Optimization of Purification Protocol for Taq DNA Polymerase used in Polymerase Chain Reaction

Submitted by
Allison Laws

Bioengineering

To
The Honors College
Oakland University

In partial fulfillment of the
requirement to graduate from
The Honors College

Mentor: Shailesh Lal, Professor and Chair of Bioengineering Department
Department of Bioengineering
Oakland University

November 30, 2021

Table of Contents

Abstract 3

Introduction 4

Drug Production and Protein Purification 4

Taq Polymerase Model 5

Protein Purification Parameters 6

Objectives 10

Materials and Methods 11

Media and Stock Solutions 11

Cell Culture and T-streaking 13

IPTG Induction 14

Gradient Column Chromatography 16

Gel Electrophoresis 18

Polymerase Chain Reaction 19

Results and Discussion 20

IPTG Concentration 20

Flow Rate 22

Enzymatic Activity 25

Conclusion 27

Acknowledgements 27

References 28

Appendix 31

Abstract

Protein purification plays a significant role in the production of therapeutic proteins, otherwise known as protein drugs. Although 380 protein drugs are currently marketed and used to treat a wide range of illnesses, the protocol for protein purification varies slightly amongst pharmaceutical companies. Thus, this study was undertaken to optimize some parameters involved in protein purification. Taq DNA polymerase was inserted into and grown in an *E. coli*. cell culture, isolated using ion gradient chromatography and a fraction collector, and assayed using polymerase chain reaction. The concentration of isopropyl-β-D-1-thiogalactopyranoside (IPTG) used to induce the bacterial culture and the column flow rate in chromatography were tested and adjusted to optimize the protein purification protocol and determine their effect on both the quantity of the purified protein collected and the enzymatic activity of the purified protein. It was discovered that an IPTG concentration of 0.1675 mM and a column flow rate of 2 mL/min resulted in the highest protein expression. It was concluded these values optimized the protein purification protocol both economically and in terms of production quantity.

Introduction

Drug Production and Protein Purification

Therapeutic proteins, or protein drugs, compose a significant percentage of the pharmaceutical market, with approximately 380 pharmaceutical protein drug products on the market. Protein drugs are highly favorable due to their versatility, extensive presence within cell membrane, specificity with desired target molecules, and low toxicity in vivo [1]. Some examples of protein drugs include insulin, bevacizumab (Avastin), adalimumab (Humira), and interferon beta-1a (Avonex) [2]. Due to the widespread need for these therapeutic drugs, drug production methods exist in order to produce large quantities of a pure, sterile, and stable form of the protein to be mixed with other ingredients for intravenous, subcutaneous, or intramuscular injection, or very rarely oral administration [3]. In general, drug production involves protein purification followed by addition of ingredients for maintaining the shelf life of the product, easy storage, and ensuring that the protein solution is of an administrable volume.

The following drug production protocol is for Avonex, an interferon beta-1a (IFNβ-1a) used in the treatment of multiple sclerosis [2]. Nonetheless, the protein purification protocol for drug production is virtually the same for all protein drugs; the main difference lies within the addition of other ingredients for packaging and storage purposes. For many protein drugs, protein purification begins with the usage of recombinant DNA technology for inserting a plasmid containing the DNA for the target protein into bacterial cells, commonly *Escherichia coli* (*E. coli.*). In the case of Avonex, the human gene coding for IFN-β was initially sourced from human white blood cells and cloned using polymerase chain reaction (PCR) in order to produce billions of copies of the gene. A plasmid containing the IFNB1 gene was inserted into the ovary cells of Chinese hamsters, which were adapted for suspension culture, through recombinant DNA technology. These cells were cultured to produce high quantities of IFNβ-1a, the produced protein

with the same amino acid sequence as the naturally occurring interferon beta, and analyzed to determine their stability, sterility, and viability to be used for drug production. The produced protein is then extracted from the Chinese hamster ovary cells, isolated from cell debris, and purified via chromatography. Any potential viruses are removed through a viral inactivation process, and the intermediate form of the final drug is transported to a different location to be mixed with other ingredients, such as albumin and sodium chloride to produce a lyophilized powder, or sodium acetate trihydrate, acetic acid, and arginine chloride to produce a prefilled syringe or pen for injection [4]. While the mentioned method is specific to Avonex, protein drug production has these general steps: recombinant DNA technology, cell culture, protein purification via chromatography, activity and viability analysis, addition of other ingredients for packaging.

Although the protein purification protocol generally remains the same for all therapeutic proteins, values of the parameters differ, and there has been little research on determining the optimal values for these parameters so that high quantities of a purified protein can be produced. Therefore, this research project was interested in determining the optimal values for two parameters that control the purification protocol. The parameters that were altered and the specific procedure for protein purification are outlined in the **Materials and Methods** section.

Taq Polymerase Model

For this project, Taq DNA polymerase (Taq) was the desired protein to be isolated. Taq DNA polymerase is a DNA polymerase enzyme that originated in the thermophilic bacterium *Thermus aquaticus* [5]. Its importance is related to its efficiency in PCR, a DNA amplification technique. In PCR, the target DNA is denatured by an increase in temperature, and thus the double helix is split into two strands through breakage of the hydrogen bonds that maintain the double

helix structure. A primer, which is a short, single-strand DNA sequence that contains the corresponding nucleic acids for the target DNA, is added to the beginning of the target DNA sequence. After addition of the primer sequence, the DNA polymerase enzyme constructs the opposite DNA strand for the two parent strands by adding the corresponding nucleic acids, thereby doubling the quantity of DNA that was originally present. PCR is repeated to produce millions or billions of copies of DNA in a short period of time [6]. PCR is a laboratory procedure used to amplify DNA sequences in a similar way to DNA replication in the body.

Since Taq is not grown in wild-type *E. coli*., recombinant DNA technology was used to insert it into *E. coli*. cells. This was achieved by introducing a plasmid vector containing the YT1 gene from *Thermus aquaticus* into an *E. coli.* cell, typically through electroporation or heat shock transformation [7,8]. *E. coli*. was used for this project, as well as in many protein purification laboratories, due to "both its relative simplicity and the ease with which it can be propagated" in laboratory research [9]. After inserting the plasmid into the bacteria, the cells can be cultured on and induced with various mediums and substances in order to promote cell growth and increase the protein yield.

Protein Purification Parameters

As mentioned, this research project was designed to optimize two parameters related to protein purification so that a high quantity of the pure, concentrated protein can be produced. The selected parameters were isopropyl β-D-1-thiogalactopyranoside (IPTG) and column chromatography flow rate.

IPTG is known to induce protein expression in bacterial cells and is a common reagent used for increasing the protein expression in protein purification protocols. IPTG acts similarly to

allolactose, which is a product of lactose metabolism and the enzyme responsible for inducing the *lactose* (*lac*) operon. In research studies, *E. coli.* is oftentimes grown on mediums containing lactose, as it causes a high expression of the enzyme β-galactosidase, which is responsible for metabolizing the disaccharide. Concurrently, β-galactosidase must be regulated; its regulation, and the regulation of galactoside permease and thiogalactoside transacetylase, in enteric bacteria was the basis for the proposal of the *lac* operon model, which describes the method by which gene expression in *E. coli.* and other bacteria is regulated. Operons consist of a regulator gene, an operator site, and a set of structural genes. So that the structural genes are not constantly undergoing transcription, the regulator gene is responsible for encoding a repressor protein that binds to the operator site DNA sequence. In the case of the *lac* operon, the *lac* repressor will bind to the *lac* operator in the absence of lactose; however, in the presence of an inducer, the operator's affinity for the repressor is significantly inhibited. Such inducers include allolactose and IPTG, which enable RNA polymerase to transcribe the structural genes of the operon and increase protein expression [10]. Despite high levels of IPTG being able induce a faster growth rate and increased protein expression, once a certain limit is reached, IPTG is known to exhibit toxicity to recombinant bacterial cells [11]. Thus, IPTG was selected as the inducer for this research project since it is commonly used for inducing gene and protein expression in bacterial cultures, and there has not been any optimal IPTG concentration determined for the protein purification protocol used in protein drug production.

The second parameter to be tested was the column chromatography flow rate. In column chromatography, the flow rate plays a significant role in the final yield of the desired protein and the enzymatic activity of that protein. Many mathematical models have been established for representing how the flow rate affects the yield of the protein. One such model was described in

the paper “Mathematical Modeling of Elution Curves for a Protein Mixture in Ion Exchange Chromatography and for The Optimal Selection of Operational Conditions.” The relationship is displayed below:

$$yield_A = \frac{\int_{t_1}^{t_2} C_A \cdot F dt}{C_{A0} V_0} \quad (1)$$

In this paper, the yield was defined as the “ratio between the mass of the target protein in the collected volume and the mass of the same protein loaded into the column,” which is related to the efficiency of the protein purification process. The variables C_A, F, C_{A0}, and V_0 refer to the concentration of Taq in the fraction collections, flow rate for the column, original concentration of Taq in the sample prior to loading the column, and the loading sample volume, respectively [12].

It is assumed that due to the high number of mathematical models for ion exchange chromatography, no optimal column flow rate has been established, especially one that is specific to Taq polymerase and the size of the columns that were used in this experiment. Regardless, it was hypothesized that a relatively low flow rate would produce the best yield, since previous studies showed that lower flow rates optimize the column efficiency by concentrating higher quantities of the isolated protein into the collected fractions [13]. Despite this knowledge, flow rates used for protein drug production differ for different drugs; thus, this project was undertaken in order to establish an optimal flow rate using Taq as a model.

Although only two parameters were tested, there are numerous parameters that control the final yield and enzymatic activity of the target protein in protein purification processes. In the initial steps of protein purification, the media used for cell culture plays a role in the total amount of protein that can be produced. *E. coli.* can grow on both minimal media and rich media, which are defined by the components of the media, yet *E. coli.* will grow more rapidly in rich media

environments. Some examples of minimal media include, M9 medium, M63 medium, and A medium, which all contain a carbon source and salts containing nitrogen, phosphorus, and trace metals. Rich media, such as H medium, Lambda broth, TNT medium, Luria-Bertani/Lysogeny (LB) medium, and terrific broth, contain amino acids, nucleotide precursors, vitamins, and other metabolites, which provide nutrients that the cell would have to synthesize itself otherwise [14]. In this project, the type of growth media was controlled; LB broth was used.

Two parameters related to the inducer, excluding the already mentioned inducer concentration, can be altered: type of inducer and time of induction. In the case of this project, IPTG was used. However, other inducers, such as allolactose, galactose, and thiomethyl galactoside (TMG), can be used to induce gene expression in *E. coli.*, as they are known to inhibit the *lac* operon repressor [10, 15, 16]. The time of induction also plays a role in the level of protein expression. The time of induction has a relatively logarithmic relationship with the optical density of the cell culture, which is proportional to the surface protein concentration [17, 18]. In order to promote protein expression, the minimum time of induction should be three hours, yet the protein concentration will continue to increase with induction time until a certain limit is reached [19, 17]. Thus, induction time can be changed and tested as a parameter in order to determine an optimal time of induction that generates the highest protein expression.

Several parameters related to ion exchange chromatography and protein collection could be tested as well. These include the salt concentration, type of salt gradient, and number of fractions collected. In terms of the salt concentration, varying values for the high and low salt concentrations could be selected. The salt concentration is related to the affinity of the salt to the binding solid, which was DE-52 resin in the case of this project. A higher salt concentration would be expected to bind more easily and closely to the negatively charged resin due to its positive

charge [20]. Changing the concentrations of the salt would impact its binding to the resin and is a possible parameter that could be tested for protein purification optimization. In addition to the salt concentration, the type of salt could be changed and tested to determine the optimal salt for protein elution in column chromatography. In this experiment, potassium chloride, KCl, was used, however salts, such as sodium chloride, ammonium sulfate, sodium sulfate, and ammonium acetate, can be used [21]. The affinity of these salts to the binding resin vary, thus making it a good parameter to test. Finally, the number of fractions collected is another parameter, yet is a weaker parameter to test due to the nature of fraction collection. When the protein sample is run through the column, it is eluted into fraction tubes that can hold a certain volume of liquid. In protein purification, a concentrated form of the target protein is desired, and this is achieved through using a low number of fractions [22]. Thus, the number of fractions is not the ideal parameter to test, since an optimal range of fractions has been established.

While many parameters can be optimized for the protein purification protocol used in protein drug production, this research project was limited to two parameters: IPTG concentration and column flow rate. Five concentrations of IPTG were tested: .08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM. Three flow rates were tested: 0.5 mL/min, 1 mL/min, and 2 mL/min. These values were selected and adjusted based on commonly used values from previous experiments.

Objectives

This study was designed to optimize the general protocol for protein purification by determining the optimal IPTG concentration and column flow rate that result in the greatest expression and enzymatic activity of Taq. Optimization of this protocol is significant, as it is used

in the production of therapeutic drugs for debilitating diseases, and can help in improving the enzymatic activity of the protein, increasing the production of the protein, and reducing economic costs. The specific objectives for this project are outlined as follows:

1. IPTG is responsible for inducing protein expression within a cell culture, so determining the optimum concentration that yields the highest Taq expression will improve the quantity of Taq produced and the protein purification process in general.
2. By determining the effect that column flow rate impacts Taq enzyme activity, a standard protocol can be established, and proteins can be isolated with fewer impurities and greater enzyme activity.
3. By optimizing the protein purification protocol, drug production companies can increase the produced quantities of protein drugs, reduce wastes, and decrease production costs.

The protein purification protocol for Taq and the optimized results are discussed in the sections below.

Materials and Methods

The general methodology for the entire experimental process is visually diagrammed in **Appendix 1**.

Media and Stock Solutions

500 mL of LB broth, which stands for Luria-Bertani or Lysogeny broth, was produced using 5.0 g of tryptone, 2.5 g of yeast extract, and 5.0 g of NaCl. The LB broth was autoclaved and mixed with 100 μg/mL ampicillin (amp). LB broth was used for *E. coli.* growth, since it is

known to induce fast growth and good growth yields in various bacterial species, as well as enables growth to an optical density of 600 nm [23]. Ampicillin was added to prevent bacterial contamination and eliminate the *E. coli.* cells that did not contain plasmids with an ampicillin resistance gene [24].

50 mL of Buffer A was made with 2.5 mL of 50 mM Tris-Cl stock with a pH of 7.9, 0.45 g of 50 mM Dextrose, and 0.2 mL of 1 mM EDTA. Distilled water was added until a final volume of 50 mL was achieved. Buffer A was added to the protein sample in order to convert the *E. coli.* culture into spheroplasts, in which the cell walls are deteriorated without harming the cytoplasmic membranes [25]. Lysozyme was added to Buffer A in order to catalyze the cell wall breakdown.

20 mL of Buffer B was created by adding 20 μL of 10 mM Tris-Cl of 7.9 pH, 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. Buffer B is also referred as a lysis immunoprecipitation buffer, and it has the main purpose of lysing cells and solubilizing proteins, such as the target protein, Taq polymerase [26].

300 mL of Buffer C was made for gradient column chromatography. Buffer C was composed of 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. A final volume of 300 mL was reached by adding 20 mM HEPES of 7.9 pH until the 300 mL mark on the flask. Buffer C was used to stabilize the proteins and prevent oxidization of the protein sample [27].

2000 mL of Storage Buffer was 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. Storage Buffer was used after fraction collection to dialyze the fractions with peak activity.

Cell Culture and T-Streaking

The strain of *E. coli.* was obtained from Dr. Shailesh Lal's laboratory at Oakland University. The *E. coli.* culture had previously been introduced with a recombinant plasmid containing Taq, which composed the Taq polymerase bacterial stock. The bacterial stock used to grow the cell culture was created by pipetting a single *E. coli.* colony into 3 mL of LB broth without amp. The bacterial stock was t-streaked onto a plate containing LB Agar and 100 μg/mL amp. The T-streak method is a type of dilution scheme and was used to reduce the cell density on the surface of the LB Agar plate so that single colonies could be isolated [28]. T-streaking involves inserting a sterilized O-ring inoculation loop into the Taq polymerase bacterial stock and streaking the loop across the plate three times. For the first streak, one quadrant of the LB Agar plate is selected, and the loop is dragged from the edge of the plate towards the center in a zig-zag motion until the quadrant is completely streaked. The loop is re-sterilized by passing it through a flame, but is not re-inserted into the bacterial stock, and then dragged lightly through the second quadrant of the LB Agar medium in the same zig-zag pattern without disturbing the previous streak. After re-sterilizing for a third time, the final streak is made in the same pattern so that the plate has been completely streaked. The figure below shows a visual depiction of the T-streak method and separation of streak sections [28].

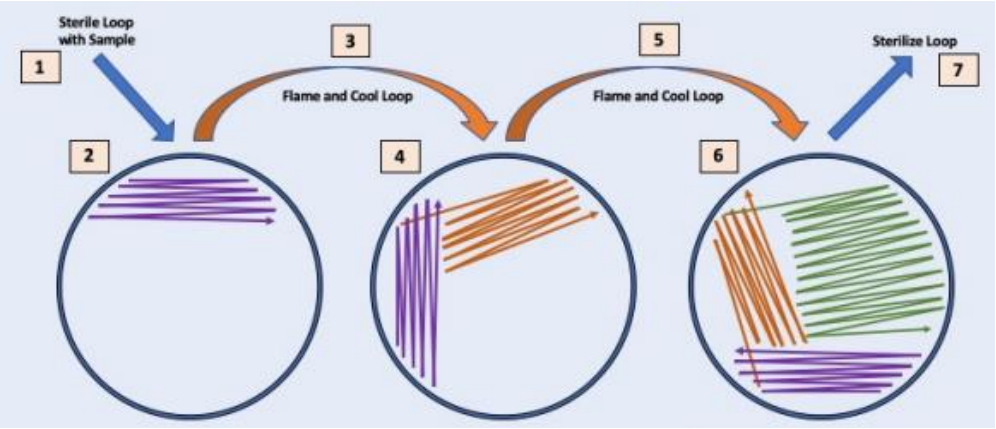

Figure 1. Visual Depiction of T-Streak Method [28]

Following t-streaking, the plate was grown overnight at 37 °C. To initially test the optimal IPTG concentration, a single *E. coli.* colony was obtained and inoculated into 3 mL of LB broth with 4 μL of amp and grown overnight. Five flasks of 50 mL LB media solution were prepared and inoculated with 50 μL of the overnight *E.coli.* solution and an additional 50 μL of amp. After 4 hours of inoculation, each flask containing the bacteria and media were induced with a different concentration of IPTG. The five concentrations were: 0.08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM. The IPTG induction process is explained in more depth in the following section.

A single colony was obtained and inoculated into 500 mL of LB broth with 0.5 mL amp, and grown on a shaker for 16 to 18 hours at 22 RPM and 37 °C. Once grown to an optical density of 600 nm, the inoculated solution was induced with five concentrations of IPTG.

IPTG Induction

As previously mentioned, five concentrations of IPTG were tested: .08375 mM, 0.1675 mM, 0.335 mM, 0.5025 mM, and 0.67 mM. These values were selected since previous experimental trials for Taq polymerase used 0.335 mM; naturally, a quarter, half, one times, 1.5 times, and twice this value were selected for testing. 0.5 M IPTG stock solution was used to do the dilutions into the five flasks containing 50 mL of the sample. After induction, the samples were grown while shaking at 220 RPM at 37 °C for 24 hours.

Initially, a calculation error was made for the buffers; the buffers were prepared for 500 mL of sample, as opposed to 50 mL of sample. Thus, the sample was 10 times more diluted than intended. For this error, 47 mL of the sample was centrifuged for 10 minutes (min) at 4 °C, and 150 mL of Buffer A and 50 mL of Buffer B were produced. After the first centrifugation, the pellet

was isolated, resuspended in 20 mL of Buffer A, homogenized, and then spun at 4 °C for 10 min. Then, 10 mL of lysozyme was added to 50 mL of Buffer A. 10 mL of Buffer A + lysozyme was added to the centrifuge tube without disrupting the pellets and incubated for 15 min, followed by addition of 10 mL of Buffer B. The solution was incubated for 1 hour at 75 °C and then centrifuged a final time for 15 min at 10,000 RPM at 4 °C. The supernatant was saved and frozen until running the gel.

Prior to running the gel, 45 µL of the sample was boiled for 30 min, then frozen for 3 min. 40 µL of each sample and 5 µL of Bio-Rad Precision Plus Protein Kaleidoscope™ Protein Ladder were loaded into separate wells in a 7.5% 15-well Bio-Rad Mini-PROTEAN® TGX Stain-Free™ Precast Gel. The Pre-stained Protein Ladder was used as a protein marker for estimating the molecular weight of the proteins separated during the SDS-PAGE gel electrophoresis, and ensured that the gel was running correctly [29]. Again, an error was made, since each well only holds 15 µL; so, some of the sample spilled over into neighboring wells, and the samples were unable to be in adjacent wells due to the spillover. The gel was run at 100 volts, and it was made sure that the power and ground wires were connected to the color coordinated inputs until the samples reached the bottom of the gel. Coomassie blue stain was used to stain the entire gel for approximately 1 hour, and then de-stained using a de-staining buffer overnight. Due to the significantly diluted sample, there was no appearance of protein expression on the gel, and the protocol was repeated, starting with a new 3 mL inoculation.

For the new trial, the same inoculation and induction steps were followed. However, only 38 mL of the sample was transferred to the centrifuge tubes and spun for 10 min at 4 °C. 50 mL of Buffer A was made, and 20 mL of Buffer B was made. After isolating the pellet, the sample was suspended in 1 mL of Buffer A, homogenized, and centrifuged for 10 min at 4 °C. 2 mL of

lysozyme was added to 10 mL of Buffer A, and then 1 mL of Buffer A + lysozyme was added to the centrifuge tube and incubated for 15 min. 1 mL of Buffer B was added and then incubated for 1 hour at 75 °C. Again, the tubes were centrifuged for 15 min at 10,000 RPM at 4 °C. The supernatant was saved and frozen.

Another gel was run for this trial, with 45 μL of the sample being boiled for 30 min and frozen for 3 min prior to loading. 15 μL of each sample and 5 μL of a Pre-stained Protein Ladder were added to individual wells. After running the gel, it was stained for an hour and de-stained overnight. The results are shown in the **Results and Discussion** section.

After inspecting the gel, it was determined that 0.1675 mM was the optimal IPTG concentration, as it showed the highest protein expression for Taq. Thus, a new 500 mL protein sample was inoculated and induced with 0.1675 mM IPTG. This corresponds to a volume of 167.5 μL of 0.5 M IPTG stock being added to the 500 mL protein sample. This sample was centrifuged and diluted with the corresponding buffers using the same process as outlined before for the 500 mL dilutions. The supernatant was saved and frozen until running the columns.

Gradient Column Chromatography

Each column required 30 mL of DE-52 resin and 70 mL of Buffer C. To prepare four columns, 120 mL of resin was mixed with 280 mL of Buffer C one day prior to packing the columns. Four columns were prepared and packed by slowly adding the resin + Buffer C mixture slowly to the column with the valve open to allow the Buffer C solution to run through the column while leaving the solid packed resin. After packing the column, 135 mL of Buffer C was run through the column for equilibration. Columns were stored in the refrigerator at 4 °C until used for flow rate testing.

To establish the salt gradient, 25 mM and 700 mM KCl solutions were used. These concentrations were made by adding 0.555 g and 15.54 g of KCl stock to 300 mL of water, respectively. 40 mL of 700 mM KCl and 25 mM KCl were added to each side of the gradient maker while the valve was closed. The 25 mM KCl was filled in the side of the gradient maker that contained the opening for the tube that attaches to the column.

As previously mentioned, three flow rates were tested: 0.5 mL/min, 1 mL/min, and 2 mL/min. Using the Pharmacia Biotech Peristaltic P-1 Pump, the flow rate was adjusted. For each flow rate, 15 mL of Buffer C was initially run through the column for re-equilibration until the meniscus of the liquid was right above the resin. Then, 8 mL of the protein sample induced with the optimal IPTG concentration, was added to the column. Due to the positively charged nature of the resin and its high binding capacity, the negatively charged Taq binds to the resin; this process is used to isolate Taq from other proteins in the solution [20, 30]. Then, 50 mL of 25 mM KCl was washed through the column to remove impurities from the column, such as undesired proteins and other substances. After washing the column, the column valve was closed and completely filled with 25 mM KCl to ensure there were no air pockets. The salt gradient, containing 700 mM KCl on one side and 25 mM KCl on the other, was attached. After attaching the gradient, the valve on the column and the valve on the gradient maker were opened and a Pharmacia Biotech RediFrac Fraction Collector was calibrated to collect 6 mL fractions. For each flow rate, 9 fractions were collected.

To explain how gradient column chromatography functions, it is a protocol used for efficiently separating desired molecules from a complex solution within a suitable time interval, which can further be used for quantitative analysis. In chromatography, components are divided into a stationary phase, mobile phase, and separated molecules. The stationary phase refers to a

solid or liquid phase coated on the surface of a solid phase, while the mobile phase refers to a liquid or gaseous phase that flows over the stationary phase. In this experiment, the stationary phase was the 8 mL protein sample, while the mobile phase was the 25 mM KCl wash buffer. For gradient chromatography, the desired protein is separated through the different affinities of the protein and salt ions for the resin. The positively charged resin has a high affinity for the negatively charged Taq, and thus Taq attaches to the resin initially. Since the affinity for the 25 mM salt concentration is lower than the affinity for Taq, the protein remains attached to the resin while the column is being washed. However, when the salt gradient is applied and the 700 mM KCl runs through the column, Taq washes off the resin and into the fraction collection tubes, since there is a higher negative charge for the 700 mM KCl as opposed to Taq [30]. This is a common method for elution of proteins from a complex protein sample.

For each flow rate tested, a gel was run for all of the collected fractions. This will be explained in the next section.

Gel Electrophoresis

7.5% 15-well Bio-Rad Mini-PROTEAN® TGX Stain-Free™ Precast Gels were once again used for determining Taq expression for each flow rate. Prior to loading the gels, 20 µL of each sample – a crude control for the 1 mL/min flow rate and fractions 1 through 9 for all flow rates – was mixed with 5 µL of NuPage LDS buffer, vortexed, and boiled for 30 minutes. 12 µL of each dyed sample was loaded into the gel electrophoresis comb. Well 1 always contained 5 µL of the Bio-Rad Precision Plus Protein Kaleidoscope™ Protein Ladder. For the 1 mL/min flow rate gel, the crude control sample was loaded into well 2 and fractions 1 through 9 were loaded into wells 3 through 11, respectively. For the 0.5 mL/min and 2 mL/min flow rates, no crude control was

used, so wells 2 through 10 were used for fractions 1 through 9. The gel ran for an hour at 100 volts, then stained with Coomassie blue stain for an hour and de-stained overnight.

The resulting gels are shown and analyzed in the **Results and Discussion** section. It was found that the optimal flow rate for isolating Taq was 2 mL/min.

Polymerase Chain Reaction

The optimized flow rate samples underwent dialysis to remove salt and reduce the salt concentration from the collected Taq solution. For the 2 mL/min optimal flow rate, fraction 4 showed the greatest banding for Taq and was selected for dialysis. A dialysis membrane cassette was used in order to filter KCl from the protein solution. The collected 6 mL of the solution from fraction 4 was injected into the cassette membrane, and any remaining air from the membrane was removed. The cassette containing the protein solution was then steeped in the Storage Buffer, a glycerol solution, overnight to remove the salt in the protein solution through osmosis. After dialysis, a total volume of 2.7 mL of pure Taq from fraction 4 was obtained.

For PCR, a PCR Master Mix was made, which contained 25 μL buffer, 25 μL dNTP, 5 μL forward primer, 5 μL backward primer, and 180 μL water. The mix was vortexed to homogenize the solution. Then, a serial dilution of Taq polymerase with H_2O was completed. It was diluted 1:10 and 1:100. First, 1 μL of Taq was added to 9 μL of water, followed by 1 μL of the diluted Taq added to 9 μL of water again. Finally, 1 μL from each dilution was tested for enzymatic activity using a Bio-Rad iCycler Thermal Cycler PCR Machine. The first well contained the pre-stained protein ladder, and wells 2 through 5 contained 1 μL of the undiluted pure Taq solution, 1:10 diluted solution, 1:100 diluted solution, and an H_2O control, respectively. The results of the gel were used to calculate the number of reactions achieved.

Results and Discussion

IPTG Concentration

Determination of the effect of IPTG concentration on Taq polymerase expression was completed through gel electrophoresis. The results of the gel are shown in **Figure 1** below.

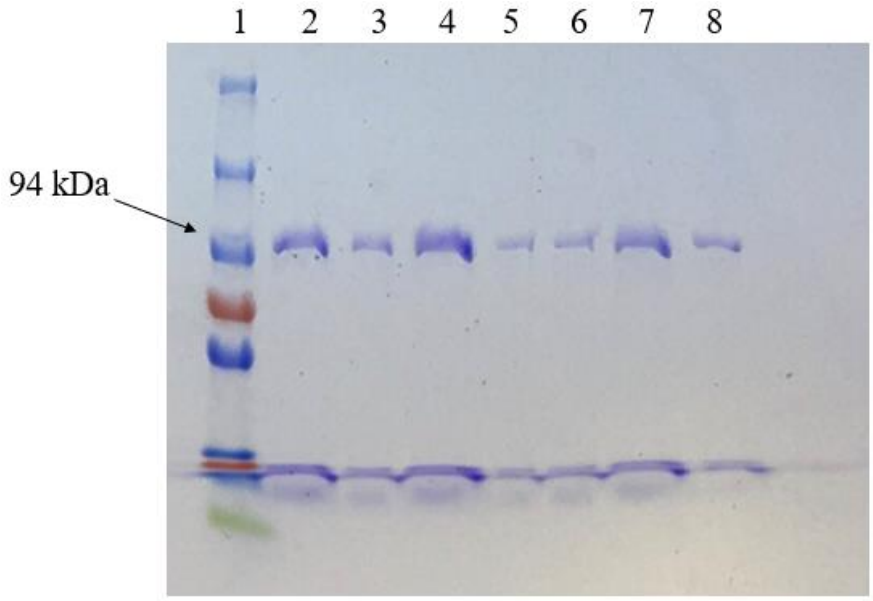

Figure 2. IPTG concentration effect on Taq polymerase gel electrophoresis. Lane 1: Pre-stained protein ladder; Lane 2: 0.08375 mM IPTG; Lane 4: 0.1675 mM IPTG; Lane 6: 0.335 mM IPTG; Lane 7: 0.5025 mM IPTG; Lane 8: 0.67 IPTG

Lanes 3 and 5 are excluded from analysis since they include spillover of the sample from lanes 2 and 4; this error in loading was mentioned in the **Materials and Methods** section. Taq bands are visible in each lane, as shown in their appearance around the 94 kDa molecular weight marker from the pre-stained protein ladder. Since Taq has a molecular weight of 94 kDa, this shows that IPTG did aid in the gene expression for Taq, and thus the overall expression of Taq. Despite the presence of banding for all IPTG concentrations, the band with the highest density appears in Lane 4, which corresponded to an IPTG concentration of 0.1675 mM. Therefore, 0.1675 mM is the optimal IPTG concentration for inducing the highest expression of Taq in a *E. coli.* culture.

This result was unexpected, as it is known that higher concentrations of IPTG induce a greater protein expression until a certain limit is reached. Additionally, this concentration is half of the expected value, as 0.335 mM was previously found to be an appropriate IPTG concentration for induction. It was expected that one of the higher concentrations of 0.335 mM, 0.5025 mM, or 0.67 mM would have induced a higher expression of Taq, and while the gel shows the presence of Taq in those protein solutions, it is significantly lower than the 0.1675 mM induced solution. This result is also significant because it economically optimizes the protein purification protocol. Half of the original volume of IPTG stock is needed to induce a greater protein expression, and thus twice the amount of *E. coli*. + Taq culture solutions, as well as more Taq overall, can be produced. Ultimately, the cost of producing Taq is reduced, as a lower volume of Taq is needed, and the cost to consumers can be reduced.

This result extends to all protein drug production. Since IPTG impacts the *lac* operon in *E. coli.*, and recombinant bacterial cells are produced and cultured in industry similar to the method used in this project, it is expected that the optimal IPTG concentration determined through this project is suitable for all protein drug production. However, the induction time can be specific to certain proteins, so the time of induction may need to be adjusted for different therapeutic proteins [32]. As mentioned for Taq, the cost of production for protein drugs can be reduced with the optimal IPTG concentration of 0.1675 mM, and thus the cost to consumers can be decreased. This would increase the accessibility of protein drugs and enable more people to receive treatment, especially because pharmaceutical companies have been increasing the prices of drugs, such as insulin. Over the past fifteen years, the price of insulin has tripled; reducing the cost of production would both increase supply and decrease the price of the drug for consumers [33]. Overall, both a

lower IPTG concentration and increased protein expression allow for increased protein drug production at a lower cost.

Flow Rate

After determining the optimal IPTG concentration of 0.1675 mM, the optimally induced protein sample was collected at three flow rates. The results of gel electrophoresis for all three flow rates are shown and discussed below.

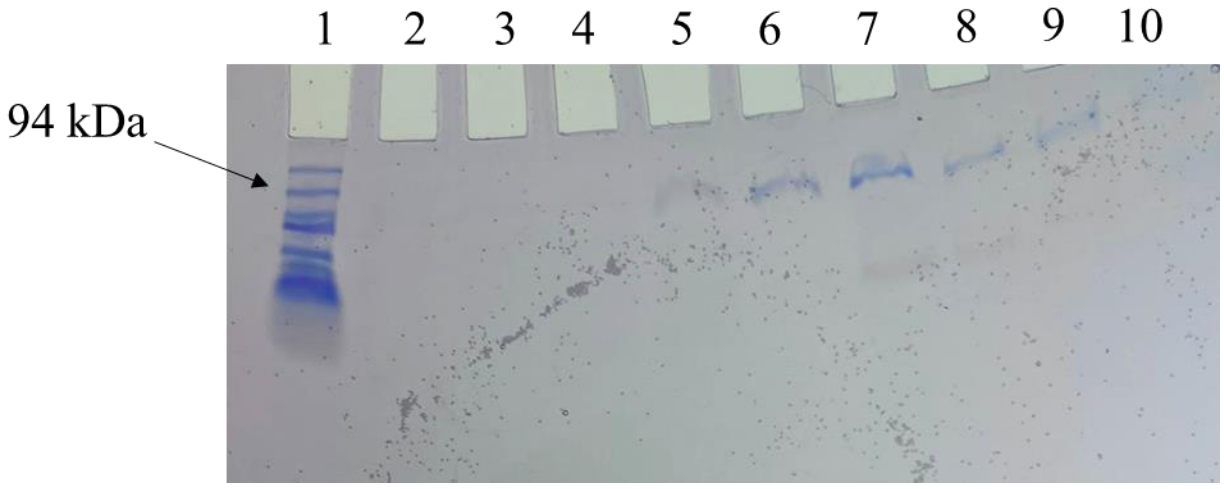

Figure 3. 0.5 mL/min flow rate Taq polymerase collection gel electrophoresis. Lane 1: Pre-stained protein ladder; Lane 2-10: Fractions 1 through 9, respectively.

As seen in **Figure 3**, fractions 4, 5, 6, 7, and 8 all produced light banding for presence of Taq with fraction 6 in Lane 7 showing the greatest banding density. This shows that Taq was the most concentrated in fraction 6 for the 0.5 mL/min flow rate. However, for protein purification, it is desired that the target protein is concentrated in a low number of fractions, typically three. In this case, Taq was collected and spread out amongst five different fractions. This result was unexpected, as a slower flow rate results in slower protein elution from the column and increased column efficiency, thus concentrating the protein into a smaller number of fractions [34].

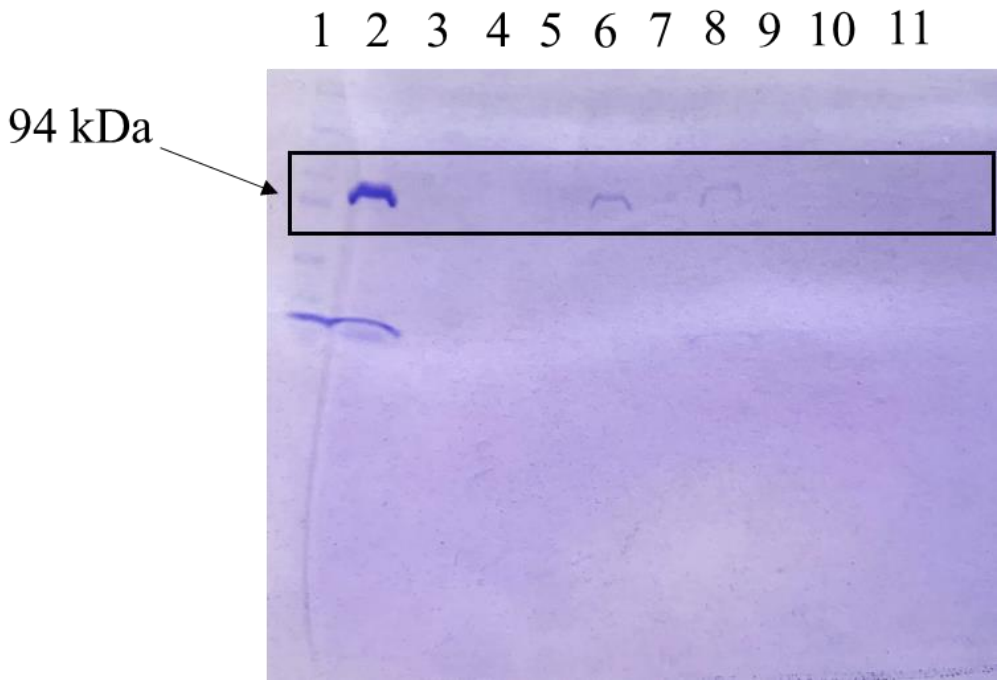

Figure 4. 1 mL/min flow rate Taq polymerase collection gel electrophoresis. Lane 1: Pre-stained protein ladder; Lane 2: Taq Polymerase Crude Sample Control; Lanes 3 through 11: Fractions 1 through 9, respectively.

Since the 1 mL/min flow rate was the first to be tested, a Taq polymerase control was inserted into well 2 to ensure that the eluted protein in the nine fractions was Taq, and that the gel was accurate. In **Figure 4**, banding for the collected Taq is visible in Lanes 6, 7, and 8, corresponding to fractions 4, 5, and 6, respectively. Thus, Taq is more concentrated for a flow rate of 1 mL/min, being present in only three of the nine fractions. However, the band density is significantly lower, showing that there is a lower expression of Taq in the three fractions compared to some of the fractions for the 0.5 mL/min flow rate. Since 1 mL/min is a commonly used flow rate for ion gradient chromatography, these results were expected to some extent. Too low of a flow rate is not beneficial to column chromatography, and 0.5 mL/min may have been too low in the case of Taq polymerase, thus making 1 mL/min a more optimal flow rate.

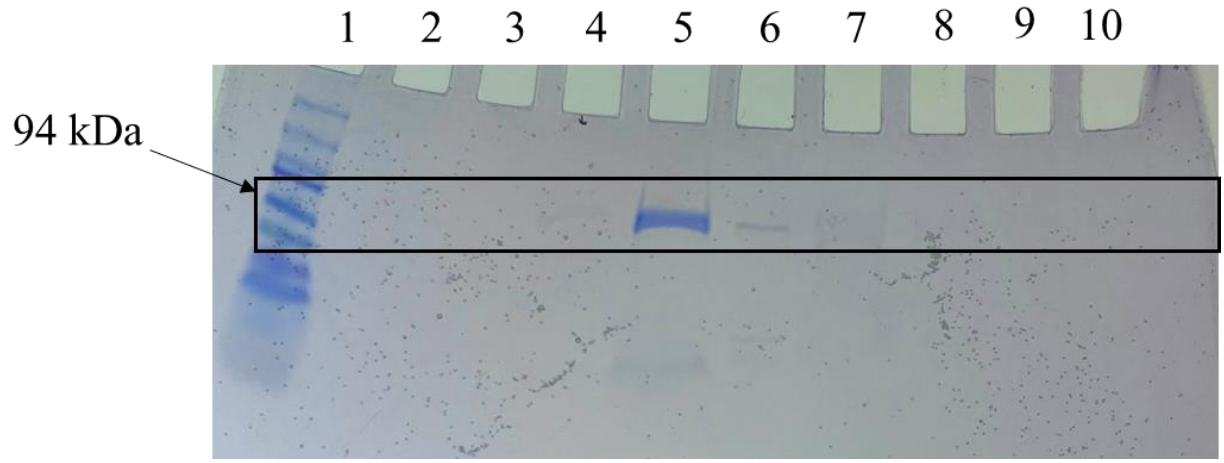

Figure 5. 2 mL/min flow rate Taq polymerase collection gel electrophoresis. Lane 1: Pre-stained protein ladder; Lane 2-10: Fractions 1 through 9, respectively.

In **Figure 5**, Taq banding is present in Lanes 5, 6, and 7, corresponding to fractions 4, 5, and 6, respectively. However, the banding for fraction 4 is greater than the other fractions and is the greatest band amongst all flow rate gels, showing that Taq has the highest expression and is the most concentrated in fraction 4 for the 2 mL/min flow rate. This gel produced a desired result, as Taq is highly expressed and concentrated in a low number of fractions. Being concentrated in a low number of fractions is crucial for drug production, as a lower volume of the collected sample needs to undergo dialysis and PCR to determine its enzymatic activity. Also, a highly concentrated protein is desired, as it is purer and encourages the protein to maintain its proper shape in the solution [35].

Therefore, it was discovered that a column flow rate of 2 mL/min is the optimal flow rate for isolating Taq polymerase. This was unexpected, since higher flow rates are known to reduce protein binding and column efficiency, since the protein elutes from the column too quickly [36]. Typically, this results in the protein being spread out amongst many fractions, yet that was not the case for Taq polymerase. Perhaps, 2 mL/min is not too high of a flow rate to negatively affect the column chromatography process, and more flow rates need to be tested to determine the limit.

Regardless, it has been concluded that 2 mL/min is the optimal flow rate for the flow rates that were tested in this project.

Once again, it is expected that this flow rate is suitable for all proteins that are purified for drug production; however, due to the different molecular weights and affinities of proteins for the DE-52 resin, different salt concentrations may need to be used in order to achieve the same results. With an optimal flow rate of 2 mL/min, drug production can be optimized, as the protein can be collected in a faster time period, thus enabling more of the drug to be produced in a shorter amount of time. Also, the protein is more concentrated, resulting in a purer and more stable form of the protein, which is required for injection into the human body.

Enzymatic Activity

After determining the optimal flow rate, it was necessary to run the collected Taq solution through PCR to verify the enzymatic activity of the collected Taq. As mentioned, PCR was run on the dialyzed Taq solution obtained from fraction 4 for the 2 mL/min column flow rate. The results of PCR show the reactivity of Taq for the undiluted solution, as well as the 1:10 and 1:100 dilutions. The resulting gel is shown below in **Figure 6**.

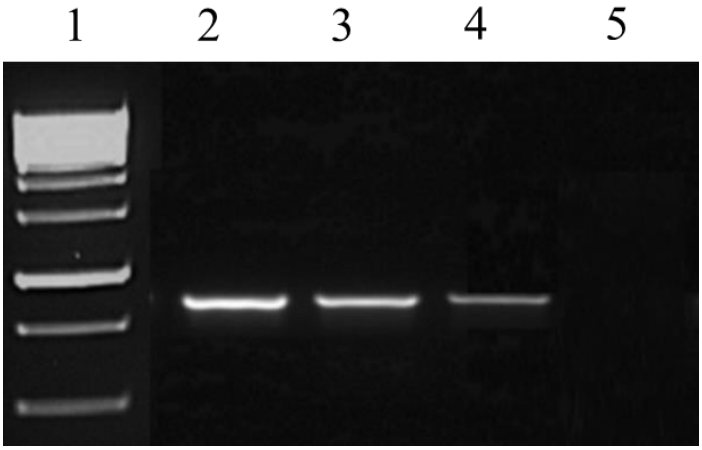

Figure 6. Polymerase Chain Reaction results for 2 mL/min flow rate. Lane 1: Pre-stained protein ladder; Lane 2: Undiluted Taq; Lane 3: 1:10 diluted Taq; Lane 4: 1:100 diluted Taq; Lane 5: H_2O control.

As seen in **Figure 6**, the brightness of the banding for Taq decreases from the undiluted Taq in Lane 2 through the 1:100 diluted Taq in Lane 4. These results are both significant and expected, as the reactivity of Taq should decrease as it is diluted more with water. It was also expected that Lane 5 would show no banding, as it was the H_2O control. This result is especially significant, because it shows that the collected Taq was pure and the collected Taq was able to achieve high reactivity.

The number of reactions that were achieved with the dialyzed Taq was calculated as well. Given that 2.7 mL of the pure Taq solution was obtained from dialysis, 1 μL was inserted into the wells for PCR, and banding was present for the 1:100 dilution, the number of reactions obtained from the pure Taq is:

$$Number\ of\ reactions = \frac{2.7\ mL}{1\ \mu L} \times \frac{1000\ \mu L}{1\ mL} \times 100 = 270{,}000\ reactions$$

This value is indicative of the importance of the optimizing the protein purification protocol. 270,000 reactions is a significantly large number of reactions to achieve from the original cell culture. This further optimizes the protein purification protocol in terms of the production cost. A high number of reactions are achieved at a low production cost using the 2 mL/min flow rate, because the protein shows activity for the 1:100 dilution. Although it was not tested, the protein could possibly still show reactivity for a 1:1000 dilution, which would increase the number of reactions and further decrease the production cost for protein drug production.

Ultimately, it has been determined that both an IPTG concentration of 0.1675 mM and column flow rate of 2 mL/min optimizes the protein purification protocol. The protocol is economically optimized as a lower IPTG concentration, and thus a lower stock volume, is needed, and is optimized in terms of drug production. These optimized parameters result in a greater expression of Taq, a purer and more concentrated form of Taq, and reduced collection time. This

not only benefits the pharmaceutical companies producing the therapeutic protein, but also the consumers, as the price of the drug can be reduced, and the drug is more viable for injection.

Conclusion

This project was successfully able to optimize the protein purification protocol in terms of IPTG concentration and flow rate for ion gradient chromatography. As mentioned, there are more parameters involved in the protein purification protocol that can be tested and optimized in order to further improve collection and isolation of proteins needed for drug production. Since a few errors were made during the research process, it would be beneficial to run through the entire methodology again in order to confirm the results of this project. Nonetheless, this project produced significant results that can be implemented in the pharmaceutical development industry. Since the methodology is not specific to Taq polymerase, the optimized parameters can be used for all protein drug production, thus establishing a standard protocol to be used in the pharmaceutical industry. As a result, both producers and consumers benefit, as the production cost is reduced, the production quantity is increased, the protein characteristics are improved, and the price to consumers is reduced.

Acknowledgements

I would like to thank my mentor, Dr. Shailesh Lal, as well as Dr. Gerard Madlambayan, fellow students of BE 4999, and the Oakland University Honors College for help in completing this research project.

References

[1] S. S. Usmani et al. “THPdb: Database of FDA-approved peptide and protein therapeutics,” *PloS One*, vol. 12, no. 7, 2017.

[2] D. Dimitrov, “Therapeutic Proteins,” *Methods in Molecular Biology*, vol. 899, pp. 1-26, 2012.

[3] B. J. Bruno, G. D. Miller, and C. S. Lim, “Basics and Recent Advances in Peptide and Protein Drug Delivery,” *Therapeutic Delivery*, vol. 4, no. 11, pp. 1443-1467, 2013.

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

[5] G. Denomme, M. Rios, and M. Reid, “Part 2 – General Considerations,” Molecular Protocols in Transfusion Medicine, Academic Press, pp. 9-18, 2000. [E-book].

[6] 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.

[7] L. I. Patrushev et al., “Klonirovanie gena termostabil'noĭ DNK-polimerazy Thermus aquaticus YT1 i ego ékspressiia v kletkakh Escherichia coli [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.

[8] EMBL Protein Expression and Purification Core Facility, “Cloning: Transformation of Competent *E. coli.* Cells with Plasmid DNA,” *European Molecular Biology Laboratory*, 2021. [Online]. Available: https://www.embl.org/sites/heidelberg/ [Accessed: Nov. 19, 2021].

[9] G. M. Cooper, The Cell: A Molecular Approach, 2nd ed. Sunderland, MA: Sinauer Associates, 2000.

[10] J. M. Berg, J. L. Tymoczko, L. Stryer, “Section 31.1, Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons,” in Biochemistry, 5th ed. New York: W H Freeman, 2002.

[11] P. R. Rizkia, S. Siliban, K. Hasan, D. S. Kamara, T. Subroto, S. Soemitro, and I. P. Maksum, “Effect of Isopropyl β-D-1-thiogalactopyranoside Concentration on Prethrombin-2 Recombinan Gene Expression in Escherichia Coli ER2566,” Procedia Chemistry, vol. 17, pp. 118-124, 2015.

[12] C. Shene, A. Lucero, B.A. Andrews, J.A. Asenjo, “Mathematical Modeling of Elution Curves for a Protein Mixture in Ion Exchange Chromatography and for the Optimal

Selection of Operational Conditions," Biotechnology and Bioengineering, vol. 95, no. 4, pp. 704-713, 2006.

[13] Cytiva, Ion Exchange Chromatography Handbook, Cytiva, 2021.

[14] K. Elbing and R. Brent, "Recipes and Tools for Culture of *Escherichia coli*," *Current Protocols in Molecular Biology*, vol. 125, no. 1, 2019.

[15] J. Xu, A. Banerjee, S. Pan, and Z. J. Li, "Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains," *Protein Expression and Purification*, vol. 83, no. 1, pp. 30-36, 2012.

[16] A. Marbach and K. Bettenbrock, "lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA," *Journal of Biotechnology*, vol. 157, no. 1, pp. 82-88, 2012.

[17] P. F. Hu et al., "Study on Characteristics of Chemokine CXCL10 Gene Cloned from cDNA Expression Library of Ujumqin Sheep," *BioMed Research International*, vol. 2013, no. 10, 2013.

[18] V. H. Rees and D. J. R. Laurence, "The Correspondence with Beer's Law for the Optical Density of Stained Protein Patterns on Filter Paper as a Function of Surface Protein Concentration," *Clinical Chemistry*, vol. 1, no. 5, pp. 329-344, 1955.

[19] Norgen Biotek Corp., Appl. Note 6.

[20] BIOpHORETICS, "DEAE 52 cellulose (DE52)," *Biophoretics*. [Online]. Available: https://biophoretics.com/ion-exchange-media/cm-deae-ion-exchange-media-/ionsep-deae-52cellulose-de52 [Accessed Nov. 16, 2021].

[21] MilliporeSigma, "Salt Selection and Buffer Preparation," *Merck KGaA*, 2021. [Online]. Available: https://www.sigmaaldrich.com/US/en [Accessed Nov. 16, 2021].

[22] D. Koehler, Thermo Fisher Scientific, "Principles of fraction collection using the Vanquish UHPLC systems," Tech. Note TN72940-EN 0519, 2019.

[23] 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.

[24] Bio-Rad, *Microbial Culturing Kit*, Bio-Rad Laboratories. [Online]. Available: https://www.bio-rad.com/webroot/web/pdf/lse/literature/1665022.pdf [Accessed: November 12, 2021].

[25] 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 [Accessed Oct. 4, 2021].

[26] H. Ji, "Lysis of Cultured Cells for Immunoprecipitation," Cold Spring Harbor Protocols, vol. 8, 2010.

[27] S. Fjelstrup et al., "The Effects of Dithiothreitol on DNA," Sensors (Basel, Switzerland), vol. 17, no. 6, 2017.

[28] C. R. Cooper, "Basic Culture Technique: Streak Plate," *Youngstown State University Microbiology Laboratory*, 2020. [Online]. Available: http://crcooper01.people.ysu.edu/microlab/streak-plate.pdf [Accessed: Nov. 12, 2021].

[29] Fisher Scientific, "How to Choose the Right Protein Ladder," *Fisher Scientific*. [Online]. Available: https://www.fishersci.com/us/en/home.html [Accessed: Nov. 12, 2021].

[30] O. Coskun, "Separation Techniques: Chromatography," *Northern Clinics of Istanbul*, vol. 3, no. 2, pp. 156-160, 2016.

[31] G-Biosciences, *Molecular Biology Handbook & Selection Guide*, G-Biosciences.

[32] Biologics International Corp, "IPTG Induction Protocol," *Biologics International Corp*, 2018. [Online]. Available: https://www.biologicscorp.com/blog/iptg-induction-protocol/#.YaA6f2DMLIU. [Accessed: Nov. 24, 2021].

[33] The Endocrine Society, "Increasing Insulin Affordability," *Endocrine Society*, Jan. 2021. [Online]. Available: https://www.endocrine.org/ [Accessed Nov. 24, 2021].

[34] R. A. Bernhard, "Effect of Flow-Rate and Sample-Size on Column Efficiency in Gas-Liquid Chromatography," *Nature*, vol. 185, pp. 311-312, 1960.

[35] B. Bibel, "Protein concentrating…and waiting…," Aug., 2020. [Online]. Available: https://thebumblingbiochemist.com/365-days-of-science/protein-concentrating-and-waiting/ [Accessed: Nov. 24, 2021].

[36] A. Astefanei, I. Dapic, and M. Camenzuli, "Different Stationary Phase Selectivities and Morphologies for Intact Protein Separations," *Chromatographia*, vol. 80, no. 5, pp. 665-687, 2017.

Appendix: Generalized Protein Purification Protocol for Taq Polymerase