Undergraduate Journal of Experimental Microbiology and Immunology (UJEMI) Vol. 29:1-13

Steps Towards Identifying the Transcription Start Site Upstream of the *brkA* Gene in pDO6935 Plasmid Using the ARF-TSS Method

Lina Anwari, Anu Chauhan, Abigail Cho, Sarah Hong

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

SUMMARY Type Va autotransporters are expressed on Gram-negative bacteria, where several of these model systems are believed to contribute to virulence. The primary virulence determinant of *Bordetella pertussis* is accounted for by the autotransporter, BrkA, which facilitates host interactions and displays resistance to the effects of serum. While the *brkA* gene expressed in the *Escherichia coli* pDO6935 plasmid has been used in studies addressing *brkA*'s functionality and characteristics, the promoter sequence responsible for *brkA*'s constitutive expression in *E. coli* remains unknown. Furthermore, elements of the Lac operon, which have been studied for their role in regulation of gene expression, are positioned upstream of the *brkA* gene within the pDO6935 plasmid and thus, may serve as the putative promoter in an *E. coli* (DH5-α) model system. Our study aims to validate procedures established in the adaptor- and radioactivity-free identification of the transcription start site (ARF-TSS) method to map the TSS of the *brkA* gene within the pDO6935 plasmid sequence. The ARF-TSS method depends on cDNA generation from target gene mRNA. The determining factor depends on reverse transcriptase used during cDNA generation, as well as temperatures used to prevent internal structure formation. To build upon previous work and attempt new temperature bounds, we aimed to introduce the SuperScript™ IV Reverse Transcriptase during cDNA generation. While unable to replicate the results of a previous study, our findings support the established usage of enzymatic lysis to generate higher RNA yields during RNA extraction, as well as indicate optimal annealing temperatures may lie within the 69-73°C temperature range for PCR following cDNA generation using the SuperScript™ IV Reverse Transcriptase. The acquired knowledge is valuable for advancing future investigations focused on identifying the promoter of the BrkA autotransporter in an *E. coli* system expressing pDO6935 using the ARF-TSS method as a reference point.

INTRODUCTION

Known autotransporters have been found to play a role in a variety of virulence determinants (1). BrkA, the *Bordetella pertussis* derived Type Va autotransporter contributing to the virulence of whooping cough, has been shown to increase host interactions via adhesion and invasion, as well as contributing to the bacterium's ability to resist the effects of serum (2, 3). In attempts to understand the BrkA protein, Oliver *et al*. constructed *Escherichia coli* plasmid pDO6935, which constitutively expresses low levels of the *brkA* gene (3). Plasmid pDO6935 is a derivative form of plasmid pRF1066 post modifications which resulted in the excision of the 476 bp AatII fragment (3). The excision of this fragment effectively removes part of the 5' promoter region of the predicted cytoplasmic membrane protein BrkB. The resultantly open reading frame (ORF) of the *brk* locus permits transcription of only the *brkA* gene, thus, allowing for aimed studies of BrkA (2). The pDO6935 plasmid has since been used in subsequent studies investigating the functionality and characteristics of the *brkA* gene within an *E. coli* model system.

While the specific promoter sequence responsible for driving the expression of constitutive low levels of *brkA* in *E. coli* remains unidentified, the pDO6935 plasmid carries regulatory elements of the Lac operon upstream of the *brkA* gene (3, 4). Given the Lac promoter's ability to independently regulate gene expression within *E. coli* (4), it may serve as the putative promoter for *brkA* in the pDO6935 plasmid. In attempts to map the promoter

Published Online: September 2024

Citation: Anwari, Chauhan, Cho, Hong. 2024. Steps towards identifying the transcription start site upstream of the *brkA* gene in pDO6935 plasmid using the ARF-TSS method. UJEMI 29:1-13

Editor: Shruti Sandilya and Ronja Kothe, University of British Columbia

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

sequence of the *brkA* gene in an *E. coli* plasmid model, research aims focused on identification of the transcription start site (TSS) upstream of the core promoter (5). Transcription of a bacterial model is initiated by the RNA polymerase holoenzyme recognizing and binding to elements within the core promoter region (5, 6). Along the nucleotide sequence of the genomic DNA, the TSS is at position +1, while other core promoter motifs are relative to this (5). Thus, the TSS provides a useful reference point in identifying the promoter elements in bacterial genomes (6, 7). While adaptors and radioactive labelling have been established as traditional methods of identifying prokaryotic gene TSSs (8), the Adaptor- and Radioactivity-Free Identification of TSS (ARF-TSS) method established by Wang *et al*. accounts for the environmental implications of those traditional methods (8).

The ARF-TSS method was employed in a study by Sidhu *et al*. where the TSS was assumed to sequence directly upstream of a 5' phosphorylated primer upon cDNA generation (9). Following procedures outlined by Wang *et al*., *brkA* mRNA was isolated from *E. coli* DH5-α cells, digested with DNase to remove plasmid and genomic contamination, and reverse transcribed with a 5' phosphorylated primer using Superscript™ II Reverse Transcriptase at 42°C to generate cDNA (8, 9). The cDNA generated by Sidhu *et al*. was then purified and ligated, followed by amplification with Q5 High Fidelity Polymerase and subsequent amplification with Taq polymerase to generate a 3' polyA overhang for cloning into a TOPO plasmid vector (8, 9). This product was then transformed into TOP10 *E. coli* cells for colony generation on selective media which were sent for nanopore sequencing (8, 9). The results of this study identified five potential regions containing the TSS of the *brkA* gene within the pDO6935 plasmid (9). Due to uncertainties regarding the varying lengths of generated cDNA, Sidhu *et al*. claimed the TSS is at least 270 bp upstream of the *brkA* gene. Conclusions were based upon cDNA read lengths which began within the *brkA* gene and transcribed toward the origin of replication with the aim of termination at the TSS (8, 9). Premature cDNA termination was associated with the high GC content upstream of the *brkA* gene, resulting in secondary structure formation within the mRNA extracted from *E. coli* DH5-α cells containing the pDO6935 plasmid. Low annealing temperatures associated with SuperScript™ II Reverse Transcriptase, which functions optimally at 42°C, may not have been high enough to resolve mRNA secondary structures during cDNA generation (9). The use of SuperScript™ IV Reverse Transcriptase, with optimal functional temperatures of 50-65°C, may thus lead to better cDNA generation and help establish the precise TSS sequence of the *brkA* gene within the pDO6935 plasmid. Furthermore, location of the TSS may determine the role of the Lac operon as the putative promoter of the *brkA* gene in an *E. coli* system and validate methods established by Wang *et al.*

METHODS AND MATERIALS

Bacterial Culture Preparation. To prepare a 5 mL stock solution of ampicillin at a concentration of 100 mg/mL, Fischer™ Ampicillin and sodium salt were dissolved in distilled water and stored at 4°C. Luria-Bertani (LB) broth was made by mixing 5 g yeast extract, 10 g NaCl, and 10 g tryptone in distilled water to yield a final volume of 1 L. LB + ampicillin (Amp) agar plates were made by adding 10 g of agar to 500 mL of LB broth. This solution was then autoclaved and cooled to a temperature below 50°C to form 1.5% agar plates. Subsequently, 100 µL of the 100 mg/mL ampicillin stock was added to generate the selective media plates. *E. coli* DH5-α cells expressing pDO6935 were cultured overnight in 5 mL LB broth with ampicillin at a final concentration of 0.1 µg/mL in the shaking incubator at 200 RPM at 37°C for 12-18 hours.

RNA Extraction and Purification. The prepared cultures of *E. coli* DH5-α cells carrying pDO6935 were diluted in LB-Amp broth. These dilutions were allowed up to an additional 1.5 hours to grow in the shaking incubator (200 RPM at 37°C) until they reached an optical density (OD_{600}) of 0.5 at 600 nm. An OD_{600} of 0.5 was aimed as it ensured that the cell culture was in the exponential growth phase, actively synthesizing cellular components, including RNA. The cell density readings were obtained using a Pharmacia Biotech Ultrospec 3000 UV-Vis Spectrophotometer. Total RNA extraction from *E. coli* DH5-α expressing pDO6935 was conducted utilizing the RNeasy Plus Mini Kit (QIAGEN, Cat. # 74134) following the

manufacturer's protocols (10). To ensure effective bacterial cell lysis and maximize RNA yield, a lysozyme solution consisting of 30 mM Tris-HCl pH 8.0, 1 mM EDTA, and 15 mg/mL lysozyme was introduced to the RNeasy samples, B1 and B2 (Table 1), prior to lysis buffer addition. For samples not treated with lysozyme, A1 and A2 (Table 1), this step was omitted. RNA concentration and purity were then assessed using a NanoDrop™ 2000 spectrophotometer. Subsequently, RNA samples underwent treatment with TURBO DNA-*free*™ Kit (Invitrogen, Cat. # AM1907) to eliminate any plasmid or genomic DNA contamination, adhering to the manufacturer's instructions for routine DNase digestion (11).

TABLE. 1 Lysozyme-dependent yield and purity of *E. coli* total RNA samples. Yield and purity were measured with a NanoDrop2000 spectrophotometer. RNA extraction was performed using QIAGEN RNeasy kits.

	RNA Sample	Yield (ng/µL)	A_{260}/A_{280}	A_{260}/A_{230}
No lysozyme	A1	13.6	1.70	0.97
	A2	9.9	1.65	0.76
With lysozyme	B1	200.5	2.14	1.49
	B2	214.9	2.14	1.07

PCR Analysis for DNase Treatment Efficacy and Plasmid DNA Contamination Check. To prevent contamination of RNA samples, it is essential to ensure the absence of pDO6935 plasmid DNA. Custom forward (C1) and reverse (C2) primers (Table 2), designed by Sidhu *et al*. using SnapGene software (v. 7.1) by Dotmatics, were tailored to bind to nucleotide positions 2197-2218 and 4486-4508 of pDO6935, respectively. This primer pair amplified a 2312 bp region, encompassing the cDNA sequence between the P1 (Table 2) sequence in the *brkA* gene and the *brkA* TSS. This amplification strategy eliminates the possibility of false positives in downstream applications. Encompassing the entire cDNA sequence between the specified regions ensures that any subsequent amplification using downstream primers would specifically target the cDNA of interest, rather than amplifying potential contaminating plasmid DNA. The primer sites were selected considering the GC-rich nature of the plasmid upstream of the *brkA* gene, and the scarcity of regions with lower GC content. PCR was conducted on RNA samples (with and without DNase treatment) with the C1 and C2 primers using the Q5® High-Fidelity DNA Polymerase (12). The thermocycler conditions involved an initial denaturation at 98°C for 30 seconds, followed by 35 cycles of denaturation at 98°C for 10 seconds, annealing/extension at 72°C for 90 seconds, and a final extension at 72°C for 5 minutes (13). Post-PCR, the amplified products were analyzed on a 1% agarose gel alongside 1 kB O'GeneRuler DNA Ladder (Thermo Scientific, Cat. # SM0312) (14).

TABLE. 2 Sequences and melting temperatures of primers used in experiments

Target	Primer Name	Sequence	T_m (°C)
pDO6935 plasmid	C1 (forward)	5' - ATTTCGCTGCCAGGCATATACG - 3'	60
	C2 (reverse)	5' - CGAACTGAGATACCTACAGCGTG - 3'	59
cDNA generation	P1	5' - phosphorylated-GAGATGTGGATGTCCTTG - 3'	51
Ligated cDNA	L1 (forward)	5' - CAACACGCTCATTGCAGTC - 3'	56
	L2 (reverse)	5' - CTGCAAGGAAGACGGACATTG - 3'	59
lacO region	O1 (forward)	5' - AAGCTTCAGGAAACAGCTATGACCATG - 3'	55
	O2 (reverse)	5' - ACGAGCCGGAAGCATAAAG - 3'	59

cDNA Generation and RNA Degradation. Sidhu *et al*. designed a custom 5'-phosphorylated primer (2 µM), P1 (Table 1), using the SnapGene software (v. 7.1) for reverse transcription (RT) of RNA templates. As the location of the transcription start site (TSS) was unknown, only one primer was designed. To ensure compatibility with the TSS, the primer targeted nucleotide positions 2773 to 2790, just downstream of the *brkA* translation start site, aiming to amplify the start of the *brkA* gene and the sequence upstream, likely containing the TSS. This primer site accounted for the plasmid's GC-rich region and was optimized to a region with restricted GC content. Reverse transcription (RT) was conducted on total RNA using SuperScript™ IV Reverse Transcriptase (ThermoScientific, Cat. # 18090050) following the manufacturer's protocol, with incubation at 57.5℃ for 10 minutes, followed by inactivation at 80℃ for 10 minutes (15). Each RT reaction contained 500 ng of total RNA, with 5% DMSO added to account for the GC-rich nature of the *brkA* mRNA. Following reverse transcription, each sample was incubated with 2 µL of 1N NaOH at 70℃ for 10 minutes to degrade any potential remaining RNA, with subsequent neutralization with 3.6 µL of 0.5M acetic acid (16).

cDNA Purification and Ligation. The cDNA fragments underwent purification utilizing the EZ-10 Spin Column PCR Products Purification Kit (Biobasic, Cat. # BS363), following the specified protocol (17). The T4 RNA Ligase 1 (ThermoFisher™, Cat. # EL0021) was then employed for cDNA ligation, adhering to the manufacturer's instructions (18). The quantity of purified cDNA (6.5 µL) used in downstream applications was determined based on the approaches described by Kreisel *et al*. (19). The reaction mixture was composed of 1 µL (10 U) of T4 RNA Ligase 1, 2.5 µL of BSA at a final concentration of 0.1 mg/mL, and 3.6 µL of the 10X T4 RNA Ligase reaction buffer. Reactions were then incubated at 37℃ for 30 minutes followed by inactivation of the T4 RNA Ligase 1 at 70℃ for 10 minutes (18).

PCR Analysis for Unligated vs Ligated cDNA. Forward (L1) and reverse (L2) primers (Table 2), were designed by Sidhu *et al*. through SnapGene software version 7.1. These primers aimed to verify the presence of the ligated cDNA fragment, which was anticipated to contain the primer (P1) and TSS sequence. PCR was conducted on both ligated and unligated cDNA samples to assess ligation efficiency, utilizing Q5® High-Fidelity DNA Polymerase following the manufacturer's protocol (12). Thermocycler conditions consisted of an initial denaturation at 98℃ for 30 seconds, followed by 35 cycles of denaturation at 98℃ for 10 seconds, annealing/extension at 68℃ for 90 seconds, and a final extension at 72℃ for 5 minutes (13). Subsequently, the amplified products underwent electrophoresis on a 1% agarose gel, alongside the Thermo Scientific O'GeneRuler DNA Ladder.

Gradient PCR for Ligated cDNA. To determine the optimal annealing temperature for PCR amplification of ligated cDNA, a PCR gradient was conducted with custom-designed primers L1 and L2 using Q5® High-Fidelity DNA Polymerase, following the manufacturer's protocol with a range of annealing temperatures (12). The thermocycler conditions involved an initial denaturation at 98℃ for 30 seconds, followed by 35 cycles of denaturation at 98℃ for 10 seconds, annealing/extension at various temperatures ranging from 64.0℃ to 72.4℃ (spanning 4℃ below and above the previously used annealing temperature of 68℃) for 90 seconds, and a final extension at 72℃ for 5 minutes (13). The resulting amplified products and controls were visualized via electrophoresis on a 1% agarose gel, alongside the Thermo Scientific O'GeneRuler DNA Ladder.

PCR Amplification of the lac Operator Region. PCR amplification targeting the lac operator region was conducted on unligated cDNA samples. The primary objective was to verify cDNA synthesis, as successful amplification of the potential operator region depends on prior cDNA generation. The amplification protocol employed Q5® High-Fidelity DNA Polymerase according to the manufacturer's instructions (12). Each reaction contained 10 ng of cDNA samples, with 5% DMSO added to destabilize secondary structures in cDNA and to promote primer annealing specificity. The forward (O1) and reverse (O2) primers (Table 2) were custom-designed by I. Aujla, G. Nikas, and J. Wong (Team 1delta from W2023/24 Term 2) using SnapGene. Thermocycler conditions included an initial denaturation at 98℃

for 30 seconds, followed by 25 cycles of denaturation at 98°C for 10 seconds, annealing at 58°C for 20 seconds, extension at 72°C for 2.2 minutes, and a final extension at 72°C for 2 minutes. The annealing time of 20 seconds at 58°C was selected based on the Q5 SDM-optimized annealing temperature recommended by NEBaseChanger for the specific primers used. The amplified products, including positive (pDO6935) and negative controls (-RT or nuclease-free water), were visualized on a 1% agarose gel, alongside the Thermo Scientific O'GeneRuler DNA Ladder.

RESULTS

RNA samples isolated from lysozyme-treated *E. coli* DH5-α cells resulted in higher yields and purities. RNA samples were extracted from *E. coli* DH5-α cells to obtain *brkA* mRNA expressed from the pDO6935 plasmid. To determine yields and purities of isolated RNA samples, concentrations (ng/µL), A_{260}/A_{280} ratios, and A_{260}/A_{230} ratios were measured using the NanoDrop2000 spectrophotometer. RNA samples were extracted with and without lysozyme solution and measured as outlined in Table 1. Lysozyme solution was added to obtain higher RNA yields as it efficiently breaks open the cell wall for more effective removal of nucleic acids (20). The addition of lysozyme solution resulted in higher yields and purities of RNA samples compared to samples extracted from *E. coli* DH5-α cells that had not been treated with lysozyme. The expected A_{260}/A_{280} absorbance ratios for pure RNA is about 2.0 and the expected A_{260}/A_{230} absorbance ratios for pure RNA is 2.0-2.2 (21). RNA samples that had been treated with lysozyme had A_{260}/A_{280} absorbance ratios close to the expected value, however, the A_{260}/A_{230} absorbance ratios were slightly lower than expected, indicating the presence of organic contaminants. The increased absorbance at 230 nm in RNA samples is most likely due to contamination from guanidine thiocyanate which is present in high concentrations in the lysis buffer for RNA isolation; however, it has been shown that this does not compromise downstream applications (22). Since our RNA samples extracted from lysozyme-treated cells resulted in higher yields and purities, we proceeded with reverse transcription to generate cDNA with these samples.

PCR check for DNA contamination shows that RNA samples are free of plasmid DNA contamination after DNase treatment. To check if DNase treatment was successful, a PCR check was performed on the DNase-treated RNA samples alongside untreated RNA samples. Since the presence of plasmid pDO6935 DNA can produce false-positive results in downstream analyses, PCR was performed on DNase-treated and untreated samples using forward primer C1 and reverse primer C2, designed by Sidhu *et al.,* which amplify a 2312 bp region of the pDO6935 plasmid (9). PCR was performed on these samples alongside a negative no template control and a positive control with the pDO6935 plasmid. PCR amplification was visualised on a 1% agarose gel as seen in Figure 1. While all lanes for RNA

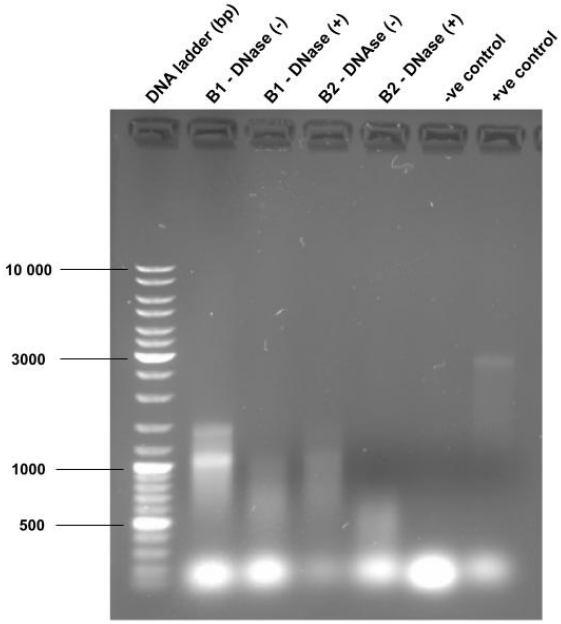

FIG. 1 RNA samples are free of plasmid DNA contamination after DNase treatment. PCR was performed on RNA samples (B1 and B2) treated with DNase (lanes 3 and 5) and without DNase (lanes 2 and 4) alongside a negative no-template control (lane 6) and a positive control using pDO6935 plasmid (lane 7). O'GeneRuler 1 kb DNA Ladder was loaded in lane 1 for reference. Primers C1 (forward) and C2 (reverse) were used to amplify a region of the pDO6935 plasmid that includes our targeted cDNA region. PCR products were run on a 1% agarose gel. Expected band sizes for samples not treated with DNase (lanes 2 and 4) were ~2312 bp. The absence of bands in the DNase-treated samples (lanes 3 and 5) indicates the absence of plasmid DNA contamination. Positive control lane reveals a band at ~3000 bp, indicating possible nonspecific amplification from primers or DNA contamination.

samples show smearing, the undigested sample lanes show clear bands that subsequently become absent after DNase treatment. Smearing the DNase-treated RNA sample lanes may be from partially degraded RNA or may indicate that the DNase treatment was not sufficient to remove additional contaminating genomic DNA material. The positive control lane reveals a band at ~3000 bp, indicating possible non-specific amplification from primers or DNA contamination. This reduces the validity of our results as this may indicate issues with the specificity of the primers used. Furthermore, the expected band sizes for the untreated lanes were ~2312 bp, however, bands at ~1500 bp were observed, which further indicate possible non-specific amplification. However, the absence of distinct bands of size ~2312 bp for plasmid DNA contamination in the DNase-treated RNA sample lanes suggest that plasmid DNA contamination present before DNase treatment had been removed and would not affect downstream cDNA analyses.

PCR check for cDNA ligation reveals non-specific amplification in ligated cDNA samples while unligated cDNA samples show lack of amplification. To check if cDNA generation and ligation were successful, a PCR check with ligated cDNA and unligated cDNA samples was performed. PCR was performed on these samples using forward primer L1 and reverse primer L2 from Sidhu *et al.*, that would amplify only the ligated cDNA (9). The orientation of these primers can be visualised in Figure 2. Figure 3 indicates smearing in lanes

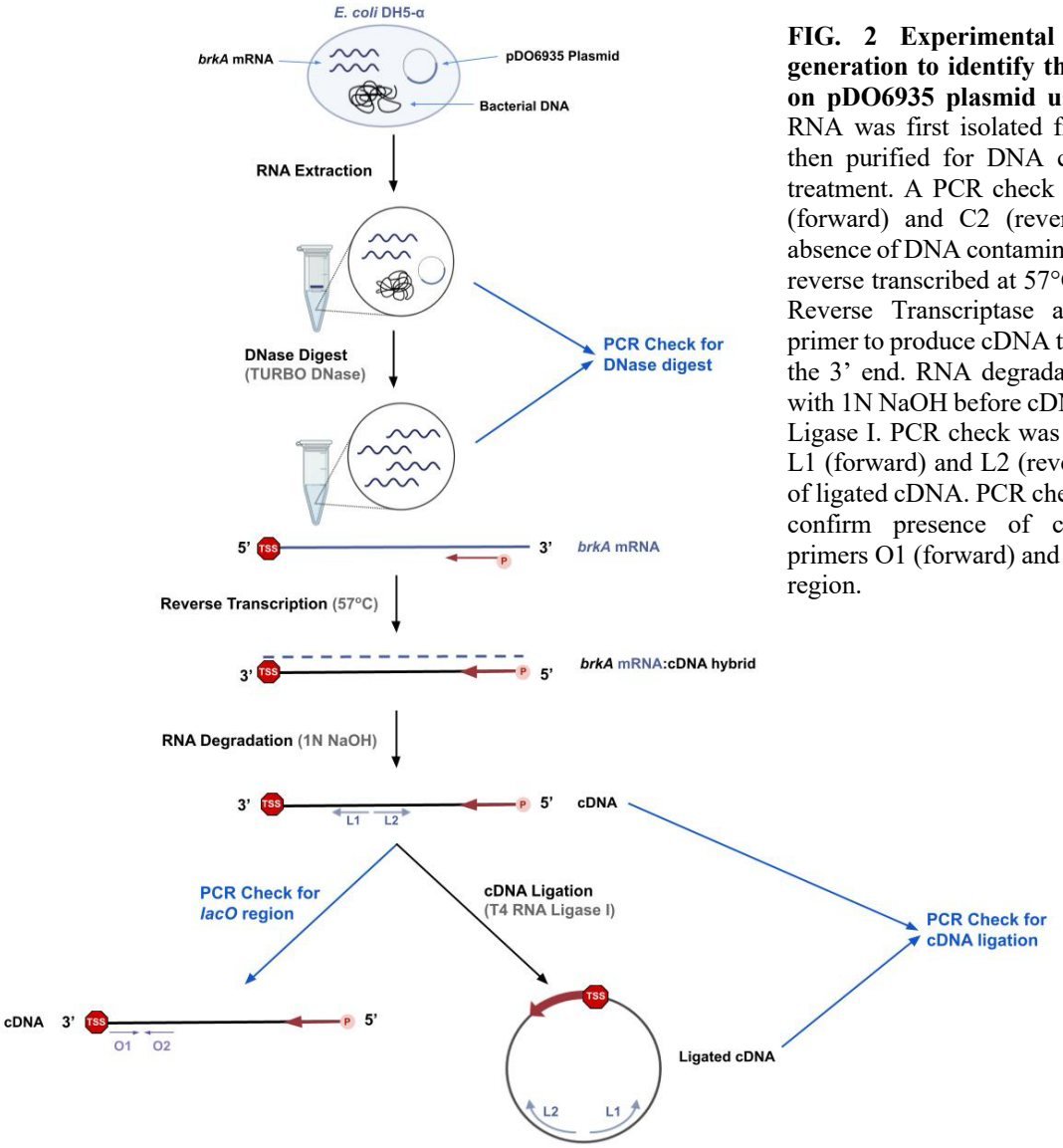

FIG. 2 Experimental workflow for cDNA generation to identify the TSS of the *brkA* gene on pDO6935 plasmid using ARF-TSS method. RNA was first isolated from *E. coli* DH5-α cells then purified for DNA contamination by DNase treatment. A PCR check was performed using C1 (forward) and C2 (reverse) primers to confirm absence of DNA contamination. RNA samples were reverse transcribed at 57°C using Superscript™ IV Reverse Transcriptase and a 5'-phosphorylated primer to produce cDNA transcripts with the TSS at the 3' end. RNA degradation was then performed with 1N NaOH before cDNA ligation with T4 RNA Ligase I. PCR check was performed using primers L1 (forward) and L2 (reverse) to confirm presence of ligated cDNA. PCR check was also performed to confirm presence of cDNA transcripts using primers O1 (forward) and O2 (reverse) for the *lacO* region.

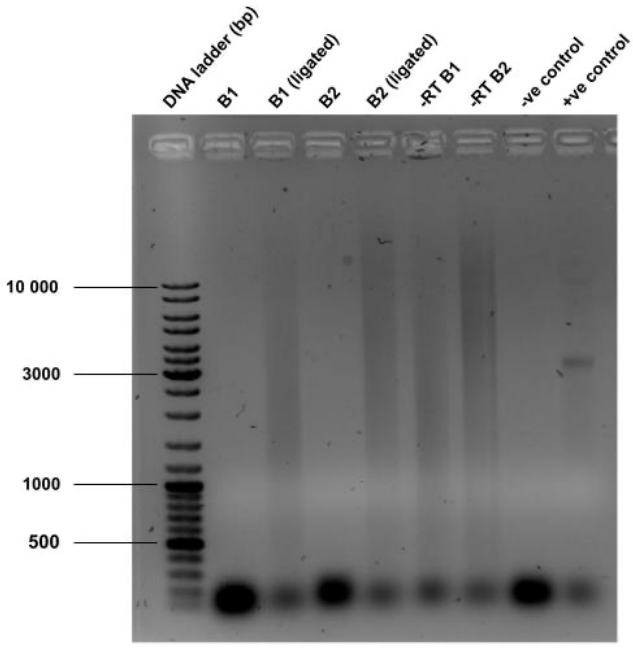

FIG. 3 PCR check for cDNA ligation shows nonspecific amplification in ligated cDNA sample lanes and no amplification in unligated cDNA sample lanes. PCR was performed on unligated (lanes 2 and 4) and ligated cDNA samples (lanes 3 and 5), alongside negative reverse transcriptase (-RT) controls containing DNase-treated RNA samples (lanes 6 and 7), a negative no-template control (lane 8), and a positive control using pDO6935 plasmid (lane 9). O'GeneRuler 1 kb DNA Ladder was loaded in lane 1 for reference. Primers L1 (forward) and L2 (reverse) were used which would amplify an 800 bp region of only the ligated cDNA. PCR products were run on a 1% agarose gel. Smearing is observed in the ligated cDNA sample lanes rather than the expected band at 800 bp. Smearing is also observed in the -RT lanes. Unligated cDNA sample lanes showed the expected absence of amplification. Positive control lane reveals a band at ~3000 bp, indicating possible contamination or nonspecific amplification.

with ligated cDNA samples, while the lanes for unligated cDNA do not show any amplification. This indicates a clear difference between the ligated and unligated samples as expected, however, it does not clearly indicate successful ligation since we do not see the expected band at 800 bp for the cDNA transcript. Additionally, the negative reverse transcriptase control (-RT) lanes show smearing, which may indicate the presence of genomic DNA contamination, consistent with the smearing observed in Figure 1, or non-specific amplification of residual RNA present in the cDNA samples. To test whether excess template DNA was causing smearing, PCR was repeated with a reduced amount of template. However, the results showed a lack of bands in all lanes including control lanes (Supplemental Figure S2). Smearing in the -RT control lanes in Supplemental Figure S2 may have also occurred for similar reasons as for Figure 1, such as insufficient DNase treatment causing genomic DNA contamination, or residual RNA in the cDNA samples.

Gradient PCR of ligated cDNA samples indicate that the optimal annealing temperature may lie within the 69-72°C range. To determine whether the smearing was caused by suboptimal PCR conditions in the PCR cDNA ligation check (Figure 3), a gradient PCR was performed on cDNA samples after ligation to find the optimal annealing temperature. A positive control using pDO6935 plasmid and a negative no template control were run alongside the ligated cDNA samples at 67.4°C. As shown in Figure 4, smearing was

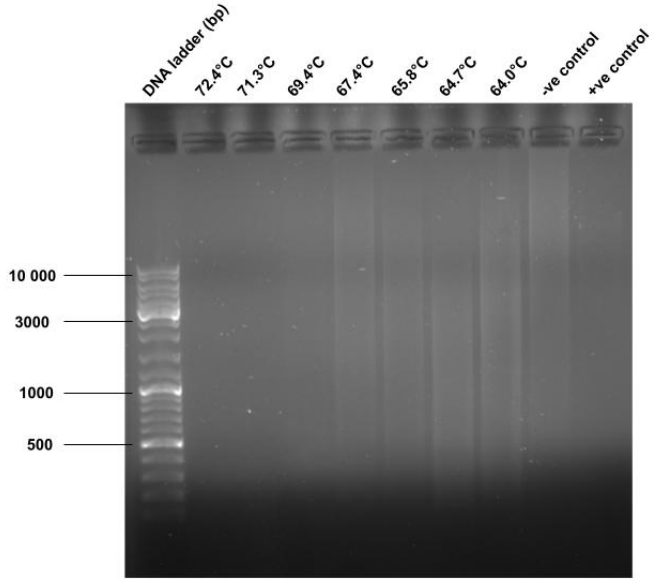

FIG. 4 Gradient PCR for ligated cDNA indicates that the optimal annealing temperature may lie within the 69-72°C range. Gradient PCR was performed on ligated cDNA samples in the temperature range of 64.0°C to 72.4°C to determine the optimal annealing temperature (lanes 2-8), alongside a negative no template control (lane 9) and a positive control using pDO6935 plasmid (lane 10) at 67.4°C. O'GeneRuler 1 kb DNA Ladder was loaded in lane 1 for reference. Primers L1 (forward) and L2 (reverse) were used to amplify an 800 bp region of the ligated cDNA. No amplification is seen in the 69.4°C to 72.4°C temperature range (lanes 2-4) while smearing is observed in the 64.0°C to 67.4°C temperature range (lanes 5-8) when PCR products were run on a 1% agarose gel. No lanes for the ligated cDNA show the expected band size at 800 bp. Positive control (lane 10) also shows lack of amplification, and negative control (lane 9) shows smearing indicating possible contamination.

observed in the temperature range from 64-67°C while no bands or smearing were observed in the temperature range from 69-72°C. Smearing at the lower temperature range could indicate non-specific amplification while the absence of bands in the higher temperature range indicates that specific target amplification also did not occur. Since smearing was consistent in the PCR check for cDNA ligation (Figure 3) and the lower temperature range (64-67°C) for gradient PCR (Figure 4), this could indicate that the optimal annealing temperature may lie within the 69-72°C temperature range as indicated by the absence of smearing and non-specific amplification. However, smearing in the negative control lane may indicate that contamination occurred during the preparation for PCR which may have also caused the smearing observed in the lower temperature ranges.

PCR check for cDNA generation did not show successful amplification of the *lacO* region. Since previous PCR analyses were unable to confirm whether cDNA had been generated, a PCR was performed using primers that amplify the lac operon operator (*lacO*) region, assuming that the *brkA* gene's promoter may be the lac operon's promoter. Confirming the generation of cDNA will help to validate our findings for ligated cDNA analyses. Assuming the *lacO* operator region would lie downstream of the TSS in our cDNA transcripts, a PCR was performed on cDNA samples using forward primer O1 and reverse primer O2 that would amplify the *lacO* region. Primers O1 and O2 were designed and provided by I. Aujla, G. Nikas, and J. Wong. Figure 5 shows absence of clear bands in all lanes, including the positive control lane. The absence of amplification in the positive control lane raises concerns regarding the reliability of the results obtained in Figure 5 as it is unclear whether the lack of amplification in cDNA sample lanes is due to experimental conditions or absence of the *lacO* region in cDNA transcripts. Additionally, smearing is observed for sample B1 in Figure 5, which suggests potential contamination in the sample or non-specific amplification due to suboptimal PCR conditions.

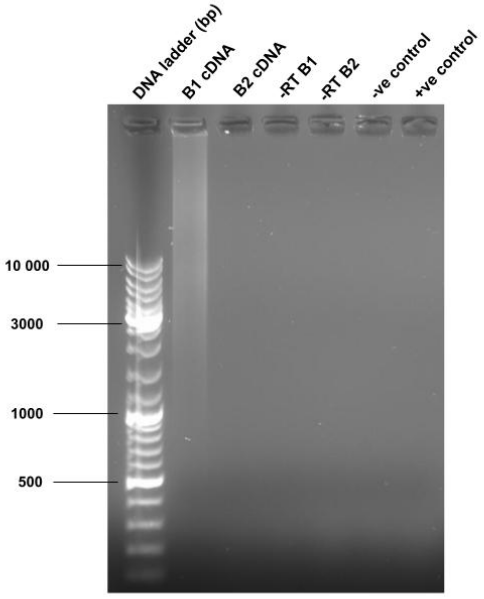

FIG. 5 PCR amplification of lacO region of cDNA transcript to confirm generation of cDNA shows inconclusive results. PCR was performed on unligated cDNA transcript samples (lanes 2 and 3) to confirm whether cDNA had been successfully generated by using primers O1 and O2 designed to amplify the *lacO* region of the lac promoter. The samples were run alongside negative reverse transcriptase controls (-RT) (lanes 4 and 5), a negative no template control (lane 6), and positive control using pDO6935 plasmid (lane 7). O'GeneRuler 1kb DNA Ladder was loaded in lane 1 for reference. PCR products show lack of successful amplification in all lanes, when run on a 1% agarose gel.

DISCUSSION

The study by Sidhu *et al*. had identified five potential transcription start sites of the *brkA* gene, with the largest mapped cDNA sequence being 565 bp upstream of the *brkA* gene (9). Given their shortest mapped cDNA fragment was 270 bp, Sidhu *et al*. suggested that the TSS is at least 270 bp upstream of the *brkA* translation start site (9). Due to the high GC content of the potential TSS region (62%), it was suggested that secondary structures may have formed in the mRNA, causing the reverse transcriptase to fall off during cDNA generation, resulting in early termination of cDNA synthesis. The formation of secondary structures was affirmed by inputting Sidhu *et al.*'s plasmid sequences into RNAfold and CoFold, which are programs that predict where secondary structures may form through calculations and analysis of the inputted RNA sequences (23, 24). The predicted results from both programs indicated the formation of secondary structures within each of the five cDNA sequences from Sidhu *et al*. (Supplemental Figure S3). To build upon their findings, we used SuperScript™ IV

Reverse Transcriptase which runs at a higher optimal annealing temperature (50-65°C) to prevent secondary structure formation and early termination of cDNA synthesis. Our study aimed to build off previous findings of Sidhu *et al.* to map the TSS of the *brkA* gene on the pDO6935 plasmid.

Enzymatic lysis with lysozyme results in higher RNA yield. During the initial attempts of RNA extraction, the manufacturer's protocol had been meticulously followed and the optional DNase steps had been executed. In the QIAGEN™ RNeasy Kit, the lysis buffer provided was used to cleave cells in order to extract cellular contents, including the RNA we aimed to isolate and use in downstream analyses. As indicated by Table 1, samples A1 and A2 had an RNA yield of 13.6 ng/μL and 9.9 ng/μL, respectively. This is lower than samples B1 and B2 which underwent enzymatic lysis using a lysozyme solution with proteinase K, resulting in an RNA yield of 200.5 ng/μL and 214.9 ng/μL, respectively. These findings demonstrate that using a lysozyme and proteinase K solution for enzymatic cell lysis yields a higher quantity of RNA. This method ensures more efficient cell lysis, thereby facilitating the extraction of RNA from each cell. Previous research states that enzymatic lysis using lysozyme provides high-yield, and high-quality RNA samples (25). Lysozyme breaks open the cell wall for easier removal of nucleic acids from cells and proteinase K aids in the inactivation of RNase and DNase, thus avoiding potential degradation of nucleic acids in solution (20). A previous study investigating the effects of a lysozyme solution on RNA yield found that there was more RNA yield per cell with enzymatic lysis when compared to the QIAGEN extraction protocol, and that 1 mg/mL of lysozyme would result in sufficient cell lysis (25). Based on previous research, we can conclude that the protocol provided by the manufacturer resulted in insufficient lysis of cells which is why RNA yield had been low in samples A1 and A2.

The smearing detected in cDNA ligated samples may be due to excess template DNA, suboptimal annealing temperatures, DNA contamination, and DNase contamination. After cDNA generation, the single stranded DNA was ligated with T4 RNA Ligase 1. As indicated by Figure 3, the PCR check for detecting ligated and unligated cDNA showed a clear difference between the two types of samples. Ligated samples and negative reverse transcriptase controls had smearing whereas unligated cDNA samples lanes did not have such smears. This differs from the expected 800 bp band anticipated in the ligated samples. It is thought that the reason for observing smearing on the gel may be due to suboptimal PCR conditions, excess template DNA, DNA contamination, or degradation of cDNA products.

One possible cause of the observed smearing in the ligated cDNA lanes in Figure 3 could be an abundance of template DNA present in the PCR samples. Based on previous literature, excessively high template concentrations can hinder the function of the polymerase due to potential carryover of inhibitors or inefficient denaturation (26). This means that in the template DNA samples, there may be inhibitors that when introduced in large quantities into PCR may influence the reaction in a negative manner. Also in the PCR reaction, a crucial step involves denaturing double-stranded DNA to facilitate the annealing of primers. As indicated previously, excess template DNA can hinder proper denaturation which may compromise the PCR reaction. To investigate if high DNA template concentrations are causing the detected smearing on the gel, a second attempt at this PCR with a lower concentration of template was performed. However, due to the lack of success seen in our positive control, we are not able to confidently conclude anything from these results (Supplemental Figure S2). A potential explanation for the failure of the positive control could stem from enzyme-related factors associated with the Q5 High-Fidelity DNA Polymerase, which is a common factor across all samples.

SuperScript™ IV Reverse Transcriptase was used to minimise the secondary structures formed in Sidhu *et al.*'s experiment, in order to obtain better results when mapping the TSS. Smearing was observed in the control lanes without reverse transcriptase in Figure 3 and Supplemental Figure S2. This indicates that there may have been DNA contamination or residual RNA that was not converted to cDNA in the samples, resulting in non-specific amplification (27). Figure 1 also had smearing in the DNase treated samples thus indicating potential DNA contamination. This suggests that smearing may have been due to inadequate DNA digestion prior to cDNA generation, or that the presence of residual RNA resulted in smearing.

Another potential explanation for the smearing is that there may have been non-specific DNA products of varying sizes being generated during PCR due to annealing temperatures not being optimised (28). The efficacy of primer annealing temperature plays a significant role in the success of PCR (29). When annealing temperatures are too low, it may result in non-specific PCR bands with variable lengths, thus producing a smearing effect on the agarose gel (30). And when temperatures are too high, it may prevent binding of primers to the template DNA (30). Gradient PCR was attempted in Figure 4, which indicates that the optimal annealing temperature may lie within the 69-72℃ range. Furthermore, this confirms that the annealing temperature of 68℃, also used by Sidhu *et al.* for the same primers, was not ideal because it is lower than this range. Therefore, smearing may be due to non-specific PCR amplified DNA due to the primers binding in unwanted regions.

Lastly, the smearing detected may also be due to degradation of cDNA products due to potential DNase contamination. This may have resulted in the degradation of DNA at various lengths resulting in a smearing effect detected on the gel. While we were conscious of potential RNase contamination, we did not take extra cautionary measures for potential DNase contamination other than using gloves, applying RNase Zap, and working in a RNase free environment. Perhaps more caution is required to account for DNase contamination to ensure the sample is not being actively degraded.

When comparing our data to Sidhu *et al*,'s data, it seems that the most likely cause of smearing would be due to DNA or residual RNA contamination as all samples have smearing present. Furthermore, this is established by the smearing present in our negative SuperScript™ IV Reverse Transcriptase controls suggesting that the issue is not caused by the enzyme, which is a major modification made in our study, but rather DNA contamination that was not effectively digested with DNase treatment or was introduced later in the process. The presence of smearing in controls and ligated lanes, regardless of which primers are used, indicates once again that it is likely due to contamination.

Gradient PCR to ascertain optimal annealing temperatures revealed smearing, indicating suboptimal PCR conditions. Determining optimal annealing temperatures is possibly the most critical component for optimising the specificity of a PCR reaction and obtaining good data (31). Therefore, performing gradient PCR with a series of temperatures allows for empirical testing to determine optimal temperatures for the annealing of primer on template DNA. The PCR annealing temperatures start about 5°C below the primer melting point (31). Figure 4 presents PCR samples on a 1% agarose gel tested with annealing temperatures of 64.0℃ - 72.4℃. While no amplification in the form of a clear band is observed, smearing was present in the PCR samples with annealing temperatures of 64.0℃ - 67.4℃. Gradient PCR should show the optimal annealing temperature indicated by a clear band at the expected region. However, a lack of a band on the gel at about 800 bp indicates that the optimal annealing temperature cannot be conclusively determined from the results of this experiment. Similar to Figure 3, plausible reasons for the smearing detected on the gel may be due to low annealing temperatures which resulted in nonspecific primer binding, or potential DNase contamination causing degradation of the DNA product, and thus resulting in DNA fragments of varying sizes. Future directions may involve performing additional gradient PCR experiments to determine optimal temperatures, and reagent concentrations to obtain the highest yield of a specific PCR product. These gradient experiments may include a magnesium chloride titration series, template DNA concentration gradient, or a denaturation and elongation temperature gradient to determine the best PCR conditions (31). Magnesium chloride titration series may assist DNA polymerases, as a cofactor, as the magnesium ions present at the active site catalyse the phosphodiester bond formation between 3'-OH of a primer and a phosphate group of a dNTP, which results in the incorporation of dNTPs during polymerization (32).

PCR amplification of *lacO* region in cDNA transcripts reveal inconclusive results. Possible explanations as to why amplification was unsuccessful include that cDNA had not been successfully generated at all, cDNA transcript that extends all the way to the *lacO* region and TSS was not generated due to premature termination of cDNA synthesis, or that the lac promoter is not the *brkA* promoter and therefore the *lacO* region was not present in the cDNA transcript.

Limitations The experiment encountered challenges most likely related to the instability of bacterial mRNA. Despite the effort to minimize compromising factors by working rapidly and using RNaseZap, the RNA samples remained vulnerable to degradation. Overnight storage at -20°C could have caused further degradation due to added freeze-thaw cycles. Furthermore, the use of Nanodrop readings for RNA quantification is not reliable, particularly when dealing with our low-concentrated samples that were near the machine's lowest detection limit, 2.5 ng/µL (33). Additionally, Nanodrop readings do not provide insight into RNA integrity, which is crucial for downstream applications. This variability in RNA concentration and integrity could have influenced the yield and quality of the resulting cDNA fragments. Omitting RNaseOUT RNase inhibitor during reverse transcription raised concerns about further RNA degradation due to contamination with RNases. This exclusion, together with overnight storage for downstream analyses, risked additional degradation of *brkA* mRNA samples from freeze-thaw cycles.

Despite the GC-rich nature of the *brkA* gene, which typically poses challenges during reverse transcription, multiple cDNA bands were still observed even with the utilization of the SuperScript™ II RT operating at a lower temperature (42°C) by Sidhu *et al*. This suggests that the high GC content might not be the sole factor responsible for the observed lack of cDNA bands. Furthermore, the absence of Ribonuclease H (RNase-H), which would typically be employed to eliminate any remaining RNA hybridized to cDNA, may have compounded the issue. This oversight in RNA contamination removal could have affected the efficiency of cDNA amplification and visualization on the gel. Other potential issues, such as insufficient RNA concentration during reverse transcription or the presence of inhibitors as previously discussed, might have also contributed to the observed challenges.

Conclusions The major modification made from Sidhu *et al.*'s experiments is the use of SuperScript™ IV Reverse Transcriptase, which runs at higher optimal temperatures, to reduce the effects of secondary structures when mapping the TSS. Although we were not able to determine the TSS of *brkA* within the pDO6935 plasmid model, we can conclude that enzymatic lysis, using lysozyme and proteinase K, results in greater RNA yield in RNA extraction. Also, possible explanations of the smearing detected on gels are suboptimal annealing temperatures, excess template DNA, DNA contamination, and DNase contamination.

Future Directions Since it was not confirmed whether cDNA had been generated in reverse transcription using Superscript™ IV Reverse Transcriptase, an assay to check the enzyme's activity could be performed. For example, the SG-PERT assay can be used to detect and quantify the activity of reverse transcriptase. This assay quantifies the amount of cDNA produced from RNA using qPCR amplification which will represent the level of the reverse transcriptase enzyme's activity (34). Additionally RNase H could be used for RNA degradation of the RNA:DNA hybrid that forms after reverse transcription. Although alkaline treatments have been used to selectively degrade RNA in RNA:cDNA hybrids, the use of the RNase H enzyme, which specifically cleaves the RNA portion of RNA:DNA hybrids, is more commonly used and may be more effective (35, 36).

Once it is confirmed that the reverse transcriptase enzyme is working and cDNA is synthesised from the *brkA* mRNA transcript, the remaining steps of the ARF-TSS method can proceed, as summarised in Figure 6. PCR reamplification with Taq polymerase would follow purification, ligation, and amplification to generate 3' polyA overhangs on the PCR products for cloning into a TOPO® vector. The vectors would then be transformed into *E. coli* TOP10 cells and plasmids would be extracted to be sent for nanopore sequencing. If successful, the sequencing results should reveal the TSS at a position directly upstream of the 5'-phosphorylated primer that was used for reverse transcription.

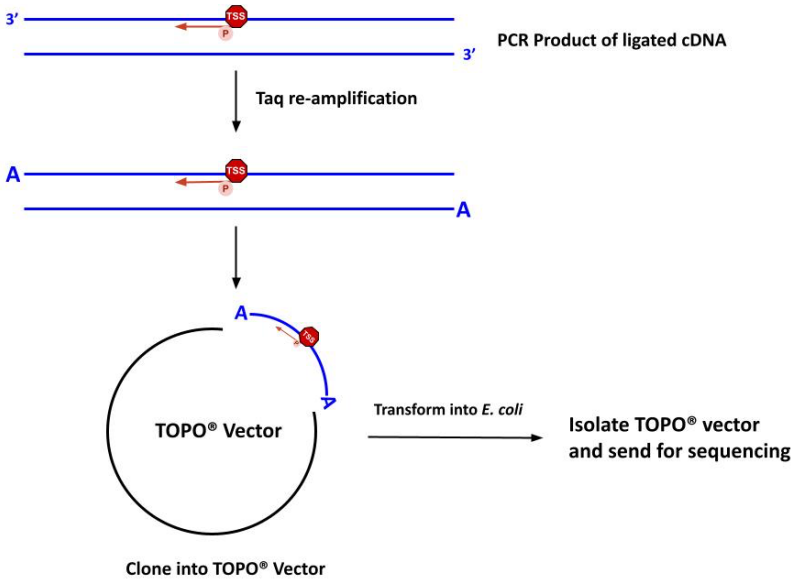

FIG. 6 Proposed experimental workflow for future directions of identifying TSS of the *brkA* gene on pDO6935 plasmid using ARF-TSS method. Following cDNA ligation, PCR reamplification should be performed using Taq polymerase to add 3' polyA overhangs for cloning into a TOPO® vector. The cloned vectors can then be transformed into *E. coli* cells, isolated, then sent for nanopore sequencing to map the TSS directly upstream of the 5'-phosphorylated primer used in reverse transcription.

ACKNOWLEDGEMENTS

We would like to thank Dr. David Oliver for his continued support and guidance in this project. We would also like to thank Jade Muileboom for her support in the lab as well as Virginie Jean Baptiste for her guidance in project design and troubleshooting. We would like to thank Sidhu *et al*. for providing the foundational design of this project as well as team 1delta for their *lacO* primers. We also acknowledge and thank the Department of Microbiology and Immunology at the University of British Columbia for funding and providing the resources for this research project.

CONTRIBUTIONS

All authors collaborated in conducting the experimental research and analyses done in the laboratory. L.Anwari contributed to writing the discussion and conclusion. A.Chauhan contributed to writing the abstract and introduction. A.Cho contributed to writing the results, future directions, and creating figures/tables. S.Hong first proposed the project idea and contributed to writing the materials/methods and limitations. All authors contributed to the final editing and revision. All authors should be considered equal for co-authorship.

REFERENCES

1. **Wells TJ, Tree JJ, Ulett GC, Schembri MA**. 2007. Autotransporter proteins: novel targets at the bacterial cell surface. FEMS Microbiol Lett **274**:163–172.
2. **Fernandez RC, Weiss AA**. 1994. Cloning and sequencing of a *Bordetella pertussis* serum resistance locus. Infect Immun **62**:4727–4738.
3. **Oliver DC, Huang G, Fernandez RC**. 2003. Identification of Secretion Determinants of the *Bordetella pertussis* BrkA Autotransporter. Journal of Bacteriology **185**:489–495.
4. **Lutz R, Bujard H**. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research **25**:1203.
5. **Siders J, Laakso M, Leung W**. TSS Module Primer: Review of Transcription, Promoter Structure, and Chromatin Packaging.
6. **Cervantes-Rivera R, Puhar A**. 2020. Whole-genome Identification of Transcriptional Start Sites by Differential RNA-seq in Bacteria. Bio Protoc **10**:e3757.
7. **Matteau D, Rodrigue S**. 2015. Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5'-Rapid Amplification of cDNA Ends (5'-RACE). Methods Mol Biol **1334**:143–159.
8. **Wang C, Lee J, Deng Y, Tao F, Zhang L-H**. 2012. ARF-TSS: an alternative method for identification of transcription start site in bacteria. Biotechniques 52.
9. **Sidhu G, Alayoubi S, Gama A, Huseynova J**. 2023. Mapping potential transcription start sites of the *brkA* gene on pDO6935 in *Escherichia coli* using the ARF-TSS method.
10. **RNeasy Plus Kits | DNAse-free RNA Isolation | QIAGEN.** https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/rna-purification/total-rna/rneasy-plus-kits.

11. **Ambion.** 2012. TURBO DNA-free™ Kit TURBO™ DNase Treatment and Removal Reagents User Guide. https://tools.thermofisher.com/content/sfs/manuals/cms_055740. pdf
12. **New England Biolabs.** 2013. PCR using Q5® high-fidelity DNA polymerase (M0491). https://www.neb.com/en/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491. Retrieved 17 February 2024.
13. **Assal N, Lin M.** 2021. PCR procedures to amplify GC-rich DNA sequences of mycobacterium bovis. J. Microbiol. Methods **181**:106121.
14. **Lee PY, Costumbrado J, Hsu C-Y, Kim YH.** 2012. Agarose Gel Electrophoresis for the Separation of DNA Fragments. J Vis Exp 3923.
15. **ThermoFisher Scientific.** 2015. SuperScript IVTM Reverse Transcriptase. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FSSIV_Reverse_Transcriptase_UG.pdf
16. **Avican K, Aldahdooh J, Togninalli M, Mahmud AK, Tang J, Borgwardt KM, *et al*.** 2021. RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. Nat. Commun. **12**:3282-3282.
17. **BioBasic.** 2019. EZ-10 Spin Column PCR Products Purification Kit Handbook. https://www.biobasic.com/us/amfilerating/file/download/file_id/25473/. Retrieved 15 February 2024.
18. **ThermoFisher Scientific.** 2012. T4 RNA Ligase Product Information. https://www.thermofisher.com/order/catalog/product/EL0021. Retrieved 19 April 2024.
19. **Kreisel K, Engqvist MKM, Clausen AR.** 2017. Simultaneous mapping and quantitation of Ribonucleotides in human mitochondrial DNA. JoVE **129**:e56551
20. **Gautam A**. 2022. DNA Isolation by Lysozyme and Proteinase K, p. 85–88. In Gautam, A (ed.), DNA and RNA Isolation Techniques for Non-Experts. Springer International Publishing, Cham.
21. **Matlock B.** Assessment of Nucleic Acid Purity.
22. What are the effects of low A260/A230 ratios in RNA preparations on downstream applications? https://www.qiagen.com/us/resources/faq?id=c59936fb-4f1e-4191-9c16-ff083cb24574&lang=en. Retrieved 11 April 2024.
23. **RNAfold web server.** http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi. Retrieved 10 April 2024.
24. **CoFold Web Server.** https://e-rna.org/cofold/. Retrieved 10 April 2024.
25. **Verbeelen T, Van Houdt R, Leys N, Ganigue R, Mastroleo F**. 2022. Optimization of RNA extraction for bacterial whole transcriptome studies of low-biomass samples. iScience **25**:105311.
26. **PCR Troubleshooting | Bio-Rad.** https://www.bio-rad.com/en-ca/applications-technologies/pcr-troubleshooting?ID=LUSO3HC4S. Retrieved 14 April 2024.
27. **Reverse Transcription troubleshooting - CA.** https://www.thermofisher.com/ca/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-troubleshooting.html. Retrieved 14 April 2024.
28. **Lorenz TC.** 2012. Polymerase chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. J Vis Exp 3998.
29. **Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M.** 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiology Ecology **60**:341-350.
30. **Working with PCR.** https://www.sigmaaldrich.com/CA/en/technical-documents/technical-article/genomics/pcr/working-with-pcr. Retrieved 14 April 2024.
31. **ThermoFisher Scientific.** PCR Setup - Six Critical Components to Consider. https://www.thermofisher.com/ca/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-component-considerations.html#:~:text=Magnesium%20ion%20(Mg2%2B),a%20dNTP%20(Figure%206). Retrieved 28 April 2024.
32. **Using Gradient PCR to Determine the Optimum Annealing Temperature.** https://www.drugdiscoveryonline.com/doc/using-gradient-pcr-to-determine-the-optimum-a-0002. Retrieved 14 April 2024.
33. **ThermoFisher Scientific.** 2012. Sample Reproducibility. https://assets.thermofisher.com/TFS-Assets/CAD/Warranties/T044-NanoDrop%20Spectrophotometers-Sample-Reproducibility-EN.pdf
34. **Vermeire J, Naessens E, Vanderstraeten H, Landi A, Iannucci V, Van Nuffel A, Taghon T, Pizzato M, Verhasselt B.** 2012. Quantification of Reverse Transcriptase Activity by Real-Time PCR as a Fast and Accurate Method for Titration of HIV, Lenti- and Retroviral Vectors. PLoS One **7**:e50859.
35. **Lemire KA, Rodriguez YY, McIntosh MT**. 2016. Alkaline hydrolysis to remove potentially infectious viral RNA contaminants from DNA. Virol J 13:88.
36. **Pallan PS, Egli M.** 2008. Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. Cell Cycle **7**:2562–2569.