University of St Andrews | FOUNDED 1413

Isolation, purification and characterisation of a novel chondroitin lyase from *Bacteroides thetaiotaomicron*.

Joanna Magnuson
jlam1@st-andrews.ac.uk
Supervisor: Dr Tracey Gloster

LAIDLAW SCHOLARS

Introduction

Within the gut, a dynamic, complex microbiome thrives in symbiosis with its host. Whilst previously neglected, this microbial ecosystem is intricately linked with our physiological processes, this obscured yet fundamental relationship unveiled by rapid advancements in whole genome sequencing and metagenomics. These bacteria not only derive nutrients from their host, but influence the immune system, mental functioning, and overall health of the host, resulting in a delicate relationship that has profound implications in human health (Wardman et al., 2022).

At the heart of this symbiosis are CAZymes, Carbohydrate Active Enzymes, defined as biological catalysts utilised by bacteria to metabolise carbohydrates, often carbohydrates the host itself is unable to degrade. This benefits the host, providing energy and nutrients from carbohydrate breakdown that would not otherwise be harvested (Garron and Henrissat, 2019). Among a myriad of CAZymes is chondroitin AC lyase (EC 4.2.2.5), an enzyme that catalyses the cleavage of a sulfate group from chondroitin sulfate. This substrate is a glycosaminoglycan that is pivotal within the human body, involved in neuronal development, inflammatory processes, and cell proliferation (Ndeh et al., 2018). The applications of chondroitin sulfate extend beyond these roles, with recent studies unveiling its potential in the treatment of inflammatory disorders such as osteoarthritis. However, the complexity in the structure of chondroitin sulfate poses challenges, potentially resolved by modifying its structure to synthesise low molecular weight chondroitin sulfate (LMWCS). This form is more readily absorbed and effective, elucidating the importance of identifying techniques to synthesise pure LMWCS in a high yield for commercial and therapeutic purposes. Enzymatic synthesis using chondroitin AC lyase may resolve the environmental and economic challenges that the chemical synthesis of LMWCS has faced (Fan et al., 2022).

This study aimed to clone and express two chondroitin AC lyases from *Bacteroides thetaiotaomicron*, using PCR to amplify the genes of interest from its genome, ligate these genes into plasmids for transformation into *E. coli*, then express and purify the two enzymes. Thin layer chromatography will be used to analyse enzymatic activity.

This essay will explore the cloning and expression of chondroitin AC lyase, delving into its potential for advancing the treatment of inflammatory conditions and developing innovative techniques to address difficulties in drug synthesis. By characterising chondroitin AC lyase, we begin to unveil the intricate nexus between microbes and human health, improving the quality of life of people globally.

Methodology

Molecular cloning

A CAZyme database (www.cazy.org, Lombard et al., 2013) was used to identify two predicted sequences encoding chondroitin AC lyase: Q8A2F5 (CL) and Q8A3J0 (CAC) (The UniProt Consortium, 2023). Primers were designed to isolate these genes from the *Bacteroides thetaiotaomicron* genome, using PCR to amplify the genes of interest (Figure 1).

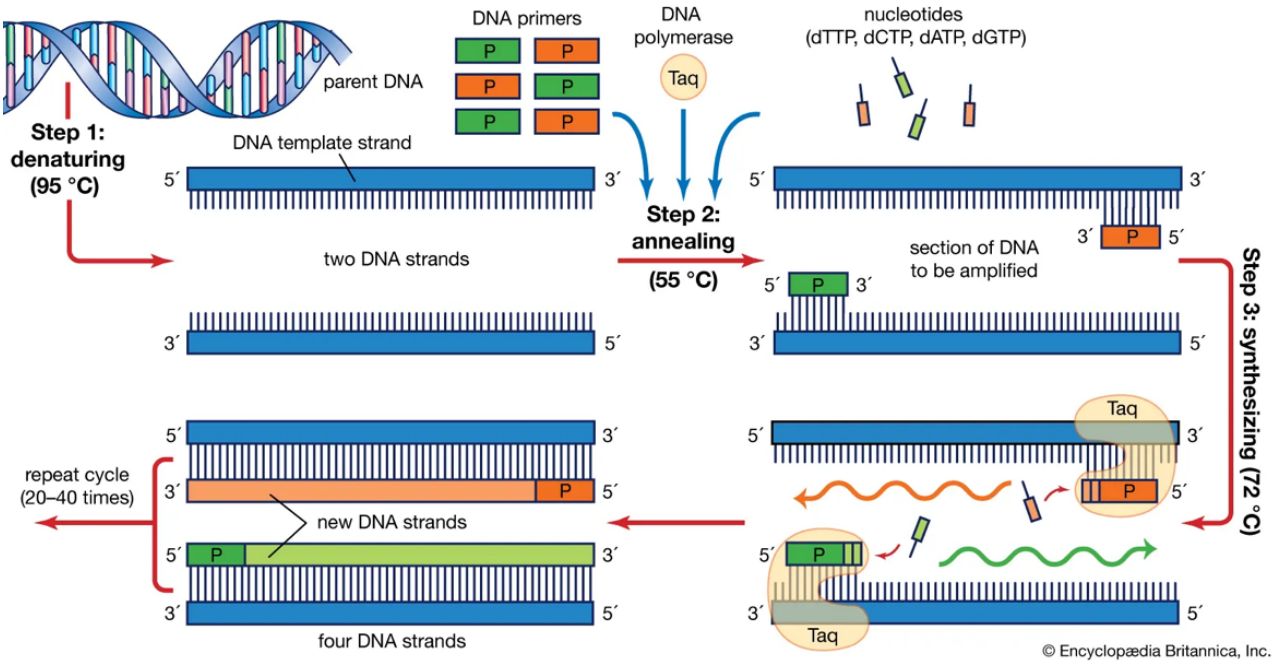

Figure 1 - Stages of Polymerase Chain Reaction. *The genome is denatured to separate the DNA strands, and the primers anneal to the complementary section of DNA. Taq polymerase extends the strand between the primers, one nucleotide at a time, to synthesise a double-stranded DNA copy of the specific section of the genome required. This cycle repeats, generating many copies of the section of DNA (Britannica, 2019).*

Once the CL and CAC genes were amplified, the genes were inserted into a plasmid backbone. These plasmids acted as a vector to deliver the genes into *E. coli* cells, enabling the cells' machinery to be exploited to make the enzyme. Two different plasmids were used, pEHISTEV to encode an enzyme with an N-terminus HIS tag, and pEBSRCTEV10HIS to encode an enzyme with a C-terminus HIS tag (Figure 2). The location of the HIS tag (a string of histidine residues that allow the enzyme to be purified once it is expressed), may affect the solubility of the protein, thus two plasmids were required to determine the HIS tag position resulting in the most soluble protein.

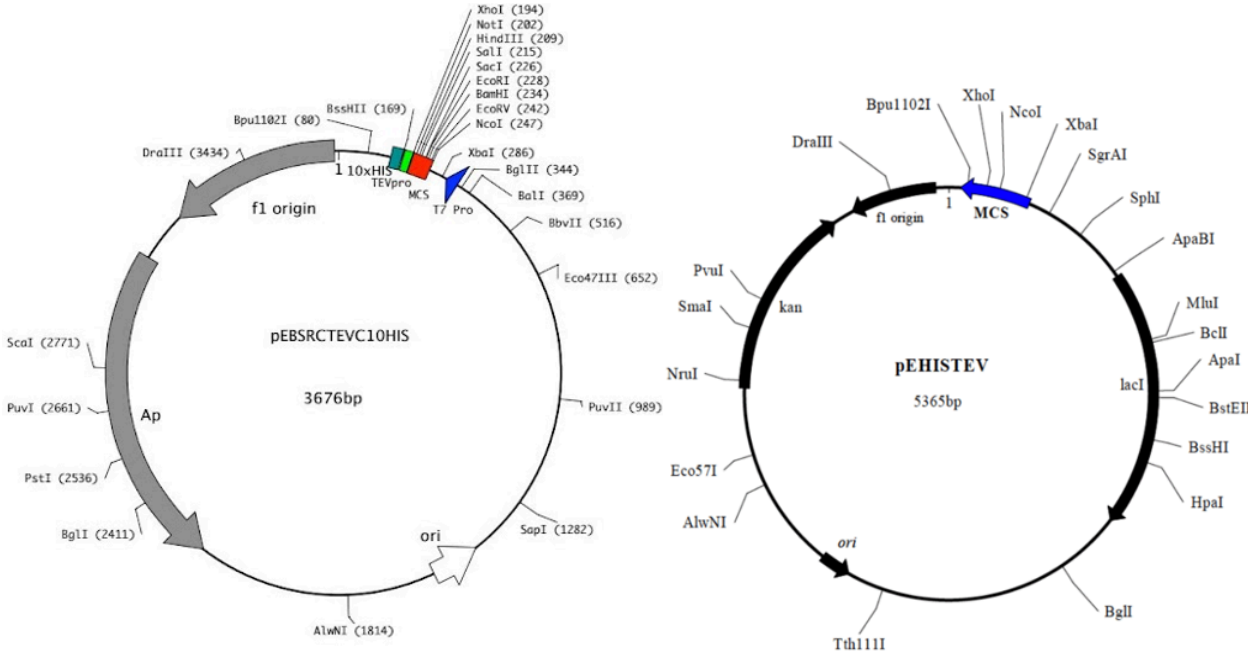

Figure 2 – The structure of plasmid vectors: pEBSRCTEV10HIS and pEHISTEV. *CL and CAC chondroitin lyase genes were inserted into these vectors to enable them to be transformed into E. coli cells for gene expression.*

To ligate the DNA into the plasmids, restriction enzymes, NcoI and XhoI, were used to digest the gene sequence and plasmids at specific points, labelled in Figure 2, to create complementary overhangs, or 'sticky ends'. The plasmids were incubated with the gene sequences, allowing the complementary sticky ends to join to form one continuous plasmid (Figure 3). Four constructs were made, CL and pEHISTEV, CL and pEBSRCTEV10HIS, CAC and pEHISTEV, and CAC and pEBSRCTEV10HIS.

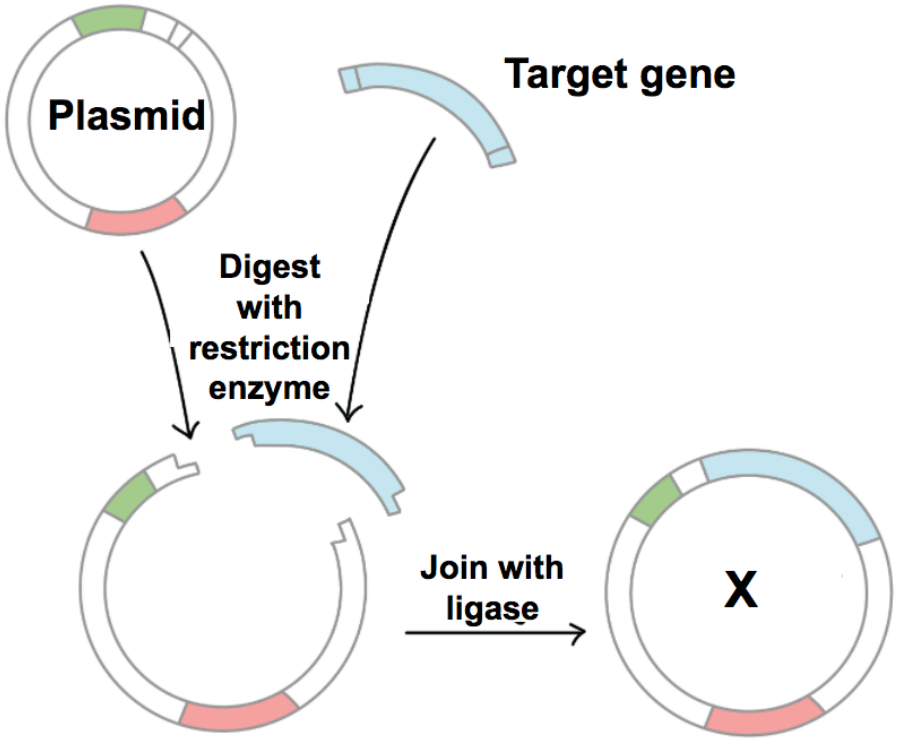

Figure 3 - Ligation of a target gene into a plasmid vector. *The plasmid backbone and target gene are ligated with specific restriction enzymes to create complementary overhangs. Ligase enzymes are used to join these overhangs, inserting the gene into the linearised plasmid backbone and forming one continuous plasmid (Kahn Academy, 2021).*

Once ligation had occurred, each plasmid was transformed into the *E. coli* genome, so that its protein synthesis machinery could be exploited to express the proteins. DH5α *E. coli* cells were used for the transformation, with each plasmid added to *E. coli* cells, incubated on ice, heat shocked in a water bath, and returned to ice, to integrate the plasmids into the *E. coli* genome. A nutrient broth was added to the cells and incubated at 37°C for an hour, to allow the cells to replicate.

To ensure the plasmids had integrated into the *E. coli* genome, the cells were plated onto agar plates with an antibiotic. As the plasmids contained a gene that conferred antibiotic resistance (Figure 2), plating the cells onto agar plates with antibiotics and incubating overnight resulted in growth of bacteria colonies possessing genomes with the integrated plasmid, indicating successful transformation. To confirm that the colonies were not contaminants, three colonies from each of the four constructs were placed in a labelled vial containing the respective antibiotic and incubated overnight at 37°C. Further growth of colonies confirmed that the plasmid had integrated into the bacterial genome, and their plasmids were isolated and digested using restriction enzymes. If the plasmids contained the target gene, restriction digests would cut out the gene, and SDS-PAGE

results would show two separate bands in a lane, corresponding with the molecular weights for the plasmid and gene. Based on the results, one colony from each of the four constructs was selected, sequenced to ensure that no mutations had occurred in the genome, and used for expression.

Expression

Once the plasmids were transformed into the bacterial genome, the *E. coli* cells were grown on a larger scale. The selected constructs were transformed into BL21 *E. coli* cells that were more suitable for a higher rate of protein expression. The transformation procedure was repeated, each culture resuspended in a greater volume of nutrient broth and antibiotic, and incubated at 37°C, shaking. The concentration of bacterial cells was measured using optical density (OD), to record the amount of light (at 600 nm) able to pass through. As the bacterial cells replicated, the broth became cloudier, and less light could penetrate. Once the OD_{600} reached 0.6 - 0.9, there were enough bacterial cells to be induced and start producing the protein.

As protein expression must be tightly regulated, the *E. coli* genome has mechanisms to control when, where, and how much protein is expressed. By adding IPTG to the *E. coli* cells, protein expression was chemically induced, initiating transcription for the synthesis of the target protein. To determine the optimal incubation conditions to produce soluble protein, small-scale expression tests were performed, by inducing each construct with IPTG in different temperatures, either for 3 - 4 hours or overnight. Each tube was centrifuged to form a pellet of bacterial cells, then were lysed to release the protein. Assuming that any soluble protein would be in the supernatant, the cell contents were purified to harvest the target protein. SDS-PAGE was used to analyse samples of the pellet, lysate and purified protein, to determine the best conditions for producing a soluble protein. These results were also used to determine which plasmid (pEHISTEV or pEBSRCTEV10HIS) resulted in the most soluble protein. Once the most appropriate plasmid and conditions were determined, further protein expression was induced to produce the proteins on a larger scale.

Purification

Following large-scale protein expression, several steps were needed to purify the target protein from other *E. coli* proteins and contaminants. The cells were centrifuged to form a pellet of bacteria, and mechanically lysed using a cell disruptor. Ni-NTA affinity chromatography was used to harvest the protein of interest from the lysate, by binding the protein's HIS tag to resin beads within an affinity chromatography column. The column was washed to remove any contaminants and unwanted proteins, leaving the protein of interest bound to the column. The proteins were then eluted, using an imidazole buffer with a higher affinity for the resin beads than the HIS tag to displace the proteins, and the eluent (containing the purified protein) was collected (Figure 4). To ensure that the imidazole buffer did not interfere with the further purification steps, a desalt step was used to exchange the imidazole buffer salt with water.

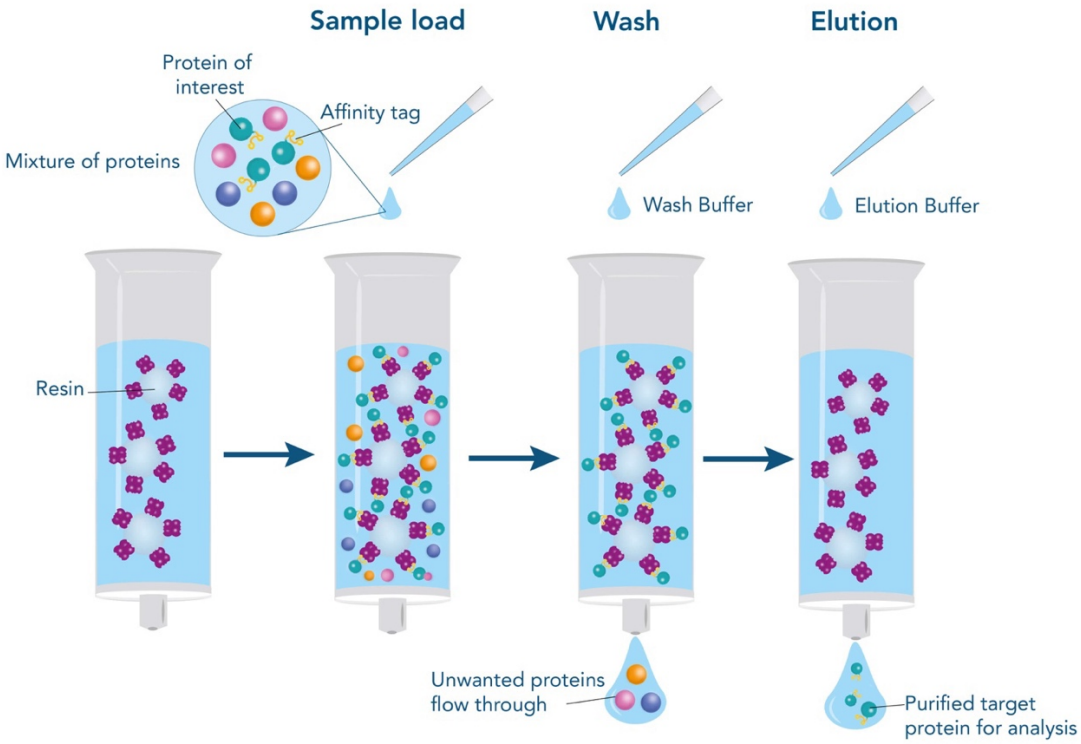

Figure 4 – Principle of Ni-NTA affinity chromatography. *Tagged proteins are bound to nickel beads on resin in a Ni-NTA column and washed to remove contaminants and untagged proteins. The purified protein of interest is then displaced with an elution buffer, therefore eluting from the column to be collected for analysis (IBA Lifesciences, n.d.).*

Once the crude protein was purified, the HIS tag was removed using TEV protease, an enzyme that binds to and cleaves a specific site between the HIS tag and the protein. TEV protease was

added to the proteins and incubated in a cold room overnight to facilitate binding. Ni-NTA affinity chromatography was repeated to separate the protein from the HIS tag and TEV, however, the protein eluted in the first wash, as it no longer had the HIS tag to bind to the resin beads. This meant that only the cleaved HIS tag and TEV (possessing a HIS tag) remained bound to the column, and this was later displaced using a high-concentration imidazole buffer and eluted. Samples from each stage of affinity chromatography were collected and analysed using SDS-PAGE, to confirm that the HIS tag was cleaved. The purified protein was concentrated, and size exclusion chromatography was performed to identify the fractions with the purest protein (where the graph peaked, Figure 5) to be used for analysis.

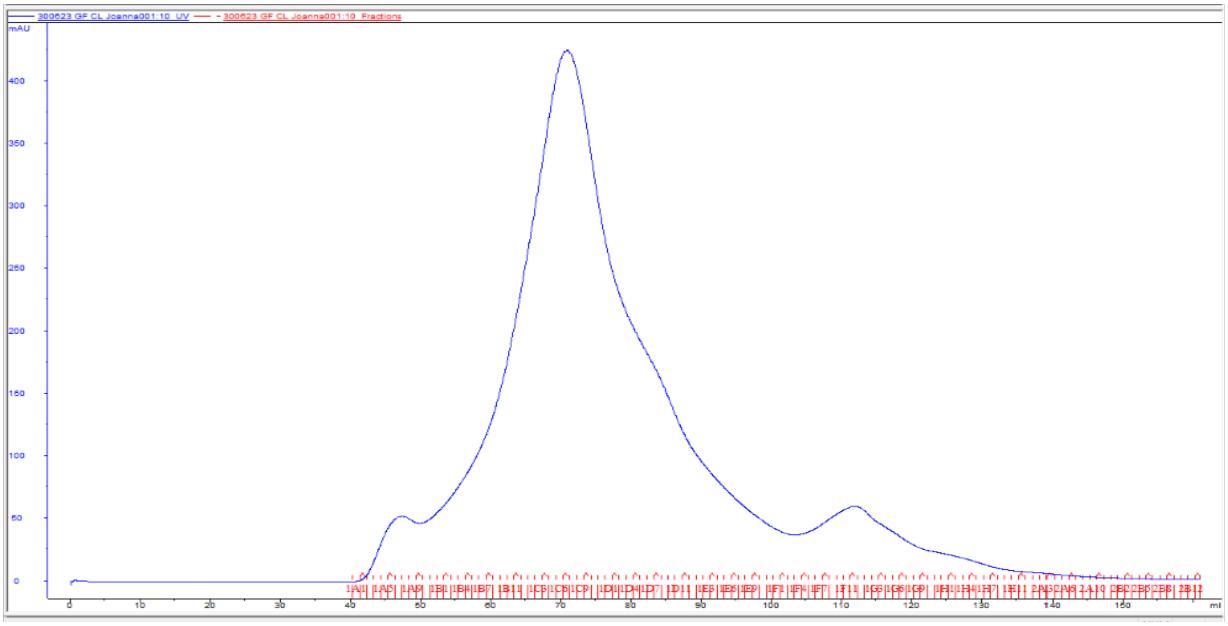

Figure 5 - Size exclusion chromatography of the CL construct. *Shown is the absorbance in mAU against volume in ml, with the peak showing the fractions with the highest concentration of pure protein.*

Analysis

Thin Layer Chromatography was used to determine whether the enzyme was active. Reactions were set up by incubating the enzyme at 37°C overnight with either 1 µM or 10 µM of substrate (chondroitin sulfate). A 2 µl spot of each reaction was pipetted on a silica plate, alongside a control of pure chondroitin sulfate (2 mg/ml), and a control of either 1 µM or 10 µM enzyme. The plate was placed in a solvent of water/acetonitrile (3:7). Different components have different affinities for the mobile phase (the solvent) and the stationary phase (the silica plate), therefore, if the

chondroitin sulfate was broken down by the enzyme, there would be more than one component in the spot, and these would migrate different distances as the solvent moved up the plate, showing two distinct spots. The spots on the plate were compared to the control (pure chondroitin sulfate), to determine whether the substrate was broken down by the enzyme.

Results

Protein expression

Once the genes of interest were isolated and amplified using PCR, gel electrophoresis was used to determine whether the PCR was successful. Figure 6 depicts a band at 2.8 kb for both CL PCR products, and a band at 2.1 kb for both CAC products, therefore, the genes of interest were successfully isolated and amplified from the *Bacteroides thetaiotaomicron* genome.

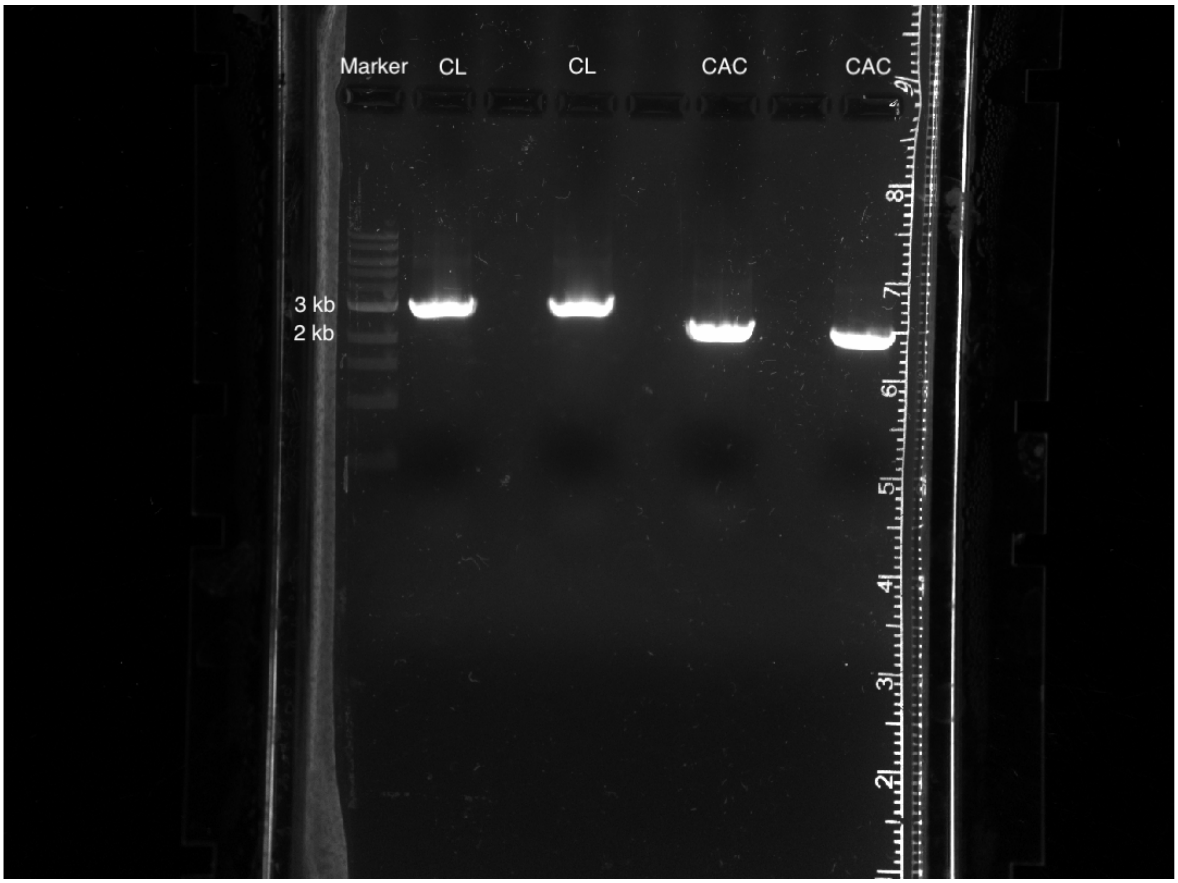

***Figure 6 - Gel electrophoresis results of PCR amplification of the CL and CAC genes from Bacteroides thetaiotaomicron**. Primers were designed to isolate the genes of interest from the Bacteroides thetaiotaomicron genome, and these were amplified to produce multiple copies of the genes.*

The CL and CAC genes were digested and ligated into one of two plasmid backbones, pEHISTEV and pEBSRCTEV10HIS. SDS-PAGE was used to determine whether these constructs were transformed into *E. coli* cells, by isolating and digesting the plasmids with restriction enzymes and comparing the molecular weights of the bands with the predicted values of the gene and plasmid backbone. As shown in lanes 1A, 1B and 1C of Figure 7, two bands were seen around 2.8 kb and 5.4 kb for the CL gene and pEHISTEV plasmid, respectively. For lanes 2A and 2B, two bands were seen at 2.8 kb for the CL gene and 3.7 kb for the pEBSRCTEV10HIS plasmid. In lanes 3A, 3B and 3C, two bands were seen at 2.1 kb and 5.4 kb, corresponding with predicted values for CAC and pEHISTEV, respectively. Therefore, the results concluded that these constructs were transformed into the *E. coli* genome. Lane 4A showed only one band, therefore this ligation had failed. Lane 4B showed two bands at 2.1 kb and 3.7 kb for CAC and pEBSRCTEV10HIS, respectively, thus, this transformation was successful. Based on these results, 1B, 2B, 3A and 4B were sequenced, confirming that no mutations had occurred, and used for further expression.

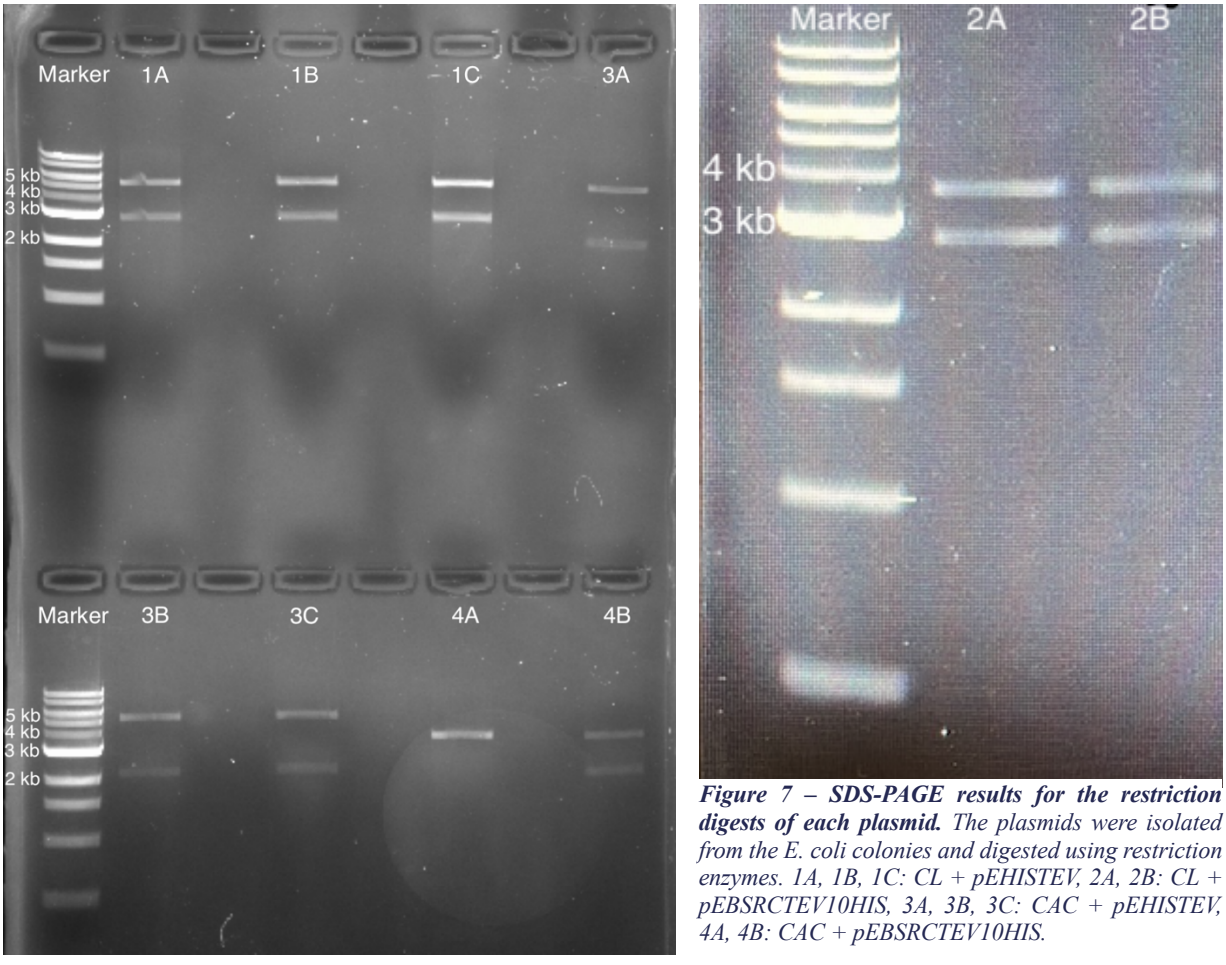

Figure 7 – SDS-PAGE results for the restriction digests of each plasmid. *The plasmids were isolated from the E. coli colonies and digested using restriction enzymes. 1A, 1B, 1C: CL + pEHISTEV, 2A, 2B: CL + pEBSRCTEV10HIS, 3A, 3B, 3C: CAC + pEHISTEV, 4A, 4B: CAC + pEBSRCTEV10HIS.*

The four selected colonies were induced under four different conditions, and the solubility of each protein was compared to determine the best condition and plasmid for the expression of a soluble protein. As shown in Figure 8, bands were seen in the elution columns for plasmids 1 and 3 (where pure, soluble protein would be expected), corresponding with the proteins expressed with a N-terminal HIS tag (using the pEHISTEV plasmid). Therefore, the position of the HIS tag on the N-terminus of the protein was crucial for the expression of a soluble protein, and further protein expression was conducted using the pEHISTEV plasmid. As indicated by the black arrows in Figure 8-1 and Figure 8-3, the clearest band in the elution column was seen in condition d (induction at 15 °C, overnight). These results indicated that these conditions produced the most soluble protein, and thus were used to express larger amounts of protein for further purification and analysis.

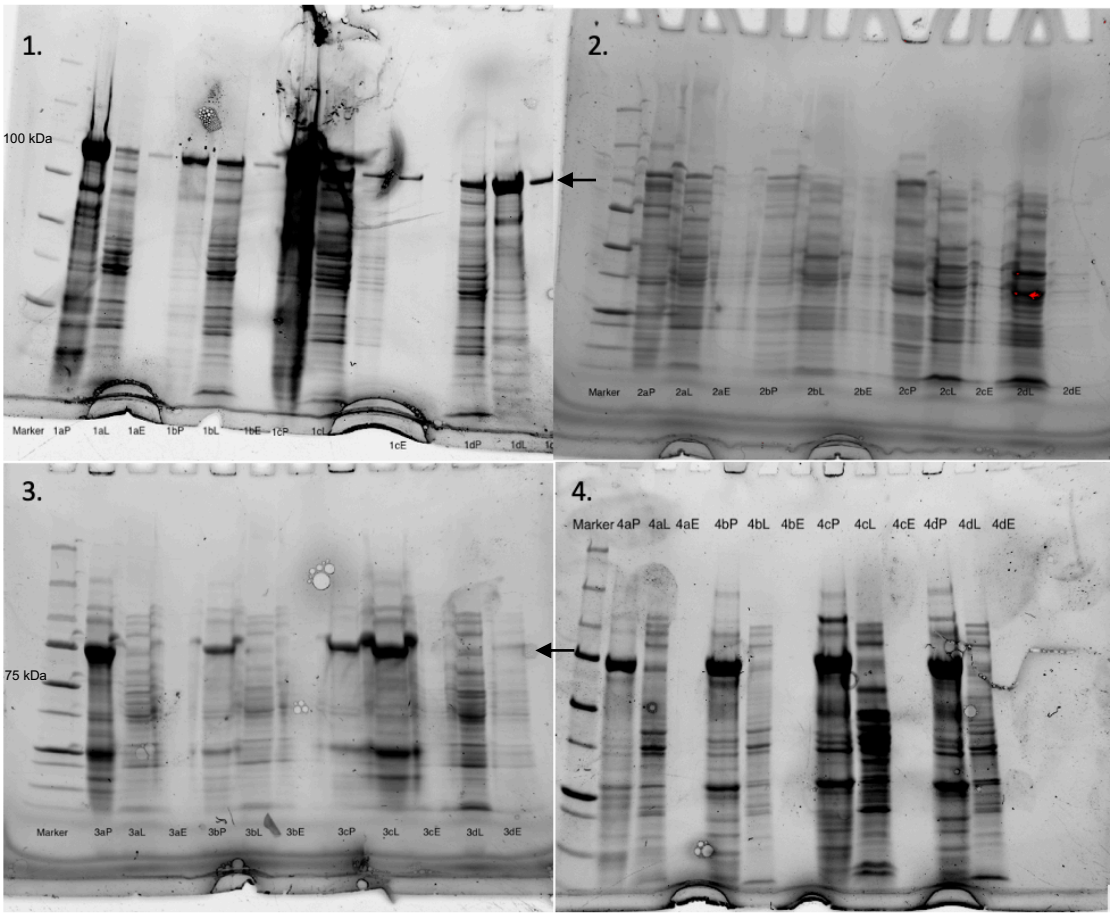

Figure 8 - SDS-PAGE of each construct following induction with IPTG in four conditions. *Each of the four plasmids (1. CL + pEHISTEV, 2. CL + pEBSRCTEV10HIS, 3. CAC + pEHISTEV, 4. CAC + pEBSRCTEV10HIS) were induced under four different conditions: a. 37 °C, 3 - 4 hours. b. 25 °C, 3 - 4 hours. c. 25 °C, overnight. d. 15 °C, overnight. Samples were taken from the bacterial cell pellet (P), lysate (L) and purified protein (E) for comparison. The black arrow indicates the eluent column with the highest concentration of pure, soluble protein.*

Purification

Once the selected constructs were expressed on a larger scale, the HIS tag was cleaved from the crude CL and CAC proteins to produce a pure protein for analysis. SDS-PAGE results for the CAC protein (Figure 9) showed a band at 82 kDa in the flow-through step, thus the pure target proteins had eluted, whilst the HIS tag had been cleaved and remained in the column. However, a band at 82 kDa was also seen in the eluent step, therefore some of the proteins had not had the HIS tag cleaved by the TEV protease. Several bands were also seen in this column, possibly due to degradation of the target protein.

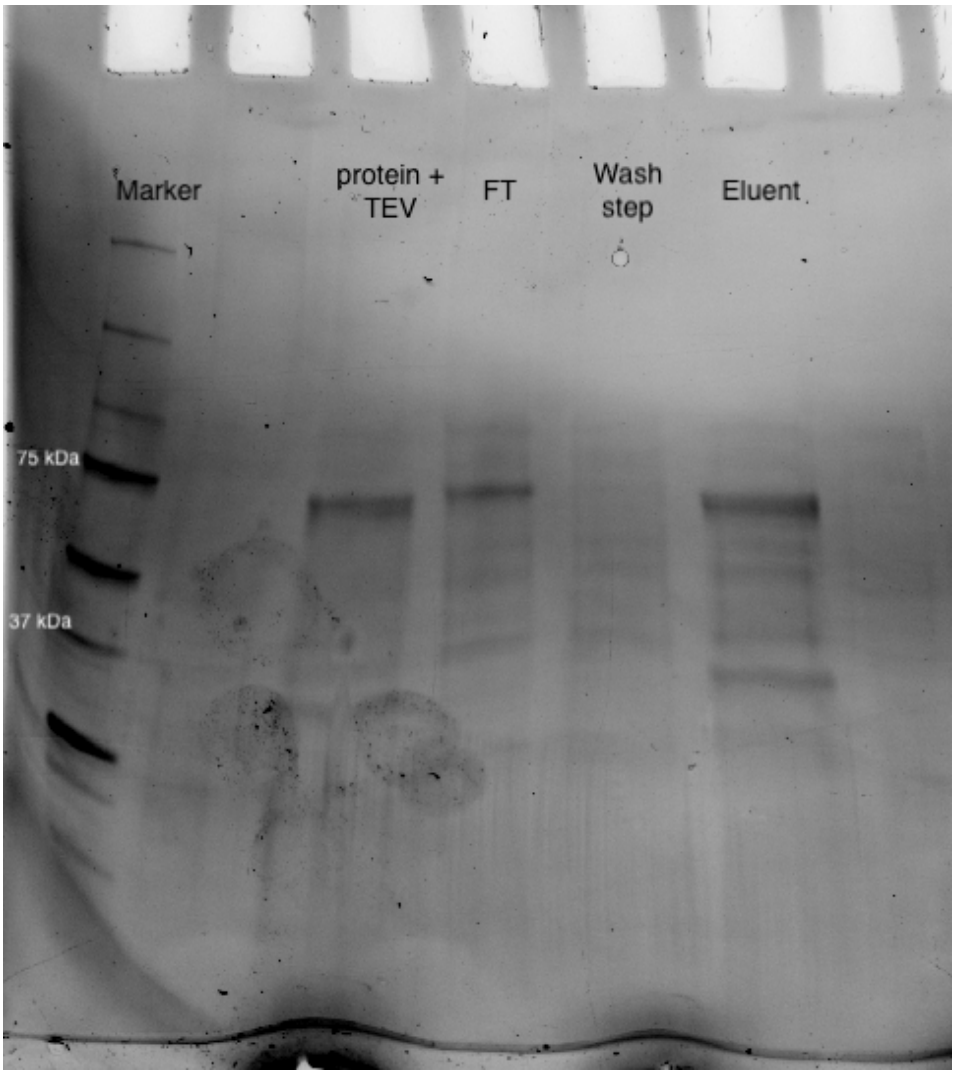

Figure 9 – SDS-PAGE results of HIS tag cleavage and purification of CAC chondroitin lyase. *A sample was taken before incubation of CAC chondroitin lyase and TEV (protein + TEV). Ni-NTA affinity chromatography was used to purify the protein, and samples were taken from the flow-through (FT), wash with a low concentration of imidazole (wash step), and wash with the elution buffer (eluent).*

The SDS-PAGE results seen in Figure 10 showed a significant amount of pure CL chondroitin lyase enzyme had been produced. A strong band around 100 kDa was seen in the flow-through and wash steps of purification, correlating with the predicted molecular weight of 108 kDa for CL. As the band seen in the wash step was significant, the products of the flow-through and wash step were combined. A band was also seen around 28 kDa in the eluent column, correlating with the molecular weight for the TEV protein which eluted when displaced with a high concentration of imidazole buffer.

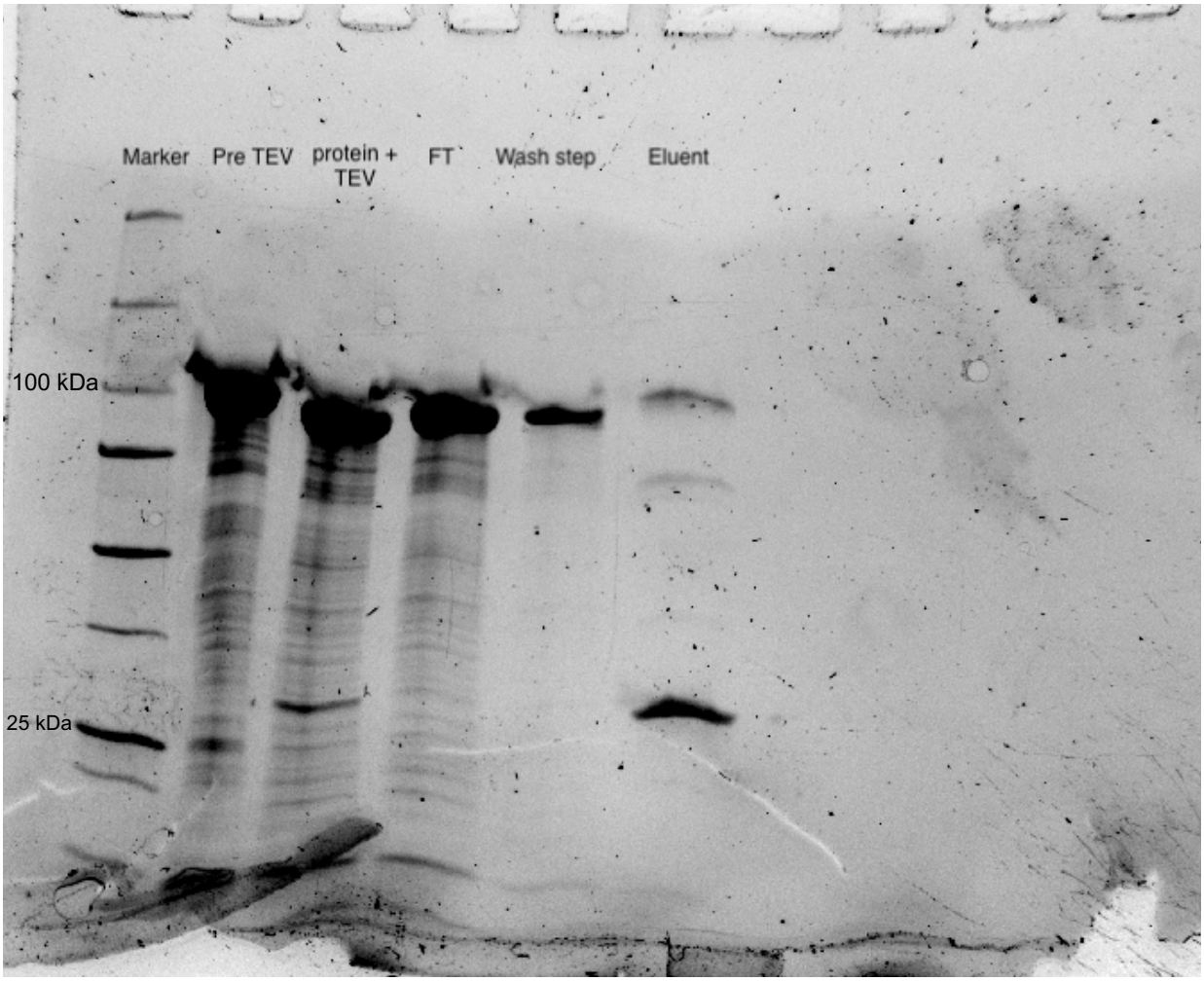

Figure 10 – SDS-PAGE analysis of HIS tag cleavage and purification of CL chondroitin lyase. *Samples were taken before TEV was added (pre-TEV) and before the protein and TEV were incubated (protein + TEV). Ni-NTA affinity chromatography was used to purify the protein, and samples were taken from the flow-through (FT), the wash with low-concentration imidazole buffer (wash step), and the wash with the elution buffer (eluent).*

Thin Layer Chromatography (TLC)

The activities of the CL and CAC chondroitin lyases were analysed using thin layer chromatography. A control of the substrate, chondroitin sulfate, was used for comparison against the reactions. As shown in Figure 11-A, two distinct spots were seen in lanes C and D, showing that chondroitin sulfate was broken down, thus CAC chondroitin lyase was active. Similarly, two distinct spots were seen in lanes 4 and 5 (Figure 11-B), showing that an active CL chondroitin lyase had degraded chondroitin sulfate.

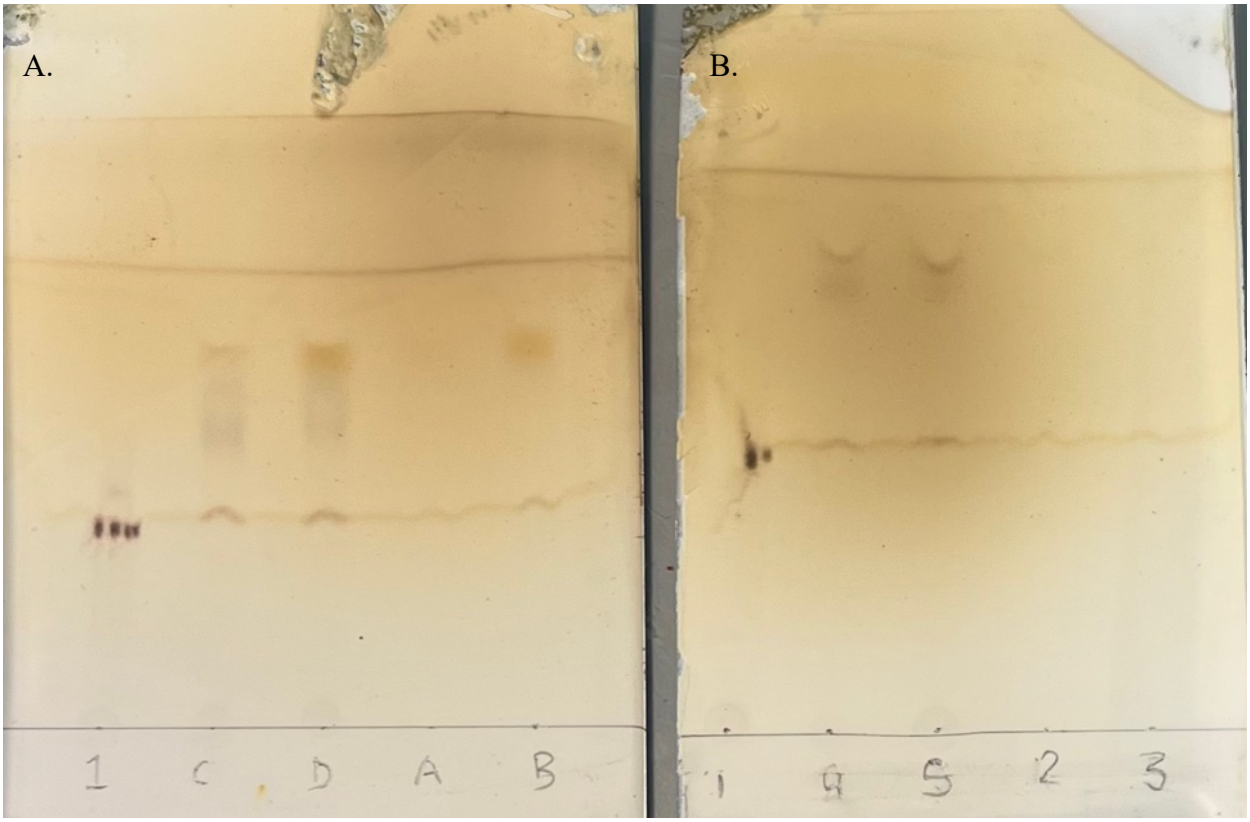

Figure 11 – TLC analysis of the degradation of chondroitin sulfate (CS) by chondroitin AC lyases CL and CAC. *A: Lane 1 - control of CS. Lane A - control of 1 µM CAC chondroitin lyase. Lane B - control of 10 µM CAC chondroitin lyase. Lane C - reaction with 1 µM CAC chondroitin lyase and CS. Lane D - reaction with 10 µM CAC chondroitin lyase and CS. B: Lane 1 - control of CS. Lane 2 - control of 1 µM CL chondroitin lyase. Lane 3 - control of 10 µM CL chondroitin lyase. Lane 4 - reaction with 1 µM CL chondroitin lyase and CS. Lane 5 - reaction with 10 µM CL chondroitin lyase and CS.*

Discussion

This project aimed to clone, express and characterise chondroitin AC lyase, by selecting two gene sequences, CL and CAC, from a database (www.cazy.org), predicted from computational analysis of the *Bacteroides thetaiotaomicron* genome. The sequences were cloned and expressed to make soluble proteins, then purified and analysed to determine whether they were active chondroitin AC lyases. As shown in Figure 6, the two genes were successfully isolated and amplified using PCR, then were ligated into one of two plasmids (Figure 2). These were transformed into *E. coli* cells, so its protein synthesis machinery could be exploited to produce soluble proteins. As seen in Figure 8, the optimal conditions for producing the soluble proteins were determined to be expressing the proteins with an N-terminal HIS tag, using the pEHISTEV plasmid, and inducing the *E. coli* cells overnight at 15 °C. Following the cleavage of the HIS tag from the proteins, and further purification, thin layer chromatography was used to confirm that both proteins were active and had specificity for chondroitin sulfate (Figure 11). Therefore, the results confirmed that these gene sequences from *Bacteroides thetaiotaomicron* coded chondroitin AC lyases.

However, it is important not to neglect the limitations of this research. Although the SDS-PAGE results of the CAC chondroitin lyase indicated the successful purification of the enzyme (Figure 9), significantly less was produced than CL chondroitin lyase (Figure 10). If repeated, further analysis should focus on whether other conditions may have been more effective in producing soluble CAC enzymes.

The findings of this research merely scratch the surface of the intricate microbiome within the gut, yet are instrumental considering the medical and biotechnological uses of this enzyme. Producing active chondroitin lyase on a commercial scale provides a more environmentally ethical, safer, and cheaper method to synthesise low molecular weight chondroitin sulfate (LMWCS), that, according to Li et al. (2016), can be used in the treatment of conditions such as osteoarthritis. By determining the optimal conditions and plasmids for making a soluble protein, this work paves the way for unravelling the enigmatic mechanisms of this enzyme. Further characterisation could be undertaken, such as by performing X-ray crystallography to determine and compare the structures of each of the chondroitin lyases or comparing the activities of the enzymes using assays.

A fusion of biochemistry, molecular techniques and scientific exploration underpins this research, exemplifying the power of persistence and innovation in unlocking new discoveries – ones that make us more equipped to navigate and shape the future of medicine and biotechnology. These qualities can be extrapolated to those displayed in a good leader, ones I hope to reflect in my leadership in action project.

My greatest thanks are extended towards Lord Laidlaw and the Laidlaw Foundation for the funding to complete this research and the opportunity to grow and develop both in and out of the lab. I am also grateful to my supervisor, Dr Tracey Gloster for the resources and support in undertaking this research.

References

Britannica (2019). polymerase chain reaction | Definition & Steps. In: *Encyclopædia Britannica*. [online] Available at: https://www.britannica.com/science/polymerase-chain-reaction.

Fan, X.-M., Huang, J.-Y., Ling, X.-M., Wei, W., Su, W.-B. and Zhang, Y.-W. (2022). A Highly Active Chondroitin Sulfate Lyase ABC for Enzymatic Depolymerization of Chondroitin Sulfate. *Polymers*, 14(9), p.1770. DOI: 10.3390/polym14091770.

Garron, M.-L. and Henrissat, B. (2019). The continuing expansion of CAZymes and their families. *Current Opinion in Chemical Biology*, 53, pp.82–87. DOI: 10.1016/j.cbpa.2019.08.004.

IBA Lifesciences. (n.d.). *Protein Affinity Chromatography*. [online] Available at: https://www.iba-lifesciences.com/applications/protein-affinity-chromatography/.

Li, L., Li, Y., Feng, D., Xu, L., Yin, F., Zang, H., Liu, C. and Wang, F. (2016). Preparation of Low Molecular Weight Chondroitin Sulfates, Screening of a High Anti-Complement Capacity of Low Molecular Weight Chondroitin Sulfate and Its Biological Activity Studies in Attenuating Osteoarthritis. *International Journal of Molecular Sciences*, 17(10), p.1685. DOI: 10.3390/ijms17101685.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42(D1), pp.D490–D495. DOI: 10.1093/nar/gkt1178.

Ndeh, D., Munoz Munoz, J., Cartmell, A., Bulmer, D., Wills, C., Henrissat, B. and Gray, J. (2018). The human gut microbe Bacteroides thetaiotaomicron encodes the founding member of a novel glycosaminoglycan-degrading polysaccharide lyase family PL29. *The Journal of Biological Chemistry*, [online] 293(46), pp.17906–17916. DOI: 10.1074/jbc.RA118.004510.

Overview: DNA cloning (article). (2021). [Online image] *Kahn Academy*. Available at: https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/biotechnology/a/overview-dna-cloning. [Accessed 6 Aug. 2023].

The UniProt Consortium. (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research*, 51, pp.D523–D531. DOI: 10.1093/nar/gky092.

Wardman, J.F., Bains, R.K., Rahfeld, P. and Withers, S.G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nature Reviews Microbiology*, 20, pp.1–15. DOI: 10.1038/s41579-022-00712-1.