Optimization of Extraction and Purification of Taq Polymerase from *E. Coli* Cells

Submitted by

Madeline Rae Trombly

Bioengineering

To

The Honors College

Oakland University

In partial fulfillment of the

requirement to graduate from

The Honors College

Mentor: Dr. Gerard Madlambayan, Professor of Engineering

Department of Biological Sciences

Oakland University

April 1, 2022

Abstract

The production of protein-based drugs rely on a basic protein purification method to extract and purify the desired protein from host cells. This research seeks to optimize the protein purification process to isolate Taq polymerase from *E. coli* host cells, which is used to model the purification of protein-based drugs such as insulin. This research identifies a purification process and subsequently explores the impact of: 1) changing the incubation time and temperature of the initial cell culture, 2) inducing protein expression with different concentrations of IPTG, and 3) changing the flow rate during ion-exchange column chromatography. These changes may introduce more time-efficient steps and/or yield a larger volume of pure, active protein than the traditional purification method. Both of these results would improve the drug industry's efficiency by allowing companies to produce a large amount of active product in a short amount of time at a low cost.

Introduction

People all over the world suffer from diabetes, a condition where their bodies are unable to produce insulin, a protein that maintains blood sugar levels. Because they lack the ability to naturally produce insulin, diabetic patients must receive insulin treatments that differ based on the type of diabetes the individual has. An unhealthy diet and lack of adequate exercise can lead to the development of insulin resistance, which results in Type 2 diabetes (Malone & Hansen, 2018). Patients with this diagnosis will eventually need to receive insulin treatment, but may have escaped requiring it for a good portion of their lives because the disease developed over time due to their lifestyle habits. Patients diagnosed with Type 1 diabetes, on the other hand, require constant insulin treatments throughout their lives because they are usually diagnosed as a child and Type 1 diabetes is considered an autoimmune disease (Norris et al., 2020).

Since insulin is so important to both types of diabetic patients, and a lifetime concern for those with Type 1 diabetes, it is essential that scientists are able to produce the drug in a way that is effective and efficient, while also taking into account the impact the production of insulin may have on other people, the environment, and the economy. One of the most recently developed methods for producing insulin grows the protein in an *E. coli* cell, then performs a purification process to extract and purify as much active insulin as possible from those cells (Zieliński et al., 2019).

There are several different processes currently used for protein purification, some of which are aqueous two-phase extraction, solid-phase extraction, isoelectric precipitation, organic solvent precipitation, salting out, dialysis, ultrafiltration, capillary electrophoresis, and a variety

of chromatography methods (Liu et al., 2020). Most protein purification and separation processes contain at least one chromatographic step, and the types of chromatography that can be used include high-performance liquid chromatography, gel chromatography, hydroxyapatite chromatography, hydrophobic interaction chromatography, affinity chromatography, supercritical fluid chromatography, and ion-exchange chromatography (Liu et al., 2020). This research will focus on ion-exchange chromatography, which uses a salt gradient and passes a solution through a column containing a resin, which first binds to the protein and then later elutes the protein as solution runs through the column; in short, the proteins are separated according to the surface charge of their molecules (Schmidt et al., 2013).

Taq polymerase is a thermostable protein that is widely used to amplify DNA segments in polymerase chain reactions (PCR) for molecular biology applications because of its ability to withstand high temperatures (Gelfand, 1989). Taq can be grown in *E. coli* cells, extracted, and purified in a manner similar to insulin (Din et al., 2020). Since these processes are so similar, Taq polymerase purification can serve as a model for insulin purification. While current methods claim to be the most cost-effective, highest-yield methods for Taq polymerase purification, this study seeks to further improve those claims by altering multiple steps of the Taq polymerase purification process. If successful in producing higher yield, the method investigated in this study can be applied to insulin production to ensure that the drug is made more easily obtainable, both cost-wise and supply-wise, for those who need it most. In addition to serving as a model for insulin purification, the method for Taq purification outlined in this research can be beneficial to the biology labs at Oakland University if it allows Taq to be produced in-house more cost-effectively than Taq can be purchased commercially.

Methods

Buffer Preparation

- Buffer A (50 mL)
    - 50 mM Tris-Cl PH 7.9 (2.5 mL)
    - 50 mM Dextrose (0.45 g)
    - 1 mM EDTA (0.2 mL)
- Buffer B (20 mL)
    - 10 mM Tris-Cl PH 7.9 (20 µL)
    - 50 mM KCl (1 mL)
    - 1 mM EDTA (40 µL)
    - 1 mM PMSF (100 µL)
    - 0.5% TWEEN 20 (400 µL)
    - 0.5% NP-40 (400 µL)
- Buffer C (200 mL)
    - 20 mM HEPES PH 7.9
    - 1 mM EDTA (0.8 mL)
    - 0.5% TWEEN 20 (1.0 mL)
    - 0.5% NP-40 (1 mL)
    - 5 mM DTT (154 µL)
- Storage Buffer (2000 mL)
    - 50 mM Tris-Cl PH 7.9 (100 mL 1M)
    - 100 mM NaCl (7.45 g)
    - 0.1 mM EDTA (400 µL 0.5M)

- 1% Triton X-100 (20.0 mL 1M)
- 1 mM DTT (0.308 mL 1M)
- In 50% glycerol

Column Preparation

- Pinch clamp is used to hold column in place
- Approximately 24 mL of DEA Resin is added to the column
- Pass 150 mL of Buffer C through the resin to prepare column for ionization
- Label finished columns and store at 4°C until ready to be used

Standard Protocol

This experiment was completed as a part of a bioengineering senior design project. The basic protocol for the purification of Taq polymerase from *E. coli* cells was provided by the advisors for the project and is listed as follows:

1. T-streak Taq Polymerase bacterial stock onto LB Agar + Amp (100 µg/mL) plate. Grow overnight in an incubator at 37°C.
2. Isolate a single colony and inoculate 50 mL of LB + amp (100 µg/mL). Grow on a shaker for 24 hours (220 RPM, 37°C).
3. Grow to an OD of 0.5 at 600 nm and induce with 335 µL IPTG (20% w/v in H_2O). Continue growing while shaking for 24 hours (220 RPM, 37°C).

4. Spin down cells for 10 minutes at 10K RPM and resuspend in 20 mL Buffer A. Pour cells into a 30 mL corex tube and spin down again for 10 minutes at 10K RPM.
5. Resuspend in 10 mL Buffer A containing 4 mg/mL lysozyme. Incubate for 15 minutes.
6. Add 10 mL Buffer B and incubate for 1 hour at 75°C.
7. Spin down for 15 minutes at 10K RPM. Save supernatant and freeze at -80°C.
8. Apply to a DE-52 column (8 mL) pre-equilibrated with Buffer C containing 25 mM KCl. Wash the column with 50 mL of equilibration buffer.
9. Now apply an 80 mL KCl gradient in Buffer C ranging from 25-700 mM. Collect 300-drop fractions (6 mL) into tubes. Taq polymerase should come off between fraction 4 and 7.
10. Test for the presence of the right protein in the fractions by running a 10% acrylamide separating gel with a 4% stacking gel. The Taq polymerase should be the first visible protein to elute off the column and it is 94 kD.
11. Dialyze the fractions with the peak activity against the Storage Buffer according to manufacturer's protocol.
12. Dialyzed fractions should be stored at -20°C or used in PCR.
13. Test for activity in the fractions by diluting with H_2O 1:10 and 1:100 and using 1 µL from each fraction as the enzyme in a standard PCR reaction.

An outline of the overall major steps of the standard protocol can be seen in Figure 1.

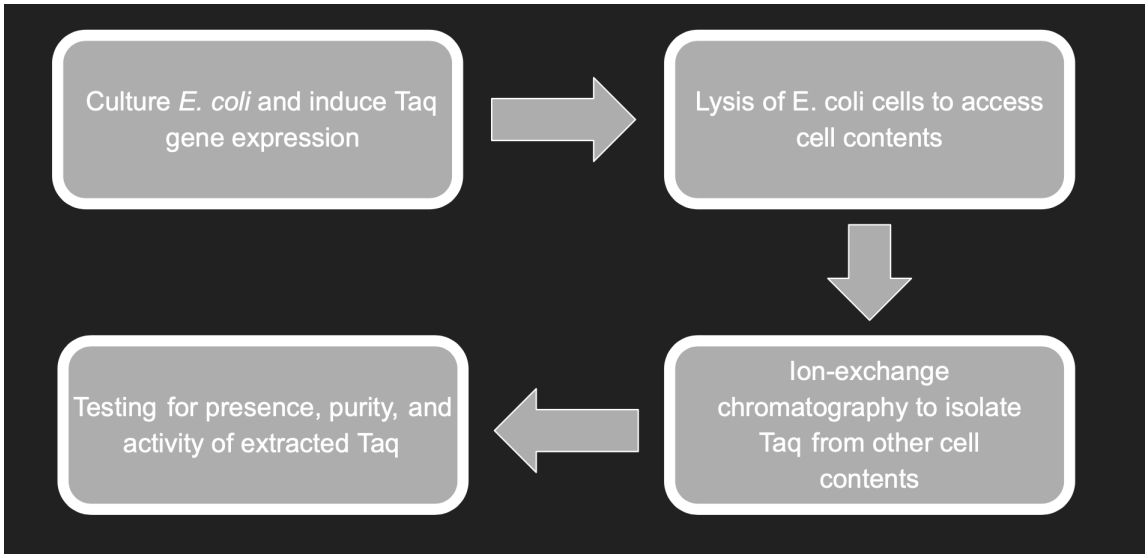

Fig 1: Major steps in basic protocol for Taq purification.

Experimental Design

This experiment was intended to determine what changes could be made to the standard protocol that would improve the efficiency of Taq production in terms of the amount of active Taq produced and the cost. In an effort to do this, changes to two main areas of the standard protocol were included in the design of the experiment: 1) the induction of gene expression in the *E. coli* cells and 2) the separation of Taq from *E. coli* cell contents during ion-exchange chromatography.

Initial Design

The presence of the lac operon in IPTG allows it to induce the expression of the Taq gene. Various concentrations of IPTG have been used in several studies and have been shown to successfully induce the expression of the gene, but it has been suggested that higher

concentrations of IPTG can be toxic because of *E. coli*'s inability to metabolize high concentrations of IPTG efficiently (Dvorak et al., 2015). To avoid the effect of IPTG toxicity on Taq production, using a more natural source for induction may prove beneficial. The lac operon necessary for induction can be found in regular cow's milk, which contains lactose. In addition to avoiding the negative effects of IPTG toxicity, using cow's milk for induction could reduce the cost of materials needed to produce Taq. Keeping in mind that the standard process is effective and the purpose of this experiment is to attempt to improve it, five samples were created: two samples where the *E. coli* was induced with a low concentration of IPTG (half the standard concentration, or 167.5 μL), two samples where it was induced with a high concentration of IPTG (double the standard concentration, or 670 μL), and one sample where the *E. coli* was induced with skim milk. The two samples each of low concentration and of high concentration were included for redundancy, and the low and high concentrations chosen to determine if the effectiveness of the skim milk as an inducing agent was better or worse than the standard concentration of IPTG by seeing if its effectiveness was closer to the low concentration or to the high concentration of IPTG. A flowchart of this step of the experimental design can be seen in Figure 2.

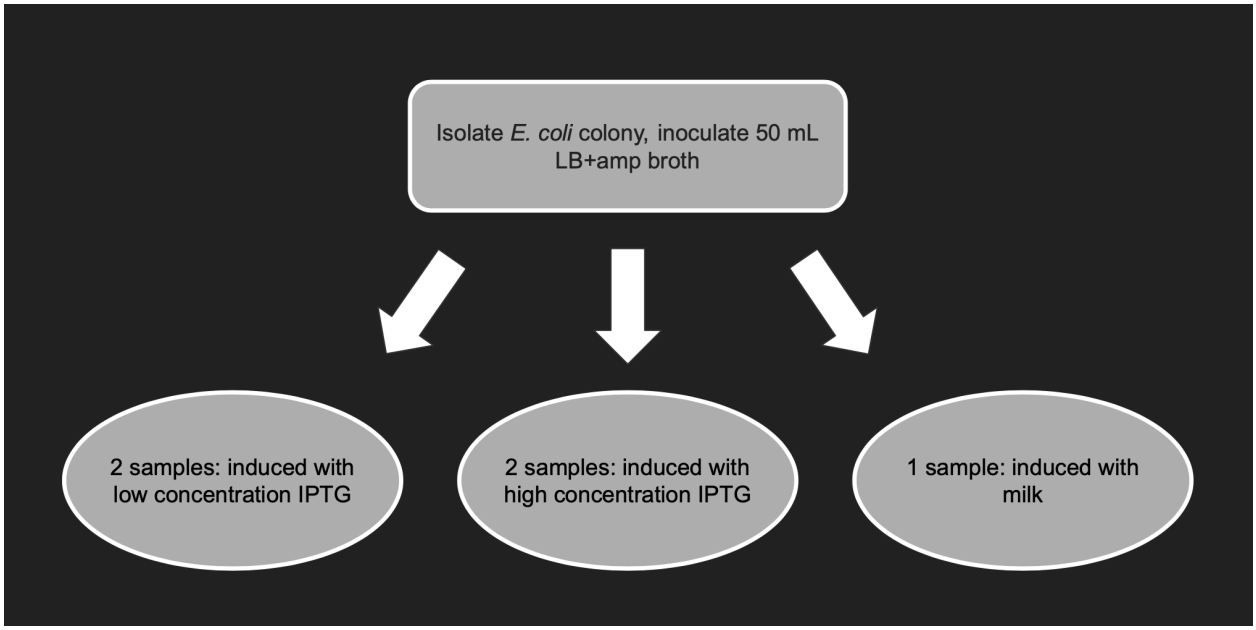

Fig. 2: Initial design of induction steps.

The second proposed change to the standard protocol involved changing the flow rates of the solution through the chromatography column. The flow rate of the solution through the column has the potential to impact the attachment of the Taq to the resin as well as its ability to elute from the column by giving it more or less time to complete either of these activities while passing through the column, therefore affecting the amount of Taq that could be successfully separated from the *E. coli* cell contents. It was determined that three flow rates should be tested: 1) a low flow rate (0.2 mL/min), 2) a medium flow rate (1.0 mL/min), and 3) a high flow rate (1.8 mL/min). If the higher flow rates proved to result in a greater amount of Taq produced than the other flow rates, it might be beneficial to ensure that the standard protocol for Taq purification includes a higher flow rate through the column because it reduces the amount of time necessary to complete the entire protocol.

Final Design

After the five media samples (two induced with low concentration IPTG, two with high concentration IPTG, and one with milk) had been created, a potential error in the experiment was realized. In the study in which a milk-derived product was used to successfully induce gene expression (Khani & Bagheri, 2020), the source of the lactase was held to standards that were not replicable within the time constraints and with the resources available for this experiment. Since milk products are variable in regards to whether they would meet certain standards or not, and since testing the skim milk intended for use in this experiment would take additional time, it was determined that such testing be eliminated from the experiment, therefore the media sample that had been induced with skim milk was discarded.

The remaining four samples were retained and repurposed. There has been evidence of a "cold-shock" treatment aiding in protein production in *E. coli*, as well as evidence of genes in other types of bacteria displaying an increase after exposure to colder temperatures (Qing et al., 2004, Larentis et al., 2014). A prior experiment in the lab in which this experiment was conducted also suggested that an overnight chilling period may impact gene expression, which could be due to post-translational activity after cell growth had been halted (Qing et al., 2004). To test this theory with the samples that had already been prepared at this point, the samples were divided into two groups: a control group that consisted of one low IPTG concentration sample and one high IPTG concentration sample, which was moved into the protein extraction steps immediately after its induction and incubation period; and the experimental group that consisted of one low IPTG concentration sample and one high concentration IPTG sample that underwent a 24-hour period of cold storage at 4°C after the samples had been induced and incubated. The extraction procedure for both groups then proceeded as outlined in the standard

protocol provided in the Methods section. A flowchart of the final design of this portion of the experiment can be seen in Figure 3. To test for the presence of the Taq protein in each of the samples after the assigned treatment of each group had been executed, the samples were loaded into an SDS-PAGE gel.

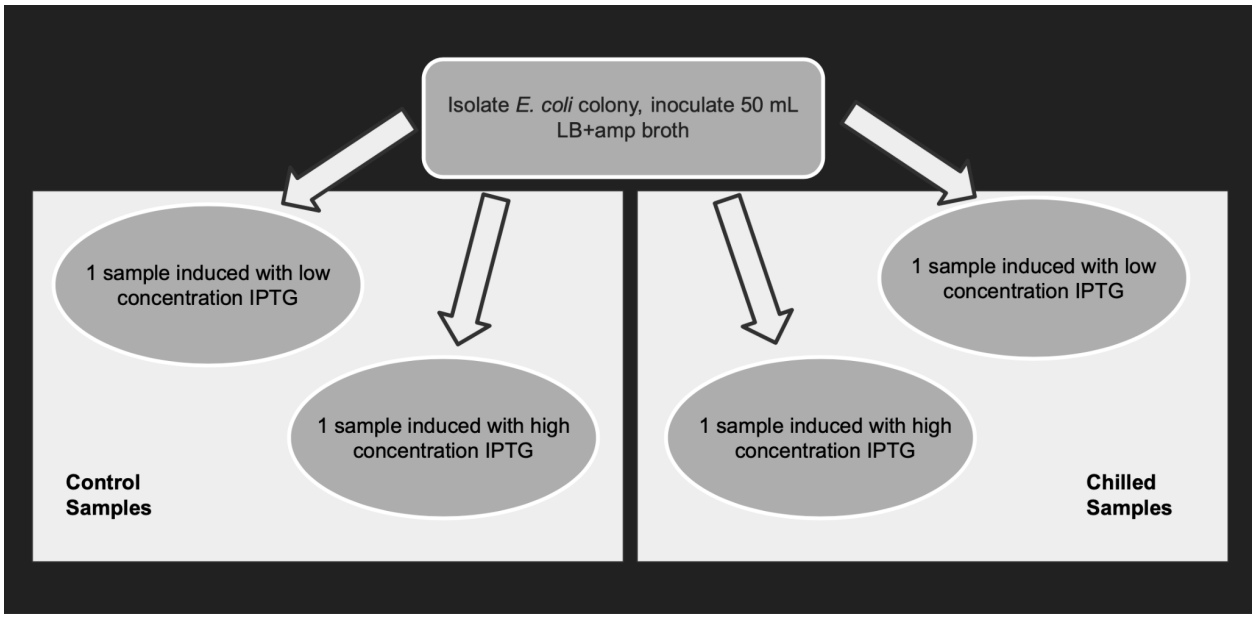

Fig. 3: Final design of induction steps

The second portion of this experiment in which the flow rates were tested for their impact on Taq production remained unchanged from the initial experimental design. Three flow rates would be tested: 1) a low flow rate (0.2 mL/min), 2) a medium flow rate (1.0 mL/min), and 3) a high flow rate (1.8 mL/min). The impact of each of these flow rates were examined using gel electrophoresis of the collected fractions from each of the flow rates.

Testing for Activity

The fractions from the flow rate that displayed the best Taq concentration were then treated and divided into four samples: 1) an undiluted Taq sample, 2) a Taq sample diluted with water at a 1:10 ratio, 3) a Taq sample diluted with water at a 1:100 ratio, and 4) a pure water sample to be used as a control. The samples were used in a PCR and then loaded into a gel to observe the Taq activity within each of the samples.

Results

Chilling Period

After the experimental group of media samples had undergone the 24-hour chilling period, those samples as well as the samples of the control group were loaded into an SDS-PAGE gel and the results can be seen in Figure 4. Comparing the sample from the control group that had been induced with high concentration IPTG to the chilled experimental sample that had been induced with high concentration IPTG, there didn't appear to be a significant difference in the sizes of the bands. The same can be said when comparing the control sample induced with low concentration IPTG and the chilled sample induced with low concentration IPTG. Since there was no significant difference between the bands, and because the chilling period therefore adds unnecessary time to the purification procedure, it was determined that the chilling period was ineffective as it did not seem to sufficiently improve the amount of Taq production. Both of the chilled samples were then discarded.

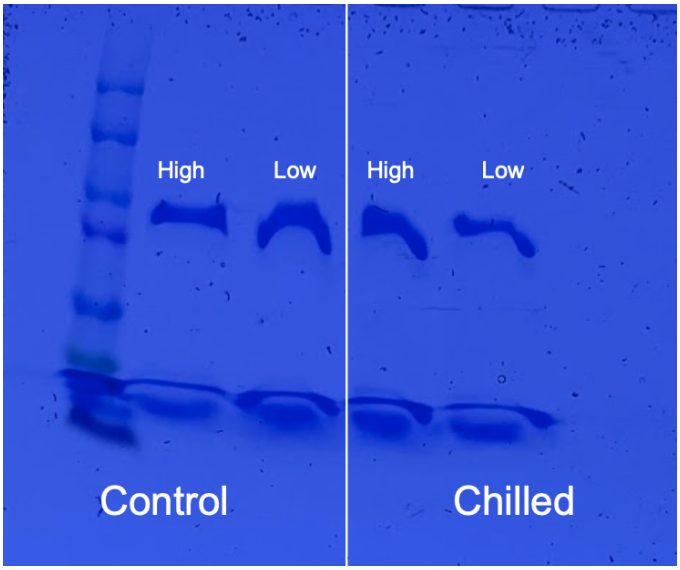

Fig. 4: Results of gel comparing control samples and chilled samples.

To move forward with the experiment, a single sample needed to be chosen. Another gel was run with only the low IPTG concentration and high IPTG concentration control samples, the results of which can be seen in Figure 5. The band corresponding to the high IPTG concentration appears to be thicker than the band for the low IPTG concentration, so the high IPTG concentration sample was selected for continuation of the experiment. The low IPTG concentration sample was discarded. A concentration of IPTG that is double the standard amount is not recommended as a change to the standard purification protocol; doing so would increase the cost of materials as more IPTG would need to be used, and this experiment does not have a control sample done with the standard concentration to determine if the yield of high concentration of IPTG is significantly higher than the yield of the standard concentration of IPTG. The samples used in this experiment were prepared as a part of the initial design for a

comparison with the effectiveness of milk as an inducing agent, and the samples were simply repurposed when the experimental design was altered to its final design.

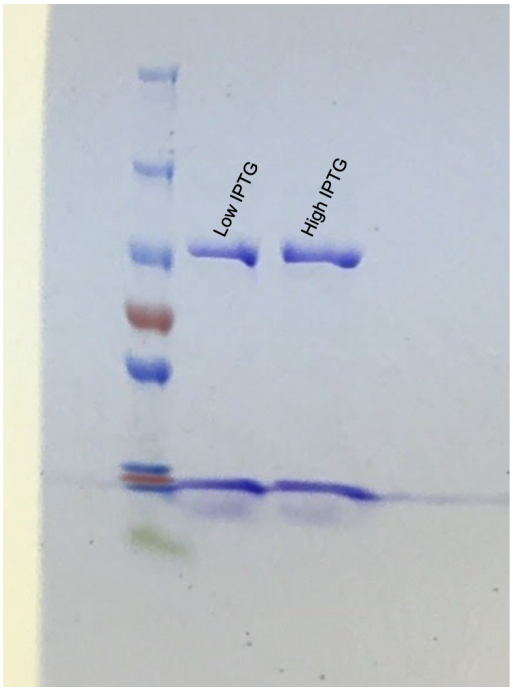

Fig. 5: Results of gel comparing high and low concentrations of control samples.

Flow Rates

Once the fractions had been collected from each of the three flow rates, they were loaded into a gel, the results of which can be seen in Figure 6. In the gel from the low flow rate, it can be seen that Taq is present in multiple fractions. The same trend can be observed in the gel from the medium flow rate, where the Taq is also dispersed throughout several fractions. In the gel from the high flow rate, however, there really only appears to be a single dark band (the darkest band on the far left of the image is a commercial Taq sample that was used for comparison). Even

though the band for the sample collected from the high flow rate is not as dark as the band from the commercial Taq, the high flow rate still concentrated more of the Taq in a single band than either of the other flow rates did. This is ideal because all of the Taq can be collected in a single, concentrated fraction. Therefore, the high flow rate is suggested as a fixture in the standard protocol to improve its time efficiency as well as the amount of concentrated Taq produced, and the high flow rate fractions were retained and prepared for a PCR to test for its activity.

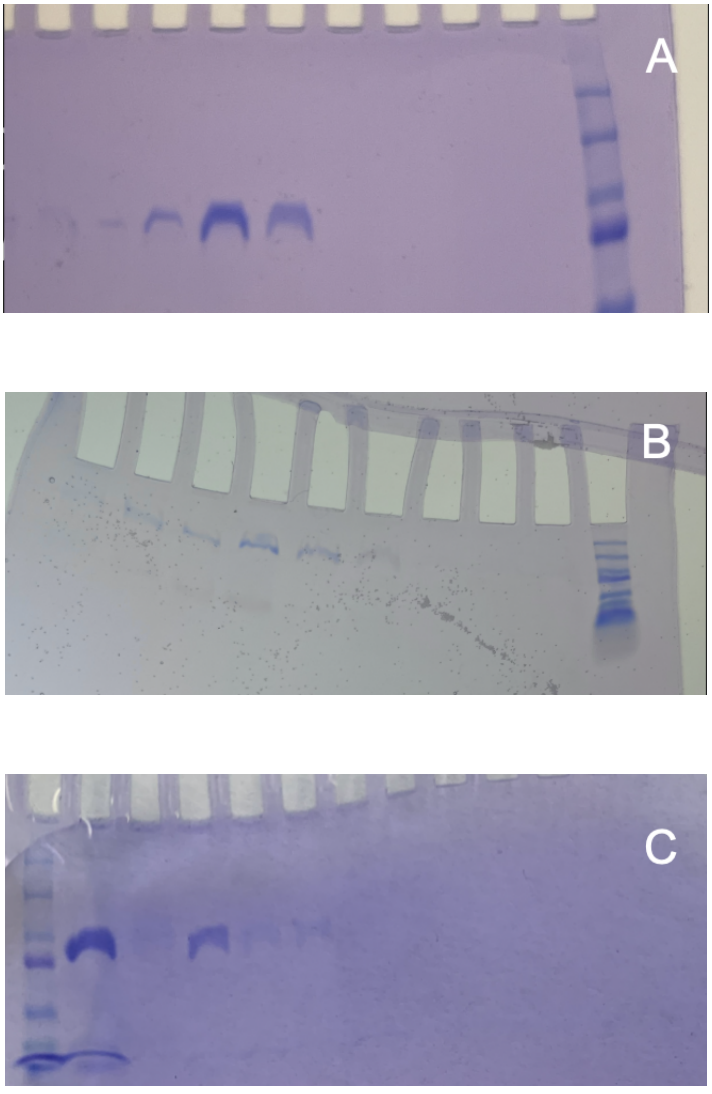

Fig. 6: a) Low flow rate, b) Medium flow rate, and c) High flow rate.

Testing for Activity

After the fractions with the highest Taq concentration were dialyzed against the storage buffer and prepared for the PCR reaction, four samples were created: 1) undiluted Taq, 2) Taq diluted with water at a 1:10 ratio, 3) Taq diluted with water at a 1:100 ratio, and 4) a pure water control sample. The samples were used in a PCR reaction and then loaded into a gel. The gel was placed under UV light, an image of which can be seen in Figure 7. No band is visible for the pure water sample, as expected, but there is also no visible band for the Taq sample that had been diluted at a 1:100 ratio. For the undiluted Taq sample, there is a clearly visible band, and for the Taq sample diluted at a 1:10 ratio the band is visible and bright. This indicates that the Taq produced in this experiment can undergo dilution at a 1:10 ratio and still be concentrated enough to be functional.

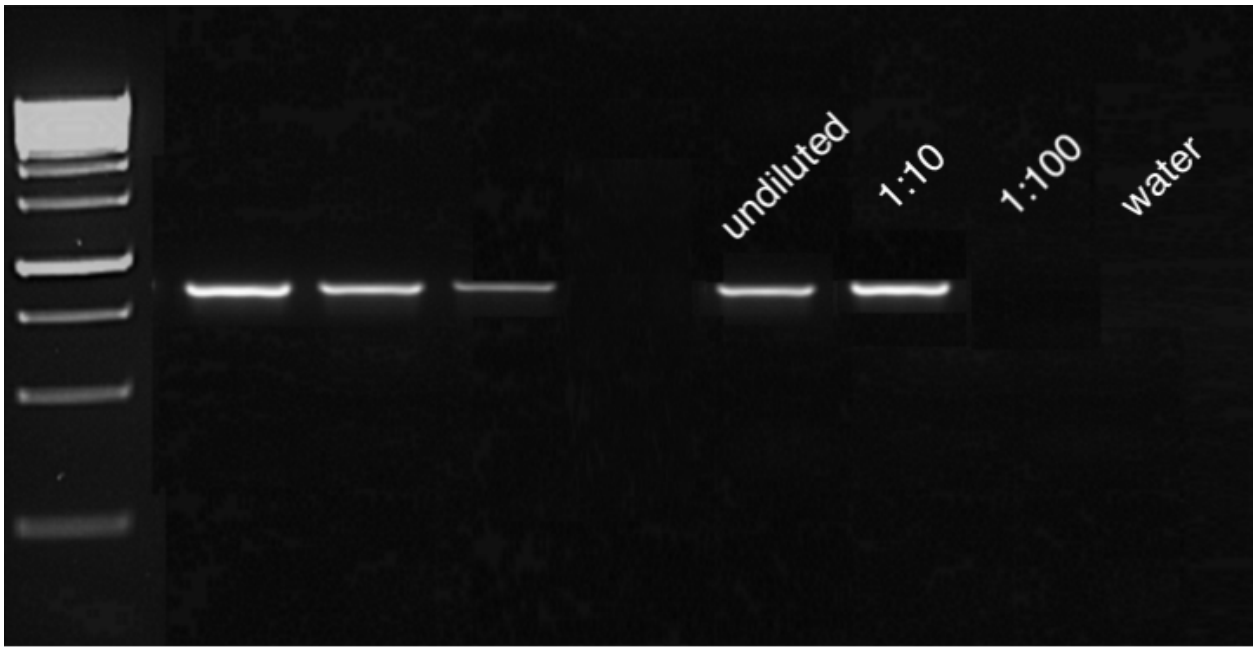

Fig. 7: Results of gel to test for Taq activity under UV light.

Cost Breakdown

The costs for each of the materials required for one run-through of the optimized protocol are listed in Table 1. The total cost of one run-through of the protocol completed in the lab was found to be $486.11.

Material	**Quantity**	**Cost**
Tris-Cl	12.44 g	$4.20
EDTA	58.6 g	$34.57
KCL	2.26 g	$0.11
DTT	4.62E-04 g	$0.01
Dextrose	0.45 g	$0.16
HEPES	1.05 g	$1.05
Lysozyme	0.01 g	$0.32
NaCl	7.45 g	$4.40
PSMF	0.12 g	$2.10
Tween 20	1.4 mL	$0.48
NP-40	1.4 mL	$4.17
Triton X-100	20 mL	$21.00
Glycerol	1000 mL	$158.20
IPTG	0.0335 g	$2.40
DEAE Resin	24 mL	$11.04

Media growth kit	1	$50
Purification column	1	$6.90
PAGE acrylamide gel	1	$150
Dialysis Slide-A-Lyzer Cassette	1	$25
PCR analysis to check activity (includes gel analysis of PCR products)	1	$10
	TOTAL:	**$486.11**

Table 1: Cost per single run-through of optimized protocol

Taking into account the amount of Taq produced in this experiment, and the amount of Taq necessary for one PCR reaction, it was determined that the total cost per PCR reaction of Taq produced using the procedure set forth in this experiment is less than 2 cents, as can be seen in Table 2. Biology labs at Oakland University purchase enough Taq for 250 PCR reactions for $401.95 per batch, which breaks down to be about $1.13 per PCR reaction, as seen in Table 3.

Total Cost per One Batch (25,000 PCR Reactions)	$486.11
Cost per PCR Reaction	**$0.019**

Table 2: Cost of Taq for proposed protocol

Total Cost per One Batch (250 PCR Reactions)	$401.95
Cost per PCR Reaction	**$1.13**

Table 3: Cost of Taq for OU biology lab

Conclusion

Since the use of milk as an inducing agent was not utilized in this experiment as planned due to the necessity of regulating the milk products, and since the chilling period did not seem to have an impact on the amount of Taq produced, this experiment does not recommend either of these steps to be included in the Taq purification protocol. However, this experiment was successful in choosing an ideal flow rate, as the high flow rate of 1.8 mL/min produced the most concentrated fractions of Taq. Therefore, the 1.8 mL/min flow rate is recommended for use in the standard purification protocol to improve its efficiency.

The optimized purification protocol presented in this experiment is also extremely cost-effective. As seen in the cost breakdown, the cost to produce Taq using the protocol in this experiment (about 2 cents per PCR reaction) is much cheaper than purchasing Taq commercially ($1.13 per PCR reaction). Such a cost difference promotes producing Taq for biology labs in-house at Oakland University rather than purchasing it commercially. Also as mentioned previously, the process used for Taq purification can be considered a model for insulin purification, therefore the findings of this experiment can have far-reaching applications in the medical industry in improving the efficiency and cost-effectiveness of protein-based drug production that would eventually lower the cost of drugs for consumers.

While using the flow rate suggested by the results of this experiment is a step towards improvement of the standard purification protocol, there are still several other parameters that

can be changed and measures that can be taken in the future to build upon this research and further improve the efficiency and cost-effectiveness of the purification process.

References

Choudhary, P., De Portu, S., Arrieta, A., Castañeda, J., & Campbell, F. M. (2019). Use of sensor‐integrated pump therapy to reduce hypoglycaemia in people with Type 1 diabetes: a real‐world study in the UK. *Diabetic Medicine*, *36*(9), 1100-1108.

Din, R. U., Khan, M. I., Jan, A., Khan, S. A., & Ali, I. (2020). A novel approach for high-level expression and purification of GST-fused highly thermostable Taq DNA polymerase in Escherichia coli. *Archives of Microbiology*, *202*(6), 1449-1458.

Dvorak, P., Chrast, L., Nikel, P. I., Fedr, R., Soucek, K., Sedlackova, M., ... & Damborsky, J. (2015). Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21 (DE3) carrying a synthetic metabolic pathway. *Microbial cell factories*, *14*(1), 1-15.

Gelfand, D. H. (1989). Taq DNA polymerase. In *PCR technology* (pp. 17-22). Palgrave Macmillan, London.

Khani, M. H., & Bagheri, M. (2020). Skimmed milk as an alternative for IPTG in induction of recombinant protein expression. *Protein Expression and Purification*, *170*, 105593.

Larentis, A. L., Nicolau, J. F. M. Q., dos Santos Esteves, G., Vareschini, D. T., de Almeida, F. V. R., dos Reis, M. G., ... & Medeiros, M. A. (2014). Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor. *BMC research notes*, *7*(1), 1-13.

Liu, S., Li, Z., Yu, B., Wang, S., Shen, Y., & Cong, H. (2020). Recent advances on protein separation and purification methods. *Advances in Colloid and Interface Science*, *284*, 102254.

Malone, J. I., & Hansen, B. C. (2019). Does obesity cause type 2 diabetes mellitus (T2DM)? Or is it the opposite?. *Pediatric diabetes*, *20*(1), 5-9.

Menzella, H. G., Ceccarelli, E. A., & Gramajo, H. C. (2003). Novel Escherichia coli strain allows efficient recombinant protein production using lactose as inducer. *Biotechnology and bioengineering*, *82*(7), 809-817.

Norris, J. M., Johnson, R. K., & Stene, L. C. (2020). Type 1 diabetes—early life origins and changing epidemiology. *The lancet Diabetes & endocrinology*, *8*(3), 226-238.

Qing, G., Ma, L. C., Khorchid, A., Swapna, G. V. T., Mal, T. K., Takayama, M. M., ... & Inouye, M. (2004). Cold-shock induced high-yield protein production in Escherichia coli. *Nature biotechnology*, *22*(7), 877-882.

Schmidt, M., Hafner, M., & Frech, C. (2014). Modeling of salt and pH gradient elution in ion‐exchange chromatography. *Journal of separation science*, *37*(1-2), 5-13.

Zieliński, M., Romanik-Chruścielewska, A., Mikiewicz, D., Łukasiewicz, N., Sokołowska, I., Antosik, J., ... & Płucienniczak, A. (2019). Expression and purification of recombinant human insulin from E. coli 20 strain. *Protein Expression and Purification*, *157*, 63-69.