Analysis of Plasmids Expressing Bacteriophage Lambda Procapsid Proteins

by

Mason Hayashi

Submitted in partial fulfillment of the requirements for graduation as an Honors Scholar at Point Loma Nazarene University, San Diego, California on 4/5, 2023.

Approved by _____
Dr. David Cummings

_____
Dr. Heidi Woelbern

_____
Dr. Kristopher Koudelka

Date April 5, 2023

Analysis of Plasmids Expressing Bacteriophage Lambda Procapsid Proteins

Mason Hayashi[1*]

[1]*Department of Biology, Point Loma Nazarene University, San Diego, CA*
[*]*Corresponding author*. E-mail address: mhayashi246@pointloma.edu

Abstract

Bacteriophage lambda is a complex double stranded DNA virus that is an ideal model to study the essential features of virus assembly. The head of this virus is called the procapsid and consists of 5 major proteins: the major capsid protein (gpE), head decoration protein (gpD), scaffolding protein (gpNu3), portal protein (gpB), and peptidase (gpC). Viral capsids are strong structures designed to transport and protect the genome when delivering viral DNA to a host cell. Because of these reasons, it is the focus of undergraduate research under Dr. Kristopher Koudelka at PLNU. In this lab, *in-vitro* synthesis of procapsids is attempted using plasmids to grow capsid proteins for combination. Previous research has not proven to be particularly productive as successful synthesis has not been achieved and for that reason, this project is aimed at verifying the identity of each gene in the plasmids as well as attempting to identify the expression vector for each. This research project had two approaches to answer this question. The first consisted of a restriction digest to verify the backbone of the gpD plasmids. This was followed by a PCR reaction to verify the presence of gpE and gpD in the control plasmid used for this lab. Primers were designed using the sequences of gpD and gpE and with these, amplicons were produced with the intention of sequencing. The second part of the experiment consisted of sequencing 13 of the 22 plasmids. Once sequenced, the gene sequences were run through Blastx to determine their identities. This determined that only 5 of the 13 sequences had all of the anticipated capsid genes present, gpNu3 was not present in pT7cap, and two plasmid sequences were not what was expected. Additionally, the backbones were run through Blastn to identify the most probable vectors. Results showed that further research can be done to identify each of the expression vectors and the larger research project should continue with the 13 verified plasmids until the rest can be sequenced and their identity validated.

Introduction

Bacteriophage Lambda is a complex double stranded DNA (dsDNA) virus that represents a plethora of genetic, biochemical, and biophysical information which makes it an ideal model to study essential features of virus assembly.[2] A major aspect of this bacteriophage is what is known as a procapsid. Viral capsids are strong structures that are designed to transport and protect the genome when delivering viral DNA to a host cell.[2] It is for these reasons that bacteriophage lambda procapsids are the focus of undergraduate research within this lab under the guidance of Dr. Kristopher Koudelka at PLNU. This

lab focuses on the *in-vitro* assembly of lambda procapsids with the goal of using these structures for targeted delivery of several potential cargos. One example is for drug delivery in applications such as chemotherapeutic cancer treatments. The Koudelka lab focuses on the selective isolation and *in-vitro* combination of the various proteins associated with the capsid head of the virus. Wild-type bacteriophage lambda procapsid utilizes five different capsid proteins for the assembly of the icosahedral structure. The major capsid protein (gpE) assembles into the icosahedral shape and the viral head decoration protein (gpD) adds to the surface of the expanded capsid lattice. In addition to these two proteins, however, there are three others that are necessary for a functioning wild-type procapsid. One is the portal protein (gpB) and another is the scaffolding protein (gpNu3). This scaffolding protein is necessary for the shell assembly and without it, the capsid proteins form into abnormal and non-functional structures. Additionally, gpC is a protease that is responsible for the degradation of gpNu3 as well as nucleation of the polymerization of the major capsid protein, gpE, into the icosahedral shape.[2]

This laboratory focused on the preliminary steps associated with *in-vitro* formation of procapsids. Transformed *E. coli* cells were utilized to individually harness the primary capsid protein gpE, the capsid decoration protein gpD, the capsid portal protein gpB, the protease gpC, and the capsid scaffolding protein gpNu3. There are 22 glycerol stocks of transformed *E. coli* cells that are used for the synthesis of these proteins and subsequent combinations. Each of these transformed cell lines hold a plasmid that contains the sequences that code for each individual protein. It is important to note that many of the plasmids are pT7cap, and pT7cap is the wild-type procapsid expression vector that expresses gpB, gpC, gpNu3, gpE, and gpD.[2]

Over the course of the past summer as well as the laboratory for the undergraduate molecular biology class, the *in-vitro* combination attempts have yielded poor results. As a result, this has led to the questioning of the validity of the identity of each of these plasmids. Therefore, the purpose of this research project is to dissect each of the plasmids in an attempt to sequence the inserts to verify their identities. During this process it was revealed that much was unknown about each of these plasmids pertaining not only to the inserts themselves but also the vectors they reside in. To explore this more, the project began with an Addgene database search to gather as much information as possible regarding the vectors of each of these plasmids. Each sample is labeled with the presumed proteins it codes for, however, the majority do not explicitly reveal the vectors. After this was completed, the project continued with a restriction digest to verify the presumed vectors were not the source of the faulty experiments and that nothing was wrong with the functionality of the backbone. This initial step was taken so that it may be confirmed that even if the genes were there, the backbone is functioning correctly to even allow for the transcription of the genes if they are present. The next step would then be to create primers to run a PCR of each of the genes. Once an amplicon is produced from the PCR, the fragment could be run on an electrophoresis gel to compare the experimental size of the amplicon with the expected size of the actual gene. If the fragments were within 100 base pairs or so of the expected length, the amplicons could then be sent off to be sequenced with reasonable confidence of their identity.

Unfortunately, research pivoted away from the PCR as an opportunity to sequence the entire plasmids was presented. The restriction digest provided inconclusive results, however, the PCR did show promising results for extended experimentation. Furthermore, a revisited literature search led to the presumed identification of 14 of the 22 samples. Therefore, the project pivoted to then send these 14 plasmids to be fully sequenced, but only 13 worked. Once the sequences had been received, a plasmid

map was produced using SnapGene. From here, the individual genes could be copied and inserted into BLAST to determine the identity of the genes in these plasmids. Additionally, the vector could also be copied to help determine the identity of backbones.

Methods

Literature Search

Before any experimentation, an Addgene database was conducted in order to determine the identities of the vectors that each of the plasmids resided in. Therefore, the initial plan was to move forward with all samples containing gpD. These included samples 3, 11, 16, 17 and 21 (Tab. 1), and it was inferred that the vectors for these 5 samples could be one of three options: gGEM, pGEX, or pETDuet. Samples 3 and 11 are pT7cap and article review showed that this is the wild-type expression vector expressing all five of the genes of interest: gpC, gpB, gpNu3, gpD, and gpE. Sample 16 is called pGEX-6P-3-"C"-gpD. Based on this it is known that the vector is pGEX-6P-3-"C" and it expresses gpD. Finally, sample 17 is gpD-2 Stock GST Construct and sample 21 is gpD-1 Stock GST Construct (Tab. 1). All experimentation then moved forward under the assumption that these five samples were one of pGEM, pGEX, or pETDuet.

As stated earlier, the first database search did not produce definitive results and additionally, subsequent experiments proved to be inconclusive as a result. Thus, a literature search was conducted using articles that referenced the samples and this proved to be much more productive. During this second review, the vector identity of 14 of the 22 plasmids was produced with reasonable confidence. It was revealed that the likely vector for samples 3, 4, 7, 8, 11, 12, 20, and 23 was pKKT7.[2,4] Then, the likely parent vector of samples 2, 13, 14, and 16 were pGEX-6P-3 as referenced in the plasmid name (Tab. 1). Finally, sample 10 was likely to be pETDuet-1 as also referenced in the plasmid name (Tab. 1). All of the 14 samples listed above and their presumed vector identities are listed in Table 1 (Tab. 1).

Number	Name
1	pH6gpNu3
2	pGex-6p-3 (empty) --> BL21 cell line
3	pT7cap+E (old)
4	pT7cap (new-carlos) *same as 7??
6	pET15b-H6gpE in BL21
7	pT7cap (new-carlos) *same as 4??
8	pT7Init_gpC (S166A) in BL21
9	pNu3_wt in BL21
10	pETDvet_gpE in BL21
11	pT7cap+E (old) 2.0
12	pT7cap - BL21 Carlos' plasmid transformation
13	pGEX-6P-3-"C"-gpE (Quikchanged)
14	pGEX-6P-3-"C"-gpNu3 (Quikchanged)
15	gpNu3-2 Stock GST Contruct
16	pGEX-6P-3-"C"-gpD (Quikchanged)
17	gpD-2 Stock GST construct
18	gpE-1 Stock GST construct
19	gpNu3-1 Stock GST construct
20	Procapsid pT7cap construct in DH5 alpha cells Dam (no D)
21	gpD-1 Stock GST constructs
22	gpE-2 Stock GST constructs
23	Procapsid pT7cap construct in BL21 cells Dam (no D)

Table 1: *List of all 22 plasmid glycerol stocks with respective names.*

Plasmid Purification

The focus of this project was to analyze the plasmid samples in our stock. In order to do this, each plasmid needed to be isolated as they were stored within transformed *E.coli* cells within a glycerol stock. These glycerol stocks are stored at -80℃. To begin the purification process, an overnight culture of the transformed *E.coli* cell stocks is necessary. Each overnight culture was prepped in 5 ml of LB

and a 1000X concentration of ampicillin as the plasmids of interest all provide ampicillin resistance. For the five gpD samples after the initial literature search, 5 ml of LB broth was added to 5 separate 15 ml conicals where 5 ul of 1M sterile filtered ampicillin was added to select for the plasmids of interest. Standard procedure was used to make the LB broth using milli-Q water. This was then autoclaved using liquid 30 to ensure sterilization. To add the transformed *E.coli* cells, each sample was taken from the -80℃ freezer and thawed on ice. Once thawed, a p20 pipette with p20 tips was used to pipette the sample up and down 5 times in order to coat the tip. Once coated and expelled of all contents, the coated tip was then ejected into the prepared 15 ml conicals. These were then placed in a shaking incubator at 37℃ and 200 rpm overnight. The samples were then checked for turbidity to ensure cell growth took place. In order to purify the plasmids from these overnight cell cultures, the Qiagen Spin Miniprep kit was used to isolate each sample. For the miniprep, 1 ml of the 5 ml overnight culture was spun down at 8000 rpm 3 times for a total of 3 ml pelleted down. From here the standard Qiagen Spin Miniprep procedure was followed with each centrifugation step conducted at 13000 rpm. Once each plasmid was miniprepped, a nanodrop reading was taken for each sample. The mini prepped samples for this part of the project were 3, 11, 16, and 21. Sample 3 had a concentration of 35.8 ng/ul, sample 11 had a concentration of 38.2 ng/ml, and sample 21 had a concentration of 104.7 ng/ul.

First Nanodrop Concentrations

Sample #	Concentration (ng/ul)
3	35.8
11	38.2
16	36
21	104.7

Table 2: *Nanodrop concentrations for samples 3, 11, 16, and 21 after initial literature search.*

It is important to note that sample 16 was repeated as the first miniprep yielded a very low concentration of plasmid. The second miniprep produced a concentration of 36.0 ng/ul. After isolating the plasmids, each prep was stored at 4℃.

After these samples were mini prepped, subsequent experimentation, explained later in the methods, did not produce reliable results. Focus was then placed on sample 3 which is the control sample for undergraduate procapsid formation. To isolate this sample, a Qiagen Midiprep Kit was used to produce a more concentrated sample for experimentation. For the midiprep, 50 ml of overnight culture was made by growing bacterial cultures in two 50 ml conicals using 25 mls of LB in each. A 1000X ampicillin concentration was also needed for this procedure, so 25 ul of 1M ampicillin was added as well. Sample 3 was thawed on ice and pipetted up and down 5 times using a p1000 with p1000 tips. Once coated, the tip was ejected in each 50 ml conical. The samples were then left in a shaking incubator at 37℃ at 200 rpm. Once grown, standard midiprep procedure was followed using the SA-600 rotor. The sample was then nano dropped and produced a concentration of 171.5 ng/ml (Tab. 3).

Midiprep Nanodrop Concentration

Sample #	Concentration (ng/ul)
3	171.5

Table 3: *Nanodrop concentration for sample 3 midi prep.*

For the second part of this project, 14 samples were miniprepped using the standard procedure. Like the prior miniprep, The 14 samples chosen were 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 16, 20, and 23. These were chosen as they had the highest confidence for vector identity after the second literature search. For the miniprep, an overnight culture of all 14 samples was carried out. Each sample was grown in 5 ml of LB with a 1000X ampicillin concentration. The stock ampicillin used for this procedure was 100 mg/ml but a concentration of 0.050 mg/ml was needed. A

stock solution was made using 70 ml of LB broth and 35 ul of the 100 mg/ml ampicillin. A coated p20 tip was made using each thawed glycerol stock and ejected into the 15 ml conicals. Each was then placed in a shaking incubator at 37℃ at 200 rpm overnight. Standard miniprep procedure was followed with 3 ml of the 5 ml each overnight culture spun down. After the samples were miniprepped, they were subsequently nano dropped to determine sample concentrations. These concentrations are listed in Table 3.

Second Nanodrop Concentrations

Sample #	Concentration (ng/ul)
2	113.3
3	23.7
4	30.4
6	26.4
7	30.3
8	18.7
10	24.4
11	23.9
12	33
13	46.8
14	76.6
16	39.1
20	38.5
23	53.6

Table 4: *Nanodrop concentrations for samples 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 16, 20, and 23.*

Restriction Digest

The next step in this research project after the initial literature search was to conduct a restriction digest with the four gpD samples that grew (Tab. 2). This was to determine the identity and functionality of the backbones for these plasmids. This ensures that even if the genes are present, the backbone is functioning to express these genes. The three presumed vectors were pGEX-6P-3 and some form of pGEM. To narrow down pGEM vectors, Addgene was utilized to filter results. Within the vector database, all pGEM vectors were filtered to a bacterial vector type. This produced 17 results. To conduct a restriction digest, restriction sites need to be identified. Therefore, a sequence needs to be present to analyze. Of the 17 bacterial pGEM vectors, 7 had sequences available, research moved forward with those 7 vectors which are listed in Table 5. Now, pGEX-6P-3, and the 7 pGEM vectors were analyzed to determine an enzyme that would cut all 8 vectors. Using the digest feature of each vector sequence, identical double digest enzymes were compiled between all 8 vectors. Then each of these enzymes were tested against the capsid genes using sequence comparison to see if they cut any of the genes. Any enzymes that cut the genes were ruled out. It was determined that PvuI was the best option. This enzyme was then purchased and prepared for the restriction digest.

pGEM Vectors

pGEM Vector
pGEM-5Zf(-)
pGEM-5Zf(+)
pGEM T
pGEM T Easy
pGEM-9Zf-
pGEM-3Zf-
pGEM-13Zf(+)

Table 5: *List of 7 plausible pGEM vectors*

The samples used for this procedure were the four nano-dropped gpD samples from the previous procedure. The PvuI enzyme used for the digest was purchased from New England Biolabs, and the standard digest procedure was used. This procedure consisted of combining 1 ug of DNA, 5 ml of 10X NEbuffer, 1 ul enzyme, and adding the difference up to 50 ul using nuclease free water. To begin, the amount of sample to add to achieve a concentration of 1 ug was calculated. To do this, 28 ul of sample 3 was used, 26 ul of sample 11 was used, 28 ul of sample 16 was used, and 10 ul of sample 21 was used. Four PCR tubes were used and filled with the 10X NEbuffer, PvuI enzyme, the calculated amount of sample and the difference up to 50 ul with nuclease free water. Once all the reagents were combined, the samples were gently mixed by pipetting up and down and then microfuged briefly.

These samples were then incubated at 37℃ for 15 minutes.

To visualize these results, a native gel was used. A 1.25% agarose gel was made by first combining 50 ml of 1X TAE buffer and 0.625 g of powdered agarose in an erlenmeyer flask. This mixture was then microwaved in bursts to dissolve all the agarose. Once dissolved and the flask was cool enough to touch with a glove, 25 ul of ethidium bromide was added for visualization. This was then poured into a gel box with a 12 well comb. Then, 300 ml of 1X TAE running buffer was used to fill the gel box and 150 ul of ethidium bromide was added to the running buffer. Once the gel was prepared, samples were prepared for loading. There were 9 wells used. The first was loaded with the ladder that was prepared using 1 ul of Invitrogen 1 Kb Plus DNA ladder, 1 ul of bluejuice, and 8 ul of nuclease free water. Then the subsequent 8 wells were loaded with the pre and post digest samples for samples 3, 11, 16, and 21. To prepare the pre-digest samples 10 ul of the sample was mixed with 2 ml of 6X loading dye. To prepare the post-digest samples, 25 ul of the digested samples was combined with 5 ul of 6X loading dye. Once each sample was prepared, the gel was loaded then ran at 100V for 50 minutes and subsequently analyzed (Fig. 3).

PCR

The results from the restriction digestion reaction were not very conclusive, nonetheless, a PCR analysis of gpE and gpD inserts was performed for sample 3, the overall control for the larger research project. To begin this process, primers were synthesized using Integrated DNA Technologies. Through this site, the FASTA sequences for gpD and gpE were inserted and then the site formulated forward and reverse primers to use for PCR amplification (Fig. 7,8). Once receiving these primers, they had to be prepped as they came dry. All primer preparation information was given via the company. For the gpD forward primer, 274 ul of TE was added, for the gpD reverse primer, 345 ul of TE was added, for the gpE forward primer 277 ul of TE was added, and for the gpE reverse primer, 273 ul of TE was added. All additions resulted in a primer concentration of 100 uM. Once the primers were prepared, sample preparation could commence. To create the PCR reaction, the midi prepped sample 3 was used. With that, for both gpE and gpD preparation, 5 ul of BioRad, 9 ul of autoclaved milli-q water, 2.5 ul of the forward primer, 2.5 ul of reverse primer and 1 ul of DNA was combined in a PCR tube and gently pippeted up and down and microfuged to mix. Three PCR reactions were performed with these samples. The first was a gradient PCR to determine the most optimal annealing temperature for these samples. This is shown in Figure 4. The PCR process used was 95℃ for 10 minutes for DNA polymerase activation and template denaturation. Then an amplification step with 95℃ for 10 seconds, gradient for 30 seconds, and 60℃ for 30 seconds all cycled 40 times. A second and third PCR was performed for gpE and gpD amplification. For the gpE amplification, the procedure was 95℃ for 10 minutes one time and 95℃ for 10 seconds, 55℃ for 30 seconds, and 60℃ for 30 seconds cycled 40 times. For the gpD amplification, the procedure was 95℃ for 10 minutes one time and 95℃ for 10 seconds, 56℃ for 30 seconds, and 60℃ for 30 seconds cycled 40 times. Now, two gels were made to visualize all 3 PCR reactions. A 1% agarose gel was used using 50 ul of 1X TBE buffer and 25 ul of ethidium bromide. Once the gels were made, samples were prepared for loading. For the experimental wells, 5 ul of PCR product was combined with 1 ul of 6X loading dye and loaded into the gel. Additionally, negative controls were made using all of the PCR reagents without the DNA, these were also run under the same PCR parameters listed above. For these samples 5 ul was combined with 1 ul of 6X loading dye. Then the Thermo Scientific GeneRuler 100 bp Plus DNA ladder was loaded. The gel was then run

in a 1X TBE running buffer at 100V for 50 minutes then subsequently analyzed (Fig. 4,5).

The PCR reaction for gpE and gpD amplification did not produce clean results, and resultantly, isolated plasmid 3 was run on a gel to check for sample purity. To visualize these results, a native gel was used. A 1.25% agarose gel was made by first combining 50 ml of 1X TAE buffer and 0.625 g of powdered agarose in an erlenmeyer flask. This mixture was then microwaved in bursts to dissolve all the agarose. Once dissolved and the flask was cool enough to touch with a glove, 25 ul of ethidium bromide was added for visualization. This was then poured into a gel box with a 12 well comb. Then, 300 ml of 1X TAE running buffer was used to fill the gel box and 150 ul of ethidium bromide was added to the running buffer. Once the gel was prepared, 3 wells were loaded, one with 5 ul of plasmid 3 and 1 ul of 6X loading dye, a negative control of 5 ul of mini prep elution buffer and 1 ul of 6X loading dye, and the final well was the Invitrogen 1 Kb Plus DNA ladder. After the gel was loaded, it was run at 100V for 50 minutes (Fig. 6).

Plasmid Sequencing

For sequencing of the plasmids, the 14 isolated plasmids from the mini prep described earlier were prepped for whole plasmid sequencing via Eurofins Genomics LLC. To begin, 20 ul of each sample were aliquoted, however, the sequencing was limited to a maximum concentration of 40 ng/ul, therefore, samples 2, 13, 14, and 23 (Tab. 4) were diluted with autoclaved milli-q to achieve a concentration of 40 ng/ul. Once all the samples were diluted and aliquoted, all 14 were sent off to Eurofins Genomics LLC for whole plasmid sequencing using NGS 3 technology. Sample 16 was not able to be sequenced meaning 13 of the 14 plasmids were successfully sequenced.

Bioinformatics

Once the sequenced plasmid files were received from Eurofins Genomics LLC, Snapgene viewer was used to create plasmid maps for the 13 successfully sequenced samples. Each plasmid map was renamed for the corresponding plasmid sample. Using the plasmid mapping, all sequences of the genes within the transcription unit of the plasmid were copied and pasted into Blastx using the top strand to determine the identity of each gene. Once established, the identified genes were annotated on the plasmid maps (Fig. 1,2). Once this concluded, the sequences of all the plasmids outside of the transcription unit were copied using the top strand and pasted into Blastn. Using these results, the most probable vector identity was noted (Tab. 6).

Results

Number	Name	Presumed Parent Vector	Top BLAST Vector	Predicted Plasmid Size	Actual Plasmid Size	# of Genes in Promoter Region	Genes Present
2	pGex-6p-3 (empty) --> BL21 cell line	pGEX-6P-3	pGEX-6P-3	4983	4993	1	GST Tag
3	pT7cap+E (old)	pKKT7	pKK223-3	7888	7960	4	gpB, gpC, gpD, gpE
4	pT7cap (new-carlos) *same as 7??	pKKT7	pKK223-3	7888	7785	4	gpB, gpC, gpD, gpE
6	pET15b-H6gpE in BL21	pET15b	pET19b_ShHTL7cpGFP(L166)LSSmOrange	6753	6726	1	gpE
7	pT7cap (new-carlos) *same as 4??	pKKT7	pKK223-3	7888	7785	4	gpB, gpC, gpD, gpE
8	pT7Init_gpC (S166A) in BL21	pKKT7	Inconclusive	8114	6351	1	gpC
10	pETDvet_gpE in BL21	pETDuet-1	pETDuet-hTERT-noMCS2	6447	6389	1	gpE
11	pT7cap+E (old) 2.0	pKKT7	pKK223-3	7888	7785	4	gpB, gpC, gpD, gpE
12	pT7cap - BL21 Carlos' plasmid transformation	pKKT7	pKK223-3	7888	7941	4	gpB, gpC, gpD, gpE
13	pGEX-6P-3-"C"-gpE (Quikchanged)	pGEX-6P-3	pGEXGSTp65	6010	6009	2	GST Tag, gpE
14	pGEX-6P-3-"C"-gpNu3 (Quikchanged)	pGEX-6P-3	pGEXGSTp65	5589	5593	2	GST Tag, gpNu3
20	Procapsid pT7cap construct in DH5 alpha cells Dam (no D)	pKKT7	pKK223-3	7555	7785	4	gpB, gpC, gpD, gpE
23	Procapsid pT7cap construct in BL21 cells Dam (no D)	pKKT7	pGEX-6P-3	7555	6001	2	GST Tag, gpE

Table 6: *List of the 13 samples used for sequencing after the second literature search that produced the presumed parent vectors for each. The table lists the presumed parent vector as well as the top result produced from the BLAST search for each sample. Additionally, lists the number of genes identified in each transcription unit as well as what those genes are. Here it shows that only two of the samples, 2 and 10, are exact matches. Samples 13 and 14 are pGEX as expected, but not the same exact vector. Samples 3, 4, 7, 11, 12, and 20 were expected to be pKKT7[3] but had very high matches as seen via the query coverage and percent identity (Tab. 7) with pKK223-3 of which pKKT7 is a derivative[3]. Sample 8 was predicted to be pKKT7[2] as well but had inconclusive BLAST results with top results having a query coverage of <60%. Sample 6 was expected to be pET15b, but proved to be a form of pET19b instead. Sample 23 was the most dramatically different as it was expected to be pKKT7 but turned out to be pGEX-6P-3. Samples 3, 4, 7, 11, 12, and 20 all had four genes which were gpB, gpC, gpD, and gpE. Samples 2, 6, 8, and 10 all had one gene in the transcription unit. Sample 2 had GST, sample 6 had gpE, sample 8 had gpC, and sample 10 had gpE. Samples 13, 14 and 23 had two genes. Sample 13 has gpE and GST, sample 14 has gpNu3 and GST, and sample 23 has gpE and GST. Additionally, this table notes the predicted sizes of the total plasmid along with the actual sizes.*

Number	Name	Top BLAST Vector	Query Cover	Percent Identity
2	pGex-6p-3 (empty) --> BL21 cell line	pGEX-6P-3	99%	99.54%
3	pT7cap+E (old)	pKK223-3	94%	99.84%
4	pT7cap (new-carlos) *same as 7??	pKK223-3	99%	99.90%
6	pET15b-H6gpE in BL21	pET19b_ShHTL7cpGFP(L166)LSSmOrange	100%	99.91%
7	pT7cap (new-carlos) *same as 4??	pKK223-3	99%	99.90%
8	pT7Init_gpC (S166A) in BL21	Inconclusive	N/A	N/A
10	pETDvet_gpE in BL21	pETDuet-hTERT-noMCS2	100%	99.94%
11	pT7cap+E (old) 2.0	pKK223-3	99%	99.90%
12	pT7cap - BL21 Carlos' plasmid transformation	pKK223-3	94%	99.89%
13	pGEX-6P-3-"C"-gpE (Quikchanged)	pGEXGSTp65	100%	99.87%
14	pGEX-6P-3-"C"-gpNu3 (Quikchanged)	pGEXGSTp65	100%	99.85%
20	Procapsid pT7cap construct in DH5 alpha cells Dam (no D)	pKK223-3	99%	99.90%
23	Procapsid pT7cap construct in BL21 cells Dam (no D)	pGEX-6P-3	99%	99.85%

Table 7: *Lists the sample number of each of 13 samples sequenced after the second literature search. Shows the top BLAST search results as well as the query cover and percent identity of each result for each respective sample. Samples 2, 4, 6, 7, 10, 11, 13, 14, 20, and 23 all have a query coverage of 99% or higher and a percent identity 99% or higher giving high confidence. Samples 3 and 12 have a query coverage of 94% and percent identities of 99% or higher giving reasonable confidence.*

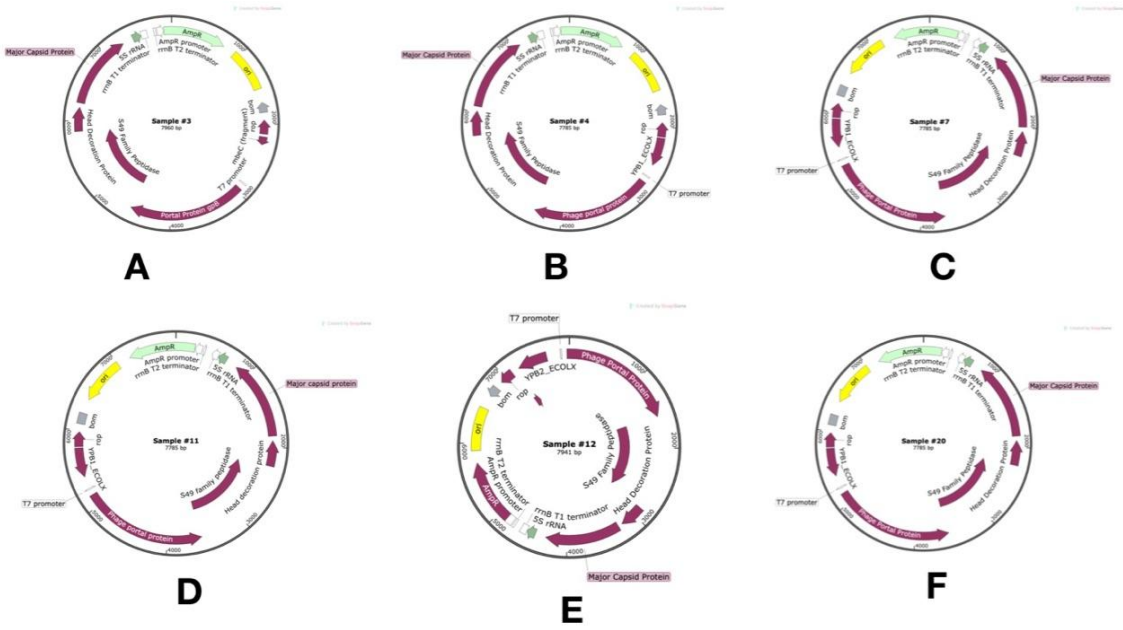

Figure 1: *Plasmid maps of samples 3, 4, 7, 11, 12, and 20. All 6 of these plasmid maps have four genes in the transcription unit being the Phage Portal Protein (gpB), S49 Family Peptidase (gpC), Head Decoration Protein (gpD), and the Major Capsid Protein (gpE). (A) Sample #3. (B) Sample #4. (C) Sample #7. (D) Sample #11. (E) Sample #12. (F) Sample #20.*

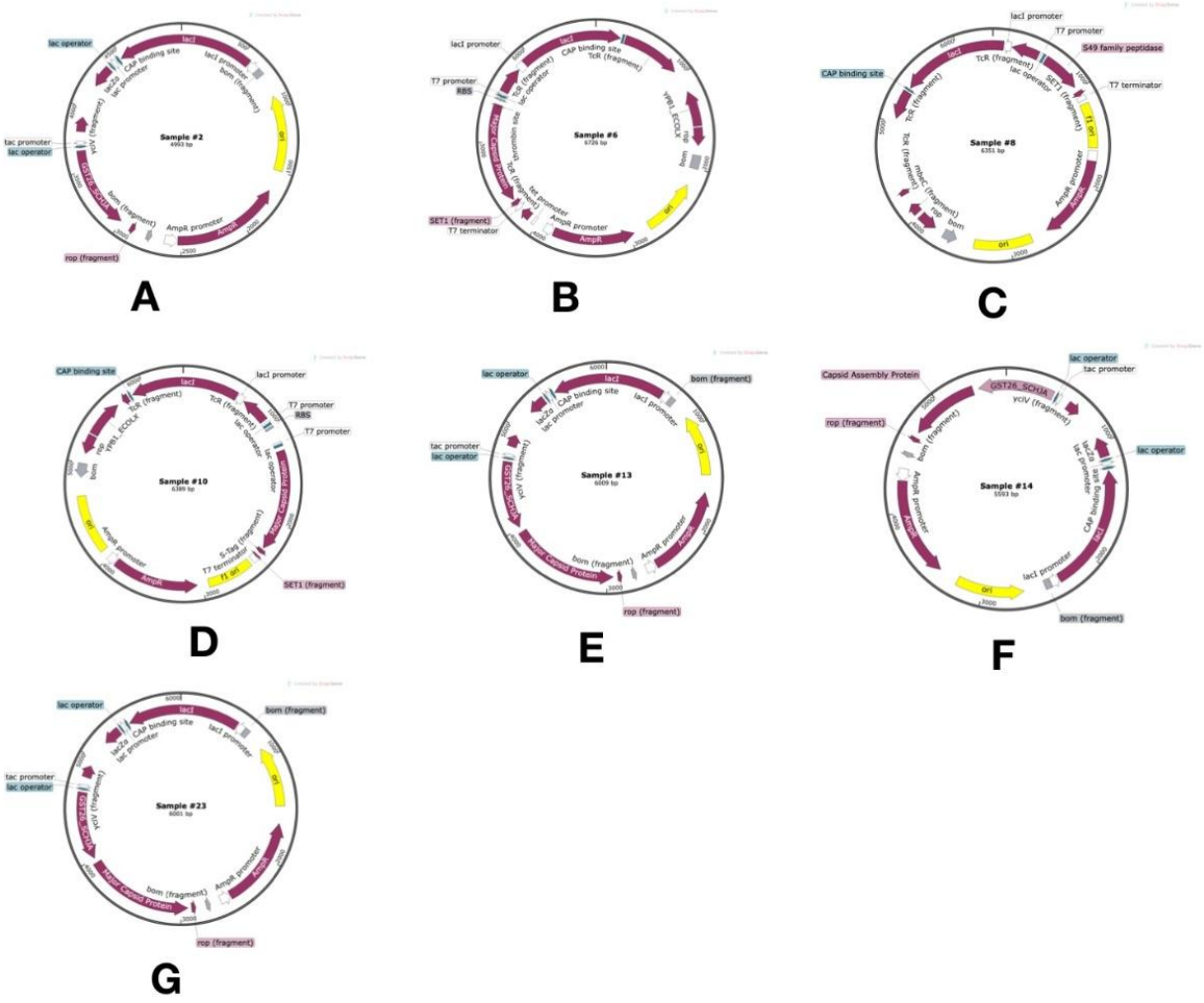

Figure 2: *Plasmid maps of samples 2, 6, 8, 10, 13, 14, and 23. Each plasmid has 2 genes or less in the transcription unit. (A) Sample #2 which is an empty pGEX-6P-3 vector and it contains one gene in its transcription unit. This gene is GST_SCHJA which is Glutathione S-transferase class-mu 26 kDa isozyme.[3] The function is conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles. (B) Sample #6 containing one gene being the Major Capsid Protein (gpE). (C) Sample #8 which contains one gene being a fragment of the S49 family peptidase (gpC). (D) Sample #10 containing one gene being gpE. (E) Sample #13 containing two genes being the gpE and GST26_SCHJA. (F) Sample #14 containing two genes being the Capsid Assembly Protein (gpNu3) and GST26_SCHJA. (G) Sample #23 containing two genes being gpE and GST26_SCHJA.*

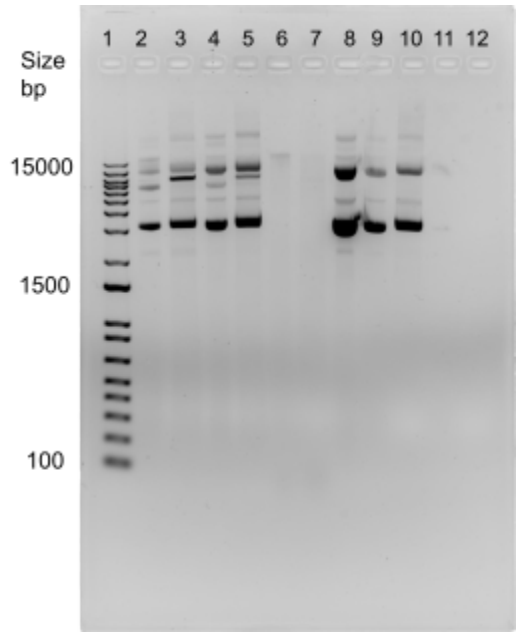

Figure 3: *Gel electrophoresis of the restriction digest experiment. Restriction digest consisted of samples 3, 11, 16, and 21. Lane 1 is the Invitrogen 1 Kb Plus DNA Ladder. Lane 2 is pre-digested sample 3 and lane 3 is post-digested sample 3. Lane 4 is pre-digested sample 11 and Lane 5 is pos-digested sample 11. Lane 6 is pre-digested sample 16 and lane 7 is post-digested sample 16. Both lanes 8 and 9 are pre-digested sample 21 as some pre-digested sample bled into well 9 when loading well 8. Lane 10 is post-digested sample 21. The results from the digest as seen in the gel are not very conclusive. The enzyme that was used was PvuI and should have yielded a double cut for each plasmid resulting in two fragments showing up in the digest lanes. The gel shows that not only are there more than two bands, the bands are significantly larger in size than anticipated. The thick bands at the larger size could represent linearization of the plasmid. Additionally, the presence of so many bands could point to an impure sample or potential spillover.*

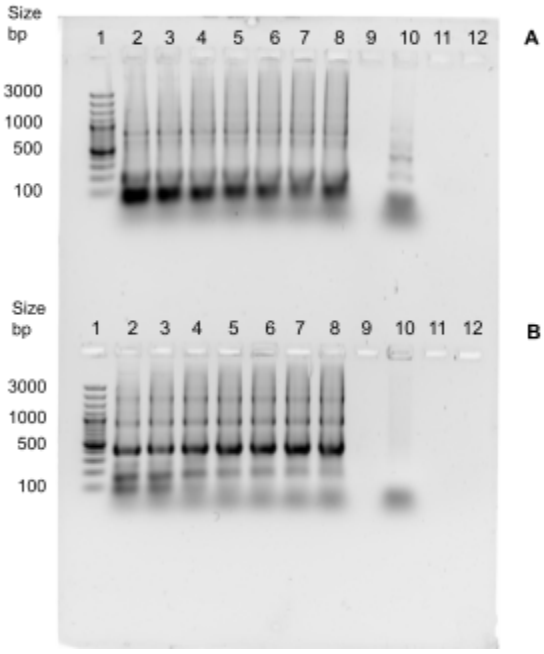

Figure 4: *Gradient PCR of sample #3, the procapsid control. This PCR procedure was carried out using primers that were specific to gpD and gpE, both of which are contained within this plasmid. (A) gpD gradient PCR. Lane 1 is the Thermo Scientific GeneRuler 100 bp Plus DNA ladder. Lanes 2-8 are the sample 3 plasmid with the gpD primers at gradient annealing temperatures. Those being 48.2℃, 50.5℃, 53.9℃, 56.7℃, 59.6℃, 61.8℃, and 63.4℃ respectively. Lane 10 is the negative control containing all reagents minus the sample 3 plasmid. This was annealed at 64.6℃. Strong banding patterns for each temperature show no significant advantage for one temperature, although the lower temperature range yields strong bands for gpD. (B) gpE gradient PCR. Lane 1 is the Thermo Scientific GeneRuler 100 bp Plus DNA ladder. Lanes 2-8 are the sample 3 plasmid with the gpE primers at gradient annealing temperatures. Those being 48.2℃, 50.5℃, 53.9℃, 56.7℃, 59.6℃, 61.8℃, and 63.4℃ respectively. Lane 10 is the negative control containing all reagents minus the sample 3 plasmid. This was annealed at 64.6℃. Strong banding patterns for each temperature show no significant advantage for one specific temperature, although a midrange temperature around 60℃ seems to yield a good banding for both gpE and gpD.*

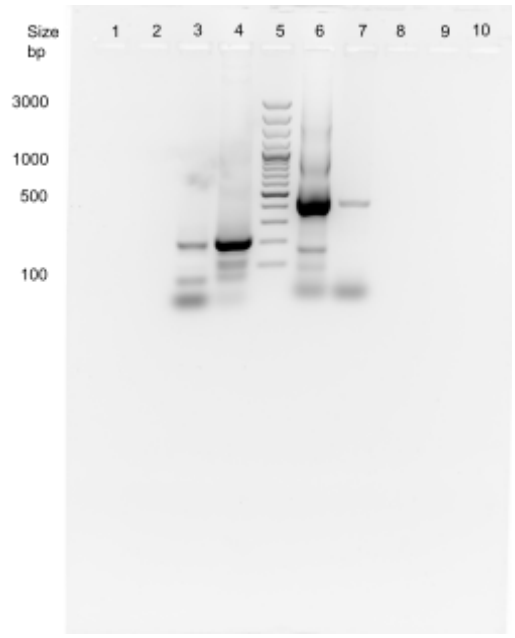

Figure 5: *PCR reaction of sample 3 using the primers from Integrated DNA Technologies. Used a forward and reverse primer for both gpD and gpE. Lane 3 is the negative control for gpD for sample 3 containing all PCR reactants minus sample 3 plasmid. This contains BioRad, nuclease-free water, and the forward and reverse primers. Lane 4 is the experimental well for the PCR reaction of gpD. Lane 5 is the Thermo Scientific GeneRuler 100 bp Plus DNA ladder. Lane 6 is the PCR product for gpE and lane 7 is the negative control for gpE containing all PCR reactants, but no sample 3 plasmid. The gpE amplicon was expected to be 426 base pairs in length (Fig. 7), and on the gpE lane there is a thick band around this size. Additionally, the gpD amplicon was expected to be 205 base pairs in length (Fig. 8) and there is a thick band at this size as well. However, the presence of bands in the negative control lanes indicates impurities and potential contamination meaning the results from this experiment are not trustworthy.*

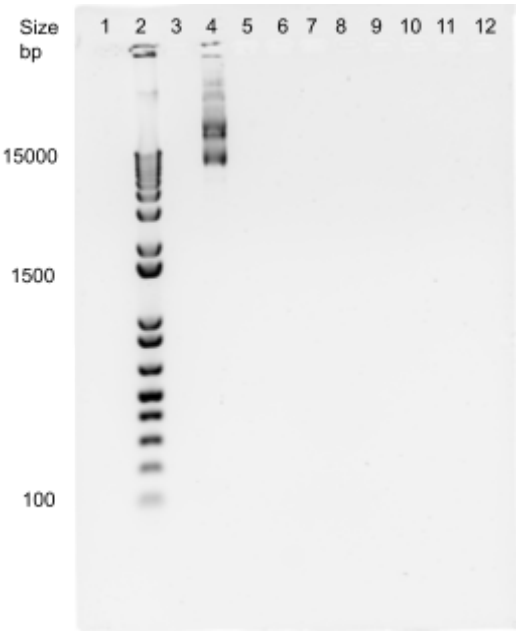

Figure 6: *Gel electrophoresis checking the purity of sample 3. Lane 2 is the Invitrogen 1 Kb Plus DNA ladder. Lane 4 is the plasmid miniprep for sample 3 and 6X loading dye. Lane 6 is the negative control containing buffer and loading dye only. Largely inconclusive.*

gpE Primer Information

PRIMER INFORMATION:
Parameter Set: General PCR (Primers only)
Sequence Name: Sequence 1
Amplicon Length: 426

	Start	Stop	Length	Tm	GC%
Forward ACAAATTCCGGGACTGGTAAA (Sense)	126	147	21	62	42.9
Reverse ATATCGTCGGTCGGGTCATA (AntiSense)	532	552	20	62	50

Figure 7: *Primer information for gpE.*

gpD Primer Information

PRIMER INFORMATION
Parameter Set: General PCR (Primers only)
Sequence Name: gpD
Amplicon Length: 205

	Start	Stop	Length	Tm	GC%
Forward CGGCTCATACCGCAACC (Sense)	53	70	17	62	64.7
Reverse ACATCCTCATAACGGAACGTG (AntiSense)	237	258	21	62	47.6

Figure 8: *Primer information for gpD.*

Discussion

There were essentially two parts to this research project with the primary focus of identifying the vectors of 22 plasmids as well as the genes that reside with them. Part one of this project was the initial plan of action, however, each step in the process yielded largely inconclusive results. After the database search, a restriction digest was carried out to test the viability of the vector backbone for samples 3, 11, 16, and 21. This experiment produced inconclusive results. It was shown that each of the experimental wells did not have clear banding so no conclusions were reached. Each well had many banding patterns that were significantly larger than expected (Fig. 3). This could point to several potential issues. The first is that the sample was contaminated and that is the reason so many bands were present in each lane. Additionally, there was zero banding for sample 16 which means the enzyme could have cut the plasmid into numerous small fragments that ran off the gel (Fig. 3). Nonetheless, it did not work and the project moved forward to identify the inserts via a PCR of the inserts within sample 3 without vector verification.

This project is tied to a larger body of work within the undergraduate research at PLNU underneath Dr. Kristopher Koudelka. Sample 3 is the plasmid used for the procapsid controls, and therefore, it was the plasmid most likely to contain the target genes for a PCR reaction. The results of the PCR reaction did show promise. The amplicon for gpD was expected to be 205 base pairs in length, and there was presence of a band this size on the gel (Fig. 5). Additionally, the expected length of the gpE amplicon was 426 base pairs long. There is also the presence of a strong band at this size (Fig. 5).

Nonetheless, the results were confounding as a similar banding pattern occurred within both of the negative controls where there should have been an empty lane (Fig. 5). This could have been a result of an impure sample or potential spillover when loading the experimental wells of the gel. A purity check of sample 3 did not yield conclusive results (Fig. 6) as there was no clear banding pattern as well as bands significantly larger than expected.

All expected capsid genes found in 5 of 13 plasmids sequenced

For 5 of the 13 plasmids sequenced using whole plasmid sequencing, all of the expected procapsid genes were present. This was true for samples 2, 6, 10, 13, and 14. Sample 2 was expected to have no procapsid genes as it was an empty pGEX-6P-3 expression vector. Sequence analysis revealed this to be true with no presence of capsid genes in the transcription unit (Fig. 2A). Sample 6 was expected to have the presence of one capsid protein, gpE, and sequencing revealed the presence of gpE alone in the transcription unit of this plasmid (Fig. 2B). Sample 10 was expected to have gpE alone in the transcription unit of this plasmid. Sequencing analysis showed that this was true with the presence of gpE alone in the tran of this plasmid (Fig. 2D). Sample 13 was expected to have one capsid protein within its transcription unit. It was expected that gpE would be the only capsid protein present in this plasmid and sequencing analysis revealed this to be true (Fig. 2E). Finally, sample 14 was expected to have the presence of gpNu3 as the only capsid protein in the transcription unit. Sequencing analysis also revealed this to be true with the capsid assembly protein (gpNu3) within the transcription unit (Fig. 2F).

Resultantly, research using these plasmids can resume with confidence in normal use as the genes present were the genes expected.

Presence of gpNu3 not seen in pT7cap

Based on the literature search of articles referencing these samples, it was revealed that pT7cap is a plasmid that should contain all five capsid protein genes.[2] Whole plasmid sequencing shows that there are only four capsid proteins present with gpNu3 being absent (Fig. 1). Each of samples 3, 4, 7, 11, 12, and 20 were labeled at pT7cap+E. This means all five proteins should be expected, yet each only has gpE, gpD, gpC, and gpB. The scaffolding protein gpNu3 is necessary for proper procapsid formation and without it, the capsid proteins form into abnormal and non-functional structures.[2] The undergraduate research has not produced expected results for the *in-vitro* formation of procapsids. The absence of gpNu3 could be part of the reason for these issues as it is necessary for proper procapsid formation.

Samples 8 and 23 yielded very unexpected results

For all of the other 11 plasmids, each contained exactly or very close to what was expected for each. However, sequencing revealed that plasmids 8 and 23 were not at all what they were presented to be. The formal name for sample 8 is pT7Init-gpC (S166A). Based on previous research, pT7Init-gpC (S166A) was said to contain gpB, gpNu3 and a mutant/nonfunctional gpC.[2] Sequencing of this plasmid shows that there is a mutated gpC present as there is only a fragmented gpC gene in the transcription unit. However, there is no gpB or gpNu3 whatsoever (Fig. 2C). Therefore, it can be concluded that this plasmid is largely of no use for current research purposes. The current research is focused on the *in-vitro* formation of procapsids for which a non-functional gpC is not needed. Furthermore, as there are no other capsid genes present on this plasmid, there is not much use for it currently. It was likely mislabeled or the lab received something that was not what it was expected to be.

Nonetheless, this could prove beneficial for future research and should be kept.

Sample 23 was the most surprising of all. The name of this plasmid is Procapsid pT7cap construct in BL21 cells Dam (no D) (Tab. 1). With this label, it is expected that there should have been gpD, gpE, gpNu3, gpC, and gpB, however, this was not the case. Sequence analysis showed that sample 23 contains only gpE and is almost an identical copy of sample 13 (Fig. 2G,E). Any research or experimentation with this plasmid would have been misguided and unlikely results are a direct byproduct of the plasmid not containing any of the genes that it says it contains. This must be noted and relabeled so that future research can be conducted with the right information at hand. Furthermore, it is extremely similar to sample 13 but they are not the same. Sample 13 was 6009 base pairs in length whereas sample 23 is 6001 base pairs in length (Tab. 6).

Further research required to verify backbones of plasmids

A BLASTn sequence analysis for all but 2 of the plasmid backbones produced matches that were not what was expected for the plasmid vectors. Samples 2 and 10 were the only backbones that aligned with the prediction (Tab. 6,7). All other sequences did not match the expected results. Results for this are all listed in the results section. Further research should be conducted to verify these backbones as it is necessary to identify the vectors to have full confidence in the plasmids used for research.

Sequencing of other 9 samples necessary to ensure confidence in plasmids

The analysis of only 13 of the 22 plasmids used for this research lab revealed that many or most of the plasmids are untrustworthy. Problems regarding the research are likely tied to misinterpretation of the plasmids and using samples that are not exactly what they are presented to be. It should be advised to continue research with the 13 sequences analyzed in this study until whole plasmid sequencing can be conducted on the remaining 9 plasmids. Entire plasmid sequencing was initially thought to be a very costly option, but whole plasmid sequencing via Eurofins Genomics LLC shows promising results at an affordable price point.

Acknowledgements

I would like to thank Dr. Kristopher Koudelka (Point Loma Nazarene University) for providing the plasmid samples for experimentation as well as the idea for this overall research project. I would like to thank Dr. David Cummings (Point Loma Nazarene University) for advising this project and providing unprecedented guidance. I would also like to thank Dr. Heidi Woelbern (Point Loma Nazarene University) for serving on my committee as well as providing unprecedented guidance. I would also like to thank my colleagues Katelyn Davis (Point Loma Nazarene University) for facilitating experimentation and data analysis and Denisse Avendano (Point Loma Nazarene University) for facilitating data analysis.

References

1. Medina, E. M., Andrews, B. T., Nakatani, E., & Catalano, C. E. (2011). The bacteriophage lambda gpNu3 scaffolding protein is an intrinsically disordered and biologically functional procapsid assembly catalyst. *Journal of molecular biology*, *412*(4), 723–736. https://doi.org/10.1016/j.jmb.2011.07.045
2. Medina, E., Wieczorek, D., Medina, E. M., Yang, Q., Feiss, M., & Catalano, C. E. (2010). Assembly and Maturation of the Bacteriophage Lambda Procapsid: gpC Is the Viral Protease. Journal of Molecular Biology, 401(5), 813–830. https://doi.org/10.1016/J.JMB.2010.06.060
3. InterPro. (n.d.). Retrieved March 5, 2023, from https://www.ebi.ac.uk/interpro/protein/UniProt/P08515/
4. Yang, Q., & Catalano, C. E. (2003). Biochemical characterization of bacteriophage lambda genome packaging in vitro. *Virology*, *305*(2), 276–287. https://doi.org/10.1006/viro.2002.1602