DECIPHERING THE TOPOLOGY OF THE *E. COLI* INNER MEMBRANE PROTEOME

By

Daniel E. Carrillo Baez

BSc in Biology & Chemistry

Thesis

Submitted to Flinders University

For the Degree of

Master of Biotechnology

College of Medicine and Public Health

November 3rd, 2023

Table of Contents

Abstract

Inner membrane proteins (IMPs) are essential components of the gram-negative bacteria inner cell membrane. They perform various functions, including nutrient uptake, energy metabolism, signal transduction, stress responses, and cell division. The number of IMPs in *E. coli* is estimated to be around 900, and collectively, they are involved in various cellular processes essential for the bacteria's survival and growth. Understanding the orientation of IMPs, i.e., whether functional domains are exposed to the cytoplasm or periplasm, is essential for elucidating their functions and the molecular mechanism of interaction with other cellular components. Thus, this thesis aims to provide a novel methodology based on dynamic location measurements that can reveal the topology of the *E. coli* IMPs.

Specifically, the current experimental topology mapping methods were expanded by (i) using a pluripotent fluorescent protein capable of expressing in the cytoplasm and periplasm and (ii) step changes in pH. sfGFP was utilized in an innovative, dynamic measurement approach, introducing pH step changes to induce variation in fluorescence intensity. These changes reported on the intracellular versus periplasmic location of sfGFP, as only in the latter scenario leads to a sustained change in fluorescence (Wilks and Slonczewski, 2007). The following steps were meticulously planned and carried out to achieve this goal. First, a new plasmid capable of expressing sfGFP was developed. Second, the replication capabilities were improved to acceptable levels. Third, a systematic exploration of HCl volume and cell numbers was developed to find the optimal conditions. Fourth, clear data measuring parameters were defined to ensure precise data analysis. Lastly, the proposed method was tested in eight test proteins with diverse topologies and lengths, similar to those found in all IMPs. This found that the method successfully predicts the orientation of the C-terminus by clearly showing a difference in fluorescence recovery between sfGFP localized in the cytoplasm and periplasm. The results of this thesis indicate that a novel method to map membrane protein topology was

successfully developed. Further work will focus on the application of this method to the entire inner proteome and correlating the obtained results to existing experimental and computational models. This method may also serve as a proxy for detecting membrane protein insertion and folding in the context of various conditions.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: Daniel Carrillo Baez

Date: 03/11/2023

Acknowledgements

I would like to begin by expressing my gratitude for the continuous support, trust, and invaluable guidance of my supervisor, Professor Harald Janovjak. Also, I would like to acknowledge everyone in the Janovjak lab for helping me transition into the laboratory and fostering meaningful relationships.

Furthermore, I would like to acknowledge my family (gracias por todo!), friends and, particularly, my partner for her unconditional support throughout this journey.

I also want to thank everyone in the Department of Biotechnology for their role in teaching me and providing essential assistance during this thesis.

Special Circumstances

While this thesis was submitted on time and is complete, I would like to declare the following Special Circumstances. In the middle of the second semester, my ability to perform experiments was limited for a period of four weeks by a need to fly back to Canada due to a family emergency.

AI

This document used AI tools (e.g., Grammarly Premium) to correct and improve sentence structure. However, Daniel Carrillo Baez wrote everything, and only some suggestions were incorporated into the text and modified to fit the narrative.

List of Figures

List of Tables

Abbreviations

AGRF	Australian Genome Research Facility
AMP	Ampicillin
BP	Base pair
C	Celsius
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
FP	Fluorescent protein
g	Grams
GFP	Green fluorescent protein
HCl	Hydrochloric acid
IDT	Integrated DNA Technologies
IM	Inner membrane
IMP	Inner membrane protein
IQR	Interquartile range
KB	Kilobase
kB	Kilobase
LB	Luria Broth
mL	Millilitre
mM	Millimolar
MP	Membrane protein
ng	Nanogram
OM	Outer membrane
O/N	Overnight

OriR	Origen of replication
PBS	Phosphate buffered saline
PBSG	Phosphate buffered saline glycerol
rpm	Revolutions per minute
sfGFP	Superfolder green fluorescent protein
TM	Transmembrane

Chapter I: Introduction

I.1 The membrane proteome of *E. Coli*

Membrane proteins are functionally essential to the cell membranes of all organisms, including *Escherichia coli (E. coli)*, a Gram-negative bacterium commonly found in the human gut. The cell envelope comprises an outer membrane (OM) and an inner membrane (IM) separated by an aqueous environment called the periplasm. The OM is an asymmetric structure with phospholipids on the inside and lipopolysaccharides on the outside (Mathelié-Guinlet et al., 2020). The IM is a complex structure surrounding the cytoplasm, critically separating the internal and periplasmic environment. It consists of a lipid bilayer with embedded proteins that perform various functions. Membrane proteins in *E. coli* can be broadly categorized into two types based on their localization within the cell membrane: integral and peripheral membrane proteins. Integral membrane proteins are embedded within the lipid bilayer and typically span the entire membrane thickness at least once. In contrast, peripheral membrane proteins are associated with the membrane but do not penetrate it per se (Yeagle & Elsevier, 2016).

IMPs are localized to the inner *E. coli* membrane exbibit hydrophobic transmembrane (TM) domains, most commonly α-helices or β-sheets, that anchor them within the lipid bilayer. In contrast, their hydrophilic domains face the cytoplasm or the periplasmic space (Luirink et al., 2012. & Strahl and Errington. 2017). The number of IMPs in *E. coli* can vary depending on the criteria for counting them and cell state. First, the definition of IMP subtype can influence their number, mainly whether it includes all inner membrane-associated proteins or only TM proteins. Additionally, factors such as different strains, growth conditions, post-translational modifications, and subcellular localization contribute to the fluctuation in the number of proteins that are expressed and membrane-localized. Two critical studies have proposed IMPs numbers. Daley et al. published in the early 2000s a paper studying 730 IMPs

in *E. coli* (Daley et al., 2005), and Orfanoudaki and Economou published a data-based in 2014 estimating the number to be 970 (Orfanoudaki & Economou, 2014).

The diversity of IMPs is reflected in their large number, form, and function. IMPs of *E. coli* are inserted into the lipid bilayer with specific topologies, and the structural characteristics of these proteins determine their orientations. As shown in **Figure 1,** there are several subclasses of orientations for IMPs; however, they can be broadly categorized into single-span and multi-span orientations (Strahl and Errington, 2017). Single-span IMPs are integral membrane proteins that cross the lipid bilayer only once. They comprise a single hydrophobic TM domain, which anchors the protein within the membrane by interacting with the hydrophobic part of the lipid bilayer. Also, the N-terminus (start of the protein) is usually located in the cytoplasm, while the C-terminus (end of the protein) faces the periplasm (Kuhn et al., 2017). These groups of proteins usually play a role in transport, receiving signals and enzymatic activity.

In contrast, multi-span IMPs cross the lipid bilayer multiple times. As in the case of single domains of single-span proteins, these domains usually comprise hydrophobic amino acids (aa) sequences. Multi-span IMPs have a more complex structure, exposing alternating regions to the cytoplasm and periplasm. The existence of multiple spans allows them to assume complex tertiary and quaternary structures to perform diverse functions, such as transport systems, signal transduction and energy production (Hedin et al., 2011). Overall, the orientation of IMPs is a tightly regulated process. The unique arrangement of TM domains and exposure of certain protein regions to diverse cellular compartments allow the proteins to perform diverse roles in various cellular processes. Understanding the orientation of IMPs is essential for elucidating their functions and the molecular mechanism of interaction with other cellular components.

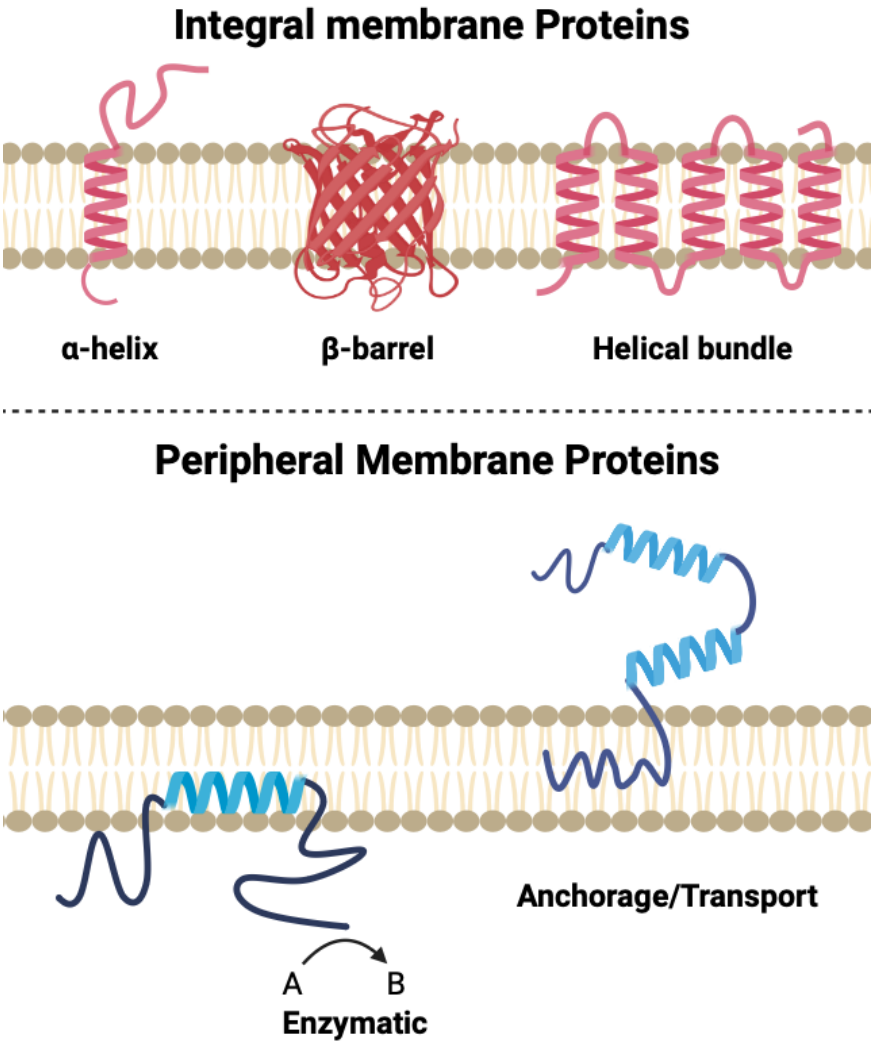

Figure 1: Cartoon depicting a comparison of integral and peripheral membrane proteins. Further, single-span vs. multi-span topologies and their structural features are explored. Imagen inspired by Hendi et al. (2011) and created in Biorender.

I.2 Biotechnological applications

The orientation of IMPs in *E. coli* is a critical aspect of cellular physiology and has an impact on numerous biotechnological applications. This is because the orientation of IMPs can determine their function, interactions with other proteins, and accessibility to drugs or other molecules. Thus, knowing the orientation of IMPs has implications for bioprocessing, biofuels production, drug development, and synthetic biology.

One of the most significant biotechnological applications of knowing the orientation of IMPs in *E. coli* is in bioprocessing. *E. coli* is a widely used platform for producing recombinant

proteins, including pharmaceuticals, enzymes, and other biologics. IMPs are essential for many of these processes, including the secretion and transport of recombinant proteins or small molecules (Luirink et al., 2012). By understanding the orientation of IMPs, researchers can harness mutations to increase binding and transport, improve the efficiency of secretion and increase the yield of the final product (Papanastasiou et al., 2012).

The production of biofuels is another example of the biotechnological advantage of understanding the orientation of IMPs in *E. coli*. From renewable sources, *E. coli* can be engineered to produce biofuels, such as ethanol, butanol, and biodiesel. IMPs are critical for transporting these fuels across the cell membrane, and their orientation can affect the efficiency of this process. By understanding the orientation of these proteins, researchers can optimize the production of biofuels by improving transport efficiency, reducing toxicity, and increasing the yield of the final product (Koppolu and Vasigala, 2016).

In addition to bioprocessing, knowing the orientation of IMPs in *E. coli* is also essential for drug development. Many drugs target IMPs, including antibiotics, antifungals, and antivirals (Yin and Flynn, 2016). Understanding the orientation of these proteins can help researchers design more effective drugs by identifying the binding sites and determining the accessibility of the drug to the target protein. For example, if a drug binds to a particular site on an IMP and that site is inaccessible, it may not treat an infection effectively. By knowing the orientation of the protein, researchers can design drugs that are more likely to be effective in treating infections or other diseases.

A further biotechnological application of knowing the IMP orientation is in synthetic biology, which is an interdisciplinary field that aims to design and construct new biological systems using engineering principles (Khalil and Collins, 2010). IMPs are critical in many systems, including synthetic metabolic pathways, biosensors, and gene circuits (Khalil and

Collins, 2010). By understanding the orientation of IMPs, researchers can design these systems to optimize their performance, improve their efficiency, and reduce their toxicity.

Overall, the biotechnological advantages of knowing the orientation of IMPs in *E. coli* are numerous and diverse. From bioprocessing to drug development to synthetic biology, understanding the orientation of IMPs can provide critical insights that can improve the efficiency, effectiveness, and safety of these processes. As the understanding of IMPs continues to evolve, more biotechnological applications will likely be discovered.

I.3 Proteome-wide analysis of IMPs

Proteome-wide analyses of IMPs in *E. coli* have been the focus of extensive research in recent years due to their vital roles and abundance. The inner membrane of *E. coli* is a highly complex and dynamic environment and understanding it has proven challenging. However, advancements in technology have enabled the creation of a few proteome-wide models that have provided valuable insights into the functions of these proteins (Kuhn et al., 2017).

Daley *et al.* developed one of the most famous proteome-wide models in 2005, which used a combination of experimental and bioinformatic approaches to predict IMP topology. They analysed the genome of *E. coli* and identified 730 open reading frames (ORFs) that encode IMPs. They used a combination of sequence-based and experimental techniques to determine the topology of these proteins, including green fluorescent protein (GFP) fusions and alkaline phosphatase (PhoA) fusions. They fused either GFP or PhoA to the C-terminus of all proteins and observed if the fusion proteins were fluorescent or enzymatically active. The underlying principle was that wild-type GFP is only fluorescent in the cytoplasm (Feilmeier et al., 2000), while PhoA is only active in the periplasm (Manoil and Beckwith. 1986). Harnessing this difference, they determined the orientation of the inner membrane proteome. This resulted

in the orientation and topology of 601 IMPs being predicted, providing a valuable resource for the scientific community.

While Daley et al.'s study provided detailed insights into the topology of IMPs on an unprecedented scale, an effect of using GFP/PhoA fusions on the orientation of the proteins and, thus the accuracy of the results cannot be excluded. Additionally, the model relies on detecting absolute intensity values, which for high or low protein expression levels of some proteins can prove challenging but is internally controlled for using two fusions tags.

I.4 pH sensitivity of fluorescent proteins

Fluorescent proteins (FPs) have revolutionized biological research by enabling scientists to visualize biological processes in real time. These proteins, originally isolated from jellyfish and other marine organisms, emit light when excited by specific wavelengths of light (Shimomura et al., 1962). FPs are widely used in bioimaging, genetic engineering, and drug discovery, among other applications. One of the unique features of FPs is their pH sensitivity, which has important implications for their use in research and biotechnology (Roberts et al., 2016).

I.4.1 Mechanism of pH sensitivity

The pH sensitivity of FPs is due to the properties of the chromophore, the part of the protein that absorbs and emits light (Shinoda et al., 2018). In the case of GFP, the chromophore is formed when three amino acid residues, serine, tyrosine, and glycine, undergo cyclization and oxidation. This process results in the forming a planar π-electron system that absorbs and emits light. The pH native intrinsic sensitivity of FPs is due to the protonation state of the chromophore. When the pH is low (acidic), the chromophore is protonated, and its properties change, resulting in different absorption and emission of wavelengths (Shinoda et al., 2018).

Conversely, when the pH is high (basic), the chromophore is deprotonated, and its properties change again.

The mechanisms of pH sensitivity vary among different types of FPs and go beyond the intrinsic mechanism. For example, GFP and its variants, such as Superfolder GFP (sfGFP), exhibit pH sensitivity due to the ionization state of the chromophore. The chromophore is protonated at low pH, and the absorption and emission wavelengths shift towards shorter wavelengths, resulting in a blue shift in fluorescence. The chromophore is deprotonated at high pH, and the absorption and emission wavelengths shift towards longer wavelengths resulting in a red shift in fluorescence (Shaner et al., 2005). Other FPs, such as pHluorins (GFP variant), exhibit pH sensitivity due to conformational changes that occur in response to changes in pH (Mahon, 2011). pHluorins are engineered to contain a pH-sensitive domain, which undergoes a structural change when the pH changes. This change results in alterations in the fluorescent properties of the protein (Mahon, 2011).

I.4.2 Applications and limitations

The pH sensitivity of FP has numerous applications in biological research and biotechnology. One of the most significant applications is in bioimaging. The pH sensitivity of FPs allows researchers to visualize changes in pH within cells and tissues. For example, pHluorin has been used to visualize changes in pH within synaptic vesicles, which play a critical role in neuronal signalling. By visualizing changes in pH, researchers can gain insights into the mechanisms of important cellular processes, such as signalling and metabolism. Additionally, this could be used to determine the protein's location inside the cell since not all the different cell regions have the same pH; thus, one could use that difference to test where a protein is inside the cell. Furthermore, with the help of specialized equipment, the difference in light wavelength emitted could be observed and converted to a location, which has not been done before.

Another application of the pH sensitivity of FPs is in drug discovery. Many drugs, including chemotherapeutic agents and antibiotics, exhibit pH-dependent activity. Understanding the pH sensitivity of FPs can help researchers design and optimize assays for drug screening. For example, researchers can use pH-sensitive FPs to measure the uptake and release of drugs within cells and monitor the pH changes that occur in response to drug treatment.

The pH sensitivity of FPs also has applications in synthetic biology. Synthetic biologists use FPs to design and construct new biological systems, such as biosensors and gene circuits. The pH sensitivity of FPs can be used to control the activity of these systems in response to changes in pH. In this way, researchers can use pH-sensitive FPs to design biosensors that detect changes in pH within the environment. These biosensors can be used to monitor environmental pollutants or detect specific pathogens (Senutovitch et al., 2015).

One of the main limitations of pH-sensitive FPs is their sensitivity to other factors, such as temperature, salt concentration, and oxygen levels (Prescher and Contag, 2010). Changes in these factors can affect the pH sensitivity of the protein, leading to inaccurate measurements. In addition, some FPs exhibit pH sensitivity over a narrow pH range, making them less useful for applications requiring measurements in multiple compartments or if the pH is unknown. For these reasons, it is important to carefully characterize the pH sensitivity of FPs under specific experimental conditions before using them for research or biotechnological applications (Roberts et al., 2016).

Another limitation of pH-sensitive FPs is their potential to interfere with cellular processes. The expression of FPs can affect the behaviour of cells, including their growth and metabolism. This interference can sometimes be minimized using low protein expression levels or inducible expression systems that allow the protein to be expressed only when needed.

Finally, it is important to note that pH-sensitive FPs are not the only tools available for measuring pH in cells and tissues. Other methods, such as pH-sensitive dyes and electrodes, have advantages and limitations (Zhang et al., 2013). The choice of method depends on the specific application and experimental conditions.

I.5 Hypothesis and Aims

Hypothesis: A new method based on dynamic location measurements can reveal the topology of the *E. coli* inner membrane proteome.

Rationale: We propose to extend topology analysis methods by using a sole pluripotent FP capable of expressing in the cytoplasm and periplasm. As shown in **Figure 2**, sfGFP will be employed in a novel dynamic measurement by applying pH step changes to induce changes in fluorescence intensity. These changes will allow distinguishing intracellular versus periplasmic location of terminus fused to sfGFP as only in the latter case a persistent change in fluorescence is produced (Dammeyer and Tinnefeld. 2012 & Roberts et al., 2016). This will allow us to determine with high precision the location of that respective terminus of the IMP.

Aims:

1. Design and test a low-copy expression system for sfGFP-tagged membrane proteins.
2. Develop a 'pH jump' method to test how sfGFP fluorescence is affected by a change in pH over time.
3. Apply the emerging localization method to ten proteins with known orientation.

Aim 1: We will obtain synthetic DNA sequences for the desired plasmid components and assemble the complete plasmid using ligation and transformation methods. Plasmid DNA will be purified and sequenced to confirm DNA sequence and integrity. Once we are confident that the plasmid is correct, we will use a fluorescent microscope to test the presence of sfGFP in *E. coli* cells under inducing conditions in the low-copy plasmid.

Aim 2: We will perform ‘pH jumps’ to test how pH affects sfGFP fluorescence. We will dilute transformed *E. coli* liquid cultures and add a mixture of buffer and HCl. Using sensitive plate reader equipment, we will measure how this change in pH affects fluorescence. Based on the literature, we should see a fast decrease in intensity initially. The intensity should then recover if the protein is located in the cytoplasm or remain reduced if it is located in the periplasm.

Aim 3: Once this method is robustly established, we will analyze ten pre-selected proteins with a known orientation. We will attach sfGFP to each of them and test them individually using the method developed in Aim 2. To ensure that the delay in measuring the first and last samples does not causing an adverse effect or result in missing the pH jump, we will add the same protein in multiple wells and expect the same results. Once all the proteins are tested, we will compare the results to the literature.

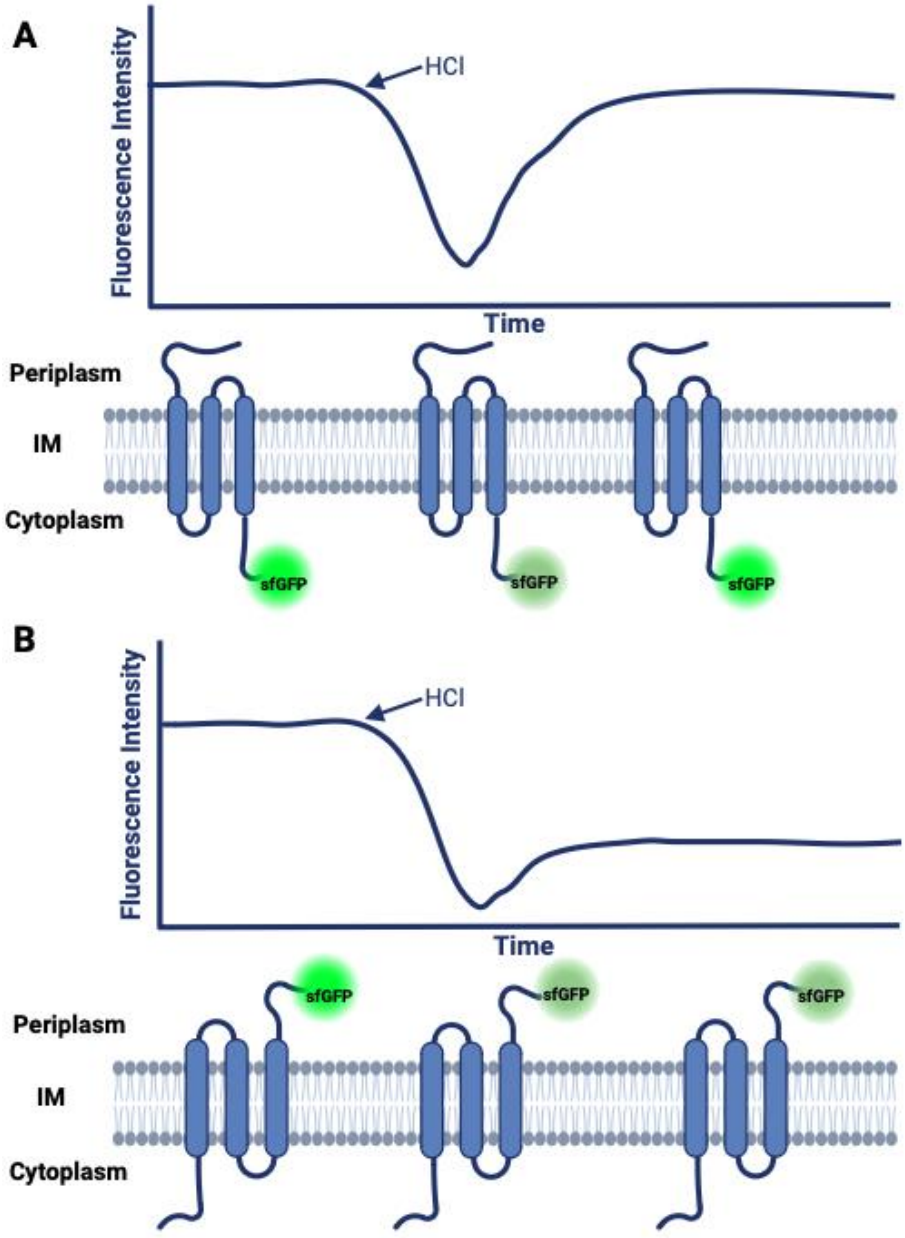

Figure 2: Cartoon showing the change in fluorescence of sfGFP when the external pH changes. A) Cytoplasm: The cell recovers from pH change. B) Periplasm: The cell is unable to recover. Created in Biorender by Daniel Carrillo Baez

Chapter II: Materials and Methods

II.1 LB Media (Miller)

LB (Miller) media was prepared by combining 3.88 g of Luria Broth (HiMedia Labs, M1725) with 250 mL Milli-Q water in a 500 mL container (Anko brand). Then, 2.375 g of sodium Chloride (Sigma, S9888) and 47.6 mg of magnesium chloride (Sigma, M8266) were added to improve plasmid DNA concentrations. The mixture was autoclaved for 20 minutes at 120°C.

II.2 Transforming and culturing bacteria for DNA amplification

NEB 5α competent *E. coli* cells (New England Biolabs, C2987H) were thawed from -80°C on ice for 10 minutes. Subsequently, the DNA was added at a volume of less than 2% of the bacteria volume cells and gently stirred in one circle; for the negative control, Milli-Q water was added instead. The cells were then further incubated for 30 to 45 minutes on ice. The transformed bacteria were streaked onto an agar plate prepared with ampicillin (INDOFINE Chemicals, A0104) with a 1:1000 ratio and incubated overnight (O/N) at 37°C. The following day, a colony was picked and transferred to a liquid culture containing a 1000:1 ratio of LB media to AMP and incubated for 12 to 16 hours at 37°C with shaking (230 rpm). Lastly, plasmid DNA was extracted from the liquid culture following the mini-prep protocol below.

II.3 Plasmid DNA Preparation

The bacteria liquid culture was centrifugated at 12,000 rpm for 2 minutes using a Fresco 17 microcentrifuge (Thermo Fisher Scientific, THR75002420). Plasmid DNA was extracted with a low-copy plasmid miniprep kit (Bio Basic, BS4149.) following the manufacturer's instructions. In summary, pelleted bacteria were lysed, and proteins and unsheared genomic DNA were removed. Subsequently, plasmid DNA was purified using an EZ-10 Column and eluted using 40 µL Milli-Q water. The DNA concentration and purity were determined using

a Nanodrop 2000C spectrophotometer (Thermo Fisher Scientific, ND-2000), and the size of the plasmid was confirmed through analytical digest and agarose gel electrophoresis. All sequencing was done by the Australian Genome Research Facility (AGRF).

II.4 Agarose gel electrophoresis

A 1% agarose gel was prepared by dissolving half a SeeGree All-in-One Agarose Tab (miniPCR, RG-1500-20) in 20mL of Milli-Q water and heating it in a microwave. Once the gel solidified, it was submerged in 1xTBE (Sigma, T4415) in a blueGel electrophoresis system (miniPCR, QP-1500-10) with a built-in transilluminator. The DNA samples were prepared in the following concentrations: Two parts DNA at a final amount of 500 ng, four parts 5x loading dye and 14 parts Milli-Q water. We used a 1 KB Plus DNA ladder (NEB, N3232) as the first and/or last lane marker. DNA was electrophoresed for 20 to 30 minutes at 90 V and 120 mA and viewed using the light source of the machine.

II.5 Restriction digest

Following the plasmid DNA preparation, a 20 µL total volume digestion was carried out as in the following: 500 ng of DNA, 0.5 to 1.0 µL of the respective restriction enzyme (NEB), 2.0 µL of 10X CutSmart buffer (NEB, B6004S) and the remaining volume was Milli-Q water. The reaction was incubated at 37°C for 15 to 90 minutes. If digestion targets the backbone for a ligation, after 1 hour of digestion, 0.8 µL of CiP (NEB, M0525) were added and then reaction incubated for an additional hour to dephosphorylate the backbone.

II.6 PCR extraction

A PCR extraction kit (Sangon LifeScience, B518191-0100) was used to purify the DNA PCR product according to the manufacturer’s instructions. In brief, this protocol uses a 5X volume

of buffer to the digestion reaction and a wash solution to remove the enzymes and buffer components. Lastly, the DNA is eluted by adding 40 µL of Milli-Q water.

II.7 Standard ligation

The total ligation volume was 20 µL, starting with 30 ng of backbone DNA and using a >3-fold molar ratio of insert to backbone. The reaction contained 1 µL of T4 ligase (Promega, M1804) and 2 µL of 10X T4 buffer (Promega), and the remaining volume was filled with Milli-Q water. To evaluate the performance of the ligation, two control reactions were included: Reaction (and control agar plate) "A" contained only backbone DNA, without ligase and insert, and reported the number of colonies from undigested plasmid. Reaction (and control agar plate) "B" contained backbone DNA ligase but no insert and reported the number of colonies from undigested plasmid and plasmid self-ligation. Reaction (and agar plate) "C" contained colonies from undigested plasmid, plasmid self-ligation, and successful ligation with insert. The ratio between control plates and plate C had to be at least threefold to proceed. Finally, six to ten colonies from plate C were picked and digested with the appropriate restriction enzyme for subsequent analysis using gel electrophoresis. If the required bands were present and around the expected size, they were prepared for sequencing to test for mutations (AGRF).

II.8 Chloramphenicol treatment

Chloramphenicol (Sigma, C0378) stocks were prepared at a concentration of 25 mg/mL and stored at -20°C. The standard LB media growth protocol was followed on the first day with AMP at a ratio of 1:1000 was added. A colony was picked and incubated O/N at 37°C. The following day, new liquid cultures were made, consisting of 80% fresh LB media with AMP and 20% of the O/N culture. Chloramphenicol was added to this liquid culture at a ratio of 2500:1 and incubated O/N (Maniatis et al., 1989 & Frenkel and Bremer, 1986). The plasmid

DNA was extracted from the samples following the standard protocol on the third day, and genomic DNA content was assessed using gel electrophoresis.

II.9 Exonuclease V treatment

The plasmid DNA preparation protocol was used to obtain a purified DNA sample. In a final reaction volume of 20 µL, 150 ng of DNA sample, 1 µL of Exonuclease V (RecBCD, New England Biolabs, M0345), 2 µL of 10 mM ATP, 2 µL of NEBuffer 4 (10X) were combined, and the remaining volume was Milli-Q water. First, the reaction was incubated at 37°C for 1 hour and then Exonuclease V heat inactivated by incubating at 70°C for 30 minutes.

II.10 Induction of sfGFP

DNA of the plasmid containing sfGFP was transformed into 50 µL of NEB 5α competent *E. coli* cells, following the standard protocol, except for an additional step. Once the O/N culture was ready, 100 µL was transferred into a new tube containing 2 mL of fresh LB media, 2 µL of AMP and 20 µL of rhamnose (Sigma, R3875) in a 0.2% w/v stock solution in water). The tubes were then incubated the tubes at 37°C with shaking (230 rpm) for 4 hours.

II.11 Cell preparation for plate reader and fluorescence microscopy

200 µL of cells from the 4-hour incubation with rhamnose were pelleted and resuspended into 500 µL PBS (Thermo Fischer Scientific, 20012050). Next, a thin agarose pad was created in a cast that was assembled using five standard glass slides. Three glass slides were placed side by side with their edges touching; the remaining two slides were positioned on top, leaving a small gap between their longer edges aligned with the middle of the centre slide below. 0.5 mL LB agarose was added into the gap and solidified. Once ready, the slides were removed and cut into small rectangles of agar (5 x 5 mm). 2 µL of bacterial solution was placed onto a slide and

carefully covered by the agarose pad. Finally, the samples were imaged using an inverted fluorescence microscope (Thermo Fischer Scientific, EvosFL) with a 40X or 60X objective.

II.12 Testing change in fluorescence over time (pH jump)

After pelleting the induced sfGFP cells at 200 G for 2 min, they were resuspended to an OD600 of 1 in PBSG. Subsequently, 90 µL of the cell suspension was added to each well of a white flat-bottom plate (Costar, 3917). The plate reader (BMG Labtech, CLARIOstar) was preheated to 30°C. The fluorescence measurement protocol was set with 20 flashes per reading, 3 second intervals, and 80 data points. At 42 seconds, 1.5 µL of HCl (concentration 85 mM) was injected using the plate reader's injectors, followed by shaking at 200 rpm for 3 seconds.

II.13 Statistical analysis

The number of samples is specified in each figure caption, along with a definition of error bars. Statistical tests (Student’s t-test) could not be performed because there were not sufficient replicates.

Chapter III: Results

III.1 Design of expression vector

The design of the expression vector began with a pre-existing plasmid in the laboratory, pUC57-AMP, featuring a high copy number of ColE1/pMB1origen of replication (OriR) and the gene for AMP resistance. To reduce the replication capacity, an essential factor for correct membrane protein expression (Schlegel et al., 2012), the OriR was replaced with colDF13 (Dietler et al., 2021). Subsequently, the *E. coli* rhaB operon promoter (of plasmid pWA21) was added to regulate gene expression in response to L-rhamnose presence (Wegerer et al., 2008). Rhamnose was chosen as the inducer to enable compatibility with other expression systems for future work (e.g., those sensitive to arabinose or Isopropyl β- d-1thiogalactopyranoside, IPTG). To ensure a proper transcription termination in the *E. coli*, the B0010 terminator, derived from the rrnB gene (iGem), was incorporated. An optimized sfGFP variant was also used, lacking the BseRI, HindIII and BbsI internal sites, thereby eliminating Type II restriction enzyme recognition sites.

The inclusion of the GFP facilitated the visualization and monitoring of gene activity in subsequent experiments. Furthermore, the start codon of the sfGFP was replaced with a NdeI site (CATATG) to allow the insertion of the IMP in later stages of the project. A TGA stop codon was used to stop the formation of the protein, and a second stop codon (TAG) was included to be used in future experiments to measure non-natural amino acid read-through with a His6-tag.

To facilitate plasmid assembly, a NheI-HF restriction site was positioned before the RhaB promoter, and a SalI-HF restriction site was added after the terminator. Finally, the plasmid was examined to ensure the presence of all necessary restriction sites only once, and no extra Type II restriction sites were present.

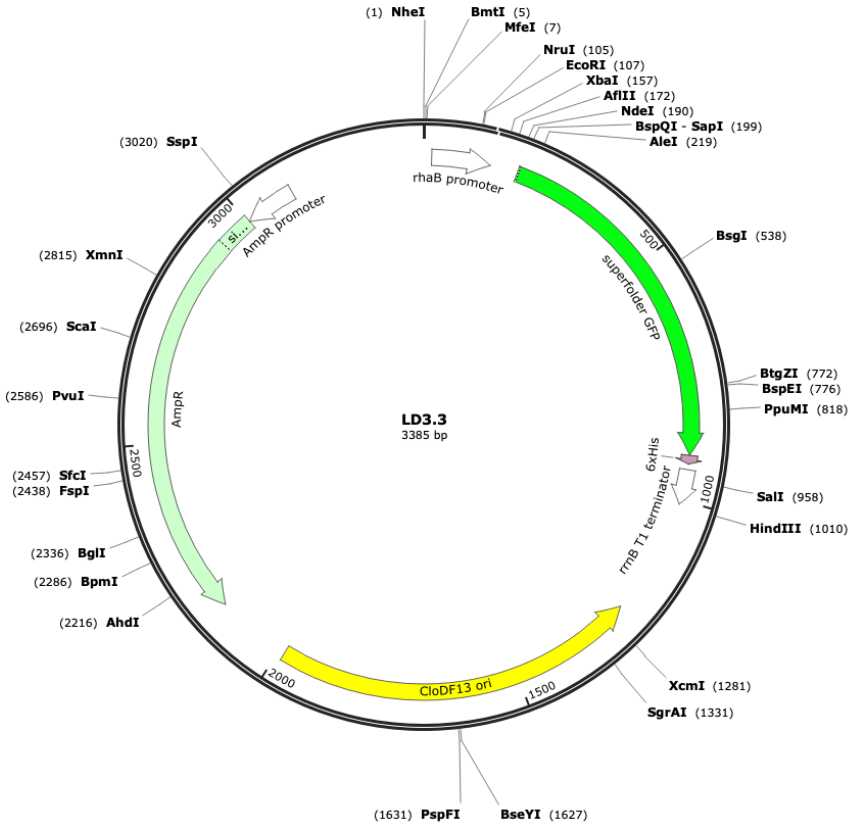

Figure 3: Plasmid map of low-copy expression plasmid LD3.3: The elements highlighted are Amp resistance, ColDF13 origin of replication, rhaB promoter, sfGFP gene and rrnB T1 terminator. Figure made using SnapGene Viewer.

III.2 Assembly of an initial low-copy vector

The complete plasmid described in Section III.1 was divided into two fragments: 7A, consisting of the rhaB promoter, sfGFP and the rrnB T1 terminator, and 7B, which contained the CloDF13 OriR, AmpR promoter and Amp resistance gene. As shown in **Figure 3**, 7A starts at the first based pair (bp) with the NheI restriction site and extends to 958 bp with the SalI restriction site, while 7B spans from 958 bp to 3385 bp.

To ensure proper plasmid assembly, both fragments were ordered as gBlocks from Integrated DNA Technologies (IDT), requiring each fragment to begin with an NheI restriction side and

end with a SalI restriction site. Additionally, 3-4 bp were added at the start and end of each gBlocks to facilitate correct binding of and cutting by the restriction enzyme.

The two gBlocks underwent double digesting with NheI-HF and SalI-HF ad ligated. To confirm the success of the ligation, the new plasmid was transformed into NEB 5α competent *E. coli* cells and plated on an agar plate containing Amp antibiotic. **Figure 4** displays the results of this agar plate, containing only the cells with the Amp resistance gene, resulting in approximately 1000 colonies. This confirms the success of the ligation and correct functionality of the Amp resistance gene. Additionally, the plasmid was sequenced to detect any potential mutations.

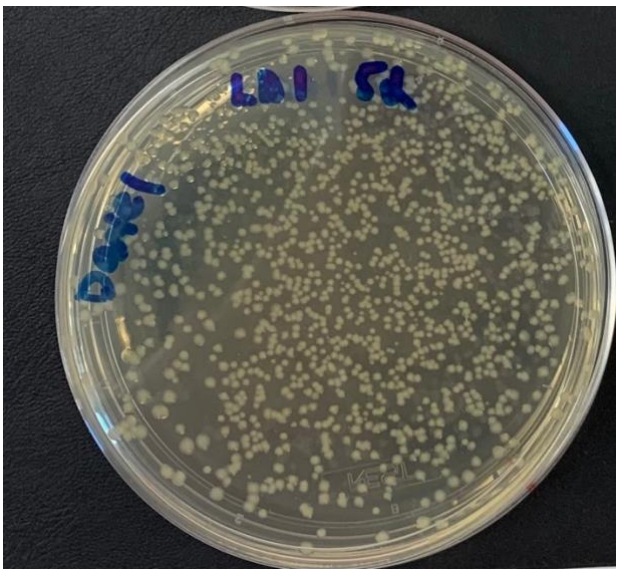

Figure 4: Agar plate with ampicillin containing NEB 5α competent *E. coli* cells with the LD3.3 plasmid containing the AMP resistance gene.

III.3 Increasing yield of low-copy vector.

Isolation of sufficient DNA yield of a low-copy vector presents a critical challenge in molecular biology, particularly when obtaining higher concentrations of plasmid DNA is necessary for future steps. Initially, the plasmid LD3.3 was subject to miniprep using the conventional kit, resulting in DNA concentrations ranging from 6.8 ng/µL to 37.5 ng/µL. However, many samples contained a substantial amount of genomic DNA, as judged using gel electrophoresis

(see below), posing challenges for future experiments. To address the issues of low DNA concentrations and high amounts of genomic DNA, three approaches were implemented: low concentrations of chloramphenicol, exonuclease V, and a specialized kit for the purification of low-copy plasmids.

In the first experiment, three samples of LD3.3 were treated with and without 2% chloramphenicol. After O/N incubation at 37°C with shaking, the samples treated with 0% chloramphenicol showed an OD600 of approximately 6.5 times higher than those treated with 2% chloramphenicol, as shown in **Table 3.** Following the miniprep of all samples, those containing no chloramphenicol displayed twice the amount of DNA compared to those treated with 2% chloramphenicol. Furthermore, gel electrophoresis analysis revealed more pronounced bands in the samples with no chloramphenicol. This indicated that while chloramphenicol was able to increase DNA per cell amounts, the reduced cell number limited the usefulness of this approach.

Table 1: First chloramphenicol experiment comparing how a 2% chloramphenicol addition impacts OD600 and DNA concentration. All samples originated from the same agar plate and were incubated at 37°C with shaking. Single representative experiment without statistical analysis.

Sample	Chloramphenicol (%)	OD600	[DNA] ng/µL
1	0	2.64	37.9
	2	0.35	15.6
2	0	3.4	44.8
	2	0.51	22.4
3	0	2.67	33.3
	2	0.48	24.6

Thus, in a second experiment, four samples were treated with and without 0.4% a reduced chloramphenicol concentration (Frenkel and Bremer, 1986). This yielded a similar trend as the

one observed in **Table 1**. The samples with no chloramphenicol had a higher OD600 value, approximately twice that of the samples treated with 0.4% chloramphenicol. however, this experiment resulted in a greater DNA concentration overall. As shown in **Figure 5,** two replicates, sample 1 and sample 2, exhibited a higher DNA concentration in the presence of 0.4% chloramphenicol. While sample 3 had the most significant difference in DNA concentrations, followed by sample 1. However, replicates sample 2 and sample 4 showed similar DNA concentrations, with the 0.4% chloramphenicol samples having a smaller DNA concentration than the 0% chloramphenicol sample. Thus, low amounts of chloramphenicol can increase the DNA concentration in some cases. However, this increase is not enough, resulting in different approaches being used.

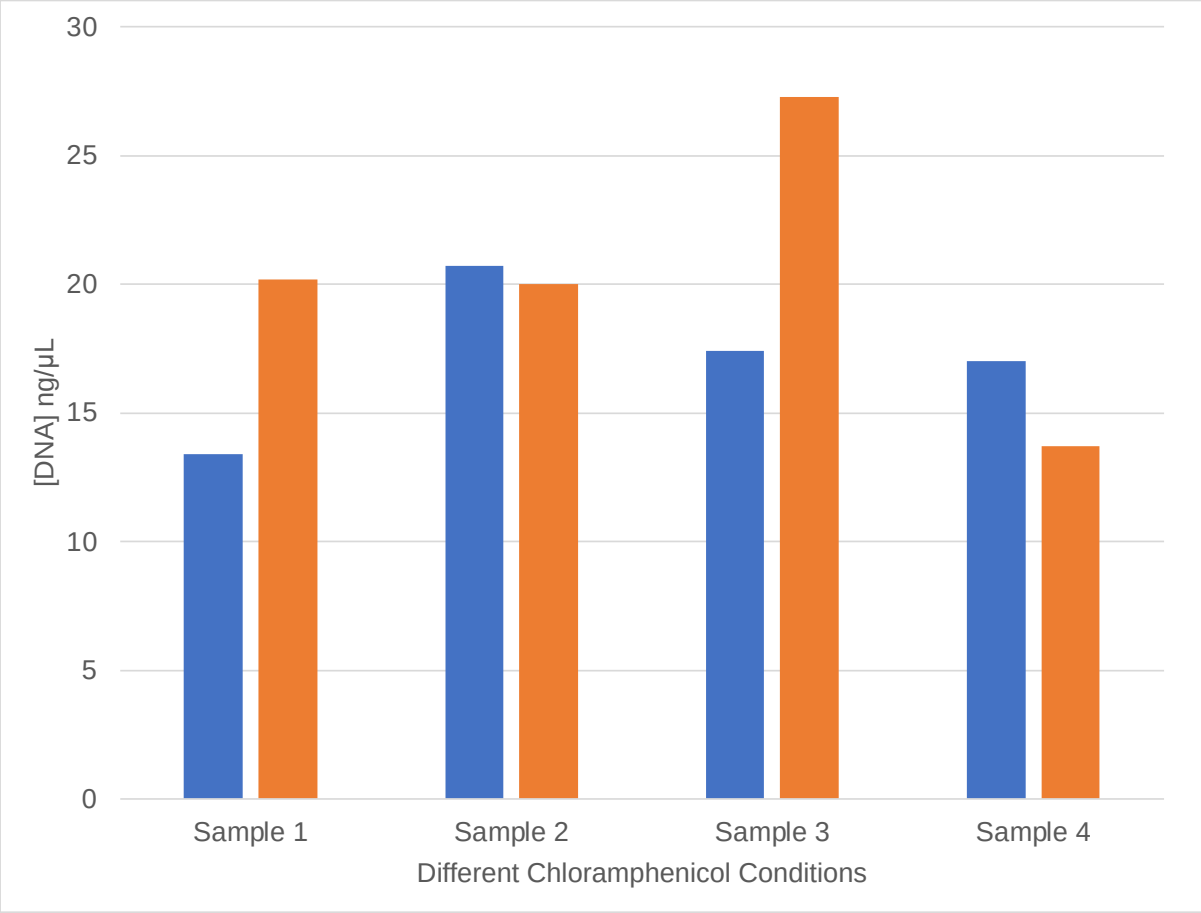

Figure 5: Impact of chloramphenicol on DNA concentrations (ng/µL). Blue bars represent no addition of chloramphenicol, while orange bars represent 0.4%. All four replicates originated

from the same O/N liquid culture and were incubated O/N with or without chloramphenicol at 37°C with shaking. Single representative experiment without statistical analysis.

The second approach to address the issues of low plasmid DNA concentrations and high amounts of genomic DNA was to employ the enzyme complex RecBCD, also known as Exonuclease V, which especially degrades genomic DNA. Four samples were selected for this experiment: two contained plasmid DNA and some genomic DNA, while the others contained only genomic DNA (prepared by selecting the worst samples from the overall sample collection).

As shown in the top gel of **Figure 6,** all four samples exhibited the presence of genomic DNA, with the first two samples displaying a band representing the LD3.3 plasmid at approximately 3.3 kb. However, after incubating the samples for one hour with Exonuclease V, the lower gel revealed a significant change. The first two samples no longer contain visible genomic DNA, whilst plasmid DNA remained intact, as evidenced by nearly identical band intensity compared to the top gel. Furthermore, the bottom gel displays no visible genomic DNA when examining the genomic DNA samples. In summary, Exonuclease V effectively degraded genomic DNA while maintaining plasmid DNA. However, this does not address the issue of low DNA concentration but successfully reduces genomic DNA which may pose a confounding factor in DNA sequencing.

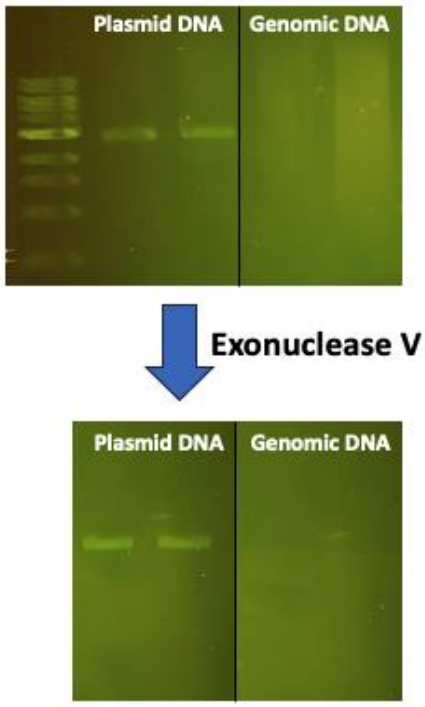

Figure 6: Gel Electrophoresis showing the before and after exonuclease V treatment. 150 ng of DNA was mixed with 1 µL of Exonuclease V (RecBCD), 2 µL of 10 mM ATP, and 2 µL of NEBuffer 4 (10X) and was incubated for 1 hour at 37°C.

To address the issue of low plasmid DNA concentrations, two purification kits were compared to determine their impact on DNA yield. Ten samples were purified using the conventional Sangon kit, while another set of 10 samples underwent purification using the low-copy specialized kit. As shown in **Figure 7,** the Sangon purification kit yielded a maximum DNA concentration of 37.5 ng/µL, a minimum value of 6.8 ng/µL, a median of 15.3 ng/µL and an average of 17.7 ng/µL. Additionally, 50% of the values fell into the range of 10.3 ng/µL to 23.2 ng/µL. In contrast, the BioBasic purification kit yielded a maximum DNA concentration of 59.6 ng/µL, a minimum value of 22.4 ng/µL, a median of 35.2 ng/µL and an average of 40.0 ng/µL. Additionally, 50% of the values fell into the range of 24.9 ng/µL to 47.8 ng/µL. It is worth noting that the BioBasic Box plot contained one outlier with a concentration of 87.0 ng/µL. This result shows an improvement of 2X more DNA concentration using the BioBasic specialized kit. However, even the final DNA concentration achieved was found to be limited when it comes to obtaining sufficient material for future cloning and sequencing reactions.

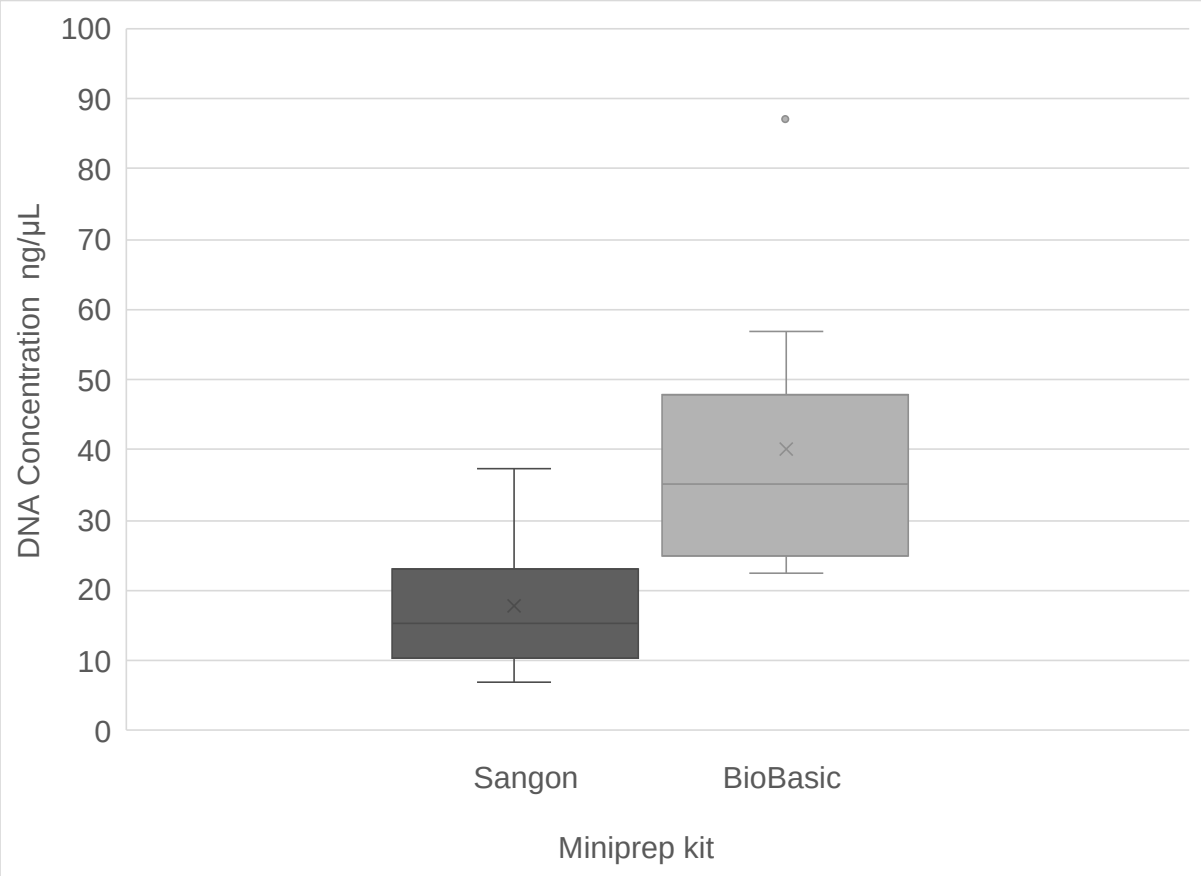

Figure 7: Box plot comparing the DNA concentration of 10 samples using the Sangon Biotech Sanprep Plasmid Extraction Kit or low-copy plasmid miniprep kit from BioBasic Inc. The line inside the box represents the sample median, and the X represents the sample average. Error bars indicate SD (n= 10).

III.4 Medium copy plasmid through modification of the origin of replication

In the next step and to increase DNA yield, the plasmid's OriR was modified, transforming it from a low-copy plasmid into a medium-copy plasmid. This significantly increased plasmid DNA concentration per preparation, yielding approximately 1.75 times more plasmid DNA. LD3.3, the original plasmid design in section III.1, featured the low-copy OriR ColDF13 from pCDF-Duet. In contrast, LD4.2 is a medium-copy plasmid containing the same components as LD3.3, except for its high-copy OriR from pUC57. However, when the new OriR was incorporated into the plasmid, it exhibited characteristics of a medium-copy plasmid, as indicated by DNA concentrations that were significantly improved compared to LD3.3 but below those normally obtained for high-copy. **Figure 8** illustrates the effect of the two

replication origins on plasmid DNA concentrations. Plasmid LD3.3 has the same values as the box plot of BioBasic from **Figure 7** because data were taken from the same experiment. In contrast, the plasmid LD4.2 yielded an enhanced performance, with a maximum DNA concentration of 97.0 ng/µL, a minimum value of 39.0 ng/µL, a median of 70.4 ng/µL and an average of 70.5 ng/µL. Furthermore, 50% of the values fell into the range of 56.7 ng/µL to 86.5 ng/µL. It is important to note that LD4.2 had an average of approximately 60% higher yield than LD3.3. While LD4.2 demonstrated increased variability in DNA concentrations compared to LD3.3, it is worth highlighting that 75% of the values of LD4.2 are higher than the maximum value of LD3.3. This transformation of the plasmid's OriR resulted in a significant improvement in plasmid DNA concentration with the desirable feature of achieving medium copy levels that should be compatible of IMP expression.

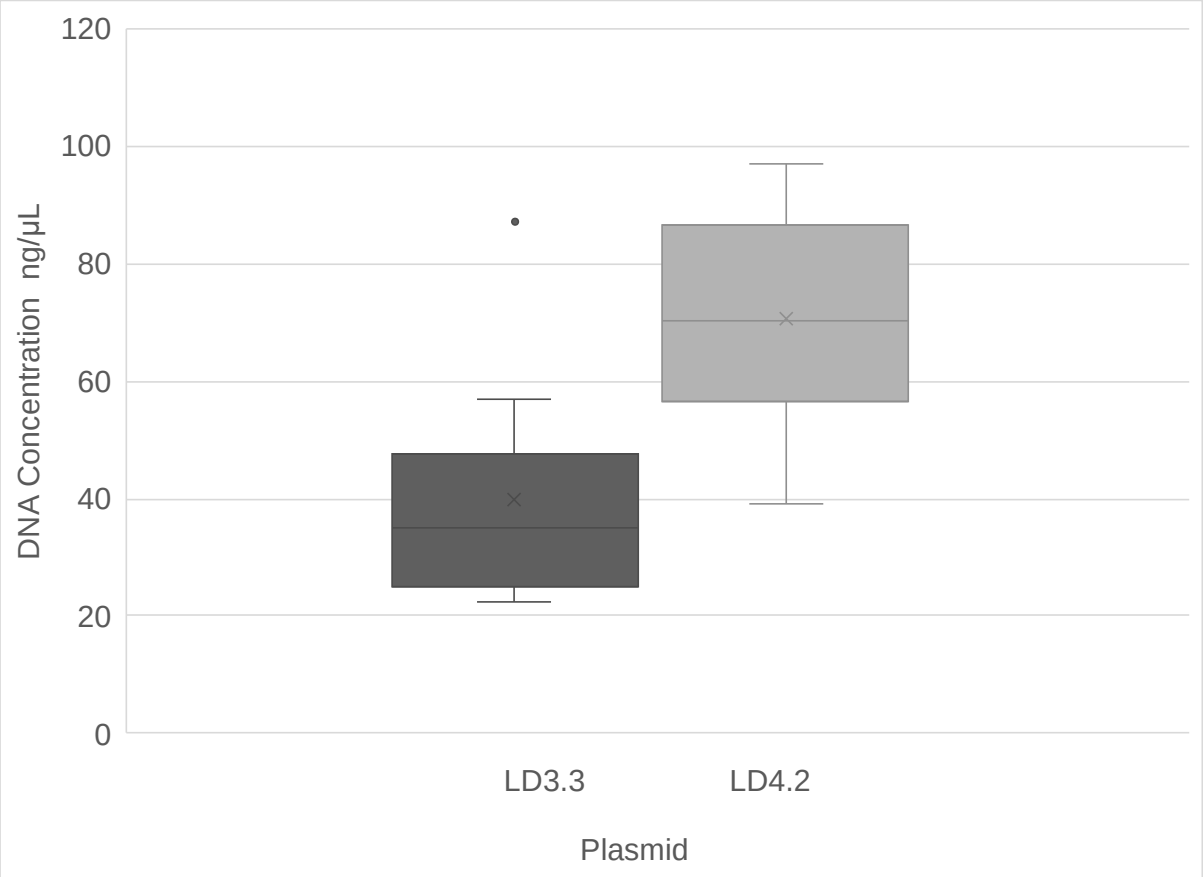

Figure 8: Box plot comparing the DNA concentrations of 10 samples containing the LD3.3 low copy plasmid and LD4.2 medium copy plasmid. All samples were miniprep using the low-

copy plasmid miniprep kit from BioBasic Inc. The line inside the box represents the sample median, and the X represents the sample average. Error bars indicate SD (n= 10).

III.5 Calibration of pH solutions

The calibration of the pH of all solutions used in this experiment is crucial to ensure reproducibility of the experimental conditions. The first step was to ensure the working concentration of HCl was 85mM. According to the literature (Wilks and Slonczewski, 2007), cells need to achieve at a minimum a pH of 5.5 to recover or not depending on the environment of sfGFP; thus, all solutions were adjusted to reach this final pH. The first solution calibrated was PBS, with a pH of 7.5. To achieve the desired pH of 5.5, 30 mL of PBS required the addition of 0.950 mL of HCl, resulting in a ratio of approximately 31.6 to 1. In preliminary experiments, it was observed that the cells required an external energy source for recovery. Therefore, 0.8% of glycerol was added to PBS, creating PBSG. This new solution had an initial pH of 7.3. To adjust it to the desired pH of 5.5, it required the addition of 0.899 mL of HCl, yielding a ratio of approximately 33.4 to 1.

The next calibration involved PBSG and cells, which may impact the effects of HCl addition on pH. 2 mL of cells with an OD600 of 3.09 were pelleted and resuspended in PBSG. This mixture had an initial pH of 7.1, and to reach the desired pH of 5.5, it required the addition of 0.860 mL of HCl, giving a ratio of approximately 34.9 to 1. Therefore, it can be concluded that when using PBSG and cells, the maximum theoretical amount of HCl that can be added to a solution of 90 µL, which is the final reaction volume in the plate reader, while allowing the cells to recover is 2.66 µL.

III.6 Parametrization of pH jump traces.

pH step traces were recorded in a microplate reader. In each experiment, a large amount of raw fluorescence data is recorded and needs to be organized and analyzed efficiently. **Figure 9**

illustrates a cartoon pH jump curve, i.e. a plot of sfGFP fluorescence intensity before and after the addition of HCl over a set time, allowing the visualize the recovery of fluorescence. Three key intervals have been highlighted in the curve: A is the average of the last five points before adding HCl (one point corresponds to three seconds), B is the average of the last five points after a set amount of time upon addition of HCl, and C is the minimum point on the curve after addition of HCl. Using those three numbers, the recovery percentage can be calculated using **Equation 1,** which states that the drop between the initial and the minimum fluorescence is divided by the gap between the initial and final fluorescence. This result is then subtracted from one and multiplied by a hundred to obtain the percentage.

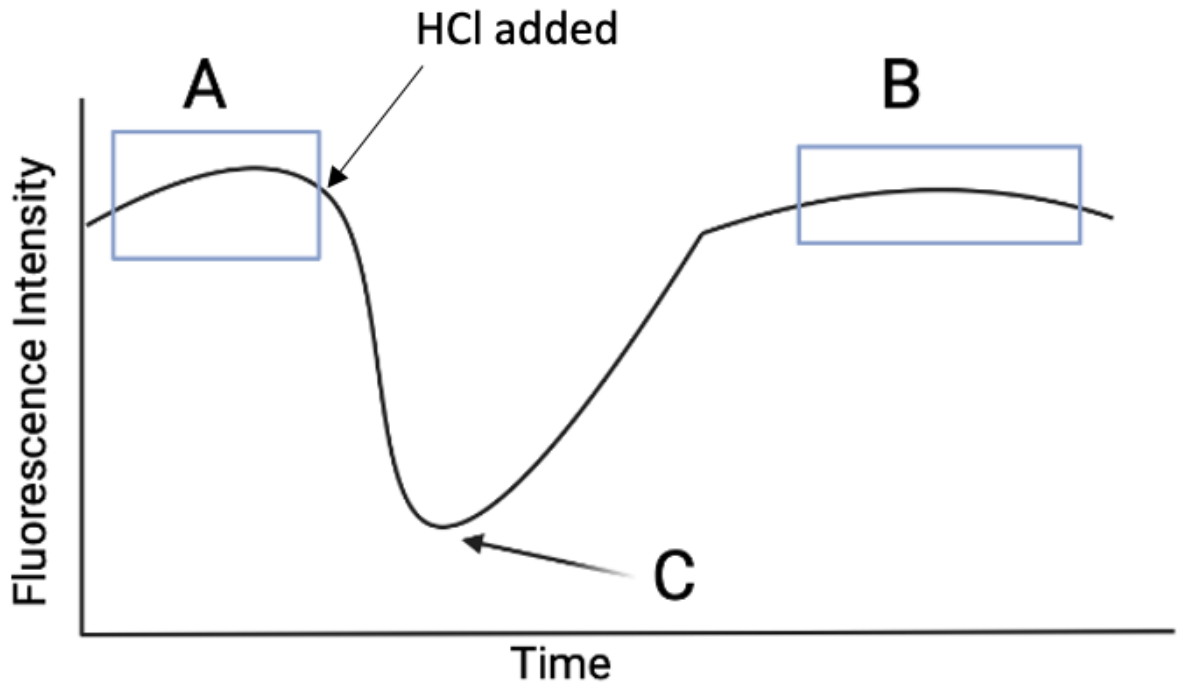

Figure 9: pH-jump curve showing the labels for the formula used to calculate fluorescence recovery. A is the average of five points before the addition of HCl. B is the average of five points after a set time post-HCl addition. C is the minimum point in the curve.

$$Recovery\ (\%) = 1 - \frac{A - C}{A - B} 100$$

Equation 1: Formula used to calculate fluorescence recovery in a pH-jump curve.

III.7 Optimization of cell number and HCl Volumes

Determining the number of cells (expressed as OD600) and amount of HCl needed to obtain the highest difference between recovery and no recovery is fundamental to obtaining precise results. These parameters need to be optimized because having consistent conditions is fundamental to ensure that all proteins undergo the exact change in pH. Since sfGFP is naturally expressed in the cytoplasm, it is very simple to test for "recovery"; however, testing for "no recovery" requires acetate (Wilks and Slonczewski, 2007). Acetate is a weak acid that, when it enters the cell, prevents the regulation of intracellular pH, thereby simulating the pH conditions found in the periplasm. The first parameter analyzed was the number of cells; after a few experiments, the best results were obtained with the OD600s of 0.5, 1.0 and 1.5. Additionally, three different HCl volumes were tested: 1.5, 1.8 and 2.0 µL. Using **Equation 1,** recovery and no recovery were calculated and subtracted from each other to give the difference between recovery and no recovery, and this is shown in **Table 2.**

Since all the cells in all the samples had the LD4.2 plasmid, they should all recover close to 100%. When looking at the recovery column in **Table 2,** it can be seen that regardless of the OD600, the addition of 2.0 µL of HCl prevented the cells from recovering. Similarly, adding 1.8 µL of HCl to cells with an OD600 of 1.0 and 1.5 also prevented the cell from recovering. Additionally, not adding enough HCl or not having enough cells might cause the No Recovery column to recover, resulting in false results. This can be evident when adding 1.5 µL of HCl to cells with an OD600 of 0.5. However, the following combinations yield a high recovery and a low no recovery: OD600 of 1.0 plus 1.5 µL of HCl had a recovery of 112% and a no recovery of 44% yielding a difference of 69% , OD600 of 1.5 plus 1.5 µL of HCl had a recovery of 102% and a no recovery of 25% yielding a difference of 77% , and OD600 of 0.5 plus 1.8 µL of HCl had a recovery of 112% and a no recovery of 50% yielding a difference of 62%. However, this last one had an outlier because the two recovery samples are very different,

making it challenging to tell which one is the actual value. Thus, adding 1.5 µL of HCl seems to give the lowest no recovery while allowing the cell to recover to its original pH. The optimal OD600 is 1.5 since it had a difference of 77%; however, an OD600 of 1.0 also had a very high difference of 69%, making both candidates for future experiments.

Table 2: Systematic determination of optimal conditions for analyzing the recovery after adding HCl. All samples used the LD4.2 plasmid, and the OD600 reading was after resuspension in PBSG. For No Recovery, 90 µL of Acetate was added to simulate periplasmic conditions. Each sample was run in duplicates (n= 2) to limit outliers.

OD600/HCl	Recovery	No Recovery	Difference
0.5+1.5	101%	68%	34%
	103%	68%	
1.0+1.5	107%	44%	69%
	117%	44%	
1.5+1.5	101%	25%	77%
	104%	25%	
0.5+1.8	137%	50%	62%
	88%	50%	
1.0+1.8	81%	45%	27%
	79%	62%	
1.5+1.8	65%	34%	27%
	64%	40%	
0.5+2.0	37%	58%	-6%
	61%	52%	
1.0+2.0	26%	46%	-16%
	22%	34%	
1.5+2.0	24%	32%	-1%
	27%	20%	

III.8 Selection of ten test proteins

The selection of the test proteins was crucial to ensure a comprehensive evaluation of the proposed method. As shown in **Table 3,** ten diverse test proteins were selected, ensuring different topologies and lengths. The selection criteria aimed to encompass various orientations to test the hypothesis rigorously. First, a protein and sequence motif that are expected to

transport sfGFP to the periplasm were selected to investigate whether fluorescence recovery would occur: ssDsbA with a length of 44 AA and MalE with a length of 409 AA (Dinh and Bernhardt, 2011). In the next step, proteins with both C- and N-termini in the periplasm were selected, such as CyoA, MthK, MarC, and YnfA (Drew et al., 2006 & Luirink et al., 2012). Additionally, proteins with both the C- and N-terminus in the cytoplasm, such as YidC and TatC, were selected. Lastly, the remaining proteins were selected to ensure that they had one terminus in the cytoplasm and one in the periplasm; NuoK has the C-terminus in the periplasm and the N-terminus in the cytoplasm, while YgjV has the C-terminus in the cytoplasm and the N-terminus in the periplasm. Ultimately, by strategically choosing these, it was ensured that various orientations and lengths were present, thus providing a comprehensive assessment as possible with the selection of this size for the different IMPs present in *E. coli.*

Table 3: Characteristics of ten test proteins selected for pH-jump analysis, including C- and N-terminus localization and their length in AA.

Name	C-terminus is…	N-terminus is…	Source	Length (AA)
CyoA	Periplasm	Periplasm	Luirink et al.	328
NuoK	Periplasm	Cytoplasm	Uniprot	113
MthK	Periplasm	Periplasmic	Koprowski et al.	349
YidC	Cytoplasm	Cytoplasm	Uniprot	561
MarC	Periplasm	Periplasm	Drew et al.	234
TatC	Cytoplasm	Cytoplasm	Drew et al.	271
YgjV	Cytoplasm	Periplasm	Drew et al.	196
YnfA	Periplasm	Periplasm	Drew et al.	121
ssDsbA	Short tag for moving sfGFP into the PP		Dinh et al.	44
MalE	Tag for moving sfGFP into the pp		Dinh et al.	409

III.9 Time courses for test proteins

Whilst the parametrization has been introduced above, the optimal time for measuring fluorescence recovery (%) after the HCl addition had yet to be determined. This parameter is essential to determine the topology of the IMPs and has practical implications for an experimental expansion of the protein scope from ten to hundreds of proteins. The literature suggests fluorescence recovery for sfGFP can take up to 5 minutes (Wilks and Slonczewski, 2007), and this may be challenging in large-scale studies that aim at analyzing the approximately 900 IMPs from *E. coli*. In these studies, a reduction of the measurement time

would allow the assay to be as efficient and fast as possible, not only to collect more data more quickly but also to avoid a long idle time of cultures in the plate reader equipment. Therefore, two shorter analysis times were tested. As shown in **Figure 10,** the fluorescence recovery compared for two and three-minute measurements follows a linear trend (slope of near unity of 0.9822), indicating no significant difference between these two analysis parameters (the R-squared (R^2) value for the line of best fit was 0.9893). Thus, a two-minute analysis was chosen for the following experiments and is optimized for a large-scale analysis as well.

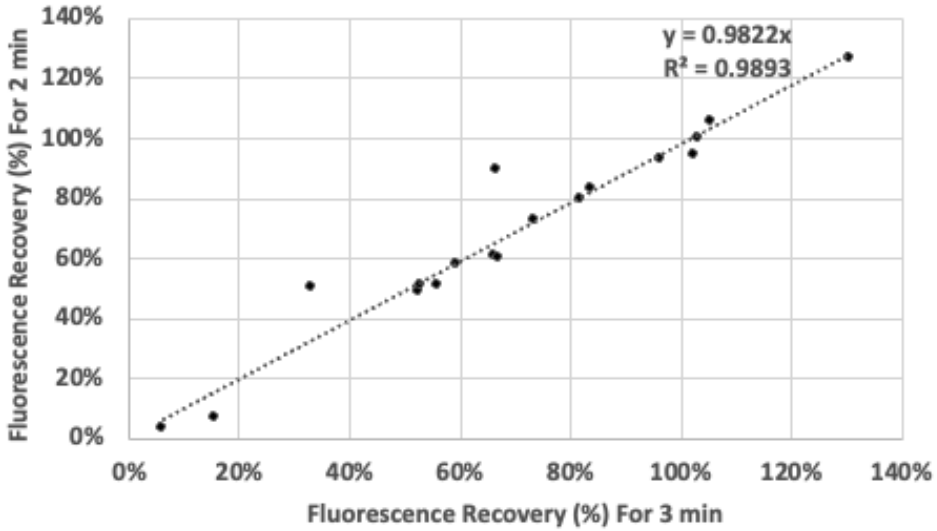

Figure 10: Comparison of sfGFP fluorescence recovery (%) at two and three minutes after the addition of 1.5 µL HCl. Each point is the average of two replicates.

Eight of the ten test proteins selected for cloning into the plasmid and testing were successfully incorporated. These were NuoK, MarC, MalE, ssDsbA, CyoA, YidC, TatC, and YgjV. Two proteins, namely MthK and YnfA, could not be successfully cloned for unclear reasons. As shown in **Table 4**, when comparing the fluorescence recovery (%) of the different proteins, it was observed that all had greater recovery at three minutes than at two minutes. As previously mentioned, this is due to the linear trend in fluorescence recovery as time increases. However, this trend stops around four minutes. Data not shown here determined that most of the proteins' fluorescence intensity reaches a plateau after four minutes of recovery. When closely examining **Table 4**, it can be seen that TatC had the highest fluorescence recovery with a value

of 118%, while NuoK displayed the smallest recovery with a value of 5%. Furthermore, the plasmid had a lower fluorescence recovery (93%) than the two proteins, YidC and TatC, inserted in the inner membrane. However, YgjV, predicted to have its C-terminus in the cytoplasm, exhibited a recovery rate of approximately 70%, putting it at the upper limit of the periplasmic proteins. Due to this, it was removed from **Figure 11** because further analysis is required to determine its actual orientation.

Table 4: Fluorescence recovery (%) for the eight test proteins and plasmid with three and two minutes of recovery time after adding HCl. C-terminus location is the predicted orientation for the protein based on the literature. Each fluorescent recovery (%) is the average of two samples (n=2).

Protein	C-terminus location	Fluorescence Recovery (%)	
		3 minutes	2 minutes
Plasmid	Cytoplasm	93%	89%
CyoA	Periplasm	74%	70%
NuoK	Periplasm	11%	5%
MalE	Periplasm	55%	50%
MarC	Periplasm	54%	51%
YidC	Cytoplasm	100%	97%
TatC	Cytoplasm	118%	116%
ssDsbA	Periplasm	63%	61%
YgjV	Cytoplasm	63%	74%

Figure 11 compares the fluorescence recovery (%) at two or three minutes after adding 1.5 µL of HCl. Part A of the figure shows the distribution after two minutes, the chosen parameter. A distinct gap is observed between the proteins localized in the periplasm and the cytoplasm. Strikingly, the cytoplasmic proteins consistently recovered above 80%, while the periplasmic proteins exhibited recoveries below this threshold. Furthermore, the periplasmic proteins displayed a broader recovery range; the recovery range was from 5% to 70% for two minutes. In contrast, the cytoplasmic proteins recovered from 89% to 116% for two minutes. Thus, a definition for recovery can be created, indicating that any fluorescence higher than 80% is

considered recovery and will ultimately indicate the presence of sfGFP in the cytoplasm. In contrast, anything below 80% indicates the contrary and confirms the presence of sfGFP in the periplasm.

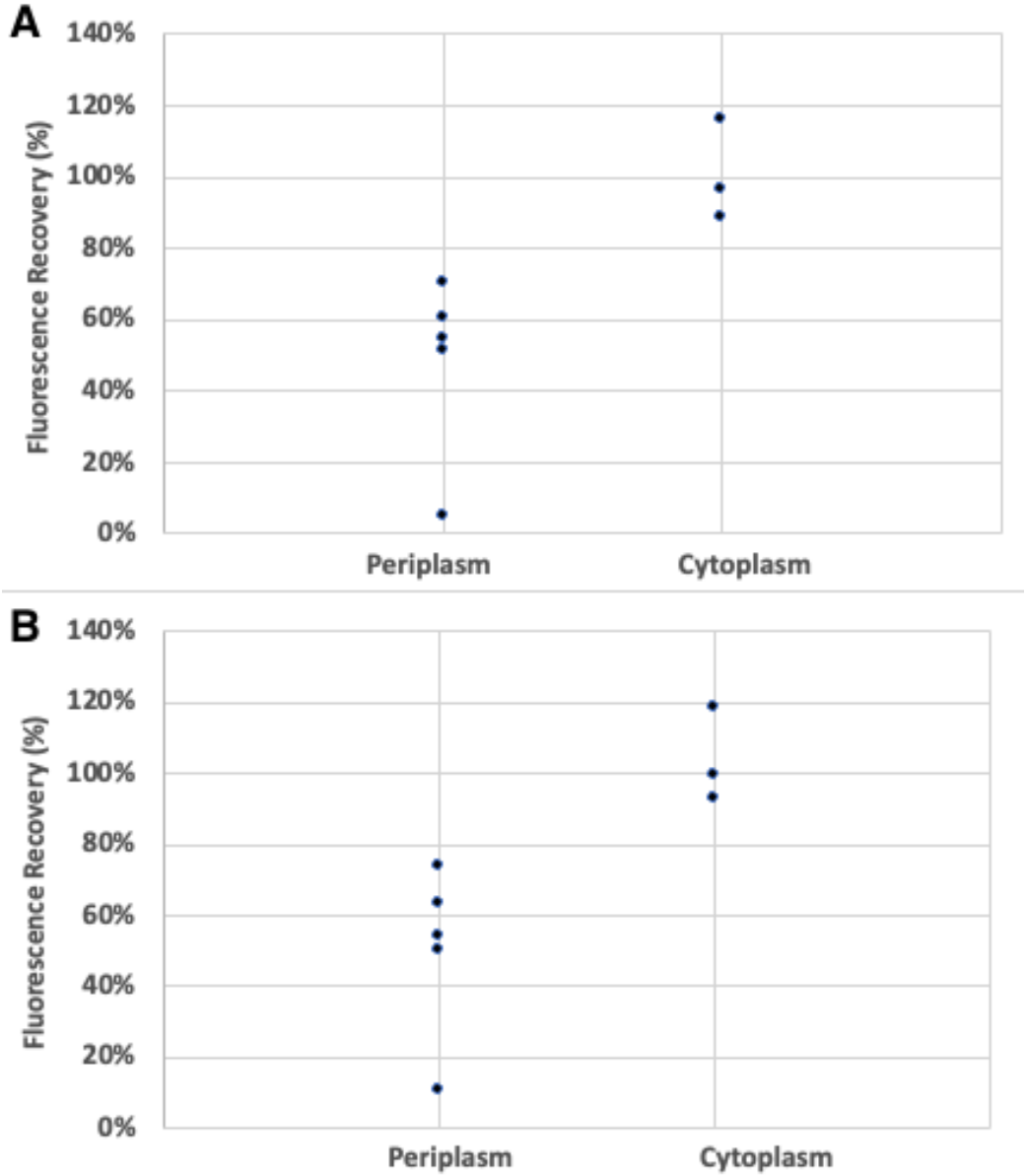

Figure 11: Comparison of sfGFP fluorescence recovery (%) for test proteins based on their know topological orientation of the C-terminus. The periplasmic proteins from the smallest recovery (%) are NuoK, MarC, MalE, ssDsbA, and CyoA. The cytoplasmic proteins from the smallest recovery (%) are the LD4.2 plasmid, YidC, and TatC. A is the average recovery after 2 minutes following HCl addition. B is the average recovery after 3 minutes following HCl addition. Each point is the average of two samples.

Figure 12 compares the fluorescence recovery of YidC, a cytoplasmic protein, and MarC, a periplasmic protein. At first glance, it can be observed that the rate of recovery is very similar for both proteins for the first few seconds. However, after approximately 25 seconds, the cytoplasmic protein (YidC) recovers quickly while the periplasmic protein (MarC) starts to plateau. Thus, when comparing different recovery times after two minutes there is no significant difference. For YidC, the A value was 1.209E6, the C value was 1.049E6, and the B value was 1.209E6 for two minutes and 1.212E6 for three minutes. While for MarC, the A value was 0.8595E6, the C value was 0.7880E6, and the B value was 0.8245 for two minutes and 0.8279E6 for three minutes. Using the chosen two-minute parameter, YidC had a fluorescence recovery of 100%, while MarC had 51%.

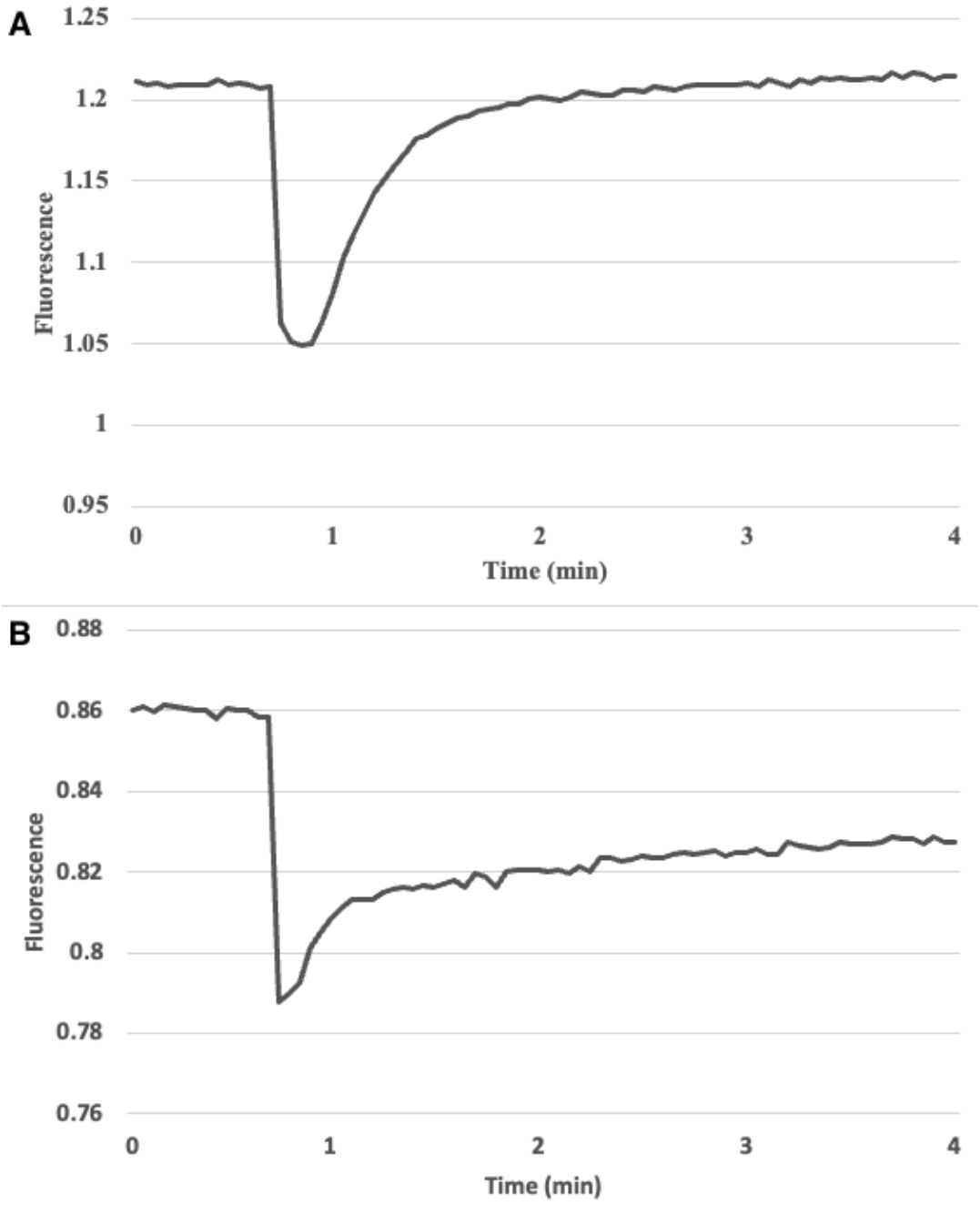

Figure 12: Fluorescence recovery (%) over time for YidC (A) and MarC (B). Each data point was taken three seconds apart, and HCl was added at 39 seconds. All fluorescent units are in millions.

Chapter IV: Discussion

IMPs are essential components of many cells, including gram-negative bacteria, where they play vital roles in a wide range of cellular processes. Understanding the topology of IMPs is important as it reveals vital insights into their arrangement within the lipid bilayer, thus providing one particular aspect required for scientist to understand their functionality. The arrangement and orientation of the TM domains dictate their roles, interactions with other molecules, participation in signal transduction, enzymatic activity, and transport of nutrients across the cell membrane (Strahl and Errington, 2017). Knowing the topology of IMPs in *E. coli* provides insight into how these proteins interact with the lipid environment in the periplasm and cytoplasm. Furthermore, it facilitates predictions of subcellular localization, accurate structural determinations and deciphering the implications of protein misfolding in diseases. These insights contribute to advances in understanding gram-negative bacteria and open new pathways for biotechnological applications.

Understanding the orientation of IMPs indeed holds several opportunities to contribute to the biotechnology field. In drug discovery and innovation, this knowledge helps develop pharmaceuticals that target IMPs, enabling the design of therapeutics to modulate their activity. Furthermore, it allows for innovative drug delivery systems that can cross the *E. coli* inner membrane, allowing for the precise and controlled delivery of drugs (Luirink et al., 2012). Additionally, the ability to engineer proteins with modified topologies enhances the creation of biotechnological tools with enhanced capabilities. Another biotechnological application is the production of biofuels, such as ethanol, butanol, and biodiesel. By understanding the orientation of the involved proteins, researchers can optimize the production of biofuels by improving transport efficiency, reducing toxicity, and increasing the yield of the final product (Koppolu and Vasigala, 2016). Therefore, the study of IMP topology in *E. coli* brings light to

our understanding of bacterial physiology and promotes innovation in biotechnology by aiding drug development, protein bioengineering, biofuel production and many more aspects.

Daley et al. developed one of the most advanced topology analysis methods for E. coli inner membrane. Their approach involved a combination of experimental and computational methods to determine the orientation of a vast number of IMPs. The study focused on 737 genes that encoded proteins longer than 100 residues with at least two predicted transmembrane helices. Using a fusion of GFP or PhoA to determine the location of the C-terminus experimentally resulted in a high-quality topology prediction for 601 proteins (Daley et al., 2005). However, one could envision extensions of some aspects of this study; for example, using GFP/PhoA fusions may have affected IMP folding due to their relatively large size. Additionally, the reliance on absolute values may reduce the accuracy and limit reproducibility because some proteins could be misinterpreted due to high or low protein expression levels (albeit the use of two fusion tags reduces the likelihood of this happening). Thus, this project aimed to improve this method by utilizing one FP and dynamic location measurements.

The pH sensitivity of FPs is a unique property that has enabled their widespread use in biological research and biotechnology. Understanding the mechanisms of pH sensitivity and the advantages and limitations of using pH-sensitive FPs is crucial for designing and interpreting experiments that utilize these proteins. Despite their intricacies, pH-sensitive FPs continue to play a central role in bioimaging, drug discovery, and synthetic biology. Advances in protein engineering and FP technology may lead to developing even more sophisticated pH-sensitive FPs, further expanding their potential applications in research and biotechnology.

The following steps were meticulously planned and conducted to establish the novel method presented in the thesis: Develop a plasmid capable of expressing sfGFP, improve the plasmid replication capabilities, systematically explore of HCl volume and cell numbers, define clear parameters for measuring the data, and test the method in pre-selected proteins.

The first step was to design a new plasmid to ensure that the OriR was low-copy, to enable transcription regulation *via* a unique inducible promoter, to incorporate an optimized version of sfGFP, and to eliminate all Type II restriction enzyme recognition sites. The use of a low-copy OriR is fundamental for the experiment because high-copy plasmids, which are present at high levels, place a significant load on the host cell. This may use cellular resources, such as metabolic energy, in plasmid replication instead of folding and inserting membrane proteins (Chen and Nielsen, 2019). High membrane protein expression levels can also lead to misfolding and aggregation (Angius et al., 2018). Specifically, the increased expression can overwhelm the cell's quality control machinery, accumulating misfolded proteins. Furthermore, some proteins with toxic domains or when overexpressed at high levels can be detrimental to the cell. A low-copy plasmid can mitigate some of the toxic effects. Lastly, low-copy plasmids tend to be more stable within the host because high-copy plasmids can be diluted out during cell division, resulting in irregular expression levels over time (Jones et al., 2000)

As the gene regulatory element, the *E. coli* rhaB operon promoter was incorporated to regulate gene expression in response to rhamnose (Giacalone et al., 2006 & Wegerer et al., 2008) and a terminator to ensure proper cessation of transcription. This inducible promoter ensures that the target gene is only expressed when needed, allowing for controlled gene expression timing and magnitude. Furthermore, in the absence of tight regulation, the expression of a foreign gene can place a burden on the host cell, affecting its growth and viability. Additionally, unintended or background expression of foreign genes can lead to experimental noise and potentially inaccurate results. Lastly, inducible promoters like the rhaB operon promoter are versatile and compatible with various host systems: The utilization of the rhamnose promoter offers an added advantage in its orthogonality to widely used protein expression systems, such as IPTG or arabinose-based systems (Zhang et al., 2015). This compatibility allows for the potential coexistence of these different systems to be used side by

side with our plasmid in future experiments, allowing the simultaneous expression of various proteins in a single study. Finally, to ensure a proper transcription termination in the *E. coli*, the terminator B0010 was incorporated, derived from the rrnB gene.

Lastly, an optimized version of sfGFP was incorporated into the plasmid. The first modification was to remove all internal Type II restriction enzyme recognition sites; these enzymes will be used in a high-throughput cloning approach for future experiments encompassing hundreds of proteins. Additionally, the start codon of the sfGFP was amended to include a NdeI site (CATATG) to allow the insertion of the small number of IMPs in the project, ensuring proper attachment at the C-terminus. sfGFP was chosen because it can fluoresce in the periplasm and cytoplasm (Wilks and Slonczewski, 2007). Furthermore, its fluorescent intensity is directly proportional to a change in pH, exhibiting its highest fluorescence at a slightly basic to neutral pH (around pH 7) and experiencing a decrease in fluoresces at a more acidic pH (around pH 5.5) (Wilks and Slonczewski, 2007 & Dammeyer and Tinnefeld, 2012). Likewise, the cell has an internal pH of around 7.5, meaning that maximum fluorescence will be observed when sfGFP is inside the cytoplasm. Since the cell must maintain a pH balance, it will always try to come back to the neutral pH. However, the pH of the periplasm is less regulated, resulting in a clear difference between compartments when an external stimulus decreases the pH (Dinh and Bernhardt, 2011). Harnessing this difference, the location of sfGFP can be determined, thus making it the perfect protein for this experiment.

While testing the first version of the plasmid, it was found that the DNA yields were very low. This posed a problem for the downstream experiments, particularly in the cloning of genes and sequencing of gene libraries. Thus, **the second step** to achieve the presented method was to increase the plasmid purification yields. Three approaches were explored to address this: low concentration of chloramphenicol, exonuclease V treatment, and modification of the

OriR. The best results were obtained by modifying the OriR to medium-copy levels. This increased the DNA concentrations to a level where the experiments could be carried out without any signs of aberrant overexpression. As mentioned before, having a high-copy plasmid is not ideal; however, having a very low-copy plasmid can likewise be an issue. Because membrane proteins require a minimum level of expression to be detectable and functional, posing a challenge for downstream assays (Jensen et al., 2017). Furthermore, low expression can result in significant cell-to-cell variation in expression levels (Wood et al., 2017). As a result, some cells may not have any plasmid or express the protein, while others might have a few plasmids and produce the protein at low levels. This variation may make it challenging to obtain consistent results. In addition to the adjustment of expression levels, it is worth noting that exonuclease V treatment significantly improved the removal of genomic DNA. Thus, this additional step will be used in preparing samples for next-generation sequencing in future experiments.

The third step to achieve the presented method was a systematic exploration of the parameter space with respect to solutions and cell density. The calibration of the pH of all solutions was a critical step to ensure the best possible measurements and reproducibility of the results. First, glycerol was added to PBS (yielding PBSG) to create a medium where the cells would have an external energy source, allowing them to recover their pH and fluorescence fully. Indeed, preliminary experiments not shown here were conducted in PBS and showed limited recovery. Then, the mixture containing PBSG and cells needed to be calibrated to a neutral pH of around 7 to ensure the first reading had maximum fluorescence since both the periplasm and cytoplasm would have the same pH. Subsequently, it was required to calibrate the solution with an exact volume of HCl that would lower the pH to a maximum change and minimum absolute value of 5.5 because this was the lowest pH the cell was believed to assume while still being able to recover internally (Wilks and Slonczewski, 2007). It was found that

the optimum theoretical amount of HCl that can be added to 90 µL of cells was 2.66 µL. However, after testing, it was determined that the optimal volume was 1.5 µL because it allowed for the most significant difference between recovery and no recovery.

In addition to the HCl volume, the exact number of cells needed to be adjusted. After preliminary experiments, it was observed that the best combination was an OD600 of 1.5 and an HCl volume of 1.5 µL. However, OD600 of 1.0 was given preference because when performing these experiments on all IMPs at a later stage, some might have different expression levels and behaviours than others, resulting in a general challenge to reach high OD600s for the entire ensemble. For instance, cultures expressing MarC and ssDsbA exhibited two-fold lower growth compared to TatC and MalE. The different rates at which IMPs are expressed can be due to several factors, such as protein structure, size, complexity, transmembrane domain architecture, or impact on cell physiology. Also, the characteristics of the gene sequence can affect the rate due to mRNA secondary structures, GC content and codon usage.

The fourth step was to define clear parameters for parametrizing the raw data that utilized the pH jump. Three key points were identified on the curve: *A*, *B* and *C*. *A* was chosen as the average of five points before adding HCl. This measures the initial fluorescence of the protein once the cells have been resuspended in PBSG and warmed up to 30°C. As mentioned earlier, at this point, maximum fluorescence intensity has been recorded since the pH inside and outside the cell is around 7.5. Then, *B* was chosen as the average of five points after two minutes of adding HCl. This time was chosen because there is no difference between two or three minutes of assessments with a clear linear relationship between them. Furthermore, saving a minute per sample could reduce the time the last sample in a 96-well plate is measured, with any delays impacting the potential to recover. Lastly, *C* was chosen as the minimum point on the graph after HCl addition. Experiments showed that the time to reach the minimum is eight seconds, making it essential to have a fast measurement frequency to avoid “missing” the

minimum. Overall, having three distinct measurements on the curve provides an advantage over relying on an absolute signal. This parametrization highlights that multiple measurements are utilized for each sample to deduce an experimental answer. In addition, and this has not been explored here, the exact time courses of recovery could be analyzed further by fitting with an exponential function.

In the fifth step, the protocol established in the above steps was applied to ten test proteins that have been studied before and have very good topological predictions (Dinh and Bernhardt, 2011 & Luirink et al., 2012 & Drew et al., 2006). The first proteins chosen were the two short tags (MalE and ssDsbA) responsible for moving sfGFP to the periplasm. These proteins were essential because they allowed us to test the behaviour of sfGFP in the periplasm, further solidifying the proposed protocol. Then, a mixture of periplasmic and cytoplasmic C-terminus proteins were chosen to test the developed protocol. This resulted in a diverse range of proteins that could be faithfully analyzed using the developed protocol. This is remarkable as the protocol was developed with soluble sfGFP, which expresses at levels much higher than IMPs. This confirms the original hypothesis that this method is signal magnitude independent for both soluble and membrane proteins. This also suggests that this method might improve comparability between experiments done in different laboratories under slightly different growth conditions. However, the sample of test proteins used was small, considering there are approximately 900 IMPs in *E. coli*.

Furthermore, each protein was only tested two times, preventing the use of statistical analysis (Student's t-test). Thus, more proteins should be tested to be able to determine the success of the proposed protocol. Additionally, more replicates for each test protein should be performed to allow the use of statistical analysis to corroborate the data. Lastly, the resulting membrane topologies must be compared to computational models of IMP structures such as Alphafold, available x-Ray or CryoEM structures, and hydrophobicity analysis.

Chapter V: Outlook

To the best of my knowledge, this was the first attempt to use pH-jump to assist with protein localization in bacteria. This method has proven functional in the framework of this thesis. However, it is not without limitations: First, it still requires the fusion of a bulky protein to an IMP. For instance, the amino acid length of sfGFP is 275 AA, which is similar in size or even longer than many IMPs (Orfanoudaki and Economou, 2014). Thus, methods which utilized smaller tags could be desirable. The smallest possible tag that one could envision are an unnatural amino acid (AUU). Adequate modifications have been made to the custom expression vector to accommodate such an experiment. For example, an extra TAG stop has been placed after sfGFP, and using a unique promoter such as rhaB allows for the regulation of gene expression in such a model orthogonally. Secondly, this approach requires a screening structure capable of measuring multiple samples quickly, limiting the number of equipment available in Australia. As added complications, measurements must include injection, shaking, and rapid measurement, adding extra time to each data point. Third, while unlikely, the stop-changing pH can alter the orientation of the protein, similar to how lipid composition and cell composition impact it. Future experiments might address this limitation, including in organisms of additional biotech importance, such as *Corynebacterium*, that serve as biological production "factories."

Additionally and importantly, computational methods such as homology modelling and molecular dynamics simulations can be used to predict the structure and behaviour of membrane proteins in *E. coli*, which can further refine our understanding of their topology. These methods can provide information on the location and orientation of specific domains within a protein and interactions between different regions. As it is often the case, combined experimental and computational analysis can yield the most insightful results.

References

Angius, F., Ilioaia, O., Amrani, A., Suisse, A., Rosset, L., Legrand, A., Abou-Hamdan, A., Uzan, M., Zito, F. and Miroux, B. (2018). A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. *Scientific Reports*, 8(1). doi:https://doi.org/10.1038/s41598-018-26668-y.

Chen, Y. and Nielsen, J. (2019). Energy metabolism controls phenotypes by protein efficiency and allocation. *Proceedings of the National Academy of Sciences*, 116(35), pp.17592–17597. doi:https://doi.org/10.1073/pnas.1906569116.

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D. and von Heijne, G. (2005). Global Topology Analysis of The Escherichia Coli Inner Membrane Proteome. *Science*, 308(5726), pp.1321–1323. doi:https://doi.org/10.1126/science.1109730.

Dammeyer, T. and Tinnefeld, P. (2012). Engineered Fluorescent Proteins Illuminate the Bacterial Periplasm. *Computational and Structural Biotechnology Journal*, 3(4), p.e201210013. doi:https://doi.org/10.5936/csbj.201210013.

Dietler, J., Schubert, R., Tobias G.A. Krafft, Meiler, S., Kainrath, S., Richter, F., Schweimer, K., Weyand, M., Harald Janovjak and Möglich, A. (2021). A Light-Oxygen-Voltage Receptor Integrates Light and Temperature. *Journal of Molecular Biology*, 433(15), pp.167107–167107. doi:https://doi.org/10.1016/j.jmb.2021.167107.

Dinh, T. and Bernhardt, T.G. (2011). Using Superfolder Green Fluorescent Protein for Periplasmic Protein Localization Studies. *Journal of Bacteriology*, 193(18), pp.4984–4987. doi:https://doi.org/10.1128/jb.00315-11.

Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. and de Gier, J.-W. (2006). Optimization of Membrane Protein Overexpression and Purification Using GFP Fusions. *Nature Methods*, [online] 3(4), pp.303–313. doi:https://doi.org/10.1038/nmeth0406-303.

Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, H. and Phillips, G.J. (2000). Green Fluorescent Protein Functions as a Reporter for Protein Localization in *Escherichia coli*. *Journal of Bacteriology*, [online] 182(14), pp.4068–4076. doi:https://doi.org/10.1128/jb.182.14.4068-4076.2000.

Frenkel, L. and Bremer, H. (1986). Increased Amplification of Plasmids pBR322 and pBR327 by Low Concentrations of Chloramphenicol. *DNA*, 5(6), pp.539–544. doi:https://doi.org/10.1089/dna.1.1986.5.539.

Giacalone, M.J., Gentile, A., Lovitt, B.T., Berkley, N.L., Gunderson, C.W. and Surber, M.W. (2006). Toxic Protein Expression in *Escherichia coli* Using a Rhamnose-Based Tightly Regulated and Tunable Promoter System. *BioTechniques*, 40(3), pp.355–364. doi:https://doi.org/10.2144/000112112.

Hedin, L.E., Illergård, K. and Elofsson, A. (2011). An Introduction to Membrane Proteins. *Journal of Proteome Research*, [online] 10(8), pp.3324–3331. doi:https://doi.org/10.1021/pr200145a.

Jensen, H.M., Eng, T., Chubukov, V., Herbert, R.A. and Mukhopadhyay, A. (2017). Improving Membrane Protein Expression and Function Using Genomic Edits. *Scientific Reports*, [online] 7(1), p.13030. doi:https://doi.org/10.1038/s41598-017-12901-7.

Jones, K.L., Kim, S.-W. and Keasling, J.D. (2000). Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria. *Metabolic Engineering*, 2(4), pp.328–338. doi:https://doi.org/10.1006/mben.2000.0161.

Joseleau-Petit, D., Vinella, D. and D'Ari, R. (1999). Metabolic Alarms and Cell Division inEscherichia coli. *Journal of Bacteriology*, 181(1), pp.9–14. doi:https://doi.org/10.1128/jb.181.1.9-14.1999.

Khalil, A.S. and Collins, J.J. (2010). Synthetic Biology: Applications Come of Age. *Nature Reviews Genetics*, [online] 11(5), pp.367–379. doi:https://doi.org/10.1038/nrg2775.

Koppolu, V. and Vasigala, V.K. (2016). Role of Escherichia coli in Biofuel Production. *Microbiology Insights*, [online] 9(24), p.MBI.S10878. doi:https://doi.org/10.4137/mbi.s10878.

Koprowski, P., Grajkowski, W., Balcerzak, M., Filipiuk, I., Fabczak, H. and Kubalski, A. (2015). Cytoplasmic Domain of Mscs Interacts with Cell Division Protein Ftsz: A Possible Non-Channel Function Of The Mechanosensitive Channel In Escherichia Coli. *PLOS ONE*, 10(5), p.e0127029. doi:https://doi.org/10.1371/journal.pone.0127029.

Kuhn, A., Koch, H.-G. and Dalbey, R.E. (2017). Targeting and Insertion of Membrane Proteins. *EcoSal Plus*, 7(2). doi:https://doi.org/10.1128/ecosalplus.esp-0012-2016.

Luirink, J., Yu, Z., Wagner, S. and de Gier, J.-W. (2012). Biogenesis of Inner Membrane Proteins in Escherichia Coli. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(6), pp.965–976. doi:https://doi.org/10.1016/j.bbabio.2011.12.006.

Mahon, M.J. (2011). pHluorin2: An Enhanced, Ratiometric, Ph-Sensitive Green Florescent Protein. *Advances in Bioscience and Biotechnology*, 02(03), pp.132–137. doi:https://doi.org/10.4236/abb.2011.23021.

Manoil, C. and Beckwith, J. (1986). A Genetic Approach to Analyzing Membrane Protein Topology. *Science*, 233(4771), pp.1403–1408. doi:https://doi.org/10.1126/science.3529391.

Mathelié-Guinlet, M., Asmar, A.T., Collet, J.-F. and Dufrêne, Y.F. (2020). Lipoprotein Lpp Regulates the Mechanical Properties of the *E. coli* Cell Envelope. *Nature Communications*, [online] 11(1), pp.1–11. doi:https://doi.org/10.1038/s41467-020-15489-1.

Orfanoudaki, G. and Economou, A. (2014). Proteome-wide Subcellular Topologies of E. coli Polypeptides Database (STEPdb). *Molecular & Cellular Proteomics*, 13(12), pp.3674–3687. doi:https://doi.org/10.1074/mcp.o114.041137.

Papanastasiou, M., Orfanoudaki, G., Koukaki, M., Kountourakis, N., Sardis, M.F., Aivaliotis, M., Karamanou, S. and Economou, A. (2012). The Escherichia coli Peripheral Inner Membrane Proteome. *Molecular & Cellular Proteomics*, 12(3), pp.599–610. doi:https://doi.org/10.1074/mcp.m112.024711.

Prescher, J.A. and Contag, C.H. (2010). Guided by the Light: Visualizing Biomolecular Processes in Living Animals With Bioluminescence. *Current Opinion in Chemical Biology*, 14(1), pp.80–89. doi:https://doi.org/10.1016/j.cbpa.2009.11.001.

Rapp, M., Drew, D., Daley, D.O., Nilsson, J., Carvalho, T., Melén, K., De Gier, J.-W. and Von Heijne, G. (2004). Experimentally Based Topology Models for E. coli Inner Membrane Proteins. *Protein Science*, 13(4), pp.937–945. doi:https://doi.org/10.1110/ps.03553804.

Roberts, T.M., Rudolf, F., Meyer, A., Pellaux, R., Whitehead, E., Panke, S. and Held, M. (2016). Identification and Characterisation of a pH-stable GFP. *Scientific Reports*, [online] 6(32). doi:https://doi.org/10.1038/srep28166.

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D.J. and de Gier, J.-W. (2012). Optimizing Membrane Protein Overexpression in the Escherichia coli strain Lemo21(DE3). *Journal of Molecular Biology*, 423(4), pp.648–659. doi:https://doi.org/10.1016/j.jmb.2012.07.019.

Senutovitch, N., Vernetti, L., Boltz, R.C., DeBiasio, R., Gough, A. and Taylor, D. (2015). Fluorescent Protein Biosensors Applied to Microphysiological Systems. *Journal of Biological Chemistry*, 240(6), pp.795–808. doi:https://doi.org/10.1177/1535370215584934.

Shaner, N.C., Steinbach, P.A. and Tsien, R.Y. (2005). A Guide To Choosing Fluorescent Proteins. *Nature Methods*, [online] 2(12), pp.905–9. doi:https://doi.org/10.1038/nmeth819.

Shen, Y., Rosendale, M., Campbell, R.E. and Perrais, D. (2014). pHuji, a pH-sensitive Red Fluorescent Protein for Imaging of Exo- And Endocytosis. *Journal of Cell Biology*, 207(3), pp.419–432. doi:https://doi.org/10.1083/jcb.201404107.

Shimomura, O., Johnson, F.H. and Saiga, Y. (1962). Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *Journal of Cellular and Comparative Physiology*, [online] 59(3), pp.223–239. doi:https://doi.org/10.1002/jcp.1030590302.

Shinoda, H., Shannon, M. and Nagai, T. (2018). Fluorescent Proteins for Investigating Biological Events in Acidic Environments. *International Journal of Molecular Sciences*, [online] 19(6), p.1548. doi:https://doi.org/10.3390/ijms19061548.

Stenberg, F., Chovanec, P., Maslen, S.L., Robinson, C.V., Ilag, L.L., von Heijne, G. and Daley, D.O. (2005). Protein Complexes of the Escherichia coli Cell Envelope. *Journal of Biological Chemistry*, 280(41), pp.34409–34419. doi:https://doi.org/10.1074/jbc.m506479200.

Strahl, H. and Errington, J. (2017). Bacterial Membranes: Structure, Domains, and Function. *Annual Review of Microbiology*, 71(1), pp.519–538. doi:https://doi.org/10.1146/annurev-micro-102215-095630.

Wegerer, A., Sun, T. and Altenbuchner, J. (2008). Optimization of an E. coli L-Rhamnose-Inducible Expression Vector: Test of Various Genetic Module Combinations. *BMC Biotechnology*, 8(1), p.2. doi:https://doi.org/10.1186/1472-6750-8-2.

Wilks, J.C. and Slonczewski, J.L. (2007). pH of the Cytoplasm and Periplasm of Escherichia coli: Rapid Measurement by Green Fluorescent Protein Fluorimetry. *Journal of Bacteriology*, 189(15), pp.5601–5607. doi:https://doi.org/10.1128/jb.00615-07.

Wood, W.N., Smith, K.D., Ream, J.A. and Lewis, L.K. (2017). Enhancing Yields of Low and Single Copy Number Plasmid Dnas from E. coli Cells. *Journal of microbiological methods*, [online] 133(3), pp.46–51. doi:https://doi.org/10.1016/j.mimet.2016.12.016.

Yeagle, P.L. and Elsevier (Amsterdam (2016). *The membranes of cells*. 3rd ed. Amsterdam I 11 Pozostałych: Elsevier/Academic Press, pp.1–25.

Yin, H. and Flynn, A.D. (2016). Drugging Membrane Protein Interactions. *Annual Review of Biomedical Engineering*, 18(1), pp.51–76. doi:https://doi.org/10.1146/annurev-bioeng-092115-025322.

Zhang, X.-X., Wang, Z., Yue, X., Ma, Y., Kiesewetter, D.O. and Chen, X. (2013). pH-Sensitive Fluorescent Dyes: Are They Really pH-Sensitive in Cells? *Molecular Pharmaceutics*, 10(5), pp.1910–1917. doi:https://doi.org/10.1021/mp3006903.

Zhang, Z., Kuipers, G., Niemiec, Ł., Baumgarten, T., Slotboom, D.J., de Gier, J.-W. and Hjelm, A. (2015). High-Level Production of Membrane Proteins in E. coli BL21(DE3) by omitting the inducer IPTG. *Microbial Cell Factories*, 14(1). doi:https://doi.org/10.1186/s12934-015-0328-z.