/L)
0	0.182	0	23.7	22.8
8.5	27.4	25	7.2	9.3
9	30.6	31.4	12.3	10.9
9.5	42.8	40.2	13.7	15.7
10	49.4	49.2	17.1	16.7
10.5	65	59	18.7	16.7
11	69.6	62.8	16.2	14.6
11.5	81.2	79.8	14.1	16.7
12	90.8	96	16.3	22.3
12.5	95.6	94	17.6	15.3
13	88	96.8	20.3	9.8
13.33*	115.6	100	31.5	11.3
13.83	114.4	101.2	22.6	15.5
14.5	105.2	96.8	11.6	19.5
15			14.7	19.6
15.5	111.6	105.6	16.6	15.9
16			26.3	25.4
16.5 hr.	105.6	99.6	28.3	23.1
17. hr.			25.9	21.2
17.5 hr.	114.8	104	22.7	18.7

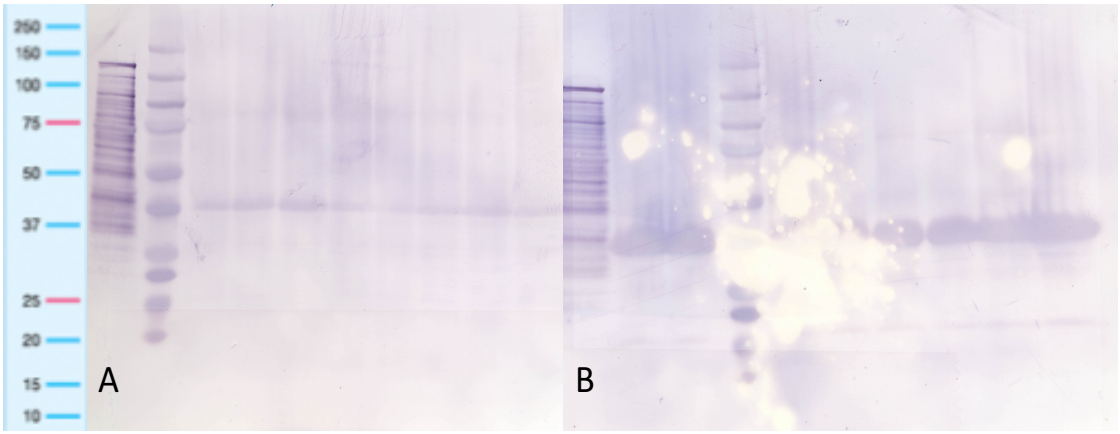

Figure 4.3 A) Western blot using His-tag antibody of the soluble fractions from 3xa and 3xc. B) Western blot using His-tag antibody of the insoluble fractions from runs 3xa and 3xc. The band at 37kDa is a known *E. coli* contaminating protein that is visualized using the mouse anti-His tag primary antibody.

suspending the bacteria from 4 hours post induction in Lysis Buffer (20mM Tris, 0.5M NaCl, 0.5% sarkosyl). Western blot analysis using an anti-His tag primary antibody repeatedly did not show the positive control. A new positive control was obtained and it was determined that the anti-His tag antibody was not detecting the protein of interest on the western blot. (Figure 4.3) To determine if the protein was not transferring to the membrane, a Coomassie blue was attempted on the gel post-transfer. (Figure 4.4 C) It appeared that several proteins were not transferring, but no bands were visualized at the proper molecular weight for the aggregate protein construct, although there appeared to be a smear at the higher molecular weight range. Exact molecular weights are not known because the ladder fully transferred from the gel. Concurrently, a western blot using a primary antibody specific to the His tag and another to the C-terminus of the protein were attempted. (Figure 4.4 A and B) Protein samples were concentrated to four times the original concentration using a Savant SpeedVac and showed a protein with the expected molecular weight of approximately 130 kDa, as well as several truncated iterations of the

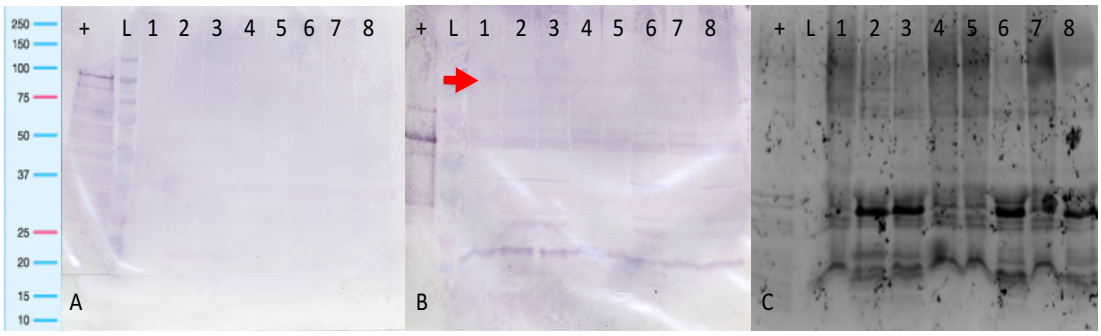

Figure 4.4 A) Western blot using His-tag antibody. B) Western blot using C-terminus antibody. A faint band at approximately 130 kDa is illustrated by the red arrow. C) Coomassie of the gel used for western post transfer. Lanes include positive control (+), soluble fraction from run 3xa (1,5), insoluble fraction from run 3xa (2,6), soluble fraction from run 3xc (4,7), and insoluble fraction from run 3xc (3,8) all at 4 hours post induction.

protein. It seemed as though there were equal amounts of the protein in both the soluble and insoluble fractions; therefore, the percentage of sarkosyl was increased for future runs.

A preliminary purification using 50g wet weight from fermentation run 3xc was attempted by suspending bacteria in 50 mL of lysis buffer (20mM Tris, 0.5M NaCl, 1% sarkosyl (w/v)). Lysate was collected after sonication for 15 minutes and heat treated at 70° C for 30 minutes, after which it was centrifuged and supernatant was collected. A portion of the supernatant was treated with 20% ethanol and the other portion was set aside in order to test both a

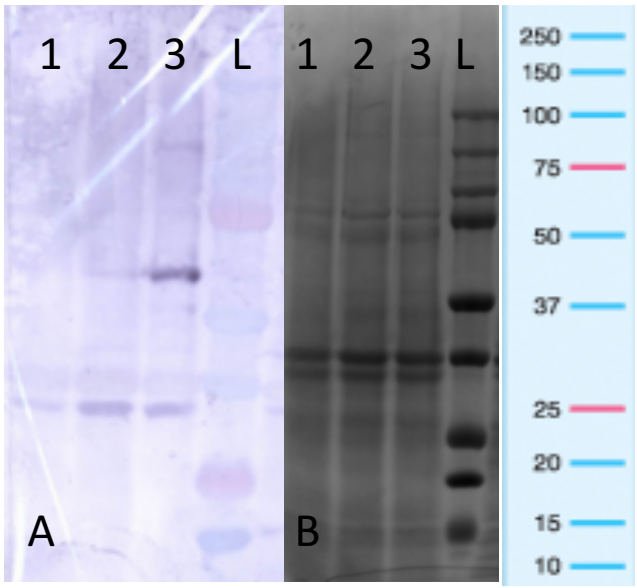

Figure 4.5 Western blot using C-terminus antibody(A) and Coomassie Blue (B) of the final pellets from 10%(1), 12.5%(2) and 15%(3) ammonium sulfate precipitations.

purification with and without ethanol at the same time. Samples were refrigerated overnight and a new precipitate was observed and collected by centrifugation. The supernatants were separated and ammonium sulfate was added to a saturation of 5%, 10%, 15%, and 20 %. Pellets were collected by centrifugation and samples were run and analyzed using Coomassie blue and western blot analysis. It was determined that the aggregate construct was found best at 10% ammonium sulfate precipitation without using the ethanol treatment. The purification was repeated using 10%, 12.5% and 15% ammonium sulfate saturations. When run next to each other, it appeared that the 15% ammonium sulfate precipitation yielded the most promising results, although the protein was still not pure (Figure 4.5). Lane 3 was used in later runs as a positive control.

AKTA His tag affinity chromatography was selected as another option for purification. Following sonication and heat treatment, the supernatant was diluted by a factor of 10 in binding buffer (.5mM Imidazole, 0.02M Tris, 0.5M NaCl) and loaded on

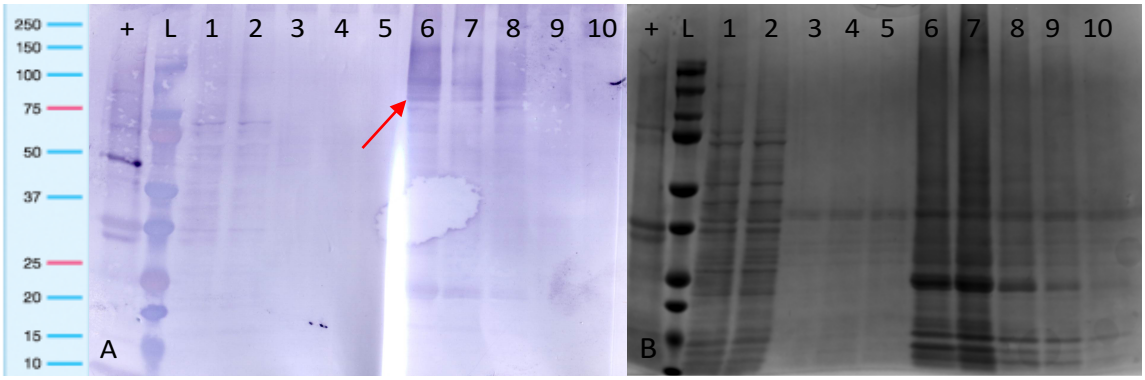

Figure 4.6 A) Western blot of AKTA run using a nickel column with an anti-C-terminus primary antibody B) Coomassie blue of the same protein samples. Lanes containing positive control (+) ladder (L), sample diluted by a factor of 10 (1) flow through (2), wash fraction A2 (3), wash fraction B1 (4), wash fraction B4 (5), elution fraction B5(6), elution fraction C1 (7), elution fraction C2 (8), elution Fraction C3 (9), and elution fraction C5 (10). Arrow indicates High molecular weight bands visualized on the western blot.

to a nickel column. Protein was eluted with elution buffer (0.5M Imidazole, 0.02M Tris, 0.5M NaCl) and fractions were collected. A peak was seen at fractions B5 and C1 and followed by a shoulder. A Coomassie blue and western blot was done on the sample and some high molecular weight bands were observed. Some of the bands were of a high enough molecular weight to indicate that there was some kind of multimerization occurring in the sample. (Figure 4.6)

Fractions B5 and C1 were combined and 300 uL ethanol was added to a 1.2 mL portion of the sample to create a 20% ethanol solution. The pellet was collected by centrifugation and was run on a gel. It appeared that the majority of our protein precipitated in the pellet, but a major contaminating protein remained in the supernatant. (Figure 4.7) It also appeared that much of the protein re-dissolved when prepping the pellet for loading on a gel by placing the pellet in a 100mM tris 20% glycerol solution. An extra western blot was run with this sample to see if anything could be seen using the

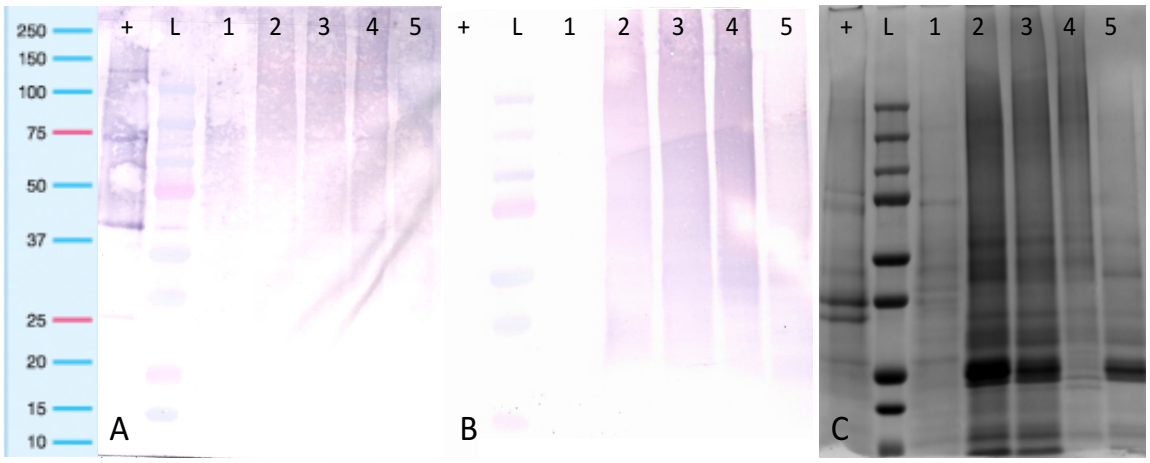

Figure 4.7 A) Western of ethanol precipitation purification with an anti C-terminus primary antibody. B) Western of ethanol precipitation purification with an anti-His tag primary antibody. C) Coomassie blue of the same protein samples. Lanes containing positive control (+) ladder (L) sample diluted by a factor of 10 from AKTA run (1), fraction B5 (2), fraction C1 (3) 20% ethanol pellet (4) 20% ethanol supernatant (5).

anti-His tag antibody. It showed that something was being pulled out of the solution with a His tag, and that it could be seen in the samples, but not on the positive control, which was the 15% ammonium sulfate precipitation from a previous experiment (Figure 4.5 lane 3). Unfortunately, no bands could be seen, just a smear.

It was then decided to try a 20% ethanol precipitation followed by a re-dissolution in an aqueous solution prior to loading the sample on the AKTA for His tag affinity chromatography. Purification proceeded as described previously, but after heat treatment ethanol was added to create a 20% solution. The pellet was collected and suspended in a solution containing 100mM tris and 20% glycerol and heated to 70°C and heated for 3 hours with agitation. The solution was centrifuged and the supernatant was collected and loaded on to a nickel column as previously described. No definitive peak was seen, but some protein came off the column during the elution step. Results were inconclusive, the C-terminus antibody had too much background, but it appeared with the His tag antibody that the protein stayed in solution with the ethanol precipitation. Additional experiments

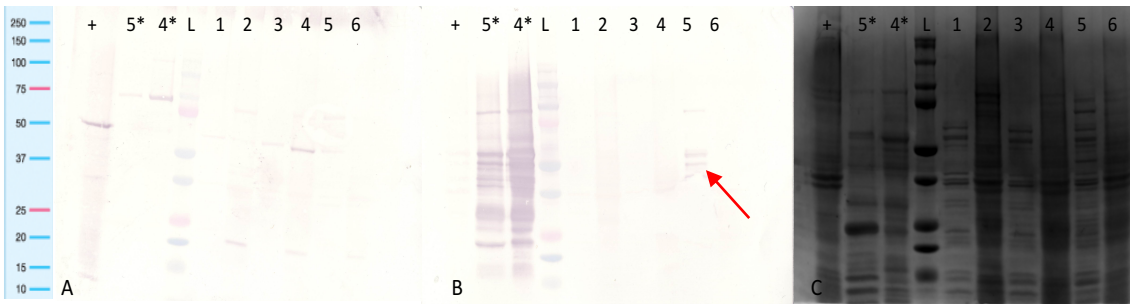

Figure 4.8 A) Western of lysis buffer experiment with an anti C-terminus primary antibody. B) Western of lysis buffer experiment with anti-His tag primary antibody C) Coomassie of the same protein samples. Lanes containing positive control (+) Lane 5 from the first ethanol experiment (5*), Lane 4 from the first ethanol experiment (4*), Ladder (L), supernatant of lysis buffer without additives (1), pellet of lysis buffer without additives (2), supernatant of lysis buffer with urea (3), pellet of lysis buffer with urea (4), supernatant of lysis buffer with sarkosyl (5), pellet of lysis buffer with sarkosyl (6). Arrow indicates aggregate protein in the supernatant.

are required to determine exactly how 20% ethanol affects the solubility of the 3x aggregate construct.

Some unanswered question were left from the original ethanol precipitation experiment. Why did the positive control give a positive result with the anti-C-terminus antibody, but not the anti-His tag antibody? And if the C-terminus is transcribed last, how can bands be seen at molecular weights lower than the full length protein? The possibility of some sort of cleavage of the protein had been considered, such that the C-terminus was transcribed and then cleaved into smaller fragments. It was also possible that the C-terminus antibody used was not monoclonal. Also, the C-terminus antibody is very old and it is possible that it is going bad. An additional experiment was completed that helped answer this question because of some of the lanes used for positive controls in this experiment.

In this experiment the issue of additives to the lysis buffer was addressed. Lysis Buffer (.5M NaCl, 20mM Tris), Lysis Buffer with Urea (.5M NaCl, 20mM Tris, .5M urea,) and Lysis Buffer with sarkosyl (.5M NaCl, 20mM Tris, 1% sarkosyl) were all used and sonicated at an amplitude of 40 for 15 minutes. The resulting solution was heat treated at 80°C for an hour before being centrifuged and the pellets collected. It was found that sarkosyl was most effective at pulling the protein into the supernatant. Laemmi Sample Buffer was obtained and used for the first time in this experiment. Using this sample buffer, the smears seen in lanes 4 and 5 from the ethanol experiment (Figure 4.7) were resolved into bands. Perhaps most importantly, we see exactly what we expect to see using both the C-terminus and His tag antibodies. Since the His tag is translated first, we would expect all iterations of this protein to contain the His tag and all

would be visible by western blot. Since the C-terminus is translated last we would expect only the full length protein to be visible by western blot. This is exactly what we see in lanes 4* and 5*. (Figure 4.8) This also reveals that the majority of the impurities in the samples are, in fact, shortened versions of the protein.

The above are all of the expression experiments attempted thus far. Additional runs should be completed to produce and purify the full length protein of each construct. It is suggested that expression be attempted in Sf9 cells, an insect line that would be more likely able to express full length protein,[2,3] and glycosylate the protein correctly.[4,5] It is also recommended to attempt dialysis or another method to remove the smaller protein fragments from the samples post AKTA to improve the purity of the protein.

Mechanical Testing

Since full length protein was in both samples and most of the impurities from the sample are in fact shortened iterations of our protein, the combined fractions B5 and C1 from the AKTA purification were frozen and lyophilized to perform a preliminary test on the adhesive properties of the protein. A 10% (w/v) solution was created and sonicated at an amplitude of 1 for two minutes and then heated in the microwave until the bottom of the vial read above 250° C by Fluke 561 IR Thermometer. Samples were then centrifuged to remove any insoluble impurities. The solution was then sprayed onto sixteen samples, four each of aluminum, polycarbonate, polyurethane, and polypropylene. This was followed by an application of 50uL of the protein solution to each sample. Samples were then adhered together and held with clips during the curing process. Samples were cured for approximately 24 hours and then tested. Only the

polyurethane, polypropylene, and one of the polycarbonate samples remained adhered after the curing process. The polycarbonate and all but one of the the polyurethane samples were broken in the process of loading the samples to be tested. One polypropylene sample was tested against the weight of its own sample, which gave the required force to break greater than 1.29KPa. One polyurethane sample was tested by free hanging weights and gave a force to break greater than 4.129 kPa and less than 7.018 kPa.

Further testing of the aggregate protein is needed. More work must be competed to provide a purer protein sample absent of truncated iterations of the protein to more effectively test the adhesive properties of the protein. In nature, the glue functions while wet, so underwater testing is another avenue that should be explored.

Other factors such as glycosylation and protein size also need to be addressed to better mimic the native protein. It is reasonable to assume that glycosylation may be a key factor to the ability of this protein to function as a glue. Multiple examples of adhesive proteins such as Mucins[6,7], or cell adhesive proteins such as Galectins[8,9] or TSLC1[10] are glycosylated in nature and it is assumed that glycosylation is important to their roles as adhesive proteins. Other adhesive proteins have been shown to require glycosylation for their adhesion[11]. Aggregate protein is unique from other spidroins because it is the only known glycosylated spidroin. Since its purpose in the web is unique, adhesion rather than structure, it can be inferred that glycosylation is a necessary difference that allows this spidroin its adhesive qualities.

In summary, the 3 repeat aggregate construct with a size of approximately 130 kDa has been successfully produced in a bacterial system. The other three constructs with 6, 9,

and 12 repeats with protein sizes of approximately 170, 210, and 250 kDa may hinder its expression in a bacterial system as bacteria natively produce proteins with sizes of around 100 kDa or less.

The synthetic 3x construct did not appear to be very adhesive. We strongly suspect this is due to lack of glycosylation and protein size. These issues can be addressed by expressing the protein in eukaryotic cells such as yeast or SF9 cells. It is recommended to use SF9 cells because they are more closely related to spider cells evolutionarily than yeast or mammalian cells and therefore more likely to glycosylate correctly.

References

(1) Jensen, E. B.; Carlsen, S. *Biotechnol. Bioeng.* **1990**, *36* (1), 1–11.
(2) Structural Genomics Consortium; China Structural Genomics Consortium; Northeast Structural Genomics Consortium; Gräslund, S.; Nordlund, P.; Weigelt, J.; Hallberg, B. M.; Bray, J.; Gileadi, O.; Knapp, S.; Oppermann, U.; Arrowsmith, C.; Hui, R.; Ming, J.; dhe-Paganon, S.; Park, H.; Savchenko, A.; Yee, A.; Edwards, A.; Vincentelli, R.; Cambillau, C.; Kim, R.; Kim, S.-H.; Rao, Z.; Shi, Y.; Terwilliger, T. C.; Kim, C.-Y.; Hung, L.-W.; Waldo, G. S.; Peleg, Y.; Albeck, S.; Unger, T.; Dym, O.; Prilusky, J.; Sussman, J. L.; Stevens, R. C.; Lesley, S. A.; Wilson, I. A.; Joachimiak, A.; Collart, F.; Dementieva, I.; Donnelly, M. I.; Eschenfeldt, W. H.; Kim, Y.; Stols, L.; Wu, R.; Zhou, M.; Burley, S. K.; Emtage, J. S.; Sauder, J. M.; Thompson, D.; Bain, K.; Luz, J.; Gheyi, T.; Zhang, F.; Atwell, S.; Almo, S. C.; Bonanno, J. B.; Fiser, A.; Swaminathan, S.; Studier, F. W.; Chance, M. R.; Sali, A.; Acton, T. B.; Xiao, R.; Zhao, L.; Ma, L. C.; Hunt, J. F.; Tong, L.; Cunningham, K.; Inouye, M.; Anderson, S.; Janjua, H.; Shastry, R.; Ho, C. K.; Wang, D.; Wang, H.; Jiang, M.; Montelione, G. T.; Stuart, D. I.; Owens, R. J.; Daenke, S.; Schütz, A.; Heinemann, U.; Yokoyama, S.; Büssow, K.; Gunsalus, K. C. *Nat. Methods* **2008**, *5* (2), 135–146.
(3) Jana, S.; Deb, J. K. *Appl. Microbiol. Biotechnol.* **2005**, *67* (3), 289–298.
(4) Vaughn, J. L.; Goodwin, R. H.; Tompkins, G. J.; McCawley, P. *In Vitro* **1977**, *13* (4), 213–217.
(5) Hollister, J. R.; Shaper, J. H.; Jarvis, D. L. *Glycobiology* **1998**, *8* (5), 473–480.
(6) Perez-Vilar, J.; Hill, R. L. *J. Biol. Chem.* **1999**, *274* (45), 31751–31754.
(7) Klinken, B. J. V.; Dekker, J.; Buller, H. A.; Einerhand, A. W. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **1995**, *269* (5), G613–G627.
(8) Colin Hughes, R. *Biochimie* **2001**, *83* (7), 667–676.

(9) Perillo, N. L.; Marcus, M. E.; Baum, L. G. *J. Mol. Med. 76* (6), 402–412.
(10) Masuda, M.; Yageta, M.; Fukuhara, H.; Kuramochi, M.; Maruyama, T.; Nomoto, A.; Murakami, Y. *J. Biol. Chem.* **2002**, *277* (34), 31014–31019.
(11) Filbin, M. T.; Tennekoon, G. I. *J. Cell Biol.* **1993**, *122* (2), 451–459.

CHAPTER 5

CONCLUSION

In this work, genetic material has been isolated for the spider aggregate glue gene with DNA fragments of greater than 20 kilobases. These fragments have yet to be cloned successfully, but the techniques used can be implemented in later work for the identification of additional spider glue genetic material.

Sequences characterized by Collin et al. have been used to create four synthetic constructs of the aggregate glue gene ranging in size from 3-12 iterations of the repetitive region. The construct with 3 repeats has been successfully expressed in bacteria and partially purified. In addition, preliminary testing of this protein as an adhesive has begun.

Aggregate spider glue represents an exciting new biomaterial with unique physical properties that could be used in many different adhesive situations. Underwater adhesives and in a surgical setting can be imagined because of its unique ability to remain sticky when wet. Since spiders cannot be farmed for the purpose of collecting this glue, expressing this glue in other organisms seems the most likely way that this biomaterial will be able to be mass produced and used in any setting other than on a spider's web.

Incomplete genetic information was first found by Choresh et al.[1] and then additional genetic information was found by Collin et al.[2] Collin et al. found that the aggregate gene contains a C-terminal region similar to other spidroins, an approximately 1Kb linker region, as well as a repetitive region consisting of approximately 300 base pair repeats.[2] Only two and a half repeats were found in this fashion; therefore, it is

important to elucidate the sequence up to the N-terminus to determine how many repeats are formed and if there are any other motifs present in the aggregate glue gene. The research in this thesis contributes to the genetic work done by others by demonstrating that 5' race has the potential to determine additional sequence data. cDNA fragments of the aggregate gene were created in this manner with a length of greater than 20kb. When using 5' RACE, care should be taken to design primers only outside the repetitive region of the gene as failure to do so will cause the primer to bind at multiple sites and give conflicting data using the 5' RACE technique.

This work has also built upon the data provided by Collin et al. to build a synthetic aggregate construct for expressing glue proteins with various lengths of the repetitive region. Constructs with 3, 6, 9, and 12 repeats have been created, and expression of these constructs has begun in *E. coli*. It is suggested that these constructs also be expressed in Sf9 cells[3] for two reasons. The first is that *E. coli* have trouble making proteins of such large size (130 kDa- 250kDa).[4,5] The eukaryotic Sf9 cells naturally produce larger proteins than *E. coli* and it therefore follows that they would not have as much trouble producing full length protein. In addition, Sf9 cells can glycosylate[6], and aggregate protein is a known glycosylated protein.[7,8] Since Sf9 cells are an insect cell line, it seems that they would be more likely to correctly glycosylate this protein than other eukaryotic cells such as yeast.

Research should continue to characterize the proteins produced by the genetic constructs created in this work. It is suggested that NMR and mass spectrometry would be likely ways to characterize this water soluble protein at the molecular level. Mechanically these proteins should be tested to determine if the adhesive properties are

similar to that of natural spider aggregate glue and if increasing the repetitive region increases the adhesive properties of the glue.

References

(1) Choresh, O.; Bayarmagnai, B.; Lewis, R. V. *Biomacromolecules* **2009**, *10* (10), 2852–2856.
(2) Collin, M. A.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y. *Sci. Rep.* **2016**, *6*, 21589.
(3) Vaughn, J. L.; Goodwin, R. H.; Tompkins, G. J.; McCawley, P. *In Vitro* **1977**, *13* (4), 213–217.
(4) Structural Genomics Consortium; China Structural Genomics Consortium; Northeast Structural Genomics Consortium; Gräslund, S.; Nordlund, P.; Weigelt, J.; Hallberg, B. M.; Bray, J.; Gileadi, O.; Knapp, S.; Oppermann, U.; Arrowsmith, C.; Hui, R.; Ming, J.; dhe-Paganon, S.; Park, H.; Savchenko, A.; Yee, A.; Edwards, A.; Vincentelli, R.; Cambillau, C.; Kim, R.; Kim, S.-H.; Rao, Z.; Shi, Y.; Terwilliger, T. C.; Kim, C.-Y.; Hung, L.-W.; Waldo, G. S.; Peleg, Y.; Albeck, S.; Unger, T.; Dym, O.; Prilusky, J.; Sussman, J. L.; Stevens, R. C.; Lesley, S. A.; Wilson, I. A.; Joachimiak, A.; Collart, F.; Dementieva, I.; Donnelly, M. I.; Eschenfeldt, W. H.; Kim, Y.; Stols, L.; Wu, R.; Zhou, M.; Burley, S. K.; Emtage, J. S.; Sauder, J. M.; Thompson, D.; Bain, K.; Luz, J.; Gheyi, T.; Zhang, F.; Atwell, S.; Almo, S. C.; Bonanno, J. B.; Fiser, A.; Swaminathan, S.; Studier, F. W.; Chance, M. R.; Sali, A.; Acton, T. B.; Xiao, R.; Zhao, L.; Ma, L. C.; Hunt, J. F.; Tong, L.; Cunningham, K.; Inouye, M.; Anderson, S.; Janjua, H.; Shastry, R.; Ho, C. K.; Wang, D.; Wang, H.; Jiang, M.; Montelione, G. T.; Stuart, D. I.; Owens, R. J.; Daenke, S.; Schütz, A.; Heinemann, U.; Yokoyama, S.; Büssow, K.; Gunsalus, K. C. *Nat. Methods* **2008**, *5* (2), 135–146.
(5) Jana, S.; Deb, J. K. *Appl. Microbiol. Biotechnol.* **2005**, *67* (3), 289–298.
(6) Hollister, J. R.; Shaper, J. H.; Jarvis, D. L. *Glycobiology* **1998**, *8* (5), 473–480.
(7) Vollrath, F.; Tillinghast, E. K. *Naturwissenschaften* **1991**, *78* (12), 557–559.
(8) Sahni, V.; Miyoshi, T.; Chen, K.; Jain, D.; Blamires, S. J.; Blackledge, T. A.; Dhinojwala, A. *Biomacromolecules* **2014**, *15* (4), 1225–1232.

APPENDICES

APPENDIX A

SUPPLEMENTAL DATA

Aggregate Genetic and Protein Sequence

Asg2 Grand Composite:
2905bp

TTGAAACCGGACCCGATGGAAAACCAAGTAAGCTCGTCGTTCCCACCACACCGAAAAGG
CCACGAGGACCAGGAGGAAGTCCTTTAGGTCCAGGAACGCAATTCCAAACTCCAGGAAC
AACTCCAACTCCCGTTCCCGGACCAGATGGTAAACCATTGCAAATTGTTCCCGCTGGTC
CAGGTACCACCCCCGGAACAGTGACCGGCCCAGACGGCAAACCAAGTAAATTCGTCGTT
CCCGACGGAGCATTCACGACACCAGGCTCAATCCCCGGTCCAGACGGAAAACCGCTCCC
AGTTGAACCAGCTGGTCCAGGCACCACTCCGGGACTCCAAACCGGACCCGATGGCAAAC
CAAGTAAACTCGTTGTTCCTACGACAACGAAGGGTCCAGGATCTGGTGTTCCTCCTGGA
TTACTTTTCCCTACTCCAGCAACTACACCAACTCCAGTTCCCGGTCCTCGTGGAGAACC
TGTTCAGATTATGGCAGCAGGTCCTGGTACTACTCCTGGTACTGTGACAGGTCCTGAGG
GTGAACCAGTTAAATTTATCGTTCCAAATGGGGCATTCTCAACTCCAGGTTCAATTCCA
GGCCCTGATGGAAAACCAATTCACGTCGAACCAGCTGGTCCTGGAACTACTCCGGGAGC
AAGGACTGGACCTGATGGAAATATAGTTAAAATATACTTGCCTTCTACTCCCAGTACCC
CTCCTCCTCCTTTGGCAACCACTCCTGCTGATGTAATGGGATCTGATGGGCAACCAATT
CTTATTTATCCTGCTGGTCCAGGAACTACTCCAGGAACAGTTACTGGTCCCGATGGGAA
GCCTACAAAGTTTATTGTTCCTTTGGGAGCCTTTACCACTCCTGGTTCTATACCAGGTC
CTGATGGAACACCAATACCTGTAGAACCAGCTGGTCCTGGAACAACTCCGGGTGTTGAA
ACAGATTCTGATGGAAACGTGAATAAGATTATTTTACCCACTACTCCAAAAAGACCATC
TCATCCTTCTCCGATGCCGCTTACAACTACTCCAATACCTAGTGATGGATCTAAACCTA
TTCAAATTGTTCCGGCTGGTCCTGGAACAACTCCAGGTACAGTCACTGGTTCTGATGGA
AAACCGACAAAATTTATTGTTCCGCAGGGAGCTTTTATAACCCCTGGTACAATTCCAGG
TCCTGACGGAAATCCTGTACCAGTTGAACCAGAAGGACCAGGAAATAGTCCCGGTGTTC
AGACTGGACCTAACGGCAATATTATTAAAATTGTAATTCCCACAACTACTCCTCTTCCA
CCACCCCCAGGACCTTTGGACCCTGCCAGTGAGCCTATTGCACCTTTTGGACCTGGTAA
TGTACCTAATTCTCCTAAATCTCCCGGCAATTATCCCGGATATTCTTTCCAATTCCCGG
GTTACCCTGATGCTCCAGGTTCCATAGGCCCTCTGGGATATTTAGATTTCAGCCAATTG
CCTAGTAGTATGTCCCCTGAAATGGAAGGCAATATTGGTTTCCTTCCGGATTTCAGTTC
TGAAATCGGAGGTCCATTCCCTGGTTTTCCACCAGGTCCAGATAACTCAGGTCCAGGGG
GATTTTTAAATGTTCATTCTCTCCCTGATTTTGTGAATCCAGGATATGGATTCCCTGGT
TCTCCTCAAGCCCCATTGGGTTTCCTAAACTTTAGTCTTTTACCGGACGATTACAATCC
AGGATTCCCTGGTCAGTTGGTTTTCCCTGGTTATCCCGGCTCTCCAGGAAGTAGTGGCC
AATTTCCCGGAGGATTCTTGAGTCTCGATGAATTACCGGAAGATGTTAGGAATATGTTG
AACAACACTTTCAGTTTACCCGAATTATTGCATTCTCTACAGCCTCTCTTCCCTGGAAG
ATCAATCAATTCTGGTGTCATTCCAAAAGACAATTTACAAAATATTCCAGGATTTAGCG
GTACTTACGATAATCTAAGACTTTCAAACATTGGAGATAACAATAACCCTACCGGAGGT
GTGTTCTACCTTCCTGAAATGGTACGACTCATTAGTTATCTTCCTGTAGGATCATTCCC
TAATGGCCCTGGAACAATCAATCAGAACGGTGGATTCGGCCATCCATTTAATTTCCCAG

GATTAAATGGCGCCCCGGGATACATTTGCGACTATCTGGATAACATCGATGTAACAGGT
GGAAGCTCAGATGACTTGGGAGGAGAAATCAGAGGAAATGACAATGGTCCATCGGGTGA
CGTAGCGGATGCGGCTCCTGGAAGTGATGTAGGTGCTCCAGCTCCTACAGGAAATACAG
CAGCTCCATCAGGGCAGAGCATGAGTTCTTCGAAACTTCAACCACCTGAAAACCAAGGT
ATGCCCGAATCTGACTGCGATGATGATGTGTTTTCCACCTTCATGAAAGCAAGATCTGC
CCTTATGGATGTATCTTCTAGTACAGGAGTCAATCCAATTAGCCAGCTAACTCAAGACA
TCATCTCTGGAATCAATCCATCTGAAGACAGTGTTGATTACAATAAATTTTTTAATAAA
CTTTCATCTCTACTTTCCCAAGTACGCTCGGGTTCTTCTGATAAACCTAATAAAGAACT
ATTATCAATCTTAATGGAAGGTTTAGTGGCCGCATTGGAAGCTCTAAACGCCGCAAAGA
TCAGTGGATTCCGAGACGACTATTATGTACCTAGCGATGTACCAGTGTATACGTCATTC
CTTTCCGAGATACTTTATTGAAGCAAACAAGCATTTGCAGTGATATAAAAATTTTTTTC
TAATGAATAACATTTTATATACATGTAAGAAACACATATATAATAGTAAAATTTTGTGC
GAACAATTCAAACTTTGTGTGATACTAAAGACTTCATAATAAAATATTTGTCGTTGTAA
AAAAACCTCGTGCC

The Blue sequence is the portion filled in by PCR of gDNA. The yellow sequences are the primers that were used for the PCR. All sequences in black are from the cDNA. The size of the total sequence, including the portion filled in from gDNA, corresponds to the size of the cDNA insert found on gels.

```
        E  T  G  P  D  G  K  P  S  K  L  V  V  P  T  T  P  K  R  P
3       gaaaccggacccgatggaaaaccaagtaagctcgtcgttcccaccacaccgaaaaggcca  62
        R  G  P  G  G  S  P  L  G  P  G  T  Q  F  Q  T  P  G  T  T
63      cgaggaccaggaggaagtcctttaggtccaggaacgcaattccaaactccaggaacaact  122
        P  T  P  V  P  G  P  D  G  K  P  L  Q  I  V  P  A  G  P  G
123     ccaactcccgttcccggaccagatggtaaaccattgcaaattgttcccgctggtccaggt  182
        T  T  P  G  T  V  T  G  P  D  G  K  P  S  K  F  V  V  P  D
183     accacccccggaacagtgaccggcccagacggcaaaccaagtaaattcgtcgttcccgac  242
        G  A  F  T  T  P  G  S  I  P  G  P  D  G  K  P  L  P  V  E
243     ggagcattcacgacaccaggctcaatccccggtccagacggaaaaccgctcccagttgaa  302
        P  A  G  P  G  T  T  P  G  L  Q  T  G  P  D  G  K  P  S  K
303     ccagctggtccaggcaccactccgggactccaaaccggacccgatggcaaaccaagtaaa  362
        L  V  V  P  T  T  T  K  G  P  G  S  G  V  P  P  G  L  L  F
363     ctcgttgttcctacgacaacgaagggtccaggatctggtgttcctcctggattacttttc  422
        P  T  P  A  T  T  P  T  P  V  P  G  P  R  G  E  P  V  Q  I
423     cctactccagcaactacaccaactccagttcccggtcctcgtggagaacctgttcagatt  482
        M  A  A  G  P  G  T  T  P  G  T  V  T  G  P  E  G  E  P  V
483     atggcagcaggtcctggtactactcctggtactgtgacaggtcctgagggtgaaccagtt  542
        K  F  I  V  P  N  G  A  F  S  T  P  G  S  I  P  G  P  D  G
543     aaatttatcgttccaaatggggcattctcaactccaggttcaattccaggccctgatgga  602
        K  P  I  H  V  E  P  A  G  P  G  T  T  P  G  A  R  T  G  P
603     aaaccaattcacgtcgaaccagctggtcctggaactactccgggagcaaggactggacct  662
        D  G  N  I  V  K  I  Y  L  P  S  T  P  S  T  P  P  P  P  L
663     gatggaaatatagttaaaatatacttgccttctactcccagtacccctcctcctcctttg  722
        A  T  T  P  A  D  V  M  G  S  D  G  Q  P  I  L  I  Y  P  A
723     gcaaccactcctgctgatgtaatgggatctgatgggcaaccaattcttatttatcctgct  782
        G  P  G  T  T  P  G  T  V  T  G  P  D  G  K  P  T  K  F  I
783     ggtccaggaactactccaggaacagttactggtcccgatgggaagcctacaaagtttatt  842
        V  P  L  G  A  F  T  T  P  G  S  I  P  G  P  D  G  T  P  I
843     gttcctttgggagcctttaccactcctggttctataccaggtcctgatggaacaccaata  902
        P  V  E  P  A  G  P  G  T  T  P  G  V  E  T  D  S  D  G  N
903     cctgtagaaccagctggtcctggaacaactccgggtgttgaaacagattctgatggaaac  962
        V  N  K  I  I  L  P  T  T  P  K  R  P  S  H  P  S  P  M  P
963     gtgaataagattatttacccactactccaaaaagaccatctcatccttctccgatgccg  1022
        L  T  T  T  P  I  P  S  D  G  S  K  P  I  Q  I  V  P  A  G
1023    cttacaactactccaatacctagtgatggatctaaacctattcaaattgttccggctggt  1082
```

```
      P  G  T  T  P  G  T  V  T  G  S  D  G  K  P  T  K  F  I  V
1083  cctggaacaactccaggtacagtcactggttctgatggaaaaccgacaaaatttattgtt  1142
      P  Q  G  A  F  I  T  P  G  T  I  P  G  P  D  G  N  P  V  P
1143  ccgcagggagcttttataacccctggtacaattccaggtcctgacggaaatcctgtacca  1202
      V  E  P  E  G  P  G  N  S  P  G  V  Q  T  G  P  N  G  N  I
1203  gttgaaccagaaggaccaggaaatagtcccggtgttcagactggacctaacggcaatatt  1262
      I  K  I  V  I  P  T  T  T  P  L  P  P  P  P  G  P  L  D  P
1263  attaaaattgtaattcccacaactactcctcttccaccacccccaggacctttggaccct  1322
      A  S  E  P  I  A  P  F  G  P  G  N  V  P  N  S  P  K  S  P
1323  gccagtgagcctattgcacctttggacctggtaatgtacctaattctcctaaatctccc  1382
      G  N  Y  P  G  Y  S  F  Q  F  P  G  Y  P  D  A  P  G  S  I
1383  ggcaattatcccggatattctttccaattcccgggttaccctgatgctccaggttccata  1442
      G  P  L  G  Y  L  D  F  S  Q  L  P  S  S  M  S  P  E  M  E
1443  ggccctctgggatatttagatttcagccaattgcctagtagtatgtcccctgaaatggaa  1502
      G  N  I  G  F  L  P  D  F  S  S  E  I  G  G  P  F  P  G  F
1503  ggcaatattggtttccttccggatttcagttctgaaatcggaggtccattccctggtttt  1562
      P  P  G  P  D  N  S  G  P  G  G  F  L  N  V  H  S  L  P  D
1563  ccaccaggtccagataactcaggtccaggggattttaaatgttcattctctccctgat  1622
      F  V  N  P  G  Y  G  F  P  G  S  P  Q  A  P  L  G  F  L  N
1623  tttgtgaatccaggatatggattccctggttctcctcaagccccattgggtttcctaaac  1682
      F  S  L  L  P  D  D  Y  N  P  G  F  P  G  Q  L  V  F  P  G
1683  tttagtcttttaccggacgattacaatccaggattccctggtcagttggttttccctggt  1742
      Y  P  G  S  P  G  S  S  G  Q  F  P  G  G  F  L  S  L  D  E
1743  tatcccggctctccaggaagtagtggccaatttcccggaggattcttgagtctcgatgaa  1802
      L  P  E  D  V  R  N  M  L  N  N  T  F  S  L  P  E  L  L  H
1803  ttaccggaagatgttaggaatatgttgaacaacactttcagtttacccgaattattgcat  1862
      S  L  Q  P  L  F  P  G  R  S  I  N  S  G  V  I  P  K  D  N
1863  tctctacagcctctcttccctggaagatcaatcaattctggtgtcattccaaaagacaat  1922
      L  Q  N  I  P  G  F  S  G  T  Y  D  N  L  R  L  S  N  I  G
1923  ttacaaaatattccaggatttagcggtacttacgataatctaagactttcaaacattgga  1982
      D  N  N  N  P  T  G  G  V  F  Y  L  P  E  M  V  R  L  I  S
1983  gataacaataaccctaccggaggtgtgttctaccttcctgaaatggtacgactcattagt  2042
      Y  L  P  V  G  S  F  P  N  G  P  G  T  I  N  Q  N  G  G  F
2043  tatcttcctgtaggatcattccctaatggccctggaacaatcaatcagaacggtggattc  2102
      G  H  P  F  N  F  P  G  L  N  G  A  P  G  Y  I  C  D  Y  L
2103  ggccatccatttaatttcccaggattaaatggcgccccgggatacatttgcgactatctg  2162
      D  N  I  D  V  T  G  G  S  S  D  D  L  G  G  E  I  R  G  N
2163  gataacatcgatgtaacaggtggaagctcagatgacttgggaggagaaatcagaggaaat  2222
      D  N  G  P  S  G  D  V  A  D  A  A  P  G  S  D  V  G  A  P
2223  gacaatggtccatcgggtgacgtagcggatgcggctcctggaagtgatgtaggtgctcca  2282
      A  P  T  G  N  T  A  A  P  S  G  Q  S  M  S  S  S  K  L  Q
2283  gctcctacaggaaatacagcagctccatcagggcagagcatgagttcttcgaaacttcaa  2342
      P  P  E  N  Q  G  M  P  E  S  D  C  D  D  D  V  F  S  T  F
2343  ccacctgaaaaccaaggtatgcccgaatctgactgcgatgatgatgtgttttccaccttc  2402
      M  K  A  R  S  A  L  M  D  V  S  S  S  T  G  V  N  P  I  S
2403  atgaaagcaagatctgcccttatggatgtatcttctagtacaggagtcaatccaattagc  2462
      Q  L  T  Q  D  I  I  S  G  I  N  P  S  E  D  S  V  D  Y  N
2463  cagctaactcaagacatcatctctggaatcaatccatctgaagacagtgttgattacaat  2522
      K  F  F  N  K  L  S  S  L  L  S  Q  V  R  S  G  S  S  D  K
2523  aaatttttaataaactttcatctctactttcccaagtacgctcgggttcttctgataaa  2582
      P  N  K  Δ  Δ  E  L  L  S  I  L  M  E  G  L  V  A  A  L  E  A  L
2583  cctaataaagaactattatcaatcttaatggaaggtttagtggccgcattggaagctcta  2642
      N  A  A  K  I  S  G  F  R  D  D  Y  Y  V  P  S  D  V  P  V
2643  aacgccgcaaagatcagtggattccgagacgactattatgtacctagcgatgtaccagtg  2702
      Y  T  S  F  L  S  E  I  L  Y  *
2703  tatacgtcattcctttccgagatactttattgaagcaaacaagcatttgcagtgatataa  2762
2763  aaatttttttctaatgaataacatttatatacatgtaagaaacacatatataatagtaa  2822
2823  aattttgtgcgaacaattcaaactttgtgtgatactaaagacttcataataaaatatttg  2882
2883  tcgttgtaaaaaaacctcgtgcc  2905
```

>Asg2_Grand_Align Translated - Frame 3
910 AA

ETGPDGKPSKLVVPTTPKRPRGPGGSPLGPGTQFQTPGTTPTPVPGPDGKPLQIVPAGP
GTTPGTVTGPDGKPSKFVVPDGAFTTPGSIPGPDGKPLPVEPAGPGTTPGLQTGPDGKP
SKLVVPTTTKGPGSGVPPGLLFPTPATTPTPVPGPRGEPVQIMAAGPGTTPGTVTGPEG
EPVKFIVPNGAFSTPGSIPGPDGKPIHVEPAGPGTTPGARTGPDGNIVKIYLPSTPSTP
PPPLATTPADVMGSDGQPILIYPAGPGTTPGTVTGPDGKPTKFIVPLGAFTTPGSIPGP
DGTPIPVEPAGPGTTPGVETDSDGNVNKIILPTTPKRPSHPSPMPLTTTPIPSDGSKPI
QIVPAGPGTTPGTVTGSDGKPTKFIVPQGAFITPGTIPGPDGNPVPVEPEGPGNSPGVQ
TGPNGNIIKIVIPTTTPLPPPPGPLDPASEPIAPFGPGNVPNSPKSPGNYPGYSFQFPG
YPDAPGSIGPLGYLDFSQLPSSMSPEMEGNIGFLPDFSSEIGGPFPGFPPGPDNSGPGG
FLNVHSLPDFVNPGYGFPGSPQAPLGFLNFSLLPDDYNPGFPGQLVFPGYPGSPGSSGQ
FPGGFLSLDELPEDVRNMLNNTFSLPELLHSLQPLFPGRSINSGVIPKDNLQNIPGFSG
TYDNLRLSNIGDNNNPTGGVFYLPEMVRLISYLPVGSFPNGPGTINQNGGFGHPFNFPG
LNGAPGYICDYLDNIDVTGGSSDDLGGEIRGNDNGPSGDVADAAPGSDVGAPAPTGNTA
APSGQSMSSSKLQPPENQGMPESDCDDDVFSTFMKARSALMDVSSSTGVNPISQLTQDI
ISGINPSEDSVDYNKFFNKLSSLLSQVRSGSSDKPNKELLSILMEGLVAALEALNAAKI
SGFRDDYYVPSDVPVYTSFLSEILY*

M indicates Omer's putative start codon.
KE the position in the sequence where Omer has a multi-repeat insert compared to our sequence.
AGPGT is a peptide Omer found on his MS analysis.

```
gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     ------------------------------------------------------------
Asg2 cDNA     ETGPDGKPSKLVVPTTPKRPRGPGGSPLGPGTQFQTPGTTPTPVPGPDGKPLQIVPAGPG

gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     ------------------------------------------------------------
Asg2 cDNA     TTPGTVTGPDGKPSKFVVPDGAFTTPGSIPGPDGKPLPVEPAGPGTTPGLQTGPDGKPSK

gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     ------------------------------------------------------------
Asg2 cDNA     LVVPTTTKGPGSGVPPGLLFPTPATTPTPVPGPRGEPVQIMAAGPGTTPGTVTGPEGEPV

gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     ------------------------------------------------------------
Asg2 cDNA     KFIVPNGAFSTPGSIPGPDGKPIHVEPAGPGTTPGARTGPDGNIVKIYLPSTPSTPPPPL

gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     ------------------------------------------------------------
Asg2 cDNA     ATTPADVMGSDGQPILIYPAGPGTTPGTVTGPDGKPTKFIVPLGAFTTPGSIPGPDGTPI

gDNA_Asg2     ------------------------------------MPLTTTPIPSDGSKPIFQIVPA
Omer Asg2     ------------------------------------MPLTTTPIPSDGSKPIFQIVPA
Asg2 cDNA     PVEPAGPGTTPGVETDSDGNVNKIILPTTPKRPSHPSPMPLTTTPIPSDGSKPI-QIVPA
                                                  *************** *****

gDNA_Asg2     GPGTTPGTVTGSDGKPTKFIVPQGAFITPGTIPGPDGNPVPVEPEGPGNSPGVQTGPNGN
Omer Asg2     GPGTTPGTVTGSDGKPTKFIVPQGAFITPGTIPGPDGNPVPVEPEGPGNSPGVQTGPNGN
Asg2 cDNA     GPGTTPGTVTGSDGKPTKFIVPQGAFITPGTIPGPDGNPVPVEPEGPGNSPGVQTGPNGN
              ************************************************************

gDNA_Asg2     IIKIVIPTTTPLPPPPGPLDPASEPIAPFGPGNVPNSPKSPGNYPGYSFQFPGYPDAPGS
Omer Asg2     IIKIVIPTTTPLPPPPGPLDPASEPIAPFGPGNVPNSPKSPGNYPGYSFQFPGYPDAPGS
Asg2 cDNA     IIKIVIPTTTPLPPPPGPLDPASEPIAPFGPGNVPNSPKSPGNYPGYSFQFPGYPDAPGS
              ************************************************************
```

```
gDNA_Asg2     IGPLGYLDFSQLPSSMSPEMEGNIGFLPDFSSEIGGPFPGFPPGPDNSGPGGFLNVHSLP
Omer Asg2     IGPLGYLDFSQLPSSMSPEMEGNIGFLPDFSSEIGGPFPGFPPGPDNSGPGGFLNVHSLP
Asg2 cDNA     IGPLGYLDFSQLPSSMSPEMEGNIGFLPDFSSEIGGPFPGFPPGPDNSGPGGFLNVHSLP
              ************************************************************

gDNA_Asg2     DFVNPGYGFPGSPQAPLGFLNFSLLPDDYNPGFPGQLVFPGYPGSPGSSGQFPGGFLSLD
Omer Asg2     DFVNPGYGFPGSPQAPLGFLNFSLLPDDYNPGFPGQLVFPGYPGSPGSSGQFPGGFLSLD
Asg2 cDNA     DFVNPGYGFPGSPQAPLGFLNFSLLPDDYNPGFPGQLVFPGYPGSPGSSGQFPGGFLSLD
              ************************************************************

gDNA_Asg2     ELPEDVRNMLNNTFSLPELLHSLQPLFPGRSINSGVIPKDNLQNIPGFSGTYDNLRLSNI
Omer Asg2     ELPEDVRNMLNNTFSLPELLHSLQPLFPGRSINSGVIPKDNLQNIPGFSGTYDNLRLSNI
Asg2 cDNA     ELPEDVRNMLNNTFSLPELLHSLQPLFPGRSINSGVIPKDNLQNIPGFSGTYDNLRLSNI
              ************************************************************

gDNA_Asg2     GDNNNPTGGVFYLPEMVRLISYLPVGSFPNGPGTINQNGGFGHPFNFPGLNGAPGYICDY
Omer Asg2     GDNNNPTGGVFYLPEMVRLISYLPVGSFPNGPGTINQNGGFGHPFNFPGLNGAPGYICDY
Asg2 cDNA     GDNNNPTGGVFYLPEMVRLISYLPVGSFPNGPGTINQNGGFGHPFNFPGLNGAPGYICDY
              ************************************************************

gDNA_Asg2     LDNIDVTGGSSDDLGGEIRGNDNGPSGDVADAAPGSDVGAPAPTGNTAAPSGQSMSSSKL
Omer Asg2     LDNIDVTGGSSDDLGGEIRGNDNGPSGDVADAAPGSDVGAPAPTGNTAAPSGQSMSSSKL
Asg2 cDNA     LDNIDVTGGSSDDLGGEIRGNDNGPSGDVADAAPGSDVGAPAPTGNTAAPSGQSMSSSKL
              ************************************************************

gDNA_Asg2     QPPENQGMPESDCDDDVFSTFMKARSALMDVSSSTGVNPISQLTQDIISGINPSEDSVDY
Omer Asg2     QPPENQGMPESDCDDDVFSTFMKARSALMDVSSSTGVNPISQLTQDIISGINPSEDSVDY
Asg2 cDNA     QPPENQGMPESDCDDDVFSTFMKARSALMDVSSSTGVNPISQLTQDIISGINPSEDSVDY
              ************************************************************

gDNA_Asg2     NKFFNKLSSLLSQVRSGSSDKPNK-----------------------------------
Omer Asg2     NKFFNKLSSLLSQVRSGSSDKPNKELLSILMEGLVVSGLGVSGSSVSGLGVSGSSVSGLG
Asg2 cDNA     NKFFNKLSSLLSQVRSGSSDKPNK-----------------------------------
              ************************

gDNA_Asg2     ------------------------------------------------------------
Omer Asg2     VSGSSVSGLGVSGSSVSGLGVSGSSVSGLGVSGSSVSGLGVSGLGVSGSSVSGLGVSGSS
Asg2 cDNA     ------------------------------------------------------------

gDNA_Asg2     ---------------------------------------------ELLSILMEGLVAALE
Omer Asg2     VSGLGVSGSSVSGLGVSGSSVSGLGVSGSSVSGLGVSGSSDKPNKELLSILMEGLVAALE
Asg2 cDNA     ---------------------------------------------ELLSILMEGLVAALE
                                                           ***************

gDNA_Asg2     ALNAAKISGFRDDYYVPSDVPVYTSFLSEILY
Omer Asg2     ALNAAKISGFRDDYYVPSDVPVYTSFLSEILY
Asg2 cDNA     ALNAAKISGFRDDYYVPSDVPVYTSFLSEILY
              ********************************
```

5’ RACE Design

Primer SP1
5’tatatcactgcaaatgcttgtttgc3’ (Primer=reverse compliment of sequence data)
5’GCAAACAAGCATTTGCAGTGATATA3’ (Coding strand sequence data)
Tm 54.7℃
GC content 36%

Primer SP2
5’cttgtttgcttcaataaagtatctcgg3’(Primer=reverse compliment of sequence data)
5’CCGAGATACTTTATTGAAGCAAACAAG3’ (Coding strand sequence data)
Tm 54.6℃
GC content 37%

Primer SP3
5’ggaatccactgatctttgcggcg3’ (Primer=reverse compliment of sequence data)
5’CGCCGCAAAGATCAGTGGATTCC3’ (Coding strand sequence data)
Tm 60.6℃
GC content 59%

Coding strand from Choresh et al.

5’TTGAAACCGGACCCGATGGAAAACCAAGTAAGCTCGTCGTTCCCACCACACCGAAAA
GGCCACGAGGACCAGGAGGAAGTCCTTTAGGTCCAGGAACGCAATTCCAAACTCCAGGA
ACAACTCCAACTCCCGTTCCCGGACCAGATGGTAAACCATTGCAAATTGTTCCCGCTGG
TCCAGGTACCACCCCCGGAACAGTGACCGGCCCAGACGGCAAACCAAGTAAATTCGTCG
TTCCCGACGGAGCATTCACGACACCAGGCTCAATCCCCGGTCCAGACGGAAAACCGCTC
CCAGTTGAACCAGCTGGTCCAGGCACCACTCCGGGACTCCAAACCGGACCCGATGGCAA
ACCAAGTAAACTCGTTGTTCCTACGACAACGAAGGGTCCAGGATCTGGTGTTCCTCCTG
GATTACTTTTCCCTACTCCAGCAACTACACCAACTCCAGTTCCCGGTCCTCGTGGAGAA
CCTGTTCAGATTATGGCAGCAGGTCCTGGTACTACTCCTGGTACTGTGACAGGTCCTGA
GGGTGAACCAGTTAAATTTATCGTTCCAAATGGGGCATTCTCAACTCCAGGTTCAATTC
CAGGCCCTGATGGAAAACCAATTCACGTCGAACCAGCTGGTCCTGGAACTACTCCGGGA
GCAAGGACTGGACCTGATGGAAATATAGTTAAAATATACTTGCCTTCTACTCCCAGTAC
CCCTCCTCCTCCTTTGGCAACCACTCCTGCTGATGTAATGGGATCTGATGGGCAACCAA
TTCTTATTTATCCTGCTGGTCCAGGAACTACTCCAGGAACAGTTACTGGTCCCGATGGG
AAGCCTACAAAGTTTATTGTTCCTTTGGGAGCCTTTACCACTCCTGGTTCTATACCAGG
TCCTGATGGAACACCAATACCTGTAGAACCAGCTGGTCCTGGAACAACTCCGGGTGTTG
AAACAGATTCTGATGGAAACGTGAATAAGATTATTTTACCCACTACTCCAAAAAGACCA
TCTCATCCTTCTCCGATGCCGCTTACAACTACTCCAATACCTAGTGATGGATCTAAACC
TATTCAAATTGTTCCGGCTGGTCCTGGAACAACTCCAGGTACAGTCACTGGTTCTGATG
GAAAACCGACAAAATTTATTGTTCCGCAGGGAGCTTTTATAACCCCTGGTACAATTCCA
GGTCCTGACGGAAATCCTGTACCAGTTGAACCAGAAGGACCAGGAAATAGTCCCGGTGT
TCAGACTGGACCTAACGGCAATATTATTAAAATTGTAATTCCCACAACTACTCCTCTTC
CACCACCCCCAGGACCTTTGGACCCTGCCAGTGAGCCTATTGCACCTTTTGGACCTGGT

AATGTACCTAATTCTCCTAAATCTCCCGGCAATTATCCCGGATATTCTTTCCAATTCCC
GGGTTACCCTGATGCTCCAGGTTCCATAGGCCCTCTGGGATATTTAGATTTCAGCCAAT
TGCCTAGTAGTATGTCCCCTGAAATGGAAGGCAATATTGGTTTCCTTCCGGATTTCAGT
TCTGAAATCGGAGGTCCATTCCCTGGTTTTCCACCAGGTCCAGATAACTCAGGTCCAGG
GGGATTTTTAAATGTTCATTCTCTCCCTGATTTTGTGAATCCAGGATATGGATTCCCTG
GTTCTCCTCAAGCCCCATTGGGTTTCCTAAACTTTAGTCTTTTACCGGACGATTACAAT
CCAGGATTCCCTGGTCAGTTGGTTTTCCCTGGTTATCCCGGCTCTCCAGGAAGTAGTGG
CCAATTTCCCGGAGGATTCTTGAGTCTCGATGAATTACCGGAAGATGTTAGGAATATGT
TGAACAACACTTTCAGTTTACCCGAATTATTGCATTCTCTACAGCCTCTCTTCCCTGGA
AGATCAATCAATTCTGGTGTCATTCCAAAAGACAATTTACAAAATATTCCAGGATTTAG
CGGTACTTACGATAATCTAAGACTTTCAAACATTGGAGATAACAATAACCCTACCGGAG
GTGTGTTCTACCTTCCTGAAATGGTACGACTCATTAGTTATCTTCCTGTAGGATCATTC
CCTAATGGCCCTGGAACAATCAATCAGAACGGTGGATTCGGCCATCCATTTAATTTCCC
AGGATTAAATGGCGCCCCGGGATACATTTGCGACTATCTGGATAACATCGATGTAACAG
GTGGAAGCTCAGATGACTTGGGAGGAGAAATCAGAGGAAATGACAATGGTCCATCGGGT
GACGTAGCGGATGCGGCTCCTGGAAGTGATGTAGGTGCTCCAGCTCCTACAGGAAATAC
AGCAGCTCCATCAGGGCAGAGCATGAGTTCTTCGAAACTTCAACCACCTGAAAACCAAG
GTATGCCCGAATCTGACTGCGATGATGATGTGTTTTCCACCTTCATGAAAGCAAGATCT
GCCCTTATGGATGTATCTTCTAGTACAGGAGTCAATCCAATTAGCCAGCTAACTCAAGA
CATCATCTCTGGAATCAATCCATCTGAAGACAGTGTTGATTACAATAAATTTTTTAATA
AACTTTCATCTCTACTTTCCCAAGTACGCTCGGGTTCTTCTGATAAACCTAATAAAGAA
CTATTATCAATCTTAATGGAAGGTTTAGTGGCCGCATTGGAAGCTCTAAACGCCGCAAA
GATCAGTGGATTCCGAGACGACTATTATGTACCTAGCGATGTACCAGTGTATACGTCAT
TCCTTTCCGAGATACTTTATTGAAGCAAACAAGCATTTGCAGTGATATAAAAATTTTTT
TCTAATGAATAACATTTTATATACATGTAAGAAACACATATATAATAGTAAAATTTTGT
GCGAACAATTCAAACTTTGTGTGATACTAAAGACTTCATAATAAAATATTTGTCGTTGT
AAAAAAACCTCGTGCC3′