Volume 10:1-9

UJEMI+

The Undergraduate Journal of Experimental Microbiology & Immunology (+Peer Reviewed)

The Distance of the *lac* Regulatory Region to the Translation Start Codon of *gfp* on Plasmid pEAH23A Does not Change GFP Expression Levels

Trushaan Bundhoo, Adam Abdirahman Hassan, Brandon Wong, Rochelle Guan

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY *Bordetella pertussis*, the causative agent of whooping cough, uses various strategies to evade host immune defenses, notably through its autotransporter protein *Bordetella* resistance to killing (BrkA) which blocks complement-mediated killing. Therefore, understanding how *brkA* expression is regulated is crucial for understanding its role in pathogenesis. Previous research has demonstrated the involvement of the *lac* regulatory region, located upstream of the *brkA* locus on the pDO6935 plasmid in *brkA* expression. When this region was cloned onto a promoterless plasmid, it was able to drive GFP expression. However, the impact of the proximity of regulatory regions on gene expression remains poorly understood. In this study, we investigated how the distance between the *lac* promoter and a reporter gene, *gfp*, on a plasmid, impacts GFP expression levels. Utilizing site-directed mutagenesis, plasmids pBRAT24C and pBRAT24D were created, with shorter distances between the *lac* promoter and *gfp* relative to pEAH23A. Plasmids were transformed into DH5α *Escherichia coli* for expression and further analysis. Quantitative GFP analysis showed that pBRAT24C and pBRAT24D induced lower levels of GFP expression than pEAH23A. This suggests that the distance between the *lac* promoter and *gfp* does not affect levels of GFP expression and that the sequences found upstream of *gfp* might affect gene expression through other mechanisms. Further study into the regulatory effects of this region can help maximize BrkA yield from pDO6935, aiding future investigations into its function, structure, and potential as a vaccine candidate.

INTRODUCTION

Bordetella resistance to killing (BrkA) is an autotransporter found in *Bordetella pertussis*, conferring resistance to the body's immune response by inhibiting complement (1). Upon export across the outer membrane, BrkA is cleaved into a 73kDa N-terminal domain and 30 kDa C-terminal domain (1). The *brkA* locus has previously been cloned onto the pDO6935 plasmid, which allows for constitutive expression of BrkA in *Escherichia coli*, enabling researchers to better study its function and important properties (2). *brkA* is known to be regulated by Bvg, a two-component system that activates the expression of virulence genes in *B. pertussis*, and until recently, the promoter driving *brkA* expression from pDO6935 in *E. coli* was unknown (1, 3).

A study previously conducted by Haniak *et al*. (2024) demonstrated that the *lac* regulatory region found upstream of *brkA* on pDO6935 may be involved in *brkA* expression, as it was able to drive green fluorescent protein (GFP) expression when cloned onto a promoterless plasmid (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). The *lac* regulatory region included the catabolite activator protein (CAP) binding site, *lac* promoter, and *lac* operator sequences (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished

Published Online: September 2024

Citation: Bundhoo, Hassan, Wong, Guan. 2024. The distance of the *lac* regulatory region to the translation start codon of *gfp* on plasmid pEAH23A does not change GFP expression levels. UJEMI+ 10:1-9

Editor: Ronja Kothe, University of British Columbia

Address correspondence to: https://jemi.microbiology.ubc.ca/

manuscript). GFP expression was observed in *E. coli* DH5α containing either plasmid pEAH23A or pEAH23B (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). Both plasmids included the *lac* regulatory region upstream of the *gfp* locus, with the only difference being the number of nucleotides between the *lac* regulatory region and *gfp* (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). pEAH23A had 807 nucleotides between the 3' end of the *lac* operator and the 5' end of *gfp*, while pEAH23B had 61 nucleotides in the same region (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). Levels of GFP expression significantly differed between the two plasmids, with pEAH23B having a 3.5-fold increase in GFP expression relative to pEAH23A (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). This suggests that the distance between the *lac* regulatory region and *gfp* might affect levels of GFP expression.

Prokaryotic gene expression is influenced by many factors working in concert with one another. DNA regions upstream of genes can directly impact transcription levels, acting as transcription enhancers (4). These enhancer regions can bind to transcription factors to exert their effects (4). Silencer regions can also exert the opposite effects, binding to regulatory proteins to inhibit transcription (5). These are examples of transcription regulation, but translational regulation might also impact gene expression. The 5' untranslated region (UTR) of prokaryotic mRNA plays an important role in protein expression through stabilizing mRNA and regulating translation (6). More specifically, previous findings have shown that the sequences of 5' UTRs were important, due to their ability to bind regulatory proteins found in *E. coli* (6). Furthermore, bacterial translation is influenced by many other factors, such as the distance between the Shine-Dalgarno sequence and a gene, and the activity of prokaryotic sigma factors (7, 8). Both transcriptional and translational regulation can depend on the structure of an mRNA molecule and the elements it interacts with, prompting prior studies to create an understanding of how the different regulatory elements interact with one another. This is typical compared to focusing on how isolated variables, such as the distance between a gene and its promoter, affect protein expression.

Given the lack of research directly assessing the relationship on how the distance between a gene and its promoter affects gene expression, in addition to the findings from Haniak *et al.* (2024), we aim to assess whether varying the distance between the *lac* promoter and *gfp* on pEAH23A would affect levels on GFP expression. As pEAH23B had a shorter distance between the *lac* promoter and *gfp* and showed higher GFP expression relative to pEAH23A, we hypothesize that the distance between the *lac* promoter and *gfp* on pEAH23A is inversely related with levels of GFP expression. We aim to test this hypothesis by decreasing the distance between the *lac* promoter and *gfp* through site-directed mutagenesis, before quantitatively analyzing GFP expression from the mutagenized plasmids using *E. coli* DH5α.

METHODS AND MATERIALS

Luria Bertani (LB) media. LB media was prepared by dissolving 5g of tryptone, 5g of sodium chloride, and 2.5g of yeast extract in 500 mL of deionized water in a glass bottle. For LB-agar, 7.5g of agar powder was added per 500 mL of LB media. For LB media requiring antibiotics, ampicillin (50 mg/mL in deionized water) or chloramphenicol (34 mg/mL in ethanol) were filter-sterilized using a 0.22 μm filter and added for final concentrations of 50 μg/mL or 25 μg/mL, respectively.

Cultivation of bacteria and plasmids. pDO6935 (Oliver *et al*., 2003), pEAH23A, and pEAH23B (Haniak *et al.,* 2024) were acquired from the MICB 471 laboratory (UBC) starter plate collection and were all transformed into DH5α *E. coli*.

Overnight culture. Cultures were grown in 5 mL of LB broth at 37ºC on an orbital shaker set to 200 rpm. When required, chloramphenicol or ampicillin was added at a working concentration of 20 μg/mL or 50 μg/mL respectively.

Plasmid extractions. All plasmid extractions were performed using the Bio Basic Incorporated EZ-10 Spin Column Plasmid DNA minipreps kit (Cat# BS413) on 1.5 mL of overnight culture and following the plasmid DNA miniprep protocol (https://www.biobasic.com/ez-10-spin-column-plasmid-dna-miniprep-kit-4155). Extracted

plasmid concentrations and purity were measured with the NanoDrop spectrophotometer (*Thermo Scientific*).

Primer design. Primers for the inverse PCR process were created for the pEAH23A plasmid utilizing the SnapGene and NEBase changer (*New England Biolabs*) software. Sequences for the primers, along with their respective melting temperatures, can be found in supplementary tables S1 and S2. All primers were obtained from Integrated DNA Technologies (IDT). Sequences from the regions of interests and plasmid maps for the pBRAT24 plasmids can be found in supplemental figure S1.

Inverse polymerase chain reaction (PCR). 25 µL PCR reactions were prepared using 5X Phusion HF or GC buffer (*Thermo Scientific*), 10 mM dNTP mix, 10 mM of forward primer, 10mM of reverse primer, and Phusion Hot Start II DNA polymerase (1 µL/2 units, *Thermo Scientific*), deionized water, and 10% DMSO. 25 ng of pEAH23A plasmid DNA was added to the reactions as the template. Annealing temperatures were set to 55°C, which was 5°C below the recommended annealing temperature. This was chosen based on the addition of DMSO to the PCR reaction mix, which lowers the DNA melting temperature during the reactions.

Agarose gel electrophoresis. Agarose gels were made using 1X TAE buffer to a concentration of 0.8% agarose for all applications. RedSafe Nucleic Acid Staining Solution (*FroggaBio)* was used and diluted to a working concentration of 1:20,000 to stain the gel. Samples were prepared by adding 6X DNA loading dye (*Thermo Scientific*) and diluted to a ratio of 1:6 dye to sample. O'GeneRuler DNA Ladder Mix (*Thermo Scientific*) was loaded in the gel in 5 µL volumes. All agarose gels were run for one hour at 120 V and visualized using the ChemiDoc Imaging System (*Bio-Rad Laboratories*) on the SYBR Safe setting.

Plasmid ligation. Ligation of the inverse PCR products was carried out in 10 µL reactions containing: T4 ligase buffer (2X, *Thermo Scientific*), T4 DNA ligase (1 µL, *Thermo Scientific*), T4 polynucleotide kinase (1 µL, *Thermo Scientific*), DpnI restriction enzyme (1 µL, *New England Biolabs*), PCR product (1 µL), and deionized water. The ligation reaction was carried out at room temperature for 1 hour.

Transformation. 50 µL aliquots of chemically competent DH5α *E. coli* were transformed with 5 µL of ligated plasmid DNA (pBRAT24C, pBRAT24D) derived from the ligation reactions. After adding the DNA, the cells were incubated on ice for 30 minutes, followed by a 30-second heat shock in a 42°C water bath, and an additional 5-minute incubation on ice. Subsequently, 950 µL of LB media was added to each tube, and cells were incubated at 37°C in a shaking incubator for 30 minutes. Cells were then centrifuged at 5500 RCF for five minutes and resuspended in 100 µL of LB. 100 µL of the transformed cells were plated onto LB-CHL plates and incubated at 37°C overnight.

Restriction digest using HindIII. All restriction digests were carried out at a volume of 20 µL with the following components added: HindIII FastDigest (*Fermentas)* (1 µL), 10X FastDigest Buffer (*Fermentas)* (1X), 1 µg of template DNA, and DNase-free water. Reactions were carried out at 37°C for 15 minutes in a heating block and if necessary, heat inactivation was performed at 80°C.

Whole plasmid sequencing. A NanoDrop spectrophotometer (*Thermo Scientific*) was used to assess isolated plasmid concentrations and purity. When required, dilutions to fulfill Plasmidsaurus whole plasmid nanopore sequencing specifications were performed. All sequencing data was analyzed with SnapGene.

Quantitative GFP expression assay. Overnight cultures were generated with *E. coli* containing the pBRAT24C, pBRAT24D, pEAH23A, pEAH23B, and pDO6935. 200 µL reactions were prepared in a black-walled 96-well plate with a clear bottom where 40 µL of overnight culture was added to 160 µL of media. pEAH23A/B and pBRAT24C/D were plated

in three wells with 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) in LB broth and chloramphenicol (20 µg/mL) and three wells with distilled water with LB broth and chloramphenicol (20 µg/mL) for uninduced conditions. IPTG was utilized for the induced conditions to allow for maximal GFP expression by preventing the repression of the *lac* promoter by LacI. For pDO6935 specifically, ampicillin (50 µg/mL) was added in both the IPTG in LB broth and distilled water with LB broth. For controls, three wells contained only LB broth. Wells were pipet mixed and the plate was inserted into the microplate reader (*BioTek*) and incubated overnight at 37ºC, measuring optical density (OD_{600}) and GFP (excitation 485 nm, emission 528 nm) every 30 minutes for 24 hours. RFU/OD_{600} was calculated by averaging the technical replicates, normalizing to a negative control in the form of an empty well (OD_{600}) or pDO6935 a plasmid vector that lacks GFP (RFU), and the adjusted averages for both RFU and OD_{600} were divided. Results were tabulated using Microsoft Excel™ and entered into the GraphPad Prism™ software to generate graphs.

RESULTS

Generation of amplicons with different base pair lengths between the *lac* regulatory region and *gfp*. We designed four plasmids: pBRAT24A, pBRAT24B, pBRAT24C, and pBRAT24D with distances of 14 base pairs (bp), 64 bp, 276bp, and 499 bp, respectively, between the *lac* regulatory region and *gfp* (Fig. 1A). To construct plasmids with these varied distances, all plasmids utilized the same reverse primer with a sequence complementary to the *lac* operator sequence but used different forward primers with sequences complementary to DNA regions upstream of *gfp* (Supplemental Table S1). Nucleotides were also added to the ends of both primers to create a HindIII cut site after ligation. Using these primers, the four pBRAT24A/B/C/D backbones were amplified with inverse PCR. Electrophoresis on a 0.8% agarose gel showed the expected band lengths for pBRAT24A, pBRAT24B, pBRAT24C, and pBRAT24D as 3153, 3199, 3420, and 3643 bp, respectively (Supplemental Table S2 and Fig. S1B). These results were consistent in 10% DMSO using both the ThermoFisher GC and HF buffers. Thus, all four amplicons were successfully amplified with inverse PCR.

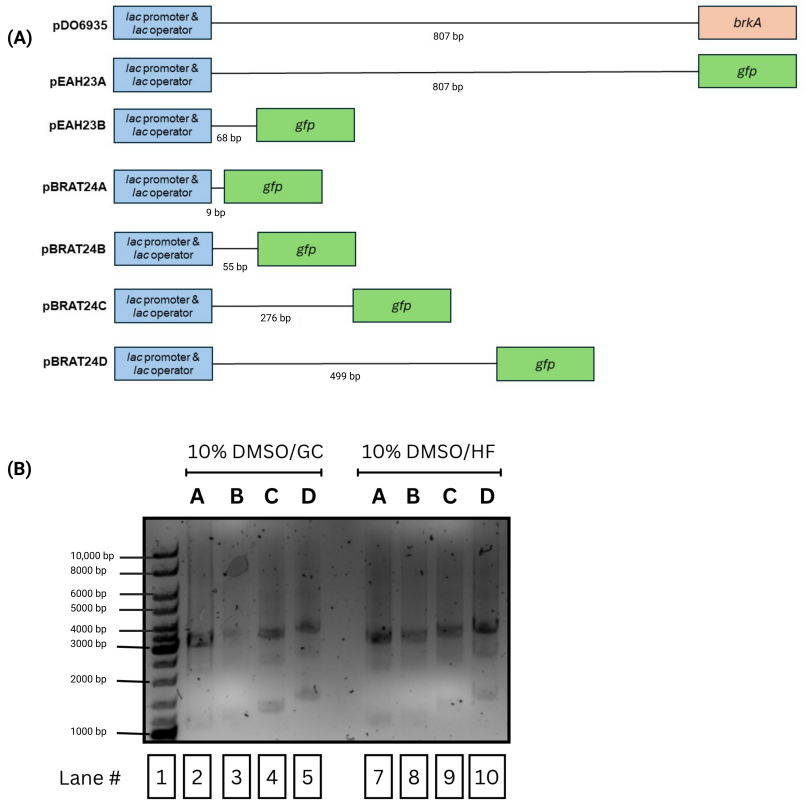

FIG. 1 Inverse Amplification of pEAH23A plasmid with deleted upstream element of *brkA* to various lengths. (A) Diagram showing the different base pair lengths between the lac regulatory region and BrkA or GFP in plasmids pDO6935, pEAH23A/B, and pBRAT24A/B/C/D. **(B)** 0.8% agarose gel from electrophoresis of inverse PCR products A-D amplified from pEAH23A using specially designed primers. Visualization was done using the ChemiDoc Imaging System (*Bio-Rad Laboratories*) and the SYBR safe setting. Inverse PCR products were run with the O'GeneRuler 10 Kb ladder (*Thermo Scientific*) and stained with RedSafe (*FroggaBio*). Lane 1 contained the ladder. Lanes 2-5 contained the inverse PCR products within 10% DMSO and 5X Phusion GC Buffer (*Thermo Scientific*). Lanes 7-10 contained the inverse PCR products within 10% DMSO and 5X Phusion HF Buffer (*Thermo Scientific*). Figure created with Biorender.com.

Amplicons for pBRAT24C and pBRAT24D were ligated then verified with HindIII digestion and whole plasmid sequencing. To ligate the plasmids, a ligation mix containing T4 DNA ligase, T4 polynucleotide kinase, and DpnI was used to produce pBRAT24C and pBRAT24D. pBRAT24A and pBRAT24B were unable to be ligated. The ligated plasmids were then transformed into DH5α *E. coli* cells and incubated overnight at 37ºC. Five clones of pBRAT24C and two clones of pBRAT24D were inoculated in LB broth with chloramphenicol and incubated overnight at 37ºC, undergoing plasmid extraction thereafter. To verify plasmid ligation, the extracted plasmids, along with pDO6935 as a positive control and pEAH23A as a negative control, underwent restriction digestion using HindIII. pDO6935 contains a HindIII cut site, while pEAH23A does not. We observed with a 0.8% agarose gel that clones C3, C4, and C5 for pBRAT24C and clones D2 and D3 for pBRAT24D had bands appear at the correct size of ~3420 and ~3643 bp, respectively (Fig. 2A and 2B). However, for pBRAT24C, we also observed smears for C1 and C2 that resemble the negative control's smeared bands. The presence of smears implied that these clones were not the plasmids that we had designed. (Fig. 2A). To further confirm ligation of pBRAT24C and pBRAT24D, plasmids were sent to Plasmidsaurus for whole plasmid nanopore sequencing. Sequences were then analyzed with SnapGene. The pBRAT24C and pBRAT24D plasmid sequences were confirmed as both had the HindIII site present, the expected distances of ~276 and ~499 bp, respectively, between the *lac* regulatory region and *gfp*, and the expected plasmid sizes of ~3420 and ~3643 base pairs, respectively (Supplemental Figure S2). Plates transformed with pBRAT24C clone 3 (C3) and pBRAT24D clone 2 (D2) were selected for subsequent experiments.

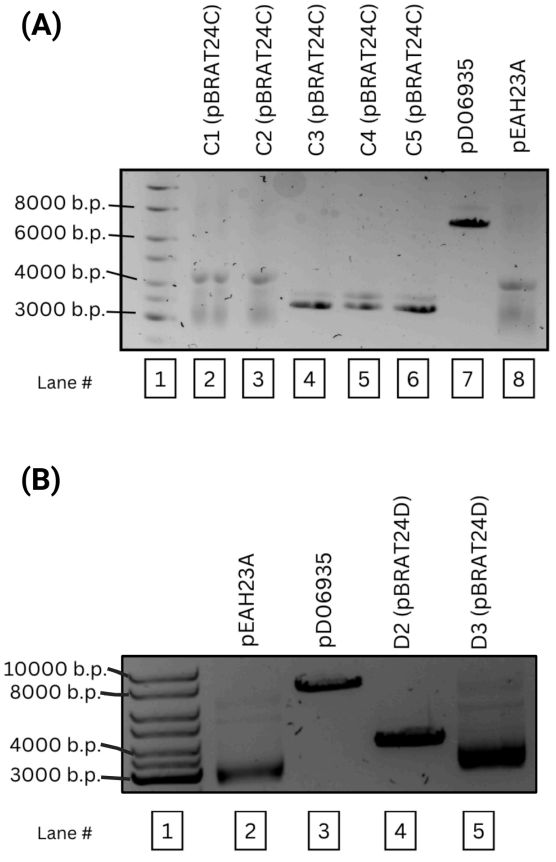

FIG. 2 Gel Electrophoresis revealed ligation of plasmid pBRAT24C/D after HindIII digest. 0.8% agarose gel from electrophoresis of amplified ligated plasmids that were all HindIII digested. Visualization was done using the ChemiDoc Imaging System (*Bio-Rad Laboratories*) and the SYBR safe setting. PCR products were run with the O'GeneRuler 10 Kb ladder (*Thermo Scientific*) and stained with RedSafe (*FroggaBio*). **(A)** Lane 1 contained the ladder. Lanes 2-6 contained five clones of pBRAT24C, referred to as C1, C2, C3, C4, and C5. Lane 6 contained pDO6935. Lane 7 contained pEAH23A. **(B)** Lane 1 contained the ladder. Lanes 2 contained pEAH23A. Lane 3 contained pDO6935. Lanes 4-5 contained two clones of pBRAT24D, referred to as D2 and D3. Figure created with Biorender.com.

Quantitative GFP assays revealed intermediate deletions of the upstream element between *lac* regulatory region and *gfp* gene led to no increase in GFP expression. To measure GFP expression, DH5α *E. coli* containing pBRAT24C, pBRAT24D, pEAH23A, and pEAH23B were incubated at 37ºC overnight in LB broth with chloramphenicol. Concurrently, an overnight culture of DH5α *E. coli* containing pDO6935 was inoculated in LB broth with ampicillin. These cultures were then plated on a 96-well plate for quantitative GFP analysis. Six wells were allocated for each plasmid, with three containing IPTG and three containing distilled water. Two plates were placed into a microplate reader (*BioTek*)

overnight at 37ºC, where OD_{600} and GFP expression were measured over time (Supplemental Figure S3). The relative fluorescence units (RFU)/OD_{600} were calculated and represented in a bar graph; maximum OD_{600} was used as cells are in the middle of their log growth phase, meaning maximum GFP expression would be present (Fig. 3). We observed that GFP expression was highest in pEAH23B, followed by pEAH23A, pBRAT24C, and pBRAT24D (Fig. 3 and Supplemental Figure S3). Furthermore, wells induced with IPTG exhibited higher GFP expression than uninduced conditions (Fig. 3 and Supplemental Figure S3).

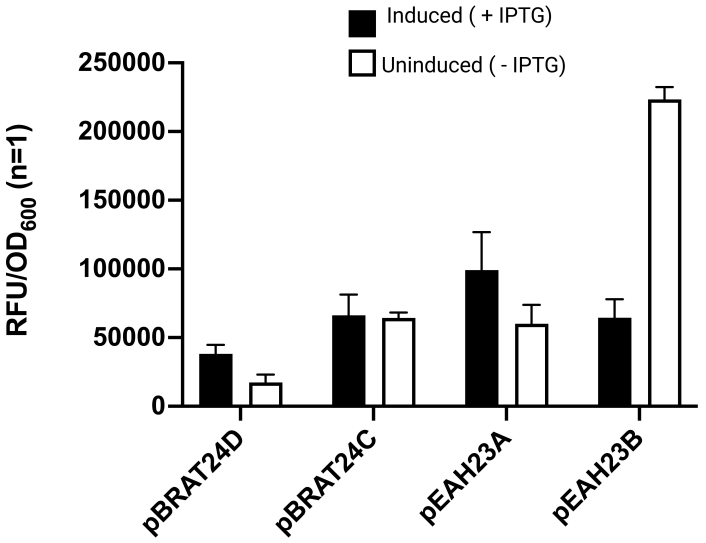

FIG. 3 Intermediate deletions of pEAH23A plasmid led to no increase in GFP expression. Results of IPTG induction of overnight liquid cultures as measured by a fluorescent plate reader (BioTek) (n=1 replicate). Colonies from pEAH23A, pEAH23B, pBRAT24C, pBRAT24D, and pDO6935 were initially grown overnight at 37°C. All overnight cultures were then seeded into a 96-well plate with either IPTG or distilled water in triplicate and grown again at 37 °C. GFP measurements were normalized to OD600 and standardized to the negative control: pDO6935. Results were plotted in GraphPad Prism™. Error bars represent standard error mean of triplicate measurements (n=1). Figure created with Biorender.com.

DISCUSSION

Previous studies performed by Haniak *et al.* (2024) shortened the distance between the 3' end of the *lac* operator and 5' end of *gfp* from 807 base pairs in pEAH23A to 61 base pairs in pEAH23B (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). With these constructs, the authors found significantly higher GFP expression in pEAH23B (E. Haniak, A. Tsoromocos, & H. Arneja, unpublished manuscript). This study aimed to investigate whether a shorter distance between the *lac* regulatory region and *gfp* in the pEAH23A plasmid would influence GFP expression. Results from this study would have implications for pDO6935, particularly regarding the relationship between the distance of the *lac* regulatory region and *brkA* and the resulting BrkA expression. GFP expression has been previously documented as an appropriate proxy for analysis of gene expression, allowing for more efficient data collection and subsequent analysis (9). Our approach consisted of site-directed mutagenesis to construct plasmids with shorter distances between the *lac* promoter and *gfp* on pEAH23A, followed by quantitative analysis of GFP expression. Our constructs aimed to shorten the distance between the 3' end of the lac promoter and the 5' end of *gfp* from 807 bp in pEAH23A to 14 bp in pBRAT24A, 64 bp in pBRAT24B, 276 bp in pBRAT24C, and 499 bp in pBRAT24D. The rationale for these deletion sizes was that the larger deletion sizes in pBRAT24C and pBRAT24D nearly halved the distance between the genetic elements of interest, and if we observed distance-dependent effects on GFP expression relative to pEAH23A, then those results would provide compelling evidence in support of our initial hypothesis. Furthermore, analysis of this region using SnapGene did not identify any genetic components that would have a known impact on *gfp* expression. The smaller distances between the 3' end of the *lac* operator and the 3' end of gfp in plasmids pBRAT24A and pBRAT24B contrasted the larger distances in pBRAT24C and pBRAT24D, to potentially remove most, if any, mRNA secondary structures that would form due to the region between the genetic elements of interest.

Four different primer pairs were used for inverse PCR, yielding four amplicons with shorter base pair lengths than pEAH23A, confirmed through gel electrophoresis (Figure 1B). It is important to note that there are some unwanted DNA bands present at around 1000 bp and around 2500 bp in length. This unwanted banding posed some difficulty during the PCR

process, caused likely by nonspecific binding of the shared pBRAT_rev primer. Nonspecific binding of that primer combined with correct binding of the forward primers would cause amplification of DNA that would match the expected banding pattern but appear lower down on the gel, consistent with Figure 1B. This issue led to a somewhat lengthy troubleshooting process, through which we found that 10% DMSO in the ThermoFisher HF buffer during PCR helped mitigate the unspecific binding to some extent. This allowed us to eventually create amplicons with our desired length, shown by the darkest bands in Figure 1B that align with the expected sizes. Subsequent ligation produced four new plasmids, but only pBRAT24C and pBRAT24D showed colony growth after the transformation of these plasmids into DH5α *E. coli*. Despite using the same ligation mix for all four plasmids, only pBRAT24C and pBRAT24D could be extracted. To address this issue, additional replicates should be performed in the future to ensure accuracy as we did not attempt to ligate pBRAT24A and pBRAT24B again due to time constraints. We utilized blunt-ended ligation at room temperature during the construction of these plasmids, which is less efficient than sticky-ended ligation (10). Previous research has shown that cycling the reaction temperatures between 10 and 30 ℃ may increase blunt-ended ligation efficiency by four to six-fold (11). Utilizing this method in the future may increase ligation efficiency for the pBRAT24A and pBRAT24B plasmids. Ultimately, pBRAT24A and pBRAT24B were excluded from subsequent experiments due to unsuccessful transformation. Restriction digests and whole-plasmid sequencing verified that pBRAT24C and pBRAT24D plasmids were correctly ligated and ready for GFP expression analysis. Subsequent quantitative GFP expression assays on DH5α *E. coli* with pBRAT24C, pBRAT24D, pEAH23A, and pEAH23B indicated that the deletions between the *lac* regulatory region and *gfp* of pBRAT24C and pBRAT24D did not increase GFP expression, as pEAH23A had a higher RFU/OD_{600} than both. This was the case in the induced condition with IPTG, and the uninduced condition without IPTG. Some expression in the uninduced condition was expected due to plasmid copy levels likely being higher than the copies of the lac repressor found in the DH5α *E. coli* genome. IPTG was still used in the induced condition to maximize *gfp* expression. This result suggests that the distance between the *lac* regulatory region and *gfp* may not be a factor that influences GFP expression. We initially hypothesized the order of highest to lowest GFP expression would be pEAH23B, pBRAT24C, pBRAT24D, and pEAH23A. This order is based on pEAH23B having the shortest distance between the *lac* promoter and *gfp*, and pEAH23A having the longest distance. However, the results suggest that pEAH23A exhibited higher GFP expression than pBRAT24C and pBRAT24D, despite pEAH23A having 526 and 308 more base pairs respectively between the *lac* promoter and *gfp*. We were unable to confirm this with correlational analyses as only one plate was analyzed.

Results from our quantitative GFP expression assay do not align with but cannot reject our original hypothesis that the distance between the *lac* regulatory region and *gfp* affects levels of GFP expression in pEAH23A. We conducted the assay with IPTG or distilled water conditions to remain consistent with the experiment performed by Haniak *et al.* (2024). pEAH23A had a higher RFU/OD_{600} than both the shorter pBRAT24C and pBRAT24D, suggesting that our deletions from pEAH23A may repress GFP expression in some way. Additionally, our results were consistent with Haniak *et al.* (2024), as pEAH23B did have much higher GFP expression than pEAH23A. Though the presence of IPTG increased RFU/OD_{600} readings, the plasmid GFP expression trends remained consistent with and without IPTG, indicating that the impact of IPTG on GFP expression was minimal.

Taken together, our findings suggest that other factors play a role in GFP expression. pBRAT24C and pBRAT24D had lower GFP expression levels than pEAH23A because potentially we may have deleted transcription factor binding sites upstream of *gfp* and the deleted regions may have led to a gene sequence that produced secondary structures such as a hairpin that may inhibit polymerase activity. Prior studies outlined the role of transcription factors for the regulation of promoter activity in the *lac* gene (12). Previous evidence has also shown that the formation of RNA hairpins can interfere with ribosome loading and affect levels of protein expression in eukaryotes (13). This remains to be proven in prokaryotes; however, it is known that RNA hairpins at the 3’ end of prokaryotic mRNAs are involved in transcription termination, highlighting the effects of nucleic acid secondary structure on gene expression (14). Possible explanations explaining the increased GFP expression in pEAH23B

might include the removal or addition of new DNA secondary structures that promote transcription, or the removal or addition of mRNA secondary structure to promote translation. As we added DMSO to the PCR reactions to melt secondary structures on the pEAH23A plasmid, it is possible that those structures somehow affect GFP expression.

One key limitation of our study is our inability to show what mRNA secondary structures are forming during *gfp* expression from the pBRAT24 plasmids. At the time of writing, we were unable to successfully use web-based mRNA secondary structure predictors, as none would provide any output when we input the pBRAT24 DNA sequences. Furthermore, we lack the equipment and programs necessary to determine mRNA secondary structures locally. An additional limitation is that GFP expression results were derived from data collected from only one 96-well plate (n=1) due to time constraints and LB broth contamination in the other two plates. This can be rectified by repeating the growth curve experiments 3 times to provide enough biological replicates for statistical testing. Overall, these findings suggest that the distance between the *lac* regulatory region and *gfp* do not explain the increased GFP expression of pEAH23B compared with pEAH23A. Other factors such as the excised DNA sequences should be considered. The same reasoning may hold true for BrkA expression in pDO6935, indicating that the distance between the *lac* regulatory region and *brkA* may not play a significant role in increasing BrkA expression. However, this remains to be tested in further experiments with the pBRAT24 plasmids and with pDO6935 directly.

Conclusions In this study, we constructed pBRAT24C and pBRAT24D, which were based on pEAH23A, with deletions between the *lac* regulatory sequence and *gfp*. The goal of this study was to evaluate how shortened distances between the *lac* regulatory sequence and *gfp* impacts GFP expression activity. We found that pEAH23A showed higher GFP expression than both pBRAT24C and pBRAT24D, suggesting that reduced distance between the promoter and the gene did not enhance gene expression. However, pEAH23B, with the largest pEAH23A deletion, exhibited the highest GFP expression. Therefore, we showed that levels of GFP expression from plasmid pEAH23A are not inversely related to the distance between the *lac* promoter and *gfp*. These results cannot be conclusively stated due to the n=1 analysis that was completed. In addition to addressing the statistical challenges of our findings, there is also a need to investigate other factors that may explain the high levels of GFP expression from plasmid pEAH23B.

Future Directions Rather than focusing on the distance between the *lac* regulatory region and *gfp*, future studies should investigate the specific gene components comprising this region. We propose three experiments to expand on our research. First, our results suggest that removing certain segments of pEAH23A decreased levels of GFP expression in pBRAT24A and pBRAT24B. Therefore, future experiments could involve creating multiple plasmids with various segments between the *lac* regulatory region and *gfp* deleted, with each deletion being the same size. For example, regions 100 bp long could be deleted, but taken from different regions within the DNA region upstream of GFP on pEAH23A. If a particular deleted segment results in higher GFP expression, then the original sequence in that region could be investigated further in a bioinformatics-based approach to assess possible binding to regulatory proteins. Second, future studies should aim to successfully construct pBRAT24A and pBRAT24B. As pBRAT24B has similar portions cut out of the plasmid as pEAH23B, GFP should be expressed at similar levels. With pBRAT24A having a larger deletion than pEAH23B, observing the GFP expression can bring valuable information. Third, we recommend looking at levels of GFP mRNA expression and GFP protein expression to assess if GFP expression is regulated at the transcriptional or translational level.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Dr. Oliver, Brynn McMillan, and Jade Muileboom from the MICB 471 teaching team for their invaluable support and guidance throughout our project. Additionally, we are grateful to the University of British Columbia's Department of Microbiology and Immunology for their funding support, which has enabled us to conduct our study, and for granting us access to their facilities and equipment. We would

also like to thank MICB 471 2023WT2 team 1iota for their support in troubleshooting our PCR reactions. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

CONTRIBUTIONS

This manuscript was collaboratively written and revised by T.B., B.W., R.G., and A.H. The original research study idea was developed by T.B., and all four members contributed equally to refining and advancing the research concept. Laboratory work was distributed among all team members.

REFERENCES

1. **Oliver DC, Huang G, Fernandez RC.** 2003. Identification of Secretion Determinants of the *Bordetella pertussis* BrkA Autotransporter. *J Bacteriol* **185**:489–495.
2. **Jain S, Van Ulsen P, Benz I, Schmidt MA, Fernandez R, Tommassen J, Goldberg MB**. 2006. Polar Localization of the Autotransporter Family of Large Bacterial Virulence Proteins. *J Bacteriol* **188**:4841–4850.
3. **Moon K, Bonocora RP, Kim DD, Chen Q, Wade JT, Stibitz S, Hinton DM**. 2017. The BvgAS Regulon of *Bordetella pertussis*. *mBio* **8**:e01526-174.
4. **Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM.** 2012. Distant Activation of Transcription: Mechanisms of Enhancer Action. *Molecular and Cellular Biology* **32**:4892–4897.
5. **Rine J**. 1999. On the mechanism of silencing in *Escherichia coli*. *Proc Natl Acad Sci USA* **96**:8309–8311.
6. **Chen F, Cocaign-Bousquet M, Girbal L, Nouaille S**. 2022. 5'UTR sequences influence protein levels in *Escherichia coli* by regulating translation initiation and mRNA stability. *Front Microbiol* **13**:1088941.
7. **Wakabayashi H, Warnasooriya C, Ermolenko DN.** 2020. Extending the Spacing between the Shine–Dalgarno Sequence and P-Site Codon Reduces the Rate of mRNA Translocation. *Journal of Molecular Biology* **432**:4612–4622.
8. **Seshasayee ASN, Sivaraman K, Luscombe NM**. 2011. An Overview of Prokaryotic Transcription Factors: A Summary of Function and Occurrence in Bacterial Genomes, p. 7–23. *In* Hughes, TR (ed.), A Handbook of Transcription Factors. Springer Netherlands, Dordrecht.
9. **Stretton S, Techkarnjanaruk S, McLennan AM, Goodman AE.** 1998. Use of Green Fluorescent Protein To Tag and Investigate Gene Expression in Marine Bacteria. Appl Environ Microbiol **64**:2554–2559.
10. **Brown TA.** 2002.Chapter 4: Studying DNA. Genomes. 2nd ed. BIOS Scientific Publ, Oxford.
11. **Lund AH, Duch M, Skou Pedersen F**. 1996. Increased Cloning Efficiency by Temperature-Cycle Ligation. Nucleic Acids Research 24:800–801.
12. **Phillips KN, Widmann S, Lai H-Y, Nguyen J, Ray JCJ, Balázsi G, Cooper TF**. 2019. Diversity in *lac* Operon Regulation among Diverse *Escherichia coli* Isolates Depends on the Broader Genetic Background but Is Not Explained by Genetic Relatedness. *mBio* **10**:e02232-19.
13. **Wang L, Wessler SR**. 2001. Role of mRNA Secondary Structure in Translational Repression of the Maize Transcriptional Activator *Lc*. *Plant Physiology* **125**:1380–1387.
14. **Wang X, N MPA, Jeon HJ, Lee Y, He J, Adhya S, Lim HM**. 2019. Processing generates 3′ ends of RNA masking transcription termination events in prokaryotes. *Proc Natl Acad Sci USA* **116**:4440–4445.