Optimizing the Production of Taq Polymerase from *E. coli*

Submitted by

Tomas Sadik

Bioengineering

To

The Honors College

Oakland University

In partial fulfillment of the

requirement to graduate from

The Honors College

Mentor: Shailesh Lal, Ph.D., Professor and Chair of Bioengineering

Department of Bioengineering

Oakland University

April 1, 2022

Table of Contents

Abstract

Protein purification is an important biotechnological process used to induce the expression of a desired protein from host cells, isolate the protein using one or more established methods, and collect it in its purified form. Applications of such exist in pharmaceutical products, clinical research, and other medicinal areas. This process describes the execution and optimization of the protein purification of Taq polymerase, a DNA polymerase enzyme natively expressed by the bacterial species *Thermus aquaticus*. Introduction of the gene into the *Escherichia coli* host cells was completed using recombinant DNA technology, and expression of the desired Taq polymerase protein was induced via isopropyl-β-D-1-thiogalactopyranoside (IPTG). Its isolation and collection were facilitated via cell lysing and ion-exchange column chromatography. The resulting Taq polymerase was successfully utilized as DNA polymerase in the polymerase chain reaction. Optimization of the process was achieved by maximizing protein yield and consolidating fraction collection into the smallest possible number of fractions. This occurred in the form of two parameter modifications: IPTG inducing concentration and chromatography flow rate. Five concentrations of IPTG were tested based on the standard protocol value of 0.335 mM, those being 0.08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM. Three flow rates were tested using peristaltic pump settings, those being 0.5 mL/min, 1 mL/min, and 2.0 mL/min. It was found that an IPTG concentration of 0.1675 mM yielded the highest protein expression, and the flow rate of 2.0 mL/min resulted in the highest Taq polymerase expression concentrated in a single fraction. These results hold value for resource and time savings of large-scale protein purification processes, and can be replicated and applied to the purification of therapeutic proteins, clinical protein study, and other fields.

Introduction

Protein purification is a crucial and innovative process that involves the induced expression of target protein(s) from host cells, the isolation of such proteins from a mixture of cell components and membranes, and their collection and purification by use of different techniques based on scientific and engineering principles. This process holds significance to multiple industries and areas of research; the collection of purified proteins allows for detailed analysis on their composition, structure and domains, function in biological processes, and interactions with other substances. This data can be used in the development of pharmaceutical products based on naturally occurring proteins [1]. For instance, the pharmaceutical product Avonex is based on the human interferon beta-1a protein, and is produced using recombinant DNA techniques that obtain the human interferon gene and insert into Chinese hamster ovary cells. This is done in order to induce its protein expression and purify the protein for use as a prescription drug for multiple sclerosis treatment [2]. Protein purification also has applications in the developing tissue engineering industry, genetic disorder research, and other areas.

This research experiment examines the protein purification of Taq polymerase (Taq), a DNA polymerase enzyme produced in nature by the bacterial species *Thermus aquaticus*. Its importance has been well-documented and researched, especially for its use in the application of the Polymerase Chain Reaction (PCR) [3]. PCR is a DNA amplification process based on principles from natural DNA replication, and is composed of three overarching, cyclical steps: denaturing, annealing, and extension/synthesis. Denaturing begins with a target double-stranded DNA (dsDNA) molecule that is heated to split the hydrogen bonds within the double helix, resulting in two separate strands of DNA. Upon cooling, annealing occurs, in which primers are

added to the single-stranded DNA (ssDNA) to signify the starting point of synthesis of the second strand. The temperature is then raised again, and the aforementioned DNA polymerase enzyme is introduced, along with dNTPs, i.e., the nucleotides added to the growing strand. DNA polymerase is crucial in this step, since it directs the addition of these dNTPs in the correct sequence to create dsDNA. Once extension is complete, two molecules of dsDNA have been created from the original dsDNA molecule [4]. This process is repeated thousands to millions of times to create substantial amounts of DNA for diagnostic purposes, such as COVID-19 testing.

The importance of DNA polymerase in PCR cannot be overstated, which indicates the value of studying and optimizing the protein purification process of Taq polymerase. The gene which codes for the production of Taq is not produced in wild-type *E. coli*, but can be introduced into its genome via recombinant plasmid technology, in which circular DNA containing the target gene is inserted into an *E. coli* cell, so that the cell culture is grown with this gene intact [3]. For the purposes of this experiment, the modified *E. coli* culture with the Taq gene will already have been created and supplied by Dr. Shailesh Lal at Oakland University.

The expression of target genes in *E. coli* occurs as the two-step process of transcription of the target DNA into single-stranded RNA and translation of this RNA into corresponding amino acids. Transcription of genes in *E. coli* is inhibited by the lac operon, which features the protein known as the *lac* repressor, LacI. This protein binds to genes that encode for proteins involved in lactose metabolism. When lactose is added to the cell and converted to allolactose, it binds to LacI and *stops* it from repressing gene expression, so the lactose metabolism enzymes are produced. Isopropyl β-D-1-thiogalactopyranoside (IPTG) acts as an inducer in the same way as allolactose – it is able to bind to LacI, allowing gene expression to occur. This is why IPTG is

used in the induction of *E. coli*, as it will allow for transcription and translation of the inserted Taq polymerase gene to occur [5].

The process of purification will continue with ion-exchange column chromatography, in which charged resin and a salt gradient are used to isolate Taq polymerase from the other cell contents, and then collecting it in its pure form. In this experiment, the output from the ion exchange columns will be collected via fraction collection via peristaltic pump, the flow rate of which can be determined and set by the user [6]. Among the resulting fractions are some which contain Taq polymerase; these will be dialyzed and then used in the aforementioned PCR reaction. This will test the function of the Taq as DNA polymerase and ensure that it remains undamaged through the purification process.

Within the process of protein purification are areas of parameter altering and optimization, in terms of yield, protein fraction purity, and enzymatic activity of the resulting proteins. Two specific areas were analyzed in this experiment, based on their ability to be altered and modeled. The first of these parameters was the concentration of IPTG to be used as an inducting substance. As discussed previously, IPTG acts as an inducer similar to allolactose, and its presence affects the process of transcription and translation of the Taq polymerase gene [7]. Its concentration, then, will have an effect on the yield of Taq polymerase from the *E. coli* culture, which will be seen via gel electrophoresis techniques. Five concentrations of IPTG will be created and tested on *E. coli* cultures; that which indicates the highest protein expression will be used for the further ion-exchange chromatography and isolation of Taq polymerase. This leads to the second opportunity for parameter optimization, that being column chromatography flow rate. Previous research on flow rate alteration has shown that slower flow rates have resulted in

higher protein yield, but there remains to be discovered a singular optimal flow rate [8]. By researching, altering, and analyzing these two parameter values, the overall product quality of the Taq polymerase will have improved, which in turn will lead to improvements in PCR. This experiment may also indicate potential improvements for other protein purification processes, such as those of other polymerases and even other pharmaceutical proteins. This study contributes to the ongoing optimization and innovation of protein purification and biotechnology, leading to positive impacts and efficiency in areas such as pharmacy, medicine, genomic research, and others.

Methods and Materials

Overview of Protocol Followed

The process of the purification of Taq polymerase can be roughly separated into five main stages. The first is genetic modification of *E. coli* to contain the desired Taq polymerase gene, which was outside the scope of this project and thus was prepared by the Bioengineering department. The next stage was growing the *E. coli* culture on a suitable medium and inducing expression with IPTG, which was the first parameter optimization action. This was followed by the third stage of buffer addition and centrifugal action to deteriorate the cells and prepare the solution for column chromatography. This leads to column chromatography being the fourth phase, in which the Taq polymerase is isolated and collected into fractions. Finally, the fifth phase was the use of PCR as a diagnostic tool to ensure that the collected Taq polymerase was functional.

Preparation of Media and Stock Solutions

Several components were needed to initiate and sustain the growth of the *E. coli* culture, and subsequently lyse cells to isolate the target Taq polymerase protein. Luria-Bertani (LB) broth was first made to act as the growth medium for the culture of *E. coli*, since it has shown an ability to induce rapid growth in bacterial species [11]. 500 mL of broth was created with 5.0 g of tryptone, 2.5 g of yeast extract, and 5.0 g of NaCl. This solution was then autoclaved for sterilization, before the antibiotic 100 μg/mL ampicillin (amp) was added to destroy *E. coli* cells that lacked the ampicillin resistance plasmid gene [12].

Buffer A was added to the solution to deteriorate the outermost protective layer of the *E. coli* cell, i.e., the cell wall, without affecting the cytoplasmic membrane [13]. 50 mL was produced from 2.5 mL of 50 mM Tris-Cl stock, 0.45 g of 50 mM Dextrose, and 0.2 mL of 1 mM EDTA, diluted in distilled water. The enzyme lysozyme was added to catalyze the breakdown of the cell wall as well.

Buffer B acted as the lysis immunoprecipitation buffer, responsible for isolating proteins from the lysed cells and making them soluble, including the target Taq polymerase protein [14]. 20 mL was produced from 20 µL of 10 mM Tris-Cl, 1.0 mL of 50 mM KCl stock, 40 µL of 1 mM EDTA, 100 µL of 1 mM PMSF, 400 µL of 0.5% Tween 20, and 400 µL of 0.5% NP-40, again diluted in distilled water.

Buffer C was necessary for the proper function of ion-exchange chromatography, for its role of stabilizing the proteins passing through the columns [15]. It was made from 1.2 mL of 1 mM EDTA, 0.75 mL of 0.5 mM PMSF, 1.5 mL of 0.5% Tween 20, 1.5 mL of 0.5% NP-40, and 231 µL of 5 mM DTT, and diluted in 20 mM HEPES until 300 mL of solution was produced.

Storage buffer was needed once the fractions of Taq polymerase were collected, to purify the fractions of peak activity via dialysis. 2000 mL were produced with 100 mL of 1 M Tris-Cl of 7.9 pH, 7.45 g of 100 mM NaCl, 400 µL of 0.5 M EDTA stock, 20.0 mL of 1 M 1% Triton X-100, and 0.308 mL of 1 M DTT stock, diluted in 50% glycerol to that final volume.

Cell Culture Growth and T-Streaking

The process of the purification of Taq polymerase from *E. coli* began with obtaining bacterial stock of *E. coli* that had the Taq polymerase gene introduced into its genome via a

recombinant plasmid. This stock was sourced from Dr. Shailesh Lal's laboratory at Oakland University. The bacterial *E. coli* stock was grown into a culture by t-streaking it onto a plate containing LB Agar and 100 µg/mL ampicillin (amp) antibiotic; t-streaking refers to the pattern of plate streaking done by imaginarily splitting the plate into three sections and streaking between them [16]. The colonies of *E. coli* were grown to a suitable optical density of 0.5 at 600 nm, and then placed in a temperature of 37°C and continued to grow overnight.

At this point, preparation of the culture for induction by IPTG was required. One colony of *E. coli* was chosen and inoculated into 3 mL of LB broth with 4 µL amp, and then grown on a shaker overnight to allow for adequate mixing. Since the IPTG concentration parameter would be tested with five different values, five flasks of 50 mL LB media solution were inoculated with 50 µL of the shaken *E. coli* solution each. An additional 50 µL of amp was added to each flask. After four hours, the flasks were suitable for IPTG induction.

IPTG Induction

As previously discussed, isopropyl β-d-1-thiogalactopyranoside (IPTG) is a reagent used to induce the gene expression, i.e., transcription and translation, of the Taq polymerase gene introduced via recombinant plasmid [5]. Five concentration values were produced and tested to induce the protein expression in the *E. coli* cultures. Since this experiment was based on previous research and protein purification trials, the five values were also based on the previous experimental concentration of 0.335 mM, taking fractions of ¼, ½, 1, 1½, and 2x this concentration, giving values of 0.08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM.

These concentrations were made by diluting 0.5 M IPTG stock solution. These solutions were added to the 50 mL solutions of *E. coli*, and grown on a shaker at 37°C for 24 hours.

At this point, 47 mL of each of the five solutions were transferred to centrifuge tubes and spun for 10 minutes at 4°C. The pellets from the centrifuge tubes were suspended in 20 mL of Buffer A, homogenized, and centrifuged again for 10 minutes at 4°C. 10 mL of Buffer A + lysozyme was then added to the centrifuge tubes, incubated for 15 minutes, and then mixed with 10 mL of Buffer B. At this point, it was discovered that there was an error in this method: the volumes of Buffers A and B used were those of the original protocol that calls for 10 times the *E. coli* solution volumes. This meant that the *E. coli* solutions were diluted in the buffers 10 times as much as they should have been. The result of this would likely end up being that there would be little to no visible protein expression because of the extreme dilution. However, for the sake of practice and completion, the protocol was still completed for these diluted solutions. The solutions were incubated for one hour at 75°C and centrifuged for 15 min at 4°C. The supernatant from the centrifugation was saved and frozen until the gel was prepared. At this point, 45 µL of each sample was boiled for 30 minutes, then frozen for 3 minutes.

Gel Electrophoresis of IPTG Induction Trial

Gel electrophoresis was performed using a 7.5% 15-well Bio-Rad Mini-PROTEAN TGX Stain-Free Precast gel. The first well was loaded with a Pre-stained Protein Ladder that would act as a guide to indicate the weight of proteins expressed in the samples [17]. For the remaining wells, 40 µL of each sample was mixed with 5 µL of Bio-Rad Precision Plus Protein Kaleidoscope Protein Ladder, and thus 45 µL of each solution was loaded into the apparatus

wells. There was an error here as well: the wells were discovered to have been designed to only hold 15 μL each, so there was spillover into adjacent wells. Regardless, the gel was run at 100 volts until completion. It was then placed into a Coomassie blue stain for an hour, and destained overnight. The staining should allow for proteins to be visualized on the gel [18]. However, the aforementioned issue proved to be true, in that the dilution of the samples in the buffers resulted in no visible protein expression. The IPTG parameter trials were run again, as discussed below.

Retrial and Corrections to IPTG Induction Parameter Trials

The same process was followed up to and including the induction of the samples with the five concentrations of IPTG, but at this point, 38 mL of each of the five solutions were initially centrifuged instead of 47 mL. The samples were suspended in 1 mL of Buffer A for homogenization and centrifugation, and then 1 mL of Buffer A + lysozyme was added to the tubes before incubation. Buffer B was then added at a volume of 1 mL to each tube as well. Gel apparatus preparation followed the same procedure, with the difference in these trials being that 15 μL of each sample was mixed with 5 μL of Pre-stained Protein Ladder, and 15 μL of each of these solutions were added to the wells. The same electrophoresis and staining procedures were taken, and the results of the induction and protein expression were now visible. The IPTG concentration resulting in the highest protein expression – that being the ½ value of 0.1675 mM – was chosen to repeat the process discussed insofar, in order to prepare for column chromatography.

Ion-Exchange Column Chromatography

The principle behind ion-exchange column chromatography is the use of charged particles to isolate different particles, that being proteins in this case, from a complex mixture, such as that of lysed cells [19]. Positively charged resin has an affinity for Taq polymerase, being that it is negatively charged, which causes the Taq to stick to the resin in the columns. By use of resin, wash buffer, and producing a salt gradient in the column, the Taq was eventually washed off the resin and collected in its pure form. This procedure is outlined in detail below.

Four columns were created to test the three different flow rates in question, plus an additional column in case of damage or clogging. Creating each column required 30 mL of DE-52 resin and 70 mL of Buffer C, so 120 mL of DE-52 resin and 280 mL of Buffer C were obtained and mixed to create a solution of 400 mL volume. Through each column, 100 mL of solution was run very slowly; by doing this, the resin remained packed in the column, while the Buffer C waste ran through the valve opening. As an equilibration technique, another 135 mL of Buffer C was run through each column once they were packed. The creation of these columns was done well before the previous step of IPTG induction, so there was not yet any *E. coli* solution to run through the columns; thus, they were stored in the refrigerator until needed.

As previously mentioned, a salt gradient was necessary to allow for the flow and collection of Taq polymerase. 25 mM and 700 mM KCl solutions were used to establish this; 40 mL of each solution was added to opposing sides of the gradient maker of the column chromatography apparatus. When the cell contents are flowed through the column, negative Taq polymerase adheres to the positively charged resin. As the 25 MM KCl solution flows through, it acts as a wash buffer – it washes the remaining cell contents through the column, but leaves the Taq adhered to the resin since the resin has a higher affinity for Taq than it does for 25 mM KCl

[20]. However, once the 700 mM KCl flows through the column, it replaces the Taq adhered to the resin, since there is a higher affinity for the KCl than the Taq. The Taq at this point is able to flow through the rest of the chromatography apparatus and into collection tubes.

The parameter being analyzed, that being pump flow rate, was tested using the Pharmacia Biotech Peristaltic Pump. The flow rates chosen to be tested were 0.5 mL/min, 1.0 mL/min, and 1.5 mL/min. The pump offers the ability to adjust flow rate of solution through the pump into the fraction collector [6]; however, the setting knob was marked with arbitrary values not corresponding to volumetric flow rates. Thus, trial and error was used to determine the relationship between the pump values and flow rate using water running through the pump before the chromatography was started.

Once the gradient was prepared and each flow rate was determined, the 0.5 mL/min flow rate was tested using 15 mL of Buffer C first being run through the column to re-equilibrate the column. 8 mL of the protein sample sourced from the optimal IPTG concentration was added to the column. The principle of affinity between the positive resin and the negative Taq was now apparent: while other proteins were washed through the column into waste collectors, the Taq remained stuck to the resin in the column. 50 mL of the aforementioned 25 mM KCl was run through the column to ensure that any other lingering proteins or debris were removed. The valve at the bottom of the column was closed and the column was filled again with 25 mM KCl. The gradient was then attached to the column and the valves on the column were opened, allowing the 700 mM KCl to flow through the column. The KCl replaces the Taq adhered to the resin, causing the Taq to wash off and flow through the pump at the determined flow rate, into the Biotech Redifrac Fraction Collector and then into nine separate fractions, i.e., miniature test

tubes. This process was run three times, once for each of the flow rates being tested, and thus nine fractions for each flow rate were collected. Analyzing these results involves determining which fractions of the nine contain Taq, and determining the optimal flow rate. This will be completed by further gel electrophoresis and finally, the PCR test reaction, which will be further discussed below.

Gel Electrophoresis of Column Chromatography Fractions

As in the gel electrophoresis of the IPTG induction results, the same 7.5% 15-well Bio-Rad Mini-PROTEAN TGX Stain-Free Precast gel was used to create the gels. Note that the results of the IPTG induction electrophoresis showed the results of *all* protein expressions of the *E. coli* prior to purification, whereas this post-chromatography electrophoresis will show the results of *only* Taq polymerase protein expression, because only purified Taq has been collected into these fractions, as intended. For each flow rate, 20 uL of each of the nine fractions was collected and mixed with 5 uL of LDS buffer, before being boiled for 30 min. The first well of the electrophoresis comb was loaded with 5 uL Bio-Rad Precision Plus Protein Kaleidoscope Protein Ladder. The rest of the wells were loaded with 12 uL of the fraction-buffer solution for each flow rate, and then electrophoresis was run for one hour at 100 volts again. The same Coomassie blue stain was applied for an hour, and then de-stain was used overnight. The optimal flow rate was shown to be 2 mL/min; detailed results of the post-chromatography electrophoresis are discussed in the Results section.

Polymerase Chain Reaction

Of the fractions collected from the optimal flow rate, fraction 4, i.e. Lane 5 on its corresponding gel electrophoresis, was used for performing PCR, as it showed the highest expression of Taq polymerase. The entire volume of the fraction was injected into a dialysis cassette membrane, and air was removed from the volume. The cassette was steeped in the Storage Buffer overnight. This was done in order to remove any KCl that may have been present in the fraction from the chromatography wash process; osmosis is the driving force removing the salt from the solution. From dialysis, a volume of 2.7 mL of protein solution was collected.

At this point, the PCR master mix was made from 25 µL buffer, 25 µL dNTP, 5 µL forward primer, 5 µL backward primer, and 180 µL water. Homogenization of the solution was done by vortexing the mix. The collected Taq was also serially diluted with water to achieve 1:10 and 1:100 concentration solutions, in order to ensure that PCR would demonstrate band intensity based on Taq concentration as expected. 1 µL of Taq was added to 9 µL of water to produce the 1:10 dilution. This was followed by 1 µL of this 1:10 solution added to 9 µL of water again, to produce the 1:100 dilution.

From each of the three solutions, 1 µL was used as the DNA polymerase enzyme in the Bio-Rad iCycler Thermal Cycler PCR Machine. Well 1 of the gel apparatus contained pre-stained protein ladder, wells 2 through 4 contained the results from the 1:1, 1:10, and 1:100 Taq solutions respectively, and well 5 contained water to act as a negative control and ensure contamination did not occur throughout the process. The results of PCR are discussed in the Results section as well.

Results

IPTG Concentration Testing

As previously mentioned, the optimal IPTG concentration was determined using gel electrophoresis, in which the gel band showing the most visible intensity would indicate the highest protein expression.The results of this gel electrophoresis are shown in Figure 1 below. Note that the 94 kDa point on the Lane 1 protein ladder indicates the size of Taq polymerase, and Lane 2, 4, 6, 7, and 8 correspond to the ¼, ½, 1x, 1½, and 2x factors of the 0.335 mM protocol concentration. Lanes 3 and 5 were disregarded because samples spilled over into these wells when the five concentrations were being loaded. Recall that these concentrations are 0.08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM, respectively.

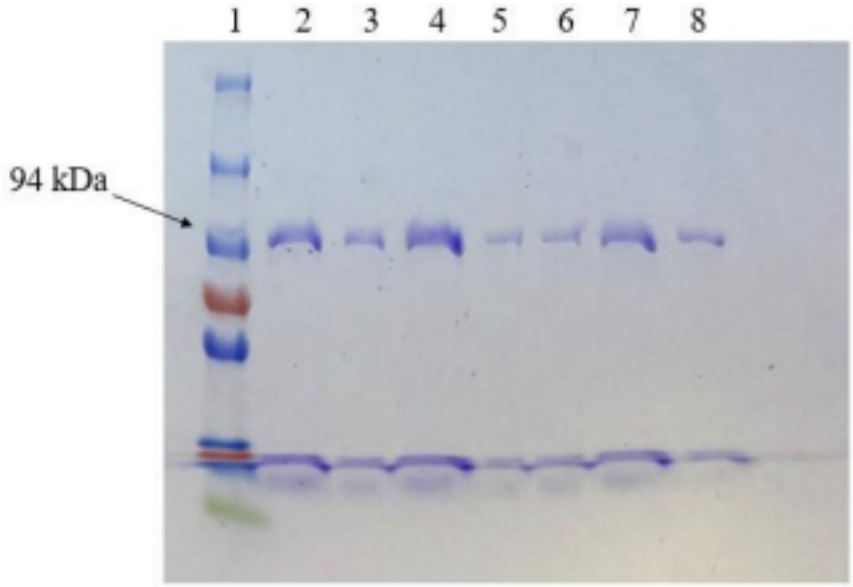

Figure 1: Gel electrophoresis results with 94 kDa protein size indicated i.e. Taq polymerase. Lane 4, the 0.1675 mM concentration, shows the highest intensity

As can be seen in the gel results, Lane 4, the 0.1675 mM IPTG concentration, shows the highest expression of Taq polymerase. This value was thus chosen to be used as the induction

concentration for the sample being used in column chromatography. These results are somewhat unexpected. Generally, one would predict that increasing IPTG concentration would increasingly induce protein expression. This would remain true up to a plateau point, before leading to toxicity to the cells, which would then lower protein expression as IPTG concentration increases [21]. Given this logic, one would expect that the original published value of 0.335 mM IPTG would have induced higher expression than the ½ factor of 0.1675 mM IPTG, and that the 1½ or 2 factor concentrations of 0.5025 mM, and 0.67 mM may have induced even more expression if cell toxicity had not yet been met. However, these results indicate that toxicity to the cells may occur earlier than expected in terms of IPTG concentration – the original 0.335 mM IPTG may have been too high and was causing some cell death, which would make 0.1675 mM IPTG closer to the absolute optimal value to be used.

These results may lead to benefits for large-scale protein production. IPTG, like any resource, has an associated cost per gram, and using ½ the stated protocol concentration would reduce the cost of IPTG by 50% for a particular amount of protein expression. Recall that IPTG acts as allolactose in regards to the lac operon, in that it induces protein expression of the genes within the operon's regulatory control. Its role in protein production is not only relevant to Taq polymerase, but to other proteins used as pharmaceutical products [22]. Should these results be replicated, large-scale processes may be modified to use less IPTG to produce more protein, potentially leading to significant cost benefits for manufacturers and, ideally, consumers and users of these products.

Flow Rate Testing

Recall that once the optimal IPTG concentration of 0.1675 mM was found, the induction and growth process was repeated using this concentration to produce another protein expression sample. This was done in order to have a sample to proceed with in column chromatography. Column chromatography was performed three times, with a column for each flow rate to be tested – 0.5 mL/min, 1.0 mL/min, and 2.0 mL/min. The optimal result for flow rate would be that which produces the highest expression of Taq polymerase within the smallest number of fractions. In achieving this, the resulting fraction(s) would be more suitable for various applications due to its higher concentration. In the context of this experiment, this was the use of Taq polymerase in PCR, but this holds true for other applications such as drug production and clinical research. The optimization of flow rate was evaluated using gel electrophoresis for each flow rate. Note that Lane 1 in each of the figures below represents the pre-stained protein ladder, with the 94 kDa Taq polymerase indicated in each.

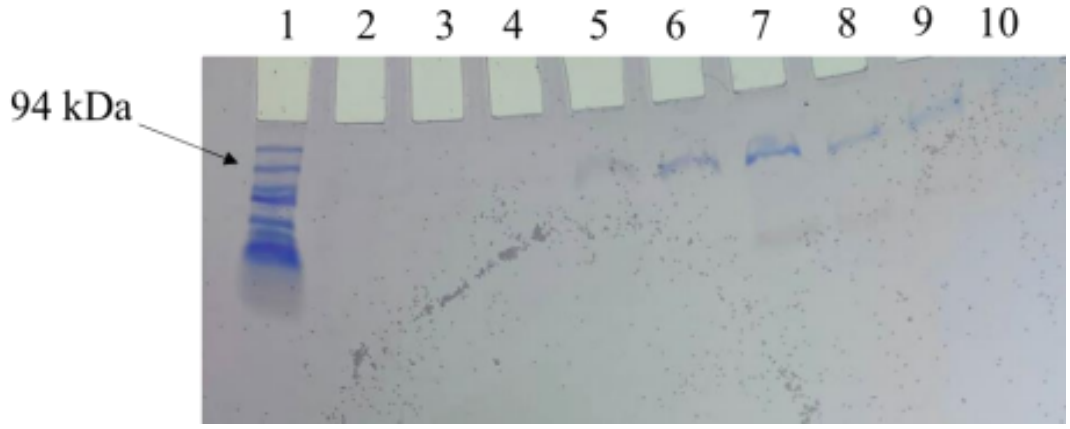

Figure 2: Gel electrophoresis for 0.5 mL/min flow rate Taq polymerase collection. Lanes 5 through 9 show presence of Taq polymerase

Figure 2 above represents the gel result for the fractions collected from the 0.5 mL/min flow rate chromatography. Collected fractions 1 through 9 correspond to Lanes 2 through 10.

Notice that Lanes 5 through 9 show the presence of Taq polymerase, and that Lane 7, i.e. fraction 6, had the highest expression of Taq. However, notice that even Lane 7 has faint expression, indicating a low concentration of Taq, and that expression is seen in five fractions, which is a relatively high number compared to those of the higher flow rate results. This is not ideal; it will be seen that the higher flow rate results have Taq expression in fewer fractions but at a higher concentration, indicated by the intensity of the banding, which is considered optimal.

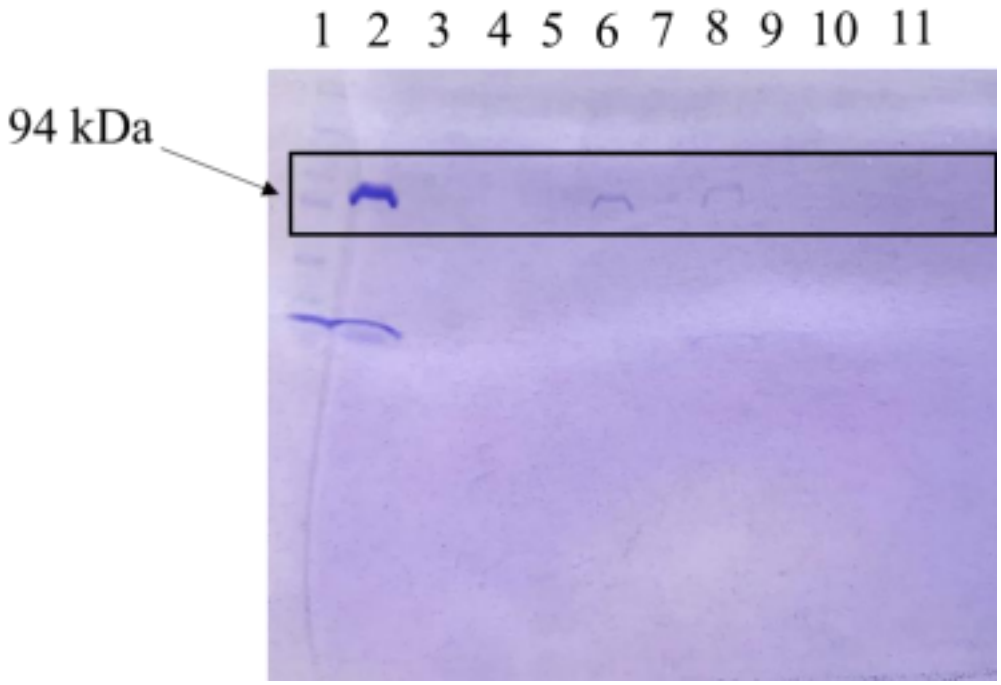

Figure 3: Gel electrophoresis results for 1.0 mL/min flow rate Taq polymerase collection. Lanes 6 through 8 show presence of Taq polymerase

Figure 3 above shows the gel result for the 1.0 mL/min flow rate testing. Lane 2 is not a fraction, but rather, a crude control of Taq polymerase to ensure that the gel electrophoresis was functioning properly, which is why it is significantly brighter. It should not be considered as part of the 1.0 mL/min results. Lanes 3 through 11 correspond to fractions 1 through 9; thus, fractions 4 through 6 demonstrate Taq expression as can be seen by the bands. The Taq is concentrated in

three fractions, as opposed to the five fractions from the 0.5 mL/min flow rate testing. These bands are fainter than that of the 0.5 mL/min flow rate, which indicates that there was less expression overall in the 1.0 mL/min, but in terms of optimization, the consolidation into a smaller number of fractions is more important. Thus, 1.0 mL/min is considered more optimal than 0.5 mL/min. This was somewhat unexpected, because previous published research indicates that a slower flow rate will allow more time for Taq polymerase to wash off the resin, resulting in a smaller number of more concentrated fractions [23].

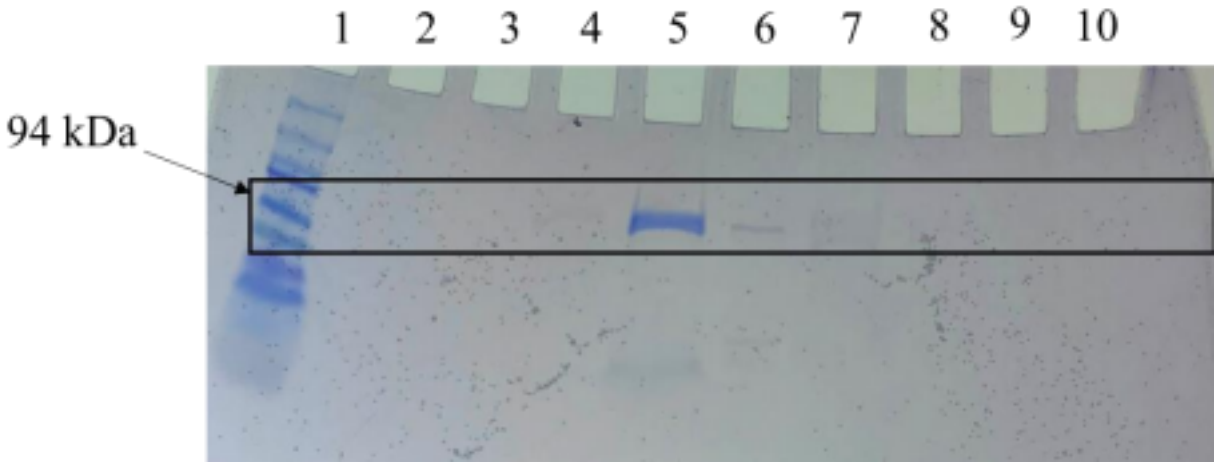

Figure 4: Gel electrophoresis results from 2.0 mL/min flow rate Taq polymerase collection. Lanes 4 through 6 show presence of Taq polymerase, with significant expression in Lane 5

As seen in Figure 4, the 2.0 mL/min flow rate chromatography resulted in Lanes 5 through 7 showing Taq polymerase expression. These lanes correspond to fractions 4 through 6. Lane 5 shows a very distinct band of expression in comparison to Lanes 6 and 7; the Taq seen in the latter lanes is nearly insignificant compared to the banding seen in Lane 5. This indicates that nearly all of the Taq was collected in one fraction, and was collected at a high concentration. Thus, of the three flow rates tested, 2.0 mL/min was determined to the optimal value in the

context of this experiment. Again, this seems to disagree with published research indicating that generally, in protein purification, a slower flow rate produces the ideal lower number of fractions that is desired. However, there is little to no published research on flow rate testing in Taq polymerase purification *specifically*, nor had research been published on this experiment being performed with these particular resin, equipment, and protocols being used. Thus, the discovery of the 2.0 mL/min flow rate testing indicates an area of further research to be done, by way of this experiment being replicated to see if this parameter optimization holds true. It would also be worthwhile to test higher flow rates, to see if there is an even more improved result that can be achieved, along with the time savings that come with a higher flow rate.

It is also important to note the potential value that comes with flow rate optimization. Applying protein purification to manufacturing processes, such as those of drug manufacturing and clinical research, requires considerations for resource and time efficiency. By using the 2.0 mL/min flow rate as opposed to the slower flow rates tested, the process is optimized in two ways: firstly, the aforementioned smaller number of fractions with higher Taq expression is seen with this flow rate. This puts the fractions in a position to be more readily used. Secondly, less time is needed to purify the Taq with this higher flow rate; this is mathematically sound, since a higher volume is flowing into the fractions per unit time, and was experimentally observed, in that the 2.0 mL/min flow rate testing took the least time compared to the slower flow rates. Applying these principles to a large-scale process could result in substantial financial savings for the administrators and companies performing the process, which would ideally be passed on to the lab performing the work or eventually, consumers who are using therapeutic proteins.

Enzymatic Activity

The final analysis performed in this experiment was the use of collected Taq polymerase in PCR. Recall that fraction 4 from the 2.0 mL/min flow rate was chosen to be the DNA polymerase enzyme in PCR, and that it was dialyzed and diluted into 1:1, 1:10, and 1:100 solutions that correspond to lanes 2, 3, and 4 on the resulting gel electrophoresis.

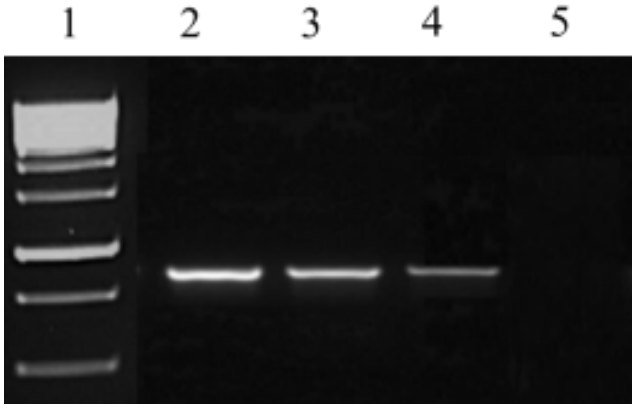

Figure 5: Gel electrophoresis results from PCR of 2.0 mL/min flow rate testing. Lane 1 shows a protein ladder, Lanes 2 through 4 show 1:1, 1:10, and 1:100 Taq serial dilutions, and Lane 5 is a water control.

As seen in Figure 5, there is an expected band intensity to concentration correlation; as the Taq becomes more diluted, it shows a less intense band of DNA amplification. It is notable that the 1:100 dilution still shows results, as this indicates that the functional ability of the collected Taq is high. Also notable is the fact that each lane shows amplification of the same target DNA, since they are all seen at the same point of the protein ladder; this was ideal and expected. The water control shows no DNA amplification, also as expected, indicating that there was no contamination of the PCR process. This demonstrates that the Taq polymerase was not damaged throughout the induction, purification, or collection process, and that it is successful in its function of acting as DNA polymerase, which was the overarching goal of the optimization of this Taq polymerase purification protocol.

Conclusion

Via the experiment that was performed, the collection of functional Taq polymerase from modified *E. coli* was successfully completed, and used to perform PCR. It was found that in the context of this experiment setup, an IPTG concentration of 0.1675 mM, i.e. ½ the original protocol value, resulted in the highest protein expression in the induced *E. coli*. It was also found that the optimal chromatography flow rate was 2.0 mL/min, indicating an anomaly to the general trend of slower flow rates being optimal. These results optimize the process both by using less IPTG than the standard protocol, concentrating the Taq into one usable fraction, and by saving time via the higher flow rate used. On a local scale, the optimization of Taq production is beneficial to the use of PCR in classes and labs, for instance, in which Taq may be used as DNA polymerase. On a larger scale, the knowledge gained from this experiment can be applied and scaled to other protein purification processes, such as those of therapeutic protein production. The resulting optimization from these parameter alterations can result in millions of dollars of savings for pharmaceutical companies, and for the consumers who require these products.

While this experiment analyzed only two parameters, there are a multitude of variables and other optimizing actions that could be taken to further improve this process. More research and repeated experimentation should be done to both validate the results of this experiment and to find further areas of improvement, but overall, the optimization process performed here contributes to the overall understanding of Taq polymerase purification, protein purification in general, and the benefits of continuing to innovate this process.

Bibliography

Citation Style: IEEE Referencing Style

[1] M. A. Valones, R. L. Guimarães, L. A. Brandão, P. R. de Souza, A. de Albuquerque Tavares Carvalho, and S. Crovela,. "Principles and applications of polymerase chain reaction in medical diagnostic fields: a review," *Brazilian Journal of Microbiology,* vol. 40, no. 1, pp. 1–11, 2009. Available: PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3768498/

[2] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, *Avonex Summary Basis for Approval*, 1996.

[3] L. I. Patrushev et al., "Cloning of the gene for thermostable Thermus aquaticus YT1 DNA polymerase and its expression in Escherichia coli," *Molecular Biology (Mosk)*, vol. 27, no. 5, pp. 1100-1112, 1993.

[4] L. Garibyan and N. Avashia, "Research techniques made simple: Polymerase chain reaction (PCR)," *The Journal of Investigative Dermatology*, vol. 133, no. 3, pp. 1-4, 2013. Available: PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4102308/

[5] J. S. Tan, R. N. Ramanan, T. C. Ling, et al. "The role of *lac* operon and *lac* repressor in the induction using lactose for the expression of periplasmic human interferon-α2b by *Escherichia coli,*" *Annals of Microbiology,* vol. 62, pp. 1427–1435, 2012. Available: BioMed Central, https://annalsmicrobiology.biomedcentral.com/articles/10.1007/s13213-011-0394-3

[6] Manuals Lib, *GE Peristaltic Pump P-1 Operating Instructions Manual*. [Online]. Available: https://www.manualslib.com/manual/1885348/Ge-Peristaltic-Pump-P-1.html

[7] J. M. Berg, J. L. Tymoczko, L. Stryer, "Section 31.1, Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons," *Biochemistry, 5th ed*. New York: WH Freeman, 2002.

[8] H. F. Liu, J. Ma, C. Winter, R. Bayer, "Recovery and purification process development for monoclonal antibody production," *MAbs*. vol. 2, no. 5, pp. 480-499, 2010. Available: PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2958570/

[9] M. Derde, V. Lechevalier, C. Guérin-Dubiard, et al. "Hen Egg White Lysozyme Permeabilizes Escherichia coli Outer and Inner Membranes," *Journal of Agricultural and Food*

Chemistry, vol. 61, iss. 41, pp. 9922-9929. Available: ACS Publications, https://pubs.acs.org/doi/10.1021/jf4029199#

[10] Y. Kamal Deepthi and R. Malothu. "Cloning, isolation and characterization of taq dna polymerase using pET expression system," *International Journal of Life Sciences Biotechnology and Pharma Research,* vol. 2, no. 1, pp. 268-283, 2013. Available: International Journal of Life Sciences Biotechnology and Pharma Research, http://new.ijlbpr.com/jlbpradmin/upload/ijlbpr_50e55dd419b17.pdf

[11] G. Sezonov, D. Joseleau-Petit, and R. D'Ari, "*Escherichia coli* Physiology in Luria Bertani Broth," *Journal of Bacteriology*, vol. 189, no. 23, pp. 8746-8749, 2007. Available: PubMed Central, https://pubmed.ncbi.nlm.nih.gov/17905994/

[12] Bio-Rad, *Microbial Culturing Kit*, Bio-Rad Laboratories. [Online]. Available: https://www.bio-rad.com/webroot/web/pdf/lse/literature/1665022.pdf

[13] University of Hohenheim, *DNA plasmid minipreps – How it works*. [Online]. Available: https://www.uni-hohenheim.de/fileadmin/einrichtungen/pflanzenphysiologie/Protokolle/II.DNA/DNA_plasmid_minireps_hintergrundinfo_en.pdf

[14] Five Photon, *Cell Lysis and Immunoprecipitation Protocol with CHAPS Buffer.* [Online]. Available: https://fivephoton.com/pdfs/CHAPS%20buffer%20IP%20Protocol.pdf

[15] M. Nagai, A. Yoshida, and N. Sato, "Additive effects of bovine serum albumin, dithiothreitol, and glycerol on PCR," vol. 44, no. 1, pp. 157-163, 1998. Available: Wiley, https://iubmb.onlinelibrary.wiley.com/doi/pdf/10.1080/15216549800201172

[16] S. Aryal, *Streak Plate Method- Principle, Methods, Significance, Limitations*, Microbe Notes. [Online]. Available: https://microbenotes.com/streak-plate-method-principle-methods-significance-limitations/

[17] R. T. Santilli, J. E. Williamson, Y. Shibata, *et al.* "The Penn State Protein Ladder system for inexpensive protein molecular weight markers," *Sci Rep*, vol. 11, iss. 16703, 2021. Available: Nature.com, https://doi.org/10.1038/s41598-021-96051-x

[18] ThermoFisher Scientific, *Coomassie Blue Staining*. [Online]. Available: https://www.thermofisher.com/us/en/home/references/protocols/proteins-expression-isolation-and-analysis/staining-protein-gels/coomassie-staining.html

[19] O. Coskun, “Separation techniques: Chromatography,” *North Clin Istanbul*, vol. 3, iss. 2, pp. 156-160, 2016. Available: PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5206469/

[20] S. Cheriyedath, *How Does Ion Exchange Chromatography Work?* News Medical, 2018. [Online]. Available: https://www.news-medical.net/life-sciences/How-Does-Ion-Exchange-Chromatography-Work.aspx

[21] F. Moazen, A. Rastegari, S. M. Hoseini, M. Panjehpour, M. Miroliaei, and H. M. M. Sadeghi, “Optimization of Taq DNA polymerase enzyme expression in Escherichia coli,” *Advanced Biomedical Research*, vol. 1, pp. 82, 2012. Available: PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3544109/

[22] K. Martin, *A Deep Dive Into Induction with IPTG*, GoldBio. [Online]. Available: https://www.goldbio.com/articles/article/a-deep-dive-into-iptg-induction

[23] R. A. Bernhard, “Effect of Flow-Rate and Sample-Size on Column Efficiency in Gas Liquid Chromatography,” *Nature*, vol. 185, pp. 311-312, 1960. Available: Nature.com, https://www.nature.com/articles/185311a0