6-2021

Non-Ribosomal In Vivo Protein Ligation

Sarah Desautel

Kenneth Joseph

Victorino Miguel Francisco

SANTA CLARA UNIVERSITY

Department of Bioengineering

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Sarah Desautel, Kenneth Joseph, Victorino Miguel Francisco

ENTITLED

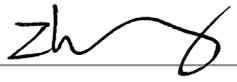
NON-RIBOSOMAL IN VIVO PROTEIN LIGATION

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF SCIENCE
IN
BIOENGINEERING

06/09/2021

Thesis Advisor — Date



06/09/2021

Department Chair — Date

NON-RIBOSOMAL IN VIVO PROTEIN LIGATION

By

Sarah Desautel, Kenneth Joseph, Victorino Miguel Francisco

SENIOR DESIGN PROJECT REPORT

Submitted to
the Department of Bioengineering

of

SANTA CLARA UNIVERSITY

in Partial Fulfillment of the Requirements
for the degree of
Bachelor of Science in Bioengineering

Santa Clara, California

2021

ABSTRACT

Protein ligation is a process integral to many biomedical applications, however the current industry standard, expressed protein ligation (EPL), is highly error prone and time-and-cost inefficient. Our team sought to improve this integral process by proving the efficacy of a new method of protein ligation using the transpeptidase Sortase A. Sortase A is known to efficiently bind two peptide fragments so long as they are attached to LPXTG motifs, which signal and activate the enzyme's ligation abilities. Previous studies affirm Sortase A's binding ability and efficacy in *in vitro* situations, however none have yet confirmed its abilities *in vivo*. Operating *in vivo* would not only allow us to take advantage of Sortase A's binding abilities, but also the mammalian expression system's own protein editing mechanisms, namely its high fidelity protein synthesis and post translational modification capabilities. This proof-of-concept project used Green Fluorescent Protein (GFP) fragments as a simple but qualitative method of measuring Sortase A's in vivo binding abilities, causing the cells to fluoresce if the introduced GFP fragments were successfully bound together. Our team found that Sortase A does in fact have the capacity to operate *in vivo*, signified by fluorescence in our test subject and further confirmed by our positive and negative controls. These findings indicate a bright future for Sortase A's role as a protein ligate, and a basis for future studies on how to further optimize and regulate its abilities to further improve protein ligation and its many applications.

ACKNOWLEDGEMENTS

We would like to thank the School of Engineering for funding our research, as well as Dr. Zhang for his financial and technical support throughout this project. We would also like to thank graduate student alumnus, Akhil Choppa, who contributed to the plasmid and protocol design the previous year, as well as current graduate students, Melina Huang and Lisa Jin, who provided extra hands in the lab. We also want to acknowledge the help of current student, Ben Tan, for his assistance in the lab during Covid closures and his contributions to the project in the previous year. Lastly, we would like to thank the lab managers, Daryn Baker and Elizabeth Johnson, for their guidance and help with inventory and lab maintenance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Expressed Protein Ligation (EPL) | Cell Wall Sorting Signal (CWSS) | Green Fluorescent Protein (GFP) | Enhanced Green Fluorescent Protein (eGFP) | Chinese Hamster Ovary (CHO) | Fetal Bovine Serum (FBS) | Phosphate-Buffered Saline (PBS) | Dulbecco's Modified Eagle Medium (dMEM) | Fluorescence Activated Cell Sorter (FACS) | Antibody Drug Conjugate (ADC)

Chapter 1: Introduction

1.1 Background and Motivation

Protein ligation is a pivotal process within a variety of biomedical applications, providing researchers a path through which they can produce or link synthetic proteins for therapeutic and research purposes. The ability to create proteins with specific functionalities is key to producing more advanced therapeutics to treat the complex diseases plaguing our society today. Additionally, the ability to link protein tags allows researchers to better track and understand the functional processes that take place within the human body, giving scientists a means through which they can explore and learn about human biology in previously impossible ways. As such an integral process to the development of medicines and understanding of the body, protein ligation methods must be as robust as possible, both to ensure a product's quality and to allow for more complex innovation. Our team identified current methods as functional, but nowhere near the level of efficiency or specificity necessary to drive forward scientific innovation, motivating us to look for improvements and alternatives to the current industry standards.

1.2 Industry Standard

The current industry standard to ligate proteins is Expressed Protein Ligation (EPL). EPL relies on inteins, which are proteins known for removing themselves from larger protein sequences and binding the adjacent protein segments together. In this way, inteins can also ligate two proteins together in vitro [1]. Though EPL is commonly used, it is flawed for several reasons. For one, there is a low final product yield after purification. This is because the intein binds proteins together but fails to fold the protein correctly, rendering it useless. The chemical mechanism of EPL also creates extra cysteine residues on the final protein, which increases the likelihood of side reactions and further decreases purity. In addition, EPL requires large amounts of reagents to complete the reaction; since inteins are not enzymes, the reaction requires a 1:1 ratio of protein to intein to occur. Lastly, EPL is only able to semi-synthesize proteins, rather than create novel proteins. This means that it is able to bind two pre-existing proteins together but cannot link smaller peptide fragments to create a new protein. Overall, EPL has much room for improvement in the areas of product yield, product purity, user ease, and sustainability [2].

1.3 Chemical Mechanism

Sortase A is a transpeptidase enzyme naturally found in gram-positive bacteria cells, originally found in *S. aureus* [3]. This enzyme has two functions: attach proteins to the peptidoglycan cell wall and create pili that project from the bacterial surface. These two reactions follow a similar process and require an LPXTG motif, part of the cell wall sorting signal (CWSS) [4]. The LPXTG motif is found in many surface proteins in these bacterial cells, making it an ideal signal for sortase A. The full CWSS contains the LPXTG motif on the N terminus, a hydrophobic domain, and a positively charged tail located inside the cell. This tail is critical for keeping the protein anchored in the cell wall [5]. The Sortase A prefers the leucine, proline, threonine and glycine in the first, second, fourth, and fifth positions respectively to optimize kinetics. Sortase A has no preference for the amino acid in the fourth position [6]. Variants of the LPXTG motif have only been recorded at the X and T positions; T has been replaced by A and X by D, E, A, N, Q, or K [7].

The Sortase A mechanism forms as a result of linking a bacterial membrane peptide to a lipid, as seen in **Figure 1.** Sortase A first recognizes the LPXTG motif on the CWSS of the surface protein by performing a nucleophilic attack to the carbonyl carbon between the threonine and glycine residues [6]. This results in the formation of a tetrahedral intermediate that quickly reduces to a thioacyl intermediate [4]. The second peptide, lipid II, is then recognized by sortase A. This substrate performs the second nucleophilic attack to the same carbonyl carbon, resulting in an unstable tetrahedral formation via the formation of an amide bond. Sortase A is then kicked out, resulting in the formation of the final product: protein - lipid II [8]. This is then incorporated into the peptidoglycan cell wall through other cell wall reactions [4].

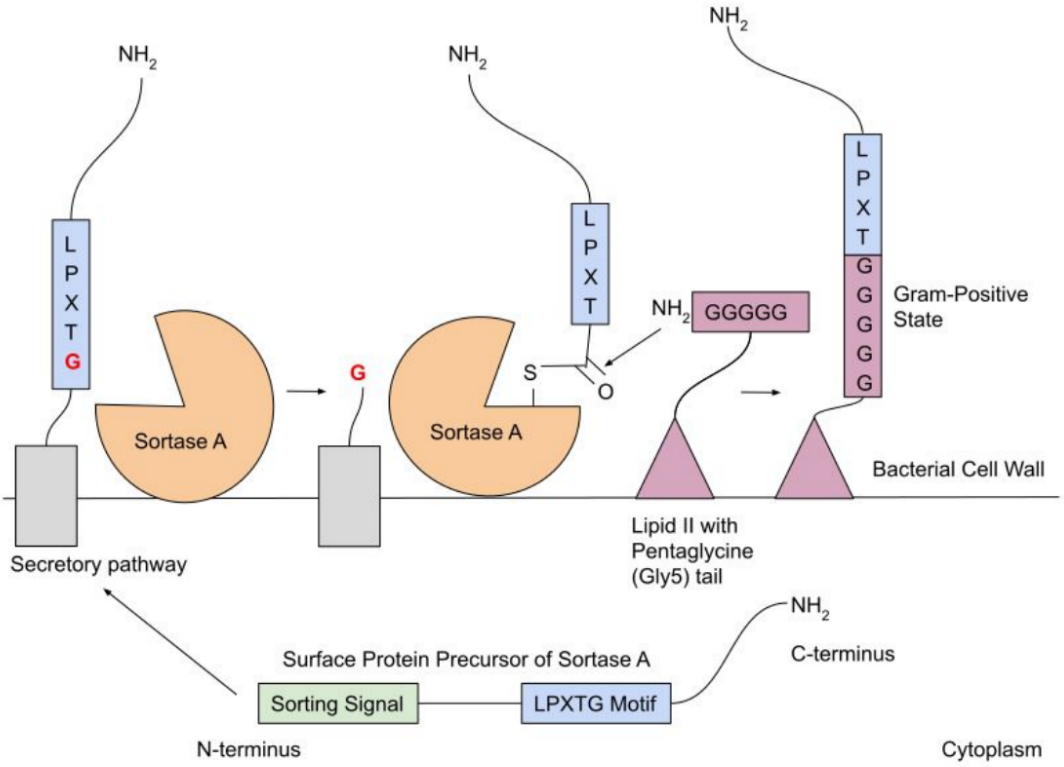


Figure 1. Sortase A mechanism in gram-positive bacteria cells (Picture obtained from Choppa 2020). The surface protein precursor of sortase A is created, and is guided to the secretory pathway using the sorting signal. The protein precursor is then embedded into the cell wall, where the LPXTG motif is recognized by Sortase A. The first nucleophilic attack results in the formation of a Sortase A - peptide intermediate, leaving the result of the peptide in the cell wall. Lipid II then performs a nucleophilic attack on the intermediate, resulting in the cross-link of a peptide-lipid that is incorporated into the bacterial cell wall [8]

1.4 Current Applications

Bioconjugation has been explored as another function of Sortase A to replace current industry standards. Sortase A has been able to conjugate two peptides, a peptide and a protein, and two proteins together *in vitro* [10]. The enzyme has been used as a cross linker in fields like tissue engineering, which require high specificity and efficiency [11]. Other non-protein molecules have been able to be conjugated as well, including vitamins such as folate through the addition of a triglycine group [10]. This functionality lends itself to applications such as molecular sensing, surface modification, and biomaterials [12]. While many studies have focused on *in vitro* applications of Sortase A, very few have achieved success with *in vivo* applications.

Sortase A has been found to tag certain proteins and substrates together *in vivo*, using a method called sortagging [13]. Sortagging is a method of labeling specific proteins using small probes of less than 2 kDa [14]. Although this use of Sortase A shows promise in understanding protein function, the applications of Sortase A can be expanded further.

1.5 Click Chemistry

Click chemistry is also known as a group of fast reactions that are easy to perform, high yielding, easily purifiable, and versatile [15, 16]. Click chemistry has many applications ranging from materials to pharmaceutical science, and also has applications in proteomics and nucleic research. One of the biggest downsides of click chemistry is most mechanisms require copper as a catalyst [17, 18]. The use of copper in click chemistry is trivial for practical applications *in vivo*, as copper can be cytotoxic and cause cell death [19]. However in applications where copper is not needed as a catalyst, click chemistry is able to offer powerful tools for labeling *in vivo* [20]. Still there is a need to overcome these limitations of click chemistry and offer an alternative method of bioconjugation *in vivo*.

1.6 Bioconjugation of Sortase A

In order for Sortase A to properly function *in vivo*, the sortase A enzyme, the LPXTG motif, and the signal probe, usually a pentaglycine, need to be expressed inside mammalian cells [9]. This mechanism will be similar to its mechanism inside bacterial cells; the reaction will need to be mediated by two nucleophilic attacks to produce a fully functional protein as seen in **Figure 2**.

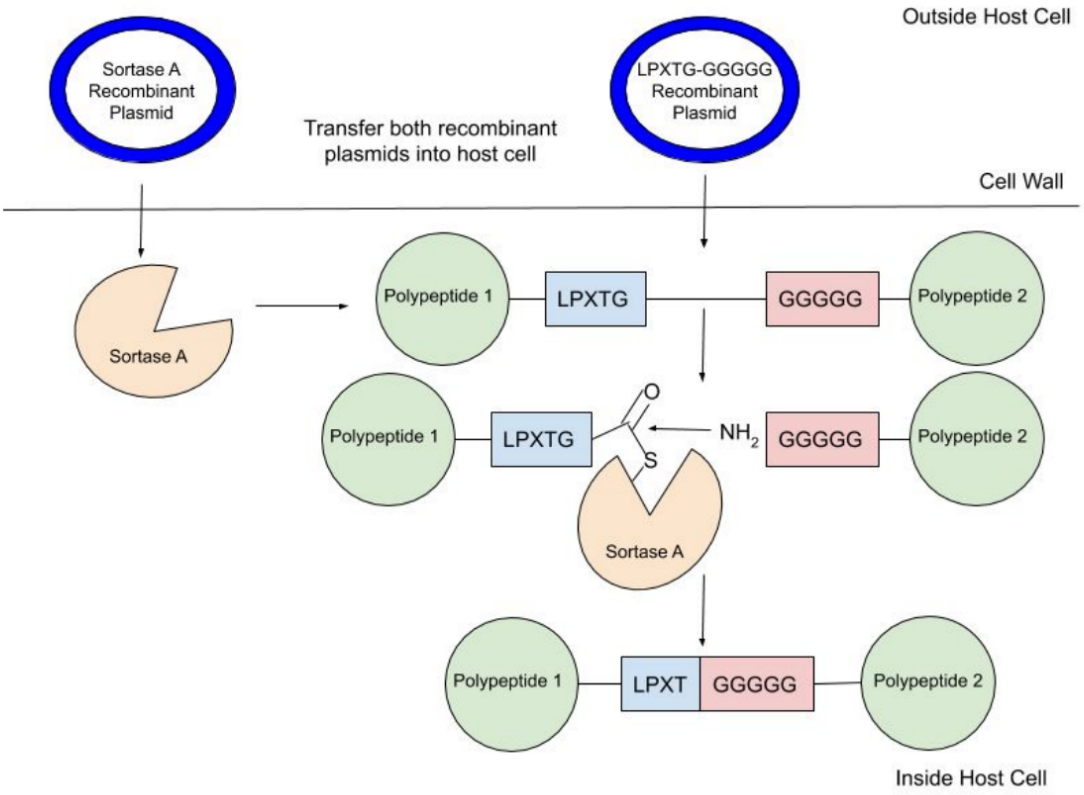


Figure 2. Sortase A mechanism *in vivo* (Picture obtained from Choppa 2020). This proposed mechanism will require two plasmids, one coding for the Sortase A enzyme, and another for the two peptide halves. Sortase A will attack the LPXTG motif of the first peptide, resulting in a peptide-sortase A intermediate. The pentaglycine on the second polypeptide will then nucleophilically attack the intermediate, resulting in the formation of a fully functional protein [8].

1.7 Project Goal

The goal of our project is to create a new method of *in vivo* protein ligation using the transpeptidase enzyme, Sortase A, to bind peptide fragments together to create a fully functional protein. This method should improve upon the current industry standards for protein ligation to create a method that is faster, less complex, and produces higher yields of product. To prove the success of this new method, a plasmid containing two nonfunctional fragments of green fluorescent protein (GFP) and a plasmid encoding for Sortase A will be introduced into the cell. Sortase A will respond to a particular motif within the GFP peptide fragments and ligate the two sequences together to create a fully functional GFP molecule. Success of the experiment will be

monitored qualitatively using a fluorescent microscope and quantitatively using a Fluorescent Activated Cell Sorter (FACS).

1.8 Objectives and Expected Results

Several experiments and cell types are needed to confirm Sortase A's ability as a protein ligase. First, the plasmids containing Sortase A, enhanced GFP (eGFP), and the GFP peptide fragments are transformed into e.coli cells. After transformation, the bacterial cells are scaled up to a larger volume. These experiments are to amplify the plasmids before mammalian cell work. Next, the plasmids are extracted from the bacterial cells and purified using a procedure called miniprep. The purity of our miniprep results is confirmed using gel electrophoresis. In this experiment, the expected results are two bands around the designed plasmid length, indicating pure plasmid DNA. If additional bands appear, this indicates the presence of contamination, such as genomic DNA or cellular debris. The next half of the experiments are performed using mammalian cells. HeLa cells are thawed and cultured to create a healthy stock of cells for the experiment. Once cells recover from thawing, they are transfected with the purified plasmid DNA. Transfected cells are monitored every 24 hours for three days using the fluorescent microscope. The well containing eGFP is expected to fluoresce, as well as the test wells containing normal to high affinity fragments of GFP and Sortase A. The wells containing isolated GFP fragments or isolated Sortase A are not expected to fluoresce, since the experiment requires the combination of peptides and Sortase. Lastly, a well containing a mix of Sortase A and low affinity GFP peptides is used as another negative control and is not expected to fluoresce.

1.9 Back-Up Plan

The original plan for the project consisted of two main parts: proving a new method of *in vivo* protein ligation and proving that Sortase A can be controlled for selective and time-dependent use. The first half of the project was expected to take the first four months to complete. First, the plasmids would be transformed into bacterial cells and amplified. Then, the plasmids would be extracted using Miniprep and confirmed pure using gel electrophoresis. Next, the plasmids would be transfected into mammalian cells and monitored for fluorescence using a fluorescent microscope and FACS machine. Lastly, a Western Blot would be used to confirm the success of the experiment and characterize the creation of GFP within the cell.

The next set of the project was expected to take the last four months to finish. After proving the success of the experiment, a new plasmid containing a genetic circuit for the regulation of the Sortase A gene in response to light would be incorporated into the cell via the same set of experiments. After incorporating the genetic circuit into mammalian cells along with the test plasmids, the cells would be exposed to light and monitored for fluorescence in a time-specific way.

Given the state of the pandemic before the project began, it was decided that the back up plan for any time constraints would be to focus on the first half of the experiment and save the genetic circuit for future iterations of the project. Due to lab closures from the Covid-19 pandemic, the expected eight to nine month research period was shortened to two months. Because of extreme loss of time, the back up plan was cut further, eliminating the opportunity for FACS and Western Blot analysis.

1.10 Significance

If this new method is proven successful, it can have many potential applications in research and drug development. In basic research, this procedure can improve upon current industry standards to create a method that is faster, simpler, and higher yielding. It can also be used to develop novel proteins and allow scientists to better understand basic protein function *in vivo*. Similarly, this method can help scientists develop non-natural protein tags containing unnatural amino acids or prosthetic groups, which cannot be encoded for in typical plasmid vectors. Lastly, if proven successful, this method can be used to ligate any biomolecule-not just proteins. DNA, lipids, and sugars are all candidate molecules that can be edited down the road for research and further development of biomolecules.

In the realm of drug design, this new method can be used to aid in selectivity. If Sortase A expression is regulated by a genetic circuit, ligation (activation) can be controlled using a stimulus like light or radiation. This method can also be used to synthesize antibody drug conjugates (ADCs) *in vivo*. In this case, Sortase A can bind a drug and antibody together, creating a molecule that is better suited to locate the proper target and eliminate off-target effects. Overall, the application of Sortase A as a protein ligase can help improve many areas of drug

development, including cancer and immuno- therapy, by creating more selective and effective drugs.

1.11 Team and Management

Our team equally shared the responsibilities of executing and optimizing experimental procedures, maintaining bacterial and mammalian cell lines, collecting and analyzing data, and communicating our conclusions with our advisor and peers. Dr. Zhang provided input and guidance during weekly meetings, and graduate students Melina Huang and Lisa Jin, as well as bioengineering student Ben Tan provided additional lab support in the execution of our bacterial cell procedures.

1.12 Budget

Our team was granted a total of $1500 to use towards research supplies. Our needed supplies exceeded the budget by approximately $100, but we were able to work around budget constraints by sharing disposables, such as petri dishes, well plates, and collection tubes, with another group from the same lab.

Table 1: Project Materials and Budget

Materials	Quantity	Price
Agarose Powder	25 g	In Storage
TAE Buffer	1 L	In Storage
Sample Loading Buffer	3 x 0.5 mL	In Storage
DNA Ladder (1 kb)	0.5 mL	In Storage
LB Broth Media Powder	500 g	In Storage
One Shot™ TOP10	11 x 50 µL/tube	$210.00

Competent Cells (10^8)		
Luria Agar	100 g	In Storage
Ampicillin	5 g	$26.00
Kanamycin	5 g	$26.00
Tetracycline	25 g	$26.00
Petri Dishes	20	$11.99
Opti-mem Reduced Serum Medium	500 mL	$52.50
HeLa Cells	1	$495.00
dMEM, High Glucose, GlutaMAX Supplement	2 x 500 mL	$65.06
FBS, Heat Inactivated	50 mL	$84.75
PBS, pH 7.4	500 mL	$32.40
Trypsin-EDTA (0.25%) Phenol Red	100 mL	$14.14
DMSO	100 mL	$52.25
Conical Vials (10 or 15 mL)	50 Tubes	$17.99
Eppendorf Tubes (2 mL)	500 Tubes	$17.45
6-Well Plate	10 Plates	$13.00

ColumnPure MiniPrep Kit	1	In Storage
Fugene6 Transfection Reagent	1 mL	$435.00
	Total Cost:	**$1579.53**

1.13 Timeline

The timeline of our experiment was greatly reduced due to the initial lab closure, as well as additional, unexpected lab closures throughout the year. This resulted in a total experiment time of two months in the wet lab. Our original plan was described in Section 1.8, however we had to use our backup plan in order to adjust to the accelerated time frame. Refer to **Appendix A** for our updated timeline.

Chapter 2: Bacterial Transformation

2.1 Design Description

Bacterial cell culture is widely used in biomedical research, ranging from vaccines to antibody generation [21,22]. Bacterial cells are often used for the production of simple, non glycosylated proteins and proteins that don't require post-translational modifications. This project relies on a mammalian system because the proteins require post-translational modifications, however, a bacterial expression system is used purely for plasmid amplification. Mammalian cell transfection requires a particular ratio of plasmid DNA to cell density, but bacterial cells only require one plasmid in order to uptake and amplify the plasmid DNA. This makes bacteria the optimal expression system to amplify plasmids prior to mammalian transfection and expression.

2.2 Key Constraints

The key constraints included optimizing our procedures and following the same protocol for each plasmid design. Antibiotic concentrations are also critical in ensuring only our transformed bacteria are able to survive the cell culture. Low yields or improper purification of the plasmids would push back mammalian transfection by a few days, which is critical as months of research time were lost from COVID-19.

2.3 Expected Results

With proper bacterial transformation, the expectation is to see healthy and isolated colonies on the agar plates, cell culture flasks with an OD600 reading greater than 0.6 after a day of growing, and two bands per plasmid after gel electrophoresis analysis. More than two bands in gel electrophoresis analysis indicates contamination.

2.4 Back-Up Plan

For plasmid purification, LAMDA Mini-Prep kits will be ordered to meet budget requirements. In case this brand of Mini-Prep kits are unsatisfactory, stored Qiagen Mini-Prep kits from 2017 are available to use.

2.5 Materials and Methods

2.5.1 Plasmid Design

The three recombinant plasmids (410, 411, 412) were designed by a previous team using ApE software as seen in **Appendix B-D** [9]. Their full plasmids can be found in **Appendix E-G.** Plasmids 88 and eGFP were previously created and stored by the Bioengineering Department as seen in **Appendix H and I**. The recombinant plasmids were designed for optimal expression in mammalian cells. Plasmids were amplified using Shot TOP 10 competent cells and purified using a Qiagen Miniprep kit. Chinese hamster ovary (CHO) cells were passaged to a confluency of 80% and transfected using Fugene6 transfection reagent. Our transfection consisted of 8 wells: 5 wells containing each plasmid separately to use as a control, and 3 wells containing a combination of Sortase A with either 410, 411, or 412. After transfection, CHO cells were observed every 24 hours for fluorescence.

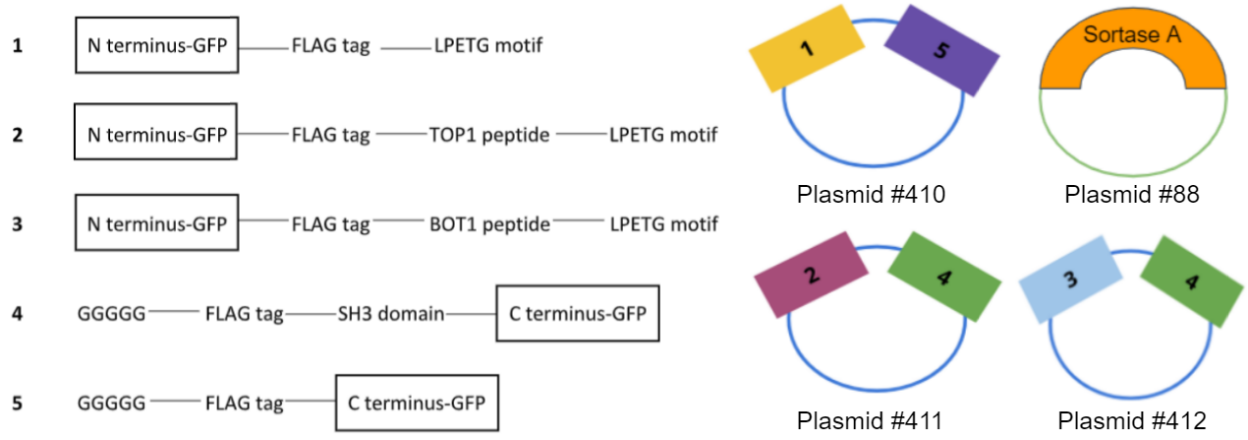


Figure 3. Simplified design of Plasmids. Plasmid 88 contains the gene that encodes for Sortase A [9].

Figure 3 shows a simplified version of the plasmids. Labeled 1 - 5 are the peptide fragments that will be included in the plasmids. The first three peptide fragments show the N-terminus of GFP. All have a FLAG tag that will aid in Western Blot characterization at the end of the experiment, as well as the LPETG motif that signals Sortase A. Fragment 2 has a TOP 1 peptide, which is an epitope designed to have high affinity binding to the Sortase A intermediate's SH3 domain. Fragment 3 has a BOT 1 peptide, which is an epitope designed to have poor binding affinity to the SH3 domain. These epitopes were found by the previous team using BioAI on an epitope

database. Plasmid 410 is used as a negative control, or basal control, for efficiency. Because it has unaltered binding, it should fluoresce at a dim level. Plasmid 411 is the test plasmid. It is expected to show medium to high fluorescence because of its high affinity binding. Plasmid 412 is another negative control and should not fluoresce due to poor binding affinity. Lastly, the plasmid 88 is the gene that encodes for Sortase A. This will allow the cell to produce the enzyme and facilitate the reaction.

2.5.2 Amplification of Recombinant Plasmids

BL21 Competent *E. coli* cells were chosen for their efficiency and resiliency. The plasmids to be amplified were sourced from last year's iteration of this project, as the previous group was able to generate and store the plasmid stock before COVID lab shutdowns.

Luria Agar plates were premade by the biology department and over-plated with antibiotics for cell selection. Any cells that do not successfully uptake the plasmids should die, while the cells that contain plasmids should live, as the plasmids contain antibiotic resistance genes. Plasmids 410, 411. 412, and Sortase A contain ampicillin resistance. The eGFP plasmids contain kanamycin resistance. To overplate, a 4 mg/mL stock solution of ampicillin and kanamycin were made by adding 0.8 grams of antibiotic powder to 200 mL LB broth. 250 uL of the correct antibiotic were added to each corresponding agar plate. The antibiotic broth was spread around the surface of the plates using sterile technique. The plates were incubated at room temperature for at least 30 minutes with the lid on until the liquid was fully absorbed.

For bacterial transformation, five BL21 competent cell vials were organized, one for each plasmid. 1 uL of plasmid 88 was added to a single vial. 1 uL of eGFP plasmid was added to another vial, 1 uL of plasmids 410, 411, and 412 were added to their respective vials, then incubated on ice for 30 minutes. The vials were heatshocked for exactly 40 seconds at 42°C without shaking, then placed back on ice for an additional 2 minutes. Following the incubation period, 250 uL of room temp LB Broth media was added to each vial, then capped tightly and placed in the incubator for 1 hour at 37°C at 225 rpm. The contents of each vial were then plated on respective Agar plates and incubated at 37°C for 12-16 hours.

The scale up procedure involved the 5 e.coli plates from transformation, micropipette and tips, plastic vials, 500 mL Erlenmeyer flasks, and LB Broth media, and was conducted entirely under

a bunsen burner using sterile technique. 1 mL of LB Broth was added to five vials, one corresponding to each *E. coli* plate/plasmid. A micropipette tip was used to remove a single colony from each *E. coli* plate and place it in the corresponding vial. The vials were then placed in the shaker incubator at 37°C and 225 rpm for 4 hours. After 4 hours, each of the 5 Erlenmeyer flasks were filled with 200 mL of LB Broth media with their respective antibiotic selection markers at a concentration of 50 ug/mL. The contents of each vial were then transferred to their corresponding Erlenmeyer flask, then placed in the shaker incubator at 37°C and 225 rpm for 4 hours. After this incubation period, the OD600 of each sample was taken by placing 1 mL of culture in a spectrophotometer. A sample of plain LB Broth media was used to set a blank for the machine. Each flask was expected to reach an OD600 reading >0.6. The OD600 reading for each flask is listed in **Table 2.**

Table 2. OD600 Readings after Final Incubation

Plasmid Number	OD600 Reading
88	1.928
410	1.304
411	0.620
412	0.736
eGFP	0.627

2.5.3 Purification of Recombinant Plasmids

To purify the plasmids, a Qiagen Miniprep kit was used. First, the Erlenmeyer flasks containing the final concentration of transformed *E. coli* were removed from the 4 °C refrigerator and swirled in the shaker incubator for two minutes. Once the cells were resuspended, each culture was harvested by transferring 1 mL into a 1.5 mL microcentrifuge tube. The cells were then

centrifuged at 8000 rpm for 3 minutes. The supernatant was removed and each bacterial pellet was resuspended in 250 µL of Buffer P1. 250 µL of Buffer P2 was added and each tube was inverted gently 8 times. 350 µL of Buffer N3 was added and each tube was inverted gently 8 times. Each resuspension was centrifuged at 12000 rpm for 10 minutes. 800 µL of the supernatant was removed and transferred to a spin column, and centrifuged at 13000 rpm for 1 minute. 750 µL of buffer PE was added to the spin column and centrifuged at 13000 rpm for 1 minute. Once again, 750 µL of buffer PE was added to the spin column and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and the spin column was centrifuged at 13000 rpm for 1 minute. Each spin column was transferred to a sterile 1.5 mL microcentrifuge tube. 50 µL of Buffer EB was added to the spin column and sat for 1 minute before being centrifuged at 13000 rpm for 1 minute. Each centrifuge step was run at 15 °C.

A larger scale miniprep procedure was used for flasks with a lower OD600 reading (below 1). The Erlenmeyer flasks containing the final concentration of transformed *E. coli* were removed from the 4 °C refrigerator and swirled in the shaker incubator for two minutes. Once the cells were resuspended, each culture was harvested by transferring 5 mL into a 10mL centrifuge tube. The cells were then centrifuged at 6000 rpm for 10 minutes. The supernatant was removed and each bacterial pellet was resuspended in 250 µL of Buffer P1. Each resuspension was transferred to its own sterile 1.5 mL microcentrifuge tube. 250 µL of Buffer P2 was added and each tube was inverted gently 8 times. 350 µL of Buffer N3 was added and each tube was inverted gently 8 times. Each resuspension was centrifuged at 12000 rpm for 10 minutes. 800 µL of the supernatant was removed and transferred to a spin column, and centrifuged at 13000 rpm for 1 minute. 750 µL of buffer PE was added to the spin column and centrifuged at 13000 rpm for 1 minute. Once again, 750 µL of buffer PE was added to the spin column and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and the spin column was centrifuged at 13000 rpm for 1 minute. Each spin column was transferred to a sterile 1.5 mL microcentrifuge tube. 50 µL of Buffer EB was added to the spin column and sat for 1 minute before being centrifuged at 13000 rpm for 1 minute. Each centrifuge step was run at 15 °C.

To check the DNA concentration of the final product from the minipreps, 1.2 µL of the final solution was added to a nanodrop device. A blank was used containing 1.2 µL of the Buffer EB.

The final concentrations of plasmid 88, plasmid 410, plasmid 411, plasmid 412, and plasmid eGFP from the Miniprep are listed in **Table 3**, and from the Large Scale miniprep in **Table 4**.

Table 3. DNA Concentrations of Plasmids after Miniprep

Plasmid Number	Trial 1 (ng/µL)	Trial 2 (ng/µL)	Trial 3 (ng/µL)
88	342.7	377.6	334.5
410	212.0	199.9	231.4
411	60.0	28.5	29.0
412	38.2	37.2	49.0
eGFP	16.2	23.0	17.9

Table 4. DNA Concentrations of Plasmids after Large Scale Miniprep

Plasmid Number	Trial 4 (ng/µL)	Trial 5 (ng/µL)
411	98.0	80.8
412	129.5	144.9
eGFP	64.1	52.3

2.6 Results and Discussion

Figure 4. Healthy Bacterial Plate. The plate was made using luria agar broth and 50 µg/mL ampicillin concentration. The plate has isolated colonies, indicating that the *e. Coli* cells have been properly transformed with the plasmid DNA and were plated properly. The other four plates displayed similar results to this plate with Plasmid 411.

After plating each of the different plasmids and incubating them overnight, each plate was covered in bacterial colonies as seen in **Figure 4**. Each plate was healthy, as colonies were circular, light yellow, and for the most part isolated. Every plate was covered in over a hundred colonies, except for plasmid 88 which only contained three colonies on the plate. The differences in these numbers can arise from a number of factors, such as differences in the vector design of Sortase A and the GFP recombinant plasmids (410 - 412).

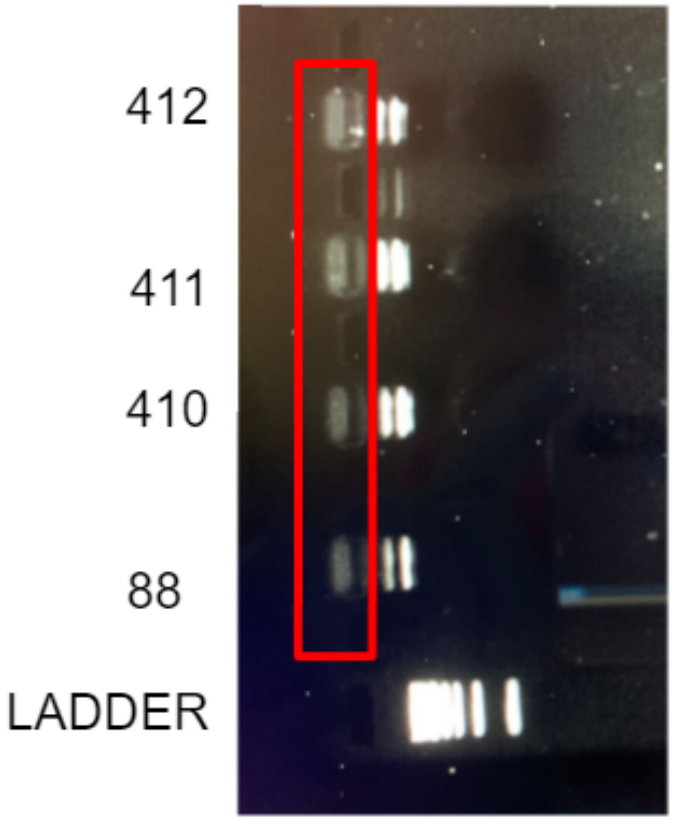


Figure 5. Gel Electrophoresis after Plasmid Purification using LAMDA Mini-Prep Kits. The lamda mini-prep kits confirmed that each of the bacterial cultures were properly transformed and contained the desired plasmid DNA. This purification also contained genomic DNA, shown by the bands in the red box.

Initial plasmid purification was done using the LAMDA Mini-Prep Kits in an effort to meet the budget requirements and acquire newer reagents. The Qiagen kits in storage were obtained back in 2017, and purchasing a new kit was significantly more expensive than other alternatives. The first trial with the LAMDA kits revealed genomic DNA, as seen in **Figure 5**. The second trial of Mini-Prep was run using an optimized LAMDA protocol run at the same time as the Qiagen kit in storage. The Qiagen kit ended up showing more promising results after gel electrophoresis analysis, and was used in further plasmid purification steps.

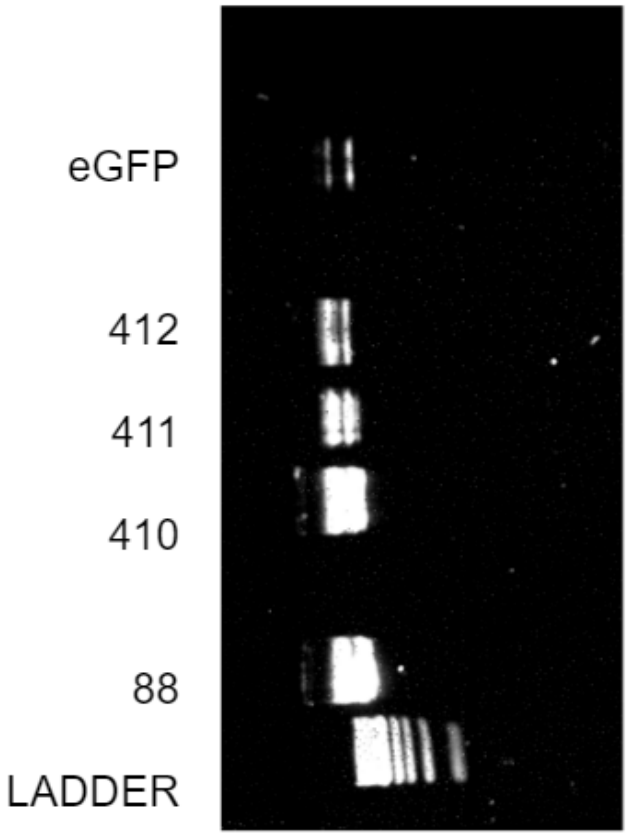


Figure 6. Gel Electrophoresis after Plasmid Purification using Qiagen Mini-Prep Kits. The Qiagen kits showed the two bands as expected from the plasmid designs, and did not show any genomic DNA after optimization.

To optimize the Qiagen protocols for plasmid purification, a second wash step was implemented. This allowed for the removal of genomic DNA from the final eluted product, as seen in **Figure 6.** These finalized protocols were implemented for three to five trials depending on the DNA yield obtained. Once enough plasmid DNA was obtained, the DNA was stored in a 4 °C refrigerator until mammalian cells were ready for transfection.

Chapter 3: Mammalian Cell Culture

3.1 Design Description

As previously noted, the majority of this project relies on a mammalian expression system due to its ability to perform post translational modifications. This is key to evaluating the efficacy of Sortase A's binding ability, as the plasmid to be reconstructed codes for a GFP protein requiring post translational modifications. Given that the plasmids were already amplified to achieve the required amounts using bacterial expression systems, the use of mammalian expression systems is simply a vessel through which Sortase A's function can be evaluated to the best accuracy and certainty.

3.2 Key Constraints

A major constraint involved optimizing the transfection protocol, especially finding the best transfection reagent and dilution for the given cell line. The original transfection reagent used was toxic to the original cell line, so both the cell line and transfection reagent were replaced. CHO cells were used instead of HeLa variants and Fugene6 transfection reagent was used instead of Lipofectamine 2000. The required ratio of transfection reagent to plasmid DNA also varies depending on the cell line and project, so additional optimization experiments are required to increase efficiency and form an official protocol. Lastly, due to COVID-related campus shutdowns, we had to coordinate with other senior design groups to maintain the CHO cell lines in preparation for future transfection trials.

3.3 Expected Results

Alongside the continued survival and proliferation of the CHO cells, the fluorescence of transfection wells 88+410, 88+411, and eGFP were expected to confirm the success of the mechanism. The combination of Sortase A with the test plasmids will confirm that Sortase A is able to ligate peptide fragments *in vivo* to create a fully functional GFP. eGFP fluorescence is used as a positive control and will confirm the success of the transfection protocol. If eGFP is unable to fluoresce, this means the transfection procedure failed. Transfection wells 88, 410, 411, 412, and 88+412 are not expected to show fluorescence. These wells are used as negative

controls to check for contamination, self-fluorescence, and Sortase A selectivity. If these wells show fluorescence, it demonstrates either a flaw in the experimental design or execution of the transfection.

3.4 Back-Up Plan

Although HeLa cells are typically resilient, lipofectamine is a toxic transfection reagent that can kill mammalian cells. In case HeLa cells do not survive the transfection process, two additional measures will be taken: CHO cells will be kept in storage as an alternative mammalian cell option and Fugene6 will be ordered as an alternative transfection reagent.

3.5 Materials and Methods

3.5.1 CHO Cell Culture

CCL2 HeLa cells were originally chosen for our mammalian expression system because of their resilience and immortality. However, the incorrect cell line was received from the manufacturer and due to time constraints, CHO cells were used. CHO cells are a commonly used cell line as they are robust and are known to produce large quantities of recombinant proteins. The CHO cells were grown in 10 cm tissue culture plates and passaged every 48-96 hours, depending on when they reached 80% confluency. Media was checked every 24 hours and changed when it turned from a bright red to a reddish-orange color. For the first 5 passages after thawing, media consisting of 15% Fetal Bovine Serum (FBS) in Dulbecco's Modified Eagle Medium (dMEM) was added to supplement the weakened cells. The following passages used media containing 10% FBS in dMEM.

To maintain cells for long-term use, a simple passaging procedure was performed. First, 10% FBS media, Trypsin, and Phosphate-Buffered Saline (PBS) reagents were placed in a water bath at 37 °C until reaching room temperature. Old media was aspirated from the plate and the cells were washed with 1.5 mL of PBS, added to the side of the plate. The PBS was then aspirated off and 2.0 mL of new PBS was added to the side of the plate as a second wash. The PBS was aspirated off the plate and 1.5 mL of Trypsin was added dropwise, directly to the plate. The cells were placed in the incubator for 2 minutes. Then, the bottom of the plate was gently tapped to detach any remaining cells. 3mL of the 10% FBS media was added to the side of the plate to

dilute the Trypsin and prevent the cells from lysing. The liquid from the plate was collected into a 15 mL centrifuge tube and placed in the centrifuge for 3 minutes at 3000 rpm. Two new petri dishes were filled with 8 mL of fresh media. For a 48 hour passage, the cell pellet was resuspended in 4 mL of 10% FBS media. 2 mL of the final solution was added dropwise in a circular motion to each plate. The plate was then gently tilted back and forth to stimulate cell movement around the plate and prevent clumping in the center. The old plate was bleached and the two new plates were stored in the incubator.

24 hours before a transfection procedure, cells were passaged in a special way and replated in 6 well plates. Cell confluency was checked to estimate cell count on the plate, ensuring at least 6 x 10^4 cells in each well for a total of 8 wells. It was assumed that an 80% confluent 10 cm plate contained 8 x 10^5 cells. Two 6 well plates were used and each well was filled with 2 mL of Opti-mem medium. Opti-mem medium was used for transfection because it does not contain serum. Cells were passaged using the protocol listed above, up until resuspension of the pellet. Cells were resuspended in 2.6 mL of Opti-mem medium. Each well received 200 µL of resuspended cells, distributed dropwise in a circular motion. Once the cells were added, the plate was gently rocked back and forth to spread the cells and placed in the incubator for 12-16 hours. Any remaining resuspended cells were bleached.

3.5.2 Plasmid Transfection

Transfection was performed 12-16 hours after plating the cells. The well lay-out was as follows: well 1 contained plasmid 88, well 2 contained plasmid 410, well 3 contained plasmid 411, well 4 contained plasmid 412, well 5 contained plasmids 88+410, well 6 contained plasmids 88+411, well 7 contained plasmids 88+412, and well 8 contained the plasmid for eGFP. The Fugene6 transfection reagent was used as it is less toxic and more effective than the traditional Lipofectamine 2000 reagent. The transfection followed a 1:3 ratio of plasmid DNA to transfection reagent. 6 µL of Fugene6 was used along with 2 µL of plasmid DNA. The total plasmid mass was calculated from the nanodrop concentration results after the Miniprep procedure. The amounts of plasmid DNA listed in **Table 4** include the tube number of the associated round.

Opti-mem media was warmed to room temperature using a 37 °C water bath. Fugene6 was left in the biohood and warmed to room temperature. After the media was warmed up, the Fugene6 reagent was gently inverted. Next, 94 ul of Opti-mem medium was pipetted into eight 1.5 mL microcentrifuge tubes. Next, 6 µL of Fugene6 reagent was added to each tube directly into the Opti-mem medium, making sure to not let the undiluted reagent make contact with the sides of the tube. Each tube was gently pipetted up and down and sat for 5 minutes at room temperature. Next, plasmid DNA was added based on **Table 5** and mixed by pipetting up and down. The plasmid DNA-transfection reagent mixture was left to incubate at room temperature for 15 minutes. 60 µL of our plasmid DNA-transfection reagent mixture was added to each corresponding well in our 6 well plate and mixed by tilting the plate back and forth. The plates were incubated for 24 hours. After 24 hours, the old media was aspirated to remove any cell debris and replaced with Opti-mem medium and 10% CaCl2 solution. A CaCl2 wash was chosen because Sortase A relies on Ca2+ as a cofactor.

Table 5. Fugene6 Transfection Trial 1 of Designed Recombinant Plasmids

Plates	Plate 1	Plate 2	Plate 3	Plate 4	Plate 5	Plate 6	Plate 7	Plate 8
Plasmids #	88	410	411	412	88 & 410	88 & 411	88 & 412	eGFP
Plasmid Amount (ug)	2	2	2	2	1+1	1+1	1+1	2
Plasmid Amount (ul)	5.30 R2	10.0 R2	20.90 R3	19.53 R3	88R2: 2.64 + 410R2: 5.0	88R2: 2.64 + 411R3: 10.45	88R2: 2.64 + 412R3: 9.77	27.82 R3
Fugene6 (µl)	6	6	6	6	6	6	6	6

3.5.3 Fluorescent Analysis

After transfection, the CHO cells were monitored for GFP expression using a Keyence Fluorescence Microscope. Cells were observed for a fluorescent signal every 24 hours after transfection for a total of 3 days. First, a magnification of 2x was used to locate major areas of fluorescence on the plate. Next, a magnification of 20x was used to take fluorescent and brightfield images of each well. These images were overlaid to visualize which cells were fluorescing. Mutated or floating cells with a fluorescent signal were not imaged or counted as true fluorescence for this experiment

3.6 Results and Discussion

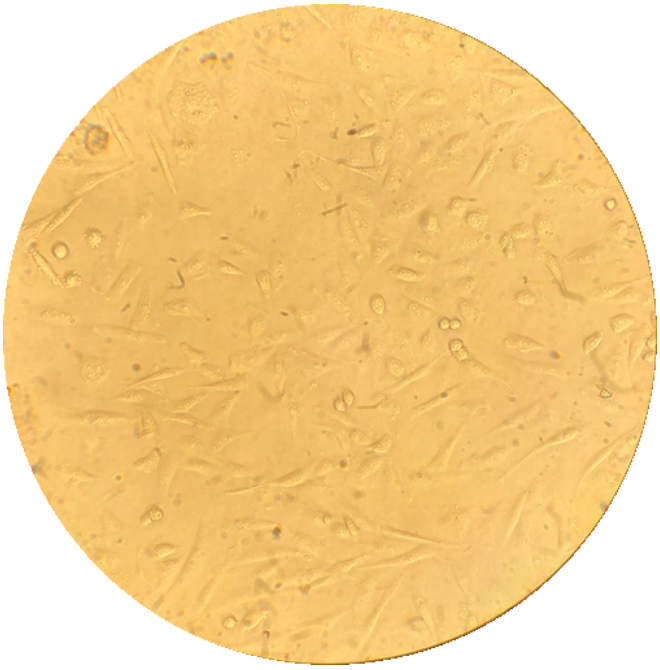


Figure 7. Mammalian CHO Cell Culture at 80% Confluency

The project was originally designed for HeLa cells because of their resilience and immortality. However, HeLa variants were received and were unable to survive our transfection and had difficulty reaching 80% confluency. **Figure 7** shows our backup cells, CHO cells. CHO cells are large, robust cells with a rod-like structure. Cells were grown to 80% confluency, which means taking up 80% of the tissue plate before they are passaged and allocated into new plates. Cells

are passaged at this confluency to ensure even distribution of media nutrients among cells and prevent overcrowding.

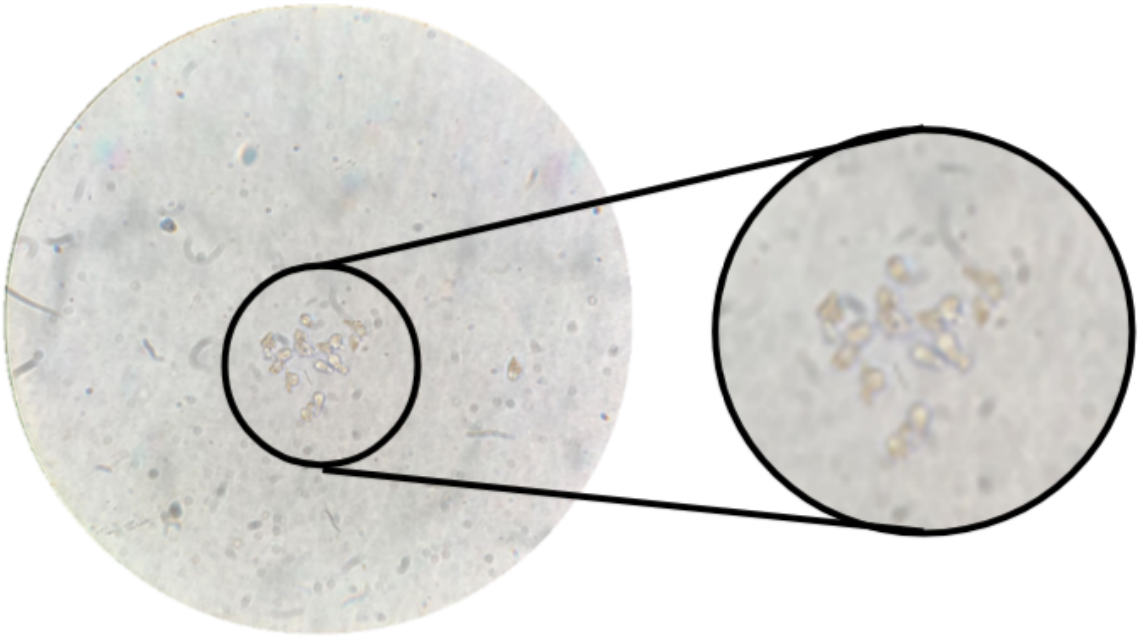


Figure 8. Transfection using Lipofectamine. Transfection resulted in cell death and mutation after using lipofectamine. No results were able to be obtained using a fluorescent microscope.

The first transfection involved HeLa variant cells and Lipofectamine 2000 transfection reagent. Though HeLa cells are typically resilient, the variant cells were unable to survive the harsh reagent after 24 hours. **Figure 8** shows an image of the cells 24 hours after transfection. The cells lost their typical fibroblast shape, which indicates they are dead or mutated. This transfection trial inspired a rewrite of the protocol, using CHO cells and nontoxic Fugene6 transfection reagent instead of HeLa variants and Lipofectamine.

The transfection trials in **Figures 9 - 15** show CHO cells that used the Fugene6 transfection reagent. Images were not taken in the same location each day and only represent a small fraction of the culture plate. Brightfield images were overlaid with their fluorescent counterparts for each figure.

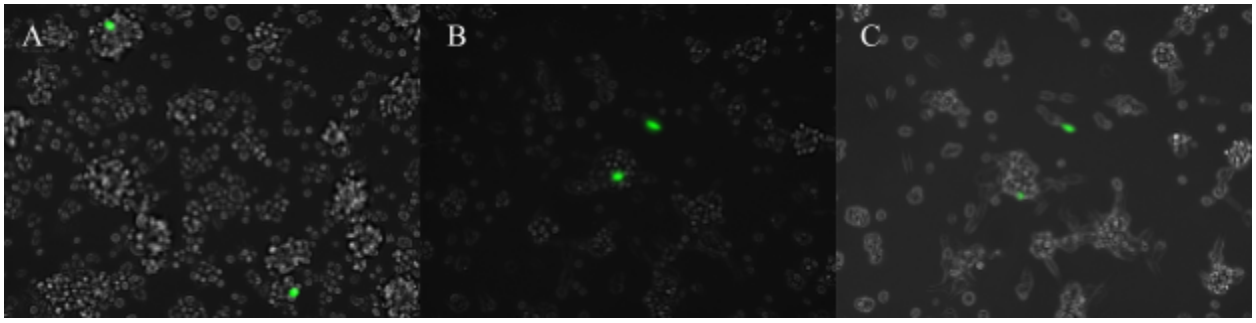


Figure 9. Fluorescent Images of eGFP at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

Figure 9 shows fluorescence, indicating a successful transformation. This plasmid acted as the positive control to confirm the protocol was valid. It is important to note that the transfection efficiency is low in these images. This is most likely due to the fact that this mechanism relies on the combination of two plasmids to fluoresce. If the plasmid concentration is too low, cells will only uptake one plasmid or the other, resulting in no fluorescence. Overall, this result is promising, but the efficiency indicates the need for further optimization trials.

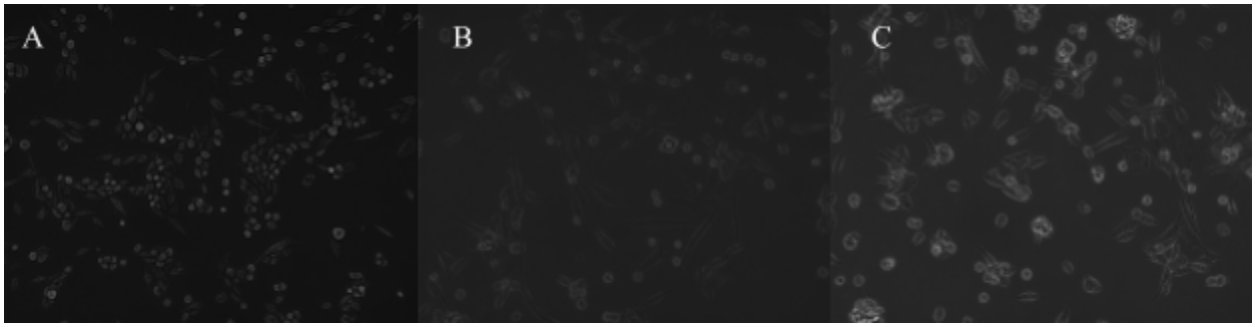


Figure 10. Fluorescent Images of Plasmid 88 at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

Figure 10 shows the results of our first negative control, the plasmid encoding for Sortase A. No fluorescence is expected from this plasmid, since the mechanism relies on the combination of both Sortase A and the test plasmid. This is a promising result, as there is no contamination of other plasmids and demonstrates that Sortase A is incapable of self-fluorescence.

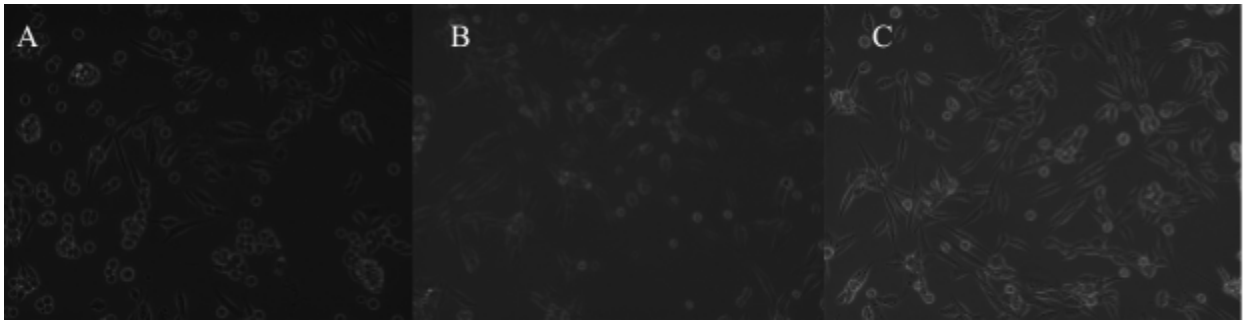


Figure 11. Fluorescent Images of Plasmid 410 at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

Figure 12. Fluorescent Images of Plasmid 411 at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

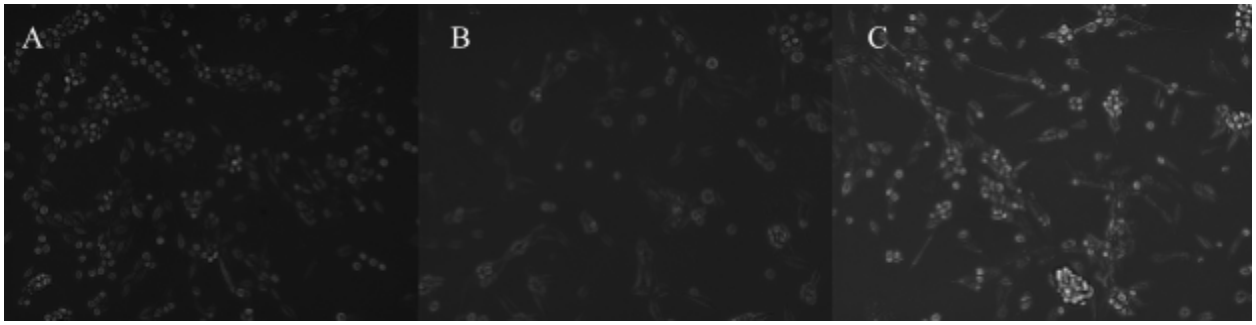


Figure 13. Fluorescent Images of Plasmid 412 at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

Figures 11 -13 show the test plasmids by themselves, without the presence of Sortase A. These are all negative controls, as the mechanism relies on the combination of Sortase A with a test plasmid. This is a promising result, as there is no fluorescence. It shows that the plasmids are incapable of self-fluorescence and that the cell is incapable of ligating the GFP fragments on its own.

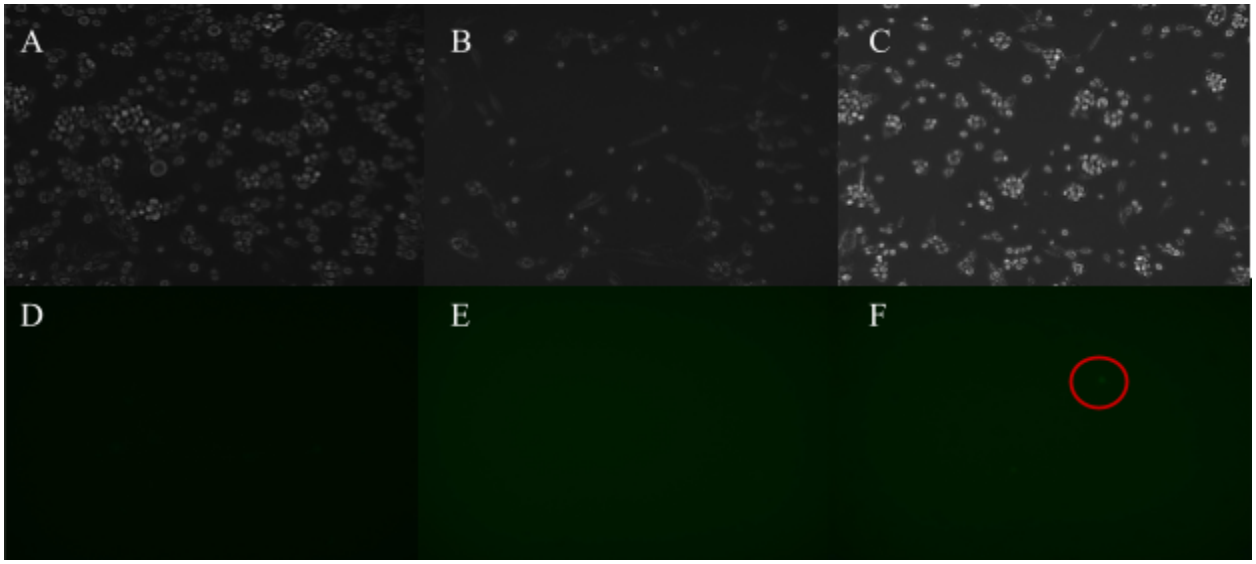


Figure 14. Fluorescent Images of Plasmids 88+410 at (A) 24 hours, (B) 48 hours, and (C) 72 hours. Fluorescent images are shown in (D) - (F).

Figure 14 shows our basal control plasmid in combination with Sortase A. A dim fluorescence is expected in this plasmid, as it was not enhanced for optimal Sortase A binding. A red circle highlights a small fluorescent signal in the original fluorescent images that is too weak to be seen with the overlay feature. Though at a low level, this signal could be significant, as the negative controls didn't show this level of dim fluorescence. This signal indicates a promising result, however, nothing can be concluded from this data until transfection efficiency is increased.

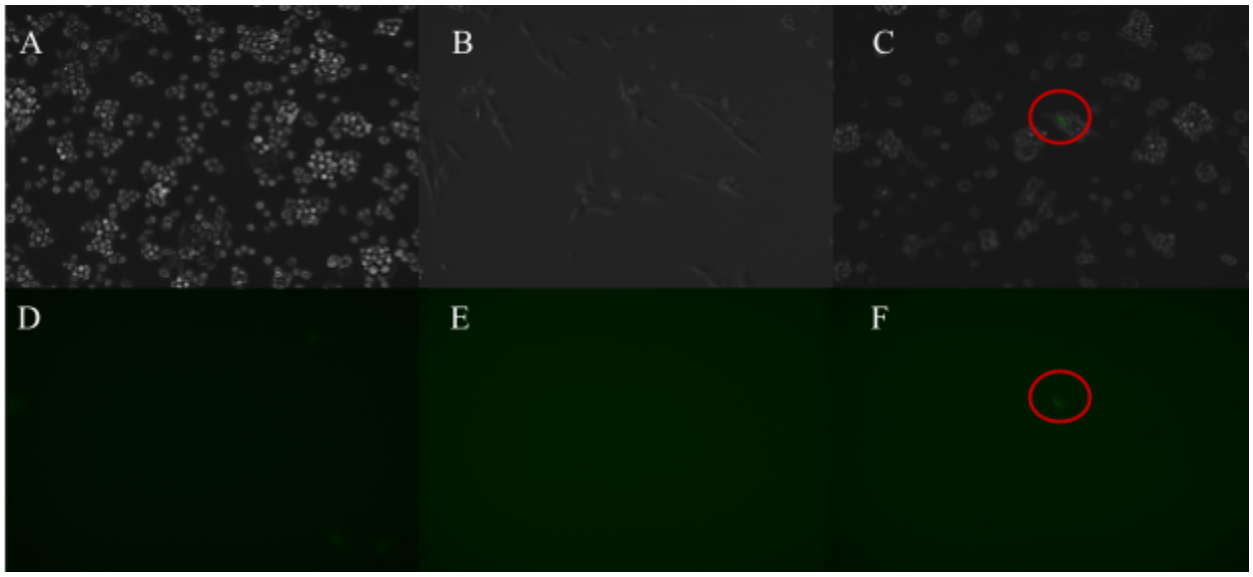


Figure 15. Fluorescent Images of Plasmids 88+411 at (A) 24 hours, (B) 48 hours, and (C) 72 hours. Fluorescent images are shown in (D) - (F).

Figure 15 shows the high affinity plasmid in combination with Sortase A. This trial is expected to show medium to high fluorescence, as the plasmid was designed for optimal Sortase A binding. A red circle highlights a fluorescent signal that is strong enough to be seen in both the fluorescent and overlay images. Though this signal is also at a low level, it could be significant, as the basal control didn't show this level of fluorescence in the overlay image. This signal indicates a promising result, however, nothing can be concluded from this data until transfection efficiency is increased.

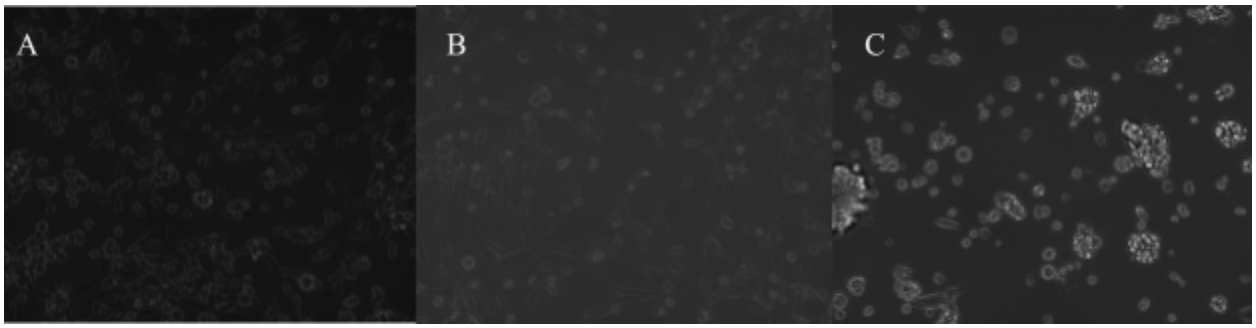


Figure 16. Fluorescent Images of Plasmids 88+412 at (A) 24 hours, (B) 48 hours, and (C) 72 hours.

Figure 16 shows the lowest affinity plasmid in combination with Sortase A. This plasmid was designed to have poor binding affinity with Sortase A and should not show fluorescence. This is a promising result, as no fluorescent signal was obtained from this trial, indicating that Sortase A can be reasonably selective and must be signaled by an optimal motif in order to ligate peptide fragments.

Chapter 4: Discussion and Conclusion

Our study demonstrates that Sortase A is a promising candidate for *in vivo* protein ligation. The transfection wells that contained negative control plasmids showed no sign of fluorescence, meaning there were no issues with the experimental design or execution. The transfection well that contained eGFP as a positive control showed fluorescence, meaning the transfection protocol was successful. However, transfection efficiency was lower than expected, indicating that further optimization studies need to be performed to increase efficiency and effectiveness of the procedure. Lastly, the test plasmids showed low signs of fluorescence. The fluorescent signals were dim but significant, as no negative controls showed signs of fluorescence. Similarly, transfection efficiency was low in these wells and transfection optimization is required to improve the results of the experiment. Overall, the results were extremely promising, but a larger transfection efficiency is needed to come to a formal conclusion about Sortase A's ability to ligate peptides *in vivo*.

4.1 Future Works

4.1.1 Transfection Optimization

Due to the time constraints from the global pandemic, Fugene6's generic protocol for transfection was used. However, Promega recommends optimizing the transfection protocol for the specific cells as each cell line responds differently to the transfection reagent. These optimization protocols are listed in **Table 6**, along with additional optimization trials specific to our experimental design. The use of two plasmids means that the transfection reagent must transfect both plasmids into the cell. The final volume, DNA amount, and reagent volume were doubled to account for the extra plasmid. If only one plasmid is able to enter the cell, there will be no GFP expression and further optimization is needed.

Table 6. Transfection Optimization Trials

Parameters	Promega Optimization	Additional Optimization

Ratio of reagent to plasmid	6:1	4:1	3:1	1.5:1	6:1	4:1	3:1	1.5:1
Final volume	100 µl	100 µl	100 µl	100 µl	200 µl	200 µl	200 µl	200 µl
DNA	2 µg	2 µg	2 µg	2 µg	4 µg	4 µg	4 µg	4 µg
Reagent	12 µl	8 µl	6 µl	3 µl	24 µl	16 µl	12 µl	6 µl

4.1.2 Flow Cytometry

Once the transfection protocols have been optimized, the visualization of cells will need to be analyzed by flow cytometry. Fluorescence should be visible in most cells, and will not be able to be visualized. Flow cytometry will be able to quantitatively analyze the efficiency of the transfection reagent as well as the different combinations of plasmid designs. With the TOP1 and SH3 domains having the highest affinity for each other, plasmids 88+411 are expected to have the highest efficiency. Plasmids 88+412 contain the BOT1 and SH3 domains which have no affinity for each other and are expected to have the lowest efficiency. Plasmids 88+410 do not contain additional domains and are expected to have a middle efficiency.

4.1.3 Western Blot

Western blot analysis is needed to characterize the fluorescent protein from the transfection trials. FLAG antibodies will be used to bind to the FLAG tag included in the original peptide sequence and confirm that Sortase A was able to create a fully functional GFP. This characterization will also prove that fluorescence was indeed due to Sortase A and not contamination or other potential outside factors. Locating a band at 27 kDa will confirm the presence of GFP.

4.1.4 Genetic Circuit

Implementation of a genetic circuit is needed to regulate the expression of Sortase A and prove its ability to ligate selectively. In future iterations of the project, a genetic circuit that responds to light will be used to activate Sortase A expression. It will be integrated into the protein using an additional plasmid vector. This experiment will follow a similar protocol, beginning with bacterial transformation and ending with fluorescent microscopy analysis. In fluorescent analysis, cells will be kept in the dark and monitored for fluorescence for 48 hours. Then, cells will be exposed to light to trigger the expression of Sortase A. After exposure, cells will be monitored every eight hours for 72 hours to determine the time-selectivity of Sortase A.

Chapter 5: Engineering Standards and Realistic Constraints

5.1 Science, Technology, and Society

If proven successful, Sortase A ligation can be used in future applications as an aid in selective drug development. When Sortase A's gene is regulated by a genetic circuit, scientists have the ability to turn its protein expression on and off. This can be used to site-selectively and time-sensitively control the formation of protein drugs within the cell. One particular application that could have a large impact on society is cancer drug development. Many current cancer treatments, like chemotherapy, are not selective drugs and kill healthy cells, causing unwanted side effects. In addition, cancer treatments like immunotherapy have no on/off switch, which sometimes causes unwanted inflammation and puts the body into unnecessary immune system overdrive. The use of Sortase A to respond to a stimulus and act site and time selectively can greatly improve the quality of life and treatment outcome of many patients worldwide. In addition, Sortase A's applications can reach far beyond protein ligation; if signaled by the LPXTG motif, Sortase A should be able to ligate other drug molecules such as lipids, DNA, sugars, and more. The use of Sortase A as a ligating molecule can greatly improve the current field of medicine and create more effective drugs.

5.2 Civic Engagement

In order to become a commercial drug in the United States, Sortase A will need to be approved by the Food and Drug Administration (FDA). The FDA has a five part process to achieve approval: discovery/concept, preclinical research, clinical research, and FDA review and post-market safety monitoring [23]. Sortase A is currently in the beginning stages of discovery and concept development. To complete this stage, further transfection optimization trials must be completed to increase transfection efficiency and confirm Sortase A's ligation ability. Next, a genetic circuit must be introduced to regulate the activity of Sortase A and prove that it can act site and time selectively upon activation. Lastly, Sortase A and a genetic circuit can be paired with a specific drug, allowing Sortase A to fully generate and regulate a protein drug within the body. The second stage involves testing the Sortase A drug on animals to ensure the effectiveness and safety of the drug. After reaching satisfactory levels in animal models, the drug

will move to clinal models in humans. Three stages of human trials will be performed: phase 1, phase 2, and phase 3. Phase 1 assesses the safety and side effects of the drug with a small pool of healthy volunteers. Phase 2 monitors the efficacy of the drug through control and blind studies. Phase 3 is large scale testing using randomized groups and is the final clinical stage of drug development. The last step in FDA approval is post-market safety monitoring. After the drug is approved for use in the United States, surveillance trials are ordered to observe efficacy, quality of life, cost-effectiveness, and compare the drug with similar market products [24].

5.3 Ethical

Drug related side effects impact many patients worldwide, making the need for specific and effective drugs extremely important. Sortase A's potential ability to ligate molecules selectively could improve patient quality of life and increase survival rates for many. This research aims to improve upon current medicines for cancer and other debilitating diseases by creating an on/off switch, which would reduce side effects and unwanted inflammation. Though this technology has the potential to improve the lives of many, it is also important to analyze the potential ethical issues with the research. One of the biggest ethical concerns of the project is the use of the controversial HeLa cell line. HeLa cells were obtained non-consensually from Henrietta Lacks, an African American woman with cervical cancer. Though many scientists worldwide use this resilient cell line, it is important to acknowledge the origin of the lines. Because this project doesn't require the specific use of HeLa cells, the cell line was switched to less controversial CHO cells.

5.4 Usability and Sustainability

This research aims to improve upon current industry standards, particularly EPL. Though EPL is effective, it is a complicated procedure that leaves room for error. This research overcomes the challenges associated with EPL by eliminating the need for excess cysteines, therefore limiting side reactions. It also has a more streamlined protocol that allows cells to do most of the work, eliminating more opportunities for human error. Once proven, this method should have a higher product yield than EPL, as it should correctly fold all proteins ligated. Lastly, this method should be a more cost-effective and sustainability approach to protein ligation; because this method uses

an enzyme to ligate fragments, little enzymes and reagents are needed to catalyze the reaction. This is an improvement from EPL's 1:1 ratio requirement of protein to intein.

5.6 Health and Safety

Though Sortase A has the potential to act as a selective drug and improve the lives of patients, its mechanism needs to be studied further before clinical or preclinical trials begin. Sortase A responds to a LPXTG motif in order to ligate molecules, however, more research needs to be done to determine what types of molecules naturally contain LPXTG. If Sortase A is used as a drug, it needs to act selectively and only ligate the desired molecules. The LPXTG motif is found naturally in bacterial cells, but more research needs to be done to ensure that this motif doesn't exist naturally in human proteins. If so, Sortase A would need to be further edited to prevent the ligation of natural proteins within the body.

References

[1] Z. A. Wang and P. A. Cole, “Methods and Applications of Expressed Protein Ligation,” in Expressed Protein Ligation: Methods and Protocols, M. Vila-Perelló, Ed. New York, NY: Springer US, 2020, pp. 1–13. doi: 10.1007/978-1-0716-0434-2_1.

[2] W. Zhao, Y. Zhang, C. Cui, Q. Li, and J. Wang, “An efficient on-column expressed protein ligation strategy: Application to segmental triple labeling of human apolipoprotein E3,” Protein Science, vol. 17, no. 4, pp. 736–747, Apr. 2008, doi: 10.1110/ps.073383708.

[3] S. K. Mazmanian, G. Liu, H. Ton-That, and O. Schneewind, “Staphylococcus aureus Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall,” Science, vol. 285, no. 5428, pp. 760–763, 1999, doi: 10.1126/science.285.5428.760.

[4] A. W. Jacobitz, M. D. Kattke, J. Wereszczynski, and R. T. Clubb, “Chapter Seven - Sortase Transpeptidases: Structural Biology and Catalytic Mechanism,” in Structural and Mechanistic Enzymology, vol. 109, T. Karabencheva-Christova, Ed. Academic Press, 2017, pp. 223–264. doi: https://doi.org/10.1016/bs.apcsb.2017.04.008.

[5] H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, and O. Schneewind, “Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif,” Proceedings of the National Academy of Sciences, vol. 96, no. 22, pp. 12424–12429, 1999, doi: 10.1073/pnas.96.22.12424.

[6] K. W. Clancy, J. A. Melvin, and D. G. McCafferty, “Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition,” Peptide Science, vol. 94, no. 4, pp. 385–396, 2010, doi: https://doi.org/10.1002/bip.21472.

[7] S. K. Mazmanian, H. Ton-That, and O. Schneewind, “MicroReview Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus,” 2001.

[8] A. M. Perry, H. Ton-That, S. K. Mazmanian, and O. Schneewind, “Anchoring of surface proteins to the cell wall of Staphylococcus aureus. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring,” Journal of Biological Chemistry, vol. 277, no. 18, pp. 16241–16248, May 2002, doi: 10.1074/jbc.M109194200.

[9] Choppa, Akhil. (2020). In Vivo Non-Ribosomal Protein Synthesis in Mammalian Cells. Santa Clara University School of Engineering, Bioengineering Master’s Theses. Santa Clara, CA.

[10] H. Mao, S. A. Hart, A. Schink, and B. A. Pollok, "Sortase-Mediated Protein Ligation: A New Method for Protein Engineering," Journal of the American Chemical Society, vol. 126, no. 9, pp. 2670–2671, Mar. 2004, doi: 10.1021/ja039915e.

[11] N. Broguiere, F. A. Formica, G. Barreto, and M. Zenobi-Wong, "Sortase A as a cross-linking enzyme in tissue engineering," Acta Biomaterialia, vol. 77, pp. 182–190, 2018, doi: https://doi.org/10.1016/j.actbio.2018.07.020.

[12] N. Voloshchuk, D. Liang, and J. Liang, "Sortase A Mediated Protein Modifications and Peptide Conjugations," Current drug discovery technologies, vol. 12, Jun. 2015, doi: 10.2174/1570163812666150903115601.

[13] K. Strijbis, E. Spooner, and H. L. Ploegh, "Protein Ligation in Living Cells Using Sortase," Traffic, vol. 13, no. 6, pp. 780–789, Jun. 2012, doi: 10.1111/j.1600-0854.2012.01345.x.

[14] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, and H. L. Ploegh, "Sortagging: a versatile method for protein labeling," Nature Chemical Biology, vol. 3, no. 11, pp. 707–708, 2007, doi: 10.1038/nchembio.2007.31.

[15] C. D. Hein, X.-M. Liu, and D. Wang, "Click chemistry, a powerful tool for pharmaceutical sciences," *Pharmaceutical research*, vol. 25, no. 10, pp. 2216–2230, Oct. 2008, doi: 10.1007/s11095-008-9616-1.

[16] "Click Chemistry Azide-Alkyne Cycloaddition," Organic Chemistry. [Online]. Available: https://www.organic-chemistry.org/namedreactions/click-chemistry.shtm. [Accessed: 09-Jun-2021].

[17] F. Alonso, Y. Moglie, and G. Radivoy, "Copper Nanoparticles in Click Chemistry," *Accounts of Chemical Research*, vol. 48, no. 9, pp. 2516–2528, Sep. 2015, doi: 10.1021/acs.accounts.5b00293.

[18] R. L. Kimber *et al.*, "Synthesis of copper catalysts for click chemistry from distillery wastewater using magnetically recoverable bionanoparticles," *Green Chemistry*, vol. 21, no. 15, pp. 4020–4024, 2019, doi: 10.1039/c9gc00270g.

[19] M. C. Cortizo and M. F. Lorenzo de Mele, "Cytotoxicity of copper ions released from metal," *Biological Trace Element Research*, vol. 102, no. 1, pp. 129–141, 2004, doi: 10.1385/BTER:102:1-3:129.

[20] K. Nwe and M. W. Brechbiel, “Growing Applications of ‘“Click Chemistry”’ for Bioconjugation in Contemporary Biomedical Research.”

[21] A. Pérez Rubio and J. M. Eiros, “Cell culture-derived flu vaccine: Present and future,” Human vaccines & immunotherapeutics, vol. 14, no. 8, pp. 1874–1882, 2018, doi: 10.1080/21645515.2018.1460297.

[22] C. Spiess et al., “Bispecific antibodies with natural architecture produced by co-culture of bacteria expressing two distinct half-antibodies,” Nature Biotechnology, vol. 31, no. 8, pp. 753–758, Jan. 2013, doi: 10.1038/nbt.2621.

[23] “FDA Approval - Process of Approving Drugs & Medical Devices,” *Drugwatch.com*. [Online]. Available: https://www.drugwatch.com/fda/approval-process/. [Accessed: 09-Jun-2021].

[24] “About Clinical Trials,” *CenterWatch*. [Online]. Available: https://www.centerwatch.com/clinical-trials/overview. [Accessed: 09-Jun-2021].

[25] UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics, “Enhanced green fluorescent protein,” *UniProt*, 07-Oct-2020. [Online]. Available: https://www.uniprot.org/uniprot/C5MKY7. [Accessed: 09-Jun-2021].

[26] “Vector Database,” *Addgene*. [Online]. Available: https://www.addgene.org/vector-database/2488/?gclid=CjwKCAjwqvyFBhB7EiwAER76VqjxHMU_IK9X5RrF7U3ClsFaj6sCbnLgzA59Q7Sv0AqoCHnNsdlgBoCMhMQAvDBwE. [Accessed: 09-Jun-2021].

Appendix

Appendix A:

	Q1			Q2			Q3		
Task	Oct	Nov	Dec	Jan	Feb	March	April	May	June
Complete lab access forms									
Receive lab training									
Create supplies list									
Write protocols									
Order supplies									
Gain lab access						*note, lab closed for most of March and April			
Bacterial transformation									
Bacterial cell culture									
Plasmid Miniprep									
Gel Electrophoresis									
Mammalian cell culture									
Plasmid transfection									
Fluorescent Microscopy									
FACS									if time
Western Blot									if time

APPENDIX B: RECOMBINANT PLASMID 410 SEQUENCE

Recombinant plasmid 410 was designed on ApE (version 2.0.61) using the pEF-GFP vector backbone from Addgene [9]. Plasmid 411 is 6353 nucleotides long.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG
AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCG
AGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC
CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC
GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT

AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAACTGCCTGAAACCGGTTAATTAATAAAGGAAATTTAT
TTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCAGGTACCGCTCCGGTGCCC
GTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTC
GGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGT
CGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGT
AGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGC
CGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGA
ATTACTTCCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGG
AAGTGGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTG
AGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTC
GCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGC
TGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACAC
TGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCAC
ATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGT
CTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGC
CCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCG
CTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCG
GGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTT
CATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCT
TTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCC
ACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCT
TGGAATTTGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGT
TCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGAGCTAGCCACCATGGGAGGTGGAG
GTGGAGACTACAAAGACGATGATGACAAGAAGAATGGAATCAAAGTTAACTTCAAA
ATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAA
TACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACA
ATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTT
TGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAAAGCGGCC
GCACTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCC

TGGCTCACAAATACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCA
TGAAGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAA
TAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATC
ATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGC
TGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCC
CCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTA
TATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTAC
TAGCCAGATTTTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTA
TGGAGATCCCTCGACCTGCAGCCCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCGGATCCGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGT
TCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGG
CCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAA
TAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTG
TCCAAACTCATCAATGTATCTTATCATGTCTGGATCCGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTG
ACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAA
GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGC
TCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA

CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACG
AAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA
TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCT
GCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATC
GTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAA
CCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG
TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG
GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGG
GTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGG

APPENDIX C: RECOMBINANT PLASMID 411 SEQUENCE

Recombinant plasmid 2 was designed on ApE (version 2.0.61) using pEF-GFP vector backbone from Addgene [9]. Plasmid 411 is 6548 nucleotides long, and includes the TOP1 and SH3 domains that increase GFP expression.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG

AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCG
AGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC
CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC
GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT
AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAAGCTCCAGCCTTTCCACCTCCTAGCCCTCCACTGCCTG
AAACCGGTTAATTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTT
TGTGTCTCTCAGGTACCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCAC
AGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGT
GGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGG

GTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACG
GGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCT
TTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACGCCCCTGGCTGCAGTACG
TGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTGCG
CTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGGCC
GCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCT
AGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTT
GTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGC
GACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCG
GCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTG
GCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCA
CCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAA
ATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAA
AGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG
TCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGG
GAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTTAG
GCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCT
TGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGTGT
CGTGAGCTAGCCACCATGGGAGGTGGAGGTGGAGACTACAAAGACGATGATGACAA
GAATGATCCTAACCTCTTTGTCGCACTGTATGATTTCGTCGCCTCTGGTGACAATACC
CTCAGCATTACTAAAGGAGAGAAGCTGAGGGTGCTGGGATACAATCACAACGGTGAG
TGGTGCGAGCCTCAGACAAAGAACGGACAGGGCTGGGTCCCTTCCAACTACAAGA
ATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAA
CTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCA
GACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAG
AGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGA
TGAACTATACAAATAAAGCGGCCGCACTCCTCAGGTGCAGGCTGCCTATCAGAAGG
TGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCT
CTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTAAT
AAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGG
AAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGA

GTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAG
GTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCT
TGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCC
TAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTCCCA
GTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGCTTGGC
GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTA
ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCC
ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGT
GAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAGC
TTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT
TTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGG
ATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA

ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT
GCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT
TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGG

Appendix D: RECOMBINANT PLASMID 412 SEQUENCE

Recombinant plasmid 412 was designed on ApE (version 2.0.61) using the pEF-GFP vector backbone from Addgene [9]. Plasmid 412 is 6548 nucleotides long, and includes the BOT1 and SH3 domains that decrease GFP expression.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG
AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC

TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGA
GGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC
CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC
GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT
AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAAGCTGCTGCAATGCAGAAACCTAGCCTGCCACTGCCT
GAAACCGGTTAATTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTT
TTTGTGTCTCTCAGGTACCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCC
ACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAG
GTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGA
GGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAA
CGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCT
CTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACGCCCCTGGCTGCAGTA
CGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTG

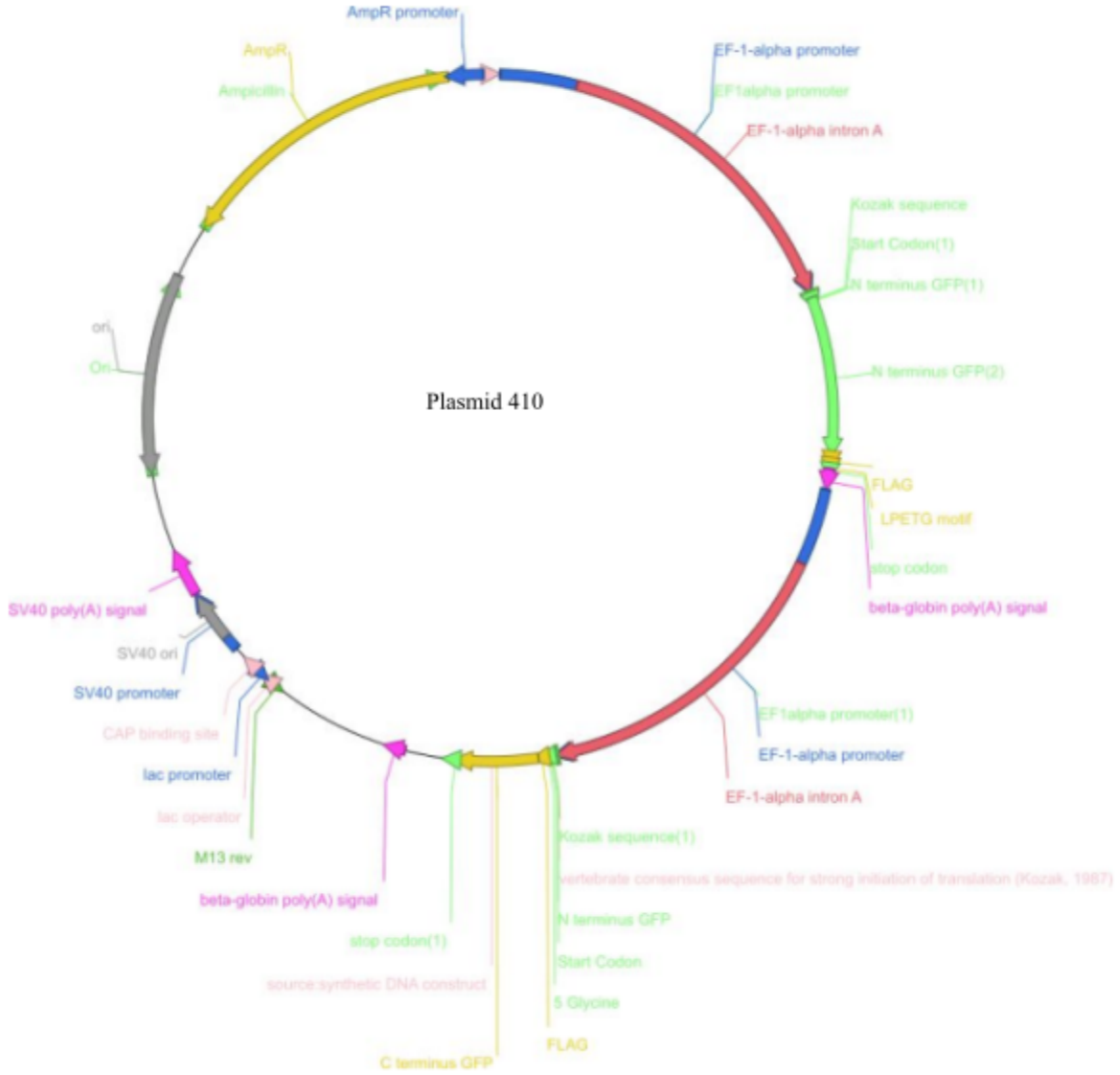
CGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGG
CCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTC
TCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGT
CTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGC
GGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGC
GCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGC
CTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGG
CACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCA
AAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGA
AAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGC
CGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG
GGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTTA
GGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGAT
CTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGT
GTCGTGAGCTAGCCACCATGGGAGGTGGAGGTGGAGACTACAAAGACGATGATGAC
AAGAATGATCCTAACCTCTTTGTCGCACTGTATGATTTCGTCGCCTCTGGTGACAATA
CCCTCAGCATTACTAAAGGAGAGAAGCTGAGGGTGCTGGGATACAATCACAACGGT
GAGTGGTGCGAGCCTCAGACAAAGAACGGACAGGGCTGGGTCCCTTCCAACTACAA
GAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTC
AACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTAC
CAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAG
AGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATG
GATGAACTATACAAATAAAGCGGCCGCACTCCTCAGGTGCAGGCTGCCTATCAGAA
GGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCC
CTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTA
ATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTC
GGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTA
GAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAG
AGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGC
CTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATC
CCTAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTCC

CAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGCTTG
GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC
ACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCC
CATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATT
TTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG
GATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAA
CGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAAT
CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT
ACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTC
GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG

AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGT
TGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT
TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGG

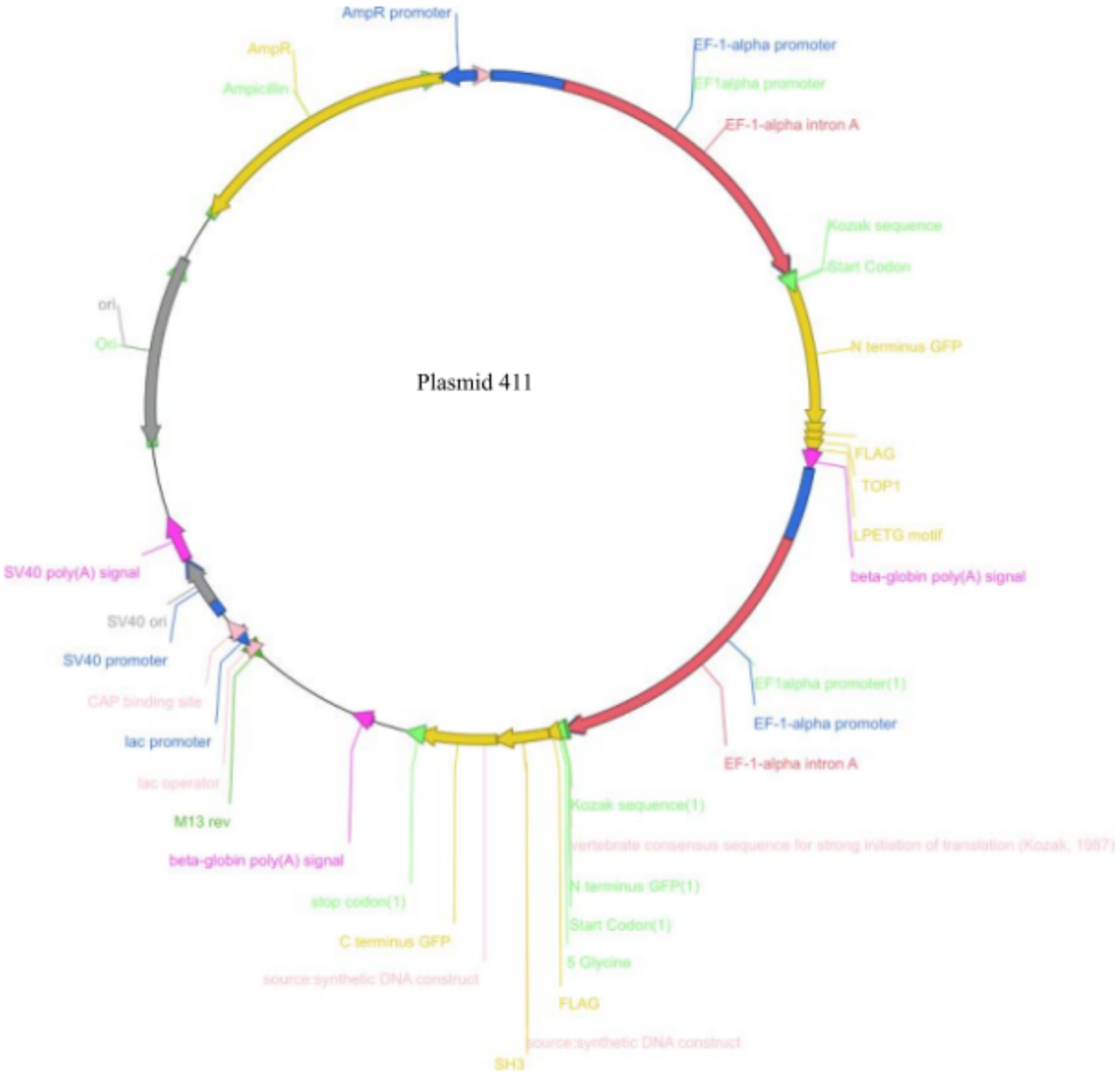
Appendix E: Plasmid 410 Vector Design.

The pEF-GFP vector plasmid was used for its small size and expression in mammalian cells. The GFP gene was spliced into two segments, with the N-terminus containing the LPETG motif, and the Gly5 chain on the C-terminus. These were 471 and 245 nucleotides respectively. The recombinant plasmid includes ampicillin for bacterial resistance, a kozak sequence for translation efficiency, and a FLAG sequence for western blot analysis. Additional elements include an EF1 alpha promoter, origin of replication, start codon, stop codon, and the polyadenylation signal [9].



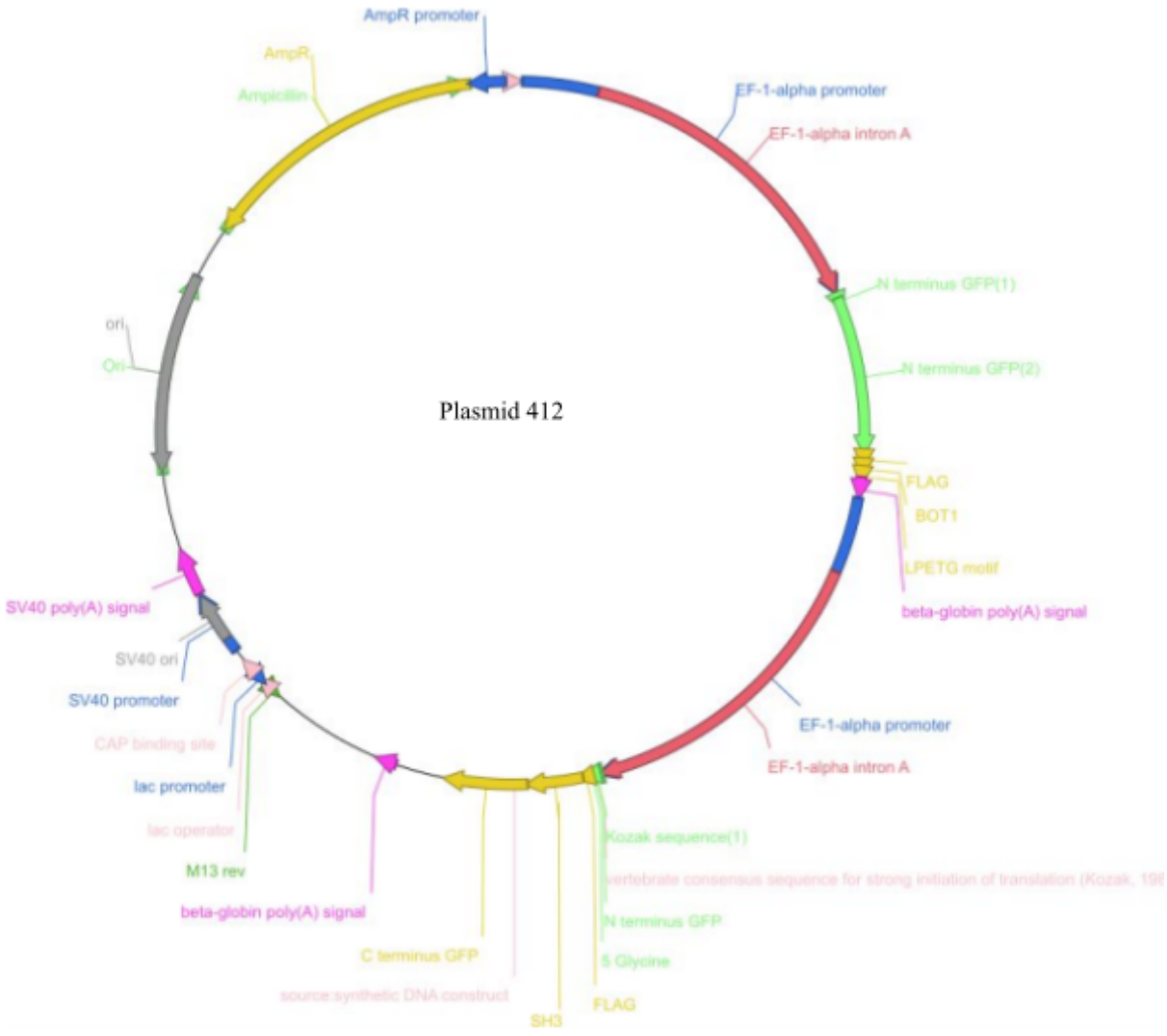
Appendix F: Plasmid 411 Vector Design.

The plasmid design has the same features as plasmid 410, with the addition of the TOP1 and SH3 domains, which have a high binding affinity for each other. This recombinant plasmid is expected to have a higher expression of GFP [9].



Appendix G: Plasmid 412 Vector Design.

The plasmid design has the same features as plasmid 410, with the addition of the BOT1 and SH3 domains, which do not have a binding affinity for each other. This recombinant plasmid is expected to have minimal expression of GFP [9].



Appendix H: RECOMBINANT PLASMID 88 SEQUENCE

Recombinant plasmid 88 encodes for Sortase A and is 710 base pairs long. The Sortase A gene is 621 base pairs with a protein sequence of 206 amino acids.

NTGTCGTGAGGATTAGCTTGGTACTAATACGACTCACTATAGGGAGACCCAAGCTGG
CTAGGTAAGCTTGGTACCGAGCTCGGATCCACTAGTATGGGCCAAGCTAAACCTCAA
ATTCCGAAAGATAAATCAAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATT
AAAGAACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAGAGGTGTAAG
CTTTGCAGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACA
CTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTA
GTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAGTATA
AGAGGTGTTAAGCCAACAGATGTAGAAGTTCTAGATGAACAAAAAGGTAAAGATAA
ACAATTAACATTAATTACTTGTGATGATTACAATGAAAAGACAGGCGTTTGGGAAA
AACGTAAAATCTTTGTAGCTACAGAAGTCAAAGAATTCTGCAGATATCCAGCACAGT
GGCGGCCGCTCGAGTCTAGAGGGCCCTTCGAACAAAAACTCATCTCAGAAGAGGAT
CTGAATATGCATACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCA
GCCTCGACTGTGCCTTCTAGTTGCCAGCCATC

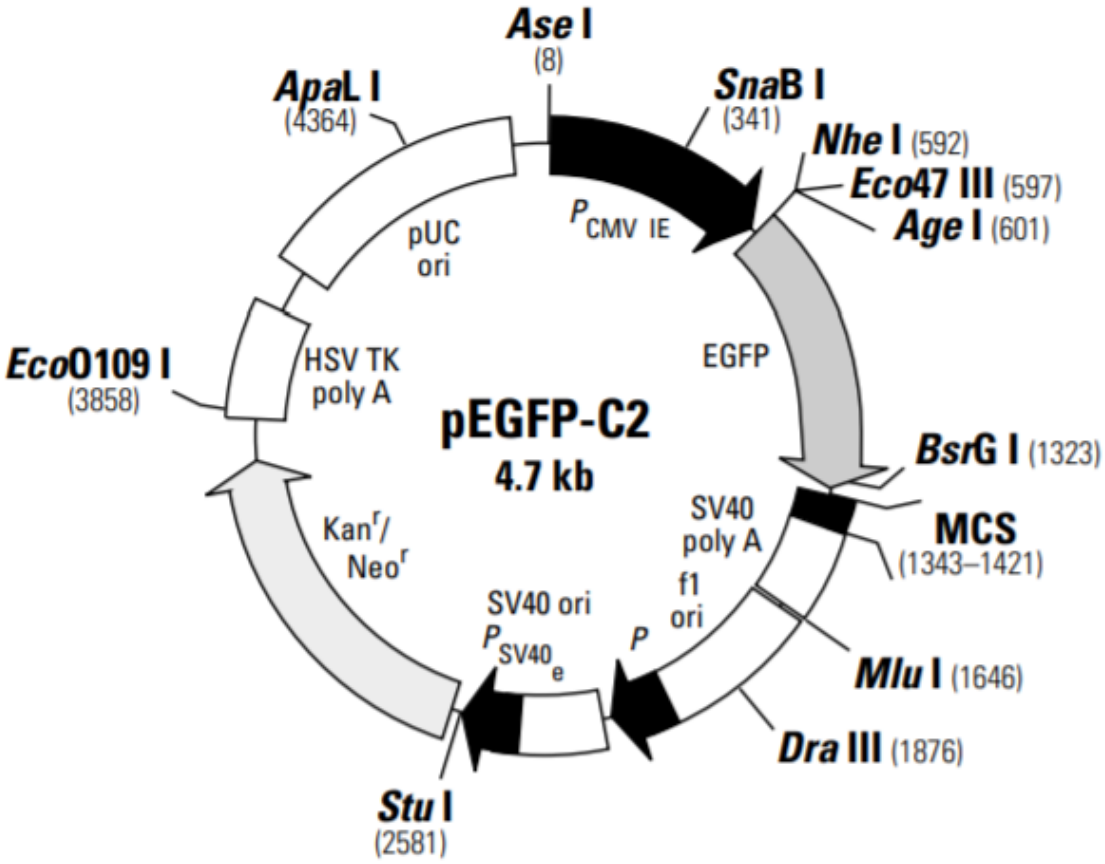
APPENDIX I: eGFP PROTEIN SEQUENCE

The DNA sequence is 717 base pairs and the protein sequence is 239 amino acids long [25].

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWP
TLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG
DTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYK

APPENDIX J: eGFP Plasmid Vector.

The pEGPF C2 plasmid used was created by BD Biosciences and contains kanamycin resistance [26].


Ase I
(8)
ApaL I
(4364)
SnaB I
(341)
Nhe I (592)
Eco47 III (597)
Age I (601)
P CMV IE
pUC
ori
EGFP
Eco0109 I
(3858)
HSV TK
poly A
pEGFP-C2
4.7 kb
BsrG I (1323)
SV40
poly A
MCS
(1343–1421)
Kanr/
Neor
f1
ori
SV40 ori
P SV40 e
P
Mlu I (1646)
Dra III (1876)
Stu I
(2581)