Optimization Protocol for Extraction and Purification of Recombinant Protein Modeled with *Taq* Polymerase

Submitted by

Lily Tir

Bioengineering

To

The Honors College

Oakland University

In partial fulfillment of the

requirement to graduate from

The Honors College

Mentors: Gerard Madlambayan Ph.D. and Shailesh Lal Ph.D.

Department of Biological Sciences and Bioengineering

Oakland University

April 2, 2021

Abstract

This experiment explores recombinant protein drug protocols and how to better optimize and purify large scale recombinant proteins expressed in bacterial cells. Using *Escherichia coli* bacteria, the gene *Taq* will be cloned into the bacterial plasmid, grown, induced, and incubated . *Taq* proteins will then be transcribed, lysed out of cells, and purified using an ion exchange chromatography column. By altering elements of this current process — varying the column flow rate and using different salt gradient concentrations for column purification — more concentrated yields were achieved with greater purity. It was found that lower flow rates of 100 mL/hour at normal salt gradient concentration of 25 to 700 mM produced the most optimal protein fractions. While *Taq* polymerase is not a recombinant protein drug, it can serve as a model to represent therapeutic protein drugs and purification optimizations that would work well for both. Utilizing these results, they can potentially provide the pharmaceutical industry novel ways to increase their production while lowering costs. Drug manufacturers would be able to produce these drugs at higher rates and more affordable prices, while saving on resources and time. Patients purchasing recombinant protein drugs would benefit from this as drugs become more widely accessible and reasonably priced. Optimization of protein purification can overall ease economic burdens, and initiate change from typical protocols.

Table of Contents

Introduction (Current Research)

There are over one hundred therapeutic protein drugs currently on the market, many of which are produced using prokaryotic bacteria cells (Tripathi & Shrivastava, 2019). This works by splicing a DNA fragment that contains a gene of interest and inserting it into a bacteria plasmid. Once it is in the plasmid — which is now known as recombinant DNA — it will be introduced back into the bacteria cell and grown in a broth medium. An inducer will then be added to help rapidly express the recombinant proteins, which then get purified and turned into a protein drug (Jia & Jeon, 2016). Many of these therapeutic protein drugs are necessary to treat various diseases: insulin for diabetes, erythropoietin for anemia, granulocyte colony stimulating factors for neutropenia, and factor VIII for hemophilia A, just to name a few. Depending on their pharmacological activity, these drugs might replace deficient or abnormal proteins, augment existing protein pathways, provide novel biological functions and activities, interfere with other biological molecules, or even transport and deliver other proteins successfully (Dimitrov, 2012).

Despite having such a high value in prolonging overall life quality, therapeutic protein drugs continue to be one of the highest priced drugs on the market. As of 2017, over $100 billion dollars in gross sales have gone to these protein drugs, which has created and continues to create a huge economic burden in the United States (Dimitrov, 2012). With high cost-sharing levels, out-of-pocket spending, and copayment levels, patients are less likely to initiate drug therapies, have low adherence rates, and are more likely to discontinue use. Much of the increased spike in pricing is due to the pharmaceutical industry's lack of transparency, however, these drugs are also difficult to produce in optimized and highly purified quantities (Hartung, 2017). To help

bring economic relief to the patient as well as to society, advancements and improvements to the recombinant protein drugs may be necessary.

Within the last six years, various tools have been used to upgrade protein production: using different *Escherichia coli (E. coli)* bacteria strains that have faster cell growth, designing the recombinant plasmid to facilitate purification of the protein, and even cultivating the bacteria cells in more closely monitored and controlled environments (Rosano et al., 2019). However, there is always opportunity for improvements in proliferation and purification to continue to optimize the drug.

For purposes of finding novel optimization routes for recombinant protein drugs, the gene *Taq* will be experimented with in the lab. It will be placed in *E. coli* plasmids, proliferated, lysed out, and purified using new protocols and methods. By altering the current methods of protein purification, higher yields may be produced with higher purity, better efficiency, and lower costs. This would not only help the pharmaceutical industry produce more drugs, but would also alleviate patient and care providers' costs. As a result, there would be less competition for the drug, fewer qualifications needed to be met which allows for easier access to protein therapies, and out-of-pocket spending would significantly decrease. Overall, this would greatly benefit many patients, help lift the economic burden, and create a backbone to better develop and optimize other recombinant protein drugs.

Materials

A multitude of buffers and media were required to carry out the experimental procedures needed to isolate and purify *Taq* polymerase. Below describes the function of each as well as how they were prepared.

Buffers

Buffer A served as an equilibrator that was used once the cells were pelletized. First, tris-Cl, or tris(hydroxymethyl) aminomethane, helps maintain a consistent pH as cells lyse (Tan, 2019). Conditions within the buffer solution change as cells break down and DNA pours out. However tris-Cl stabilizes the pH to prevent destruction of DNA. It also interacts with the lipopolysaccharide to destabilize the membrane, assisting EDTA, or ethylenediaminetetraacetic acid, in reducing the integrity of the cellular membrane. EDTA is a metal-chelating agent which interacts with the calcium and magnesium ions in the cellular membrane to form chelates. This destabilizes the membrane but also inhibits metallo protease activity (Mónico et al., 2017). Protease is an enzyme that hydrolyzes peptide bonds. Preventing protease activity allows proteins within the cell to be maintained (López-Otín & Bond, 2008). Dextrose helps stabilize lysosomal membranes and also assists in reduction of protease production and activity. Buffer A was mixed with the specific concentrations and amounts of the solutions listed below in Table 1. Then, 47.3 mL of deionized water was added to reach a total volume of 50 mL.

Table 1: *Constituents of Buffer A (50 mL)*

Solution	Concentration	Amount
Tris-Cl pH 7.9	50 mM	2.5 mL
Dextrose	50 mM	0.9 g
EDTA	1 mM	0.2 mL

Buffer B primarily served to wash, stabilize, and straighten the proteins. As explained above in Buffer A, tris-Cl and EDTA continue to maintain pH of the solution and inhibit protease

activity, respectively. PMSF, or phenylmethylsulfonyl fluoride, is also a protease inhibitor that specifically prevents the activity of serine proteases (“Protein Purification,” n.d.). Tween 20, or polyoxyethylene sorbitan monolaurate, is a non-ionic surfactant used as a water-in-oil emulsifier (Eskandani et al., 2013). It is a mild detergent with the ability to break protein-lipid and lipid-lipid associations without denaturing or breaking down proteins (Johnson, 2020). NP-40, or nonidet P-40, is a non-ionic, non-denaturing, lipophilic detergent which helps to solubilize membrane proteins and isolate cytoplasmic proteins (Sandlin et al., 2011). Finally, KCl increases DNA polymerase activity allowing for increased transcription and translation of genes and proteins. Buffer B was mixed at the specific concentrations and amounts of solution listed below in Table 2. Then, 18.04 mL of deionized water was added to reach a total volume of 20 mL.

Table 2: *Constituents of Buffer B (20 mL)*

Solution	Concentration	Amount
Tris-Cl	10 mM	20 µL
KCl	50 mM	1.0 mL
EDTA	1 mM	40 µL
PMSF	1 mM	100 µL
Tween 20	0.5%	400 µL
NP-40	0.5%	400 µL

Buffer C served as an equilibrator during column chromatography. The constituents EDTA, PMSF, Tween 20, and NP-40 maintain the same functions and qualities as stated above in Buffer B. DTT, also known as dithiothreitol or Cleland’s reagent, is a reducing agent that helps

prevent oxidation of intracellular proteins (Wingfield, 2001). This also prevents protein aggregations and helps stabilize enzymes ("Protein Purification," n.d.). HEPES, or 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), is one of Good's zwitterionic buffers which helps control and maintain pH of the solution ("National Center for Biotechnology," 2020). The HEPES solution was created first to ensure it had the designated pH of 7.9 utilizing pH meter. Buffer C was then mixed at the specific concentrations and amounts of solution listed below in Table 3.

Table 3: *Constituents of Buffer C (200 mL)*

Solution	Concentration	Amount
HEPES pH 7.9	20 mM	200 mL
EDTA	1 mM	0.8 mL
PMSF	0.5 mM	0.5 mL
Tween 20	0.5%	1.0 mL
NP-40	0.5%	1.0 mL
DTT	5 mM	0.154 mL

Running buffer was used during electrophoresis with the SDS-PAGE gel. It provides a liquid medium for pipetting samples into the wells, and a liquid buffer during the electrophoresis. Glycine is an amino acid that is normally negatively charged in the slightly basic running buffer, causing glycinate anions to form (Archuleta, 2020). As the electric field is applied, the glycinate anions meet the slightly acidic gel layer and become neutrally charged glycine zwitterions. Since these ions are neutral, they move slowly through the first layer of the gel and help line up all the

proteins. Then once the zwitterions reach the second slightly basic layer of gel, they become negatively charged again and move quickly down the gel. This helps the protein move through the SDS-PAGE gel and they stop once the gel pores are too small for them to go through. Tris base is used as a good biologically neutral buffer ("Tris Buffer", n.d.). SDS, or sodium dodecyl sulfate, is used to help unfold proteins into more linear structures so they move down the gel more easily (Archuleta, 2020). Additionally, it binds to all the positive charges on the protein to create an overall net negative charge to also help the protein move through the gel. The running buffer was prepared by adding the glycine and tris base, as described below in Table 6, to 1000 mL of deionized water in a large flask and put on a stir plate to dissolve. Once dissolved, an additional 1000 mL of deionized water was added alongside the SDS, as described below in Table 4. This was once again mixed until fully dissolved.

Table 4: *Constituents of Running Buffer (2000 mL)*

Material	Amount
Glycine	28.8 g
Tris Base	6.05 g
SDS	2.0 g

Dialysis buffer was used to dialyze the final fractions to perform PCR. The tris-Cl and EDTA maintain the same functions and qualities as described in Buffer A, and similarly for DTT as described in Buffer C. The NaCl provides a small salt concentration in the buffer to ensure osmosis does not occur too quickly in the chances that the sample has a high salt concentration ("Dialysis Methods", n.d.). Triton X-100 is used as a detergent, similar to NP-40, to help

solubilize the membrane proteins while still maintaining protein activity (“Remove Detergent”, n.d.). Lastly, the glycerol is added as a low weight osmotic agent to help transport small solutes and decrease protein loss (Smit et al., 2000). Dialysis buffer was made by combining all of the solutions in the concentrations and amounts described below in Table 5, however only adding 1000 mL of glycerol, into a large flask. This was then put on a stir plate to mix until everything was combined. Then the additional 1000 mL of glycerol was added and again stirred until combined.

Table 5: *Constituents of Dialysis Buffer (2000 mL)*

Solution	Concentration	Amount
Tris-Cl	1 M	100.0 mL
NaCl	100 mM	1.0 mL
EDTA	0.1 mM	1.0 mL
1% Triton X-100	1M	20.0 mL
DTT	1M	0.308 mL
Glycerol	--	2000 mL

Broth

Luria Broth (LB broth) was used as the growth medium for the bacteria cells. Tryptone is an enzymatic digest of bovine casein which provides a nitrogen source for higher growth rates (“Tryptone”, n.d.). Yeast extract is a hydrolysate of yeasts which provides essential growth components: nitrogen, carbon, sulfur, vitamin B complex, and other growth factors (“Yeast Extracts”, 2020). NaCl, or sodium chloride, provides the broth with sodium ions that can be

used for transport and osmotic balance (“LB Media”, n.d.). It was prepared by combining the Tryptone, Yeast Extract, and NaCl in the amounts shown below in Table 6. Deionized water was then added to get the solution to a total volume of 500 mL. Once mixed, the solution was autoclaved for 20 minutes. After coming out of the autoclave and cooling down, ampicillin — an antibiotic — was added to a final concentration of 100 μg/mL.

Table 6: *Constituents of LB Broth (500 mL)*

Material	Amount
Tryptone	5.0 g
Yeast Extract	2.5 g
NaCl	5.0 g

Slurry

Diethylaminoethyl-cellulose (DEAE-cellulose) was used in the ion exchange chromatography column. It is a positively charged resin material that allows for binding and releasing of charged proteins. The cellulose material needed to be rehydrated for column loading. To do so the DEAE-cellulose was placed into a flask with Buffer C at the specific amounts listed below in Table 7. Then it was put on a stir bar for overnight mixing.

Table 7: *Constituents of DEAE-cellulose Slurry (100 mL)*

Material	Amount
DEAE-cellulose	30 mL
Buffer C	70 mL

Methodology

In general, there was a basic eight-step process involved in recombinant protein production and purification: growth, induction, pelletization, isolation, column purification, elution, measuring presence, and testing for enzymatic activity. The steps of the default procedure are fairly straightforward with what a typical protocol follows, however many alterations can be made to optimize the procedure and produce the most desired results. For the purposes of this experiment with lab and time constraints, two variables were altered with a total of five different runs.

The process began with growth of a recombinant plasmid transformed colony of *E. coli.* This occured on an agar plate where the selected *E. coli* culture was placed and T-streaked using an inoculating loop within 1 foot of a bunsen burner to sterilize the environment. Agar plates were then incubated overnight at 37 ℃. A single colony was then selected from a streaked plate and inoculated into 500 mL of LB broth and 0.5 mL of ampicillin to ensure enough cells were present to produce the recombinant protein. The resultant media was subjected to growth at 37℃ overnight in an incubator shaker at 225 rotations per minute (rpm).

After the *E. coli* grew overnight, induction occurred to enhance transcription of the recombinant plasmid at the location where *Taq* polymerase is encoded for. Using a micropipette, 335 μL of IPTG was added into the flask. IPTG is a non-metabolizable form of lactose which unbinds a lac repressor from the *E. coli* genome to allow transcription of T7 RNA polymerase (Martin, n.d.). This T7 RNA polymerase specifically binds to the T7 promoter region of the plasmid, which is the location right before the target gene, *Taq* polymerase, is spliced into the plasmid. As a result, after the IPTG was added, induction of transcription began and *Taq*

polymerase proteins were formed. The flask was then put into a shaking incubator at 220 rpm to continue growth for 24 hours at 37°C.

After, the cells were placed in a 300 mL centrifuge bottle to be pelletized in a centrifuge at 10,000 rpm for 10 minutes at 4℃ to begin concentrating the *E. coli* cells from the LB broth. The excess liquid was poured off and the pellet was resuspended in 20 mL of Buffer A. Once again this was spun down for 10 minutes at 10,000 rpm to finally concentrate all the *E. coli* cells. The excess liquid was poured off again and the centrifuge bottle with the pellet was placed in the freezer to be used when ready.

Once the pellet was ready to be used, it was resuspended in 10 mL of Buffer A, with 4 mg/mL of added lysozyme. Incubation for 15 minutes allowed the lysozyme to lyse the cells, or break down the cell wall of the bacterial cells. 10 mL of Buffer B was then added and the cells were further incubated for 1 hour at 75°C. During this incubation period, Buffer B worked to stabilize the solution. After 1 hour, the solution was spun down for 15 minutes at 10,000 rpm in order to separate the protein from the other components released from the broken down cells. The resulting supernatant was the raw protein lysate, containing all the proteins from the lysed cells. At this point the protein was isolated in the supernatant and was saved and kept on ice from this point on.

An anion exchange chromatography column was used for purification. The specific column used was a DE-52 column, containing the DEAE-cellulose. The column was packed with approximately 25 mL of DEAE-cellulose slurry and washed with 590 mL of Buffer C containing 25 mM KCL to equilibrate. Next, the supernatant was added and allowed to run through the column, allowing any negatively charged proteins to bind to the Cl^- counter ion within the

DEAE-cellulose. Washing again with Buffer C allowed excess unbound proteins to run out, leaving only negatively charged proteins and the DEAE-cellulose in the column.

Elution then began with an 80 mL KCl salt gradient in Buffer C flowing into the column. Using a two chamber gradient maker, incorporation of the KCl created a linear gradient that increased the ionic strength within the column. This increased strength causes bound proteins to be in competition with K^+ ions— both of which will try to bind to the Cl^- counter ion. Once enough KCl was added and *Taq* polymerase's specific net charge was reached, the *Taq* polymerase unbound from the column and eluted out into 15 collection tube fractions, each containing 6 mL. The 15 collected fractions contained some *Taq* polymerase as well as other proteins that were eluted off of the column. Cleaning of the column occurred after each run to ensure the column would be ready for the next run. 50 mL of NaOH was pumped through the column to help unbind, solubilize, and wash out any excess proteins ("Use of Sodium Hydroxide", 2001). Then an additional 50 mL of Buffer C was pumped through to re-equilibrate the column.

In order to test for the presence of *Taq* polymerase in each fraction, a 10% polyacrylamide separating gel with a 4% stacking gel was run. In order to do this, fractions 4-12 were prepared by combining the sample, DTT, LSB, and deionized water to make a 30 µL solution. The mixtures were then incubated at 100°C for 5 minutes in a thermocycler. A polyacrylamide gel was then prepared with running buffer and placed into an electrophoresis gel box. 20 µL of each sample was pipetted into wells 2-10 and a 5 µL of a ladder was pipetted into well 1. The electrophoresis machine was then run at 150 Volts for 30 minutes. After running, the

SDS-PAGE gel was imaged using a Chemi-doc under the stain-free blot setting. The dark bands representing 94 kilodaltons showed presence of *Taq* polymerase in that sample.

After all trials were run, the fractions with the most concentrated protein from the most successful trial were selected to be dialyzed. Dialysis was done to remove low molecular-weight contaminants, exchange buffers, desalt, and to concentrate the sample ("Slide-A-Lyzer," n.d.). To do this, a Slide-A-Lyzer Cassette was prepared by hydrating the membrane in dialysis buffer for 2 minutes. After initial hydration, 12 mL of the selected fractions were added into the Slide-A-Lyzer Cassette using a syringe needle. The cassette was then put into a buoy to float in dialysis buffer for 4 hours at 4°C with a change in the dialysis buffer midway. After dialysis, the sample was removed and tested for enzymatic activity via PCR.

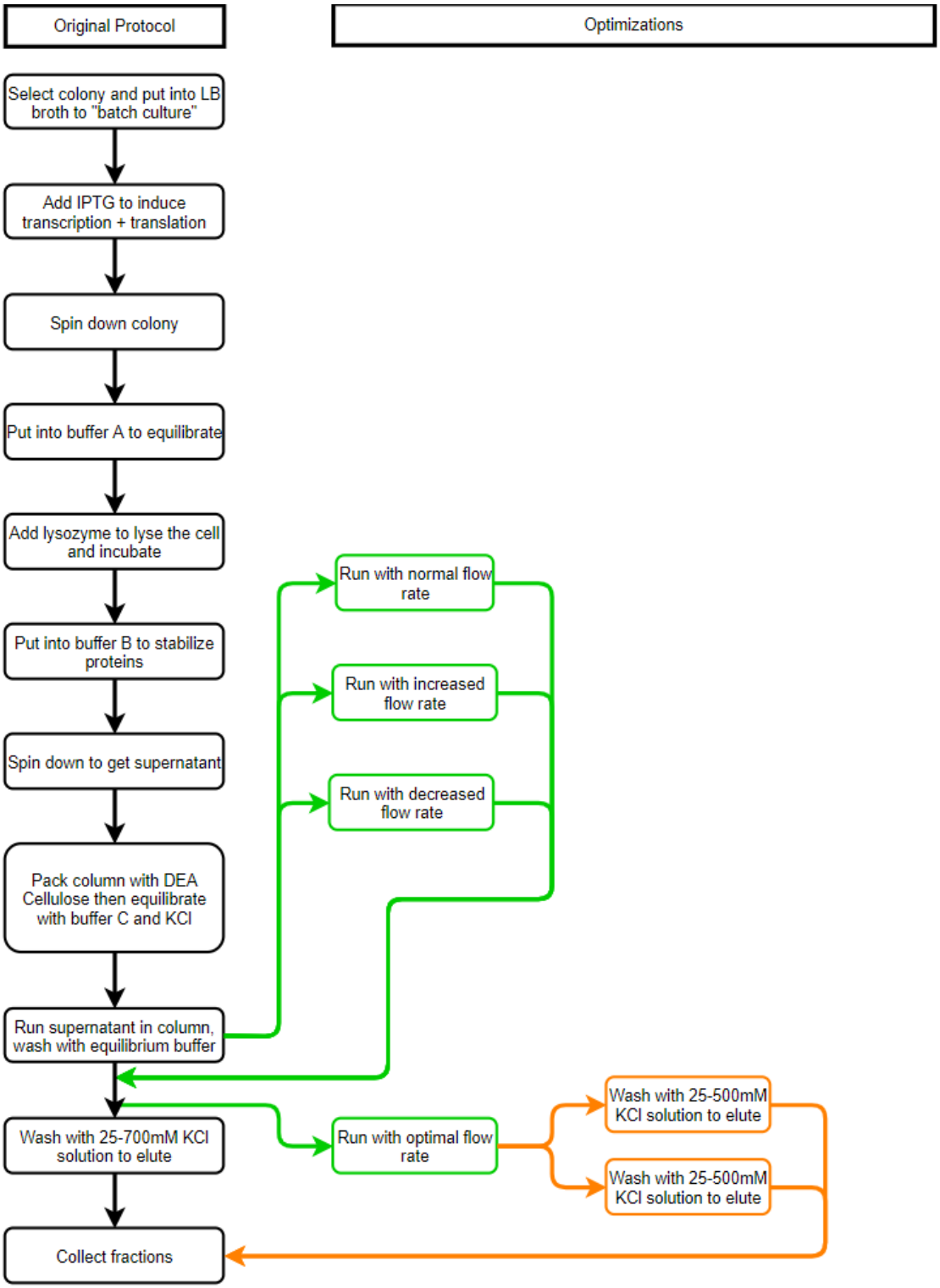
During the elution in step 6, two variables were altered — the flow rate and the KCl salt gradient going into the column. Flow rate can affect the separation efficiency of the column based on the van Deemter equation (Libretexts, 2020). Since the optimal flow rate was not known, 3 different rates — one that is set as a default, one at an increased rate, and one at a decreased rate — were tested. The salt concentration can also affect protein elution. Concentrations too low may not unbind proteins and therefore produce low yield, concentrations too high may denature proteins and cause them to be inactive, however at the right concentration, proteins will be more stable and elute at higher rates (Din et al., 2020). Once the optimal flow rate was found with a typical salt concentration, two other concentration gradients were tested to find the combination that is most ideal. A summary of the experiments conducted are shown below in Table 7 and a flow chart protocol is shown below in Figure 1.

Table 7: *Number of Experiments Corresponding to Variable Changes in Protocol*

	Experiments				
Variable	**1**	**2**	**3**	**4**	**5**
Flow Rates	Normal	Increased	Decreased	Optimal	Optimal
Salt Gradient	25 mM - 700 mM	25 mM - 700 mM	25 mM - 700 mM	25 mM - 500 mM	25 mM - 900 mM

Note. Normal flow rate, increased flow rate, and decreased flow rate are 200 mL/hour, 250 mL/hour, and 100 mL/hour, respectively.

Figure 1: Flow chart for normal protocol and optimizations made in test trials



Results

The outcomes of SDS-PAGE gels and PCR gels provide qualitative results on which trials and optimizations produced the best protein purification protocol.

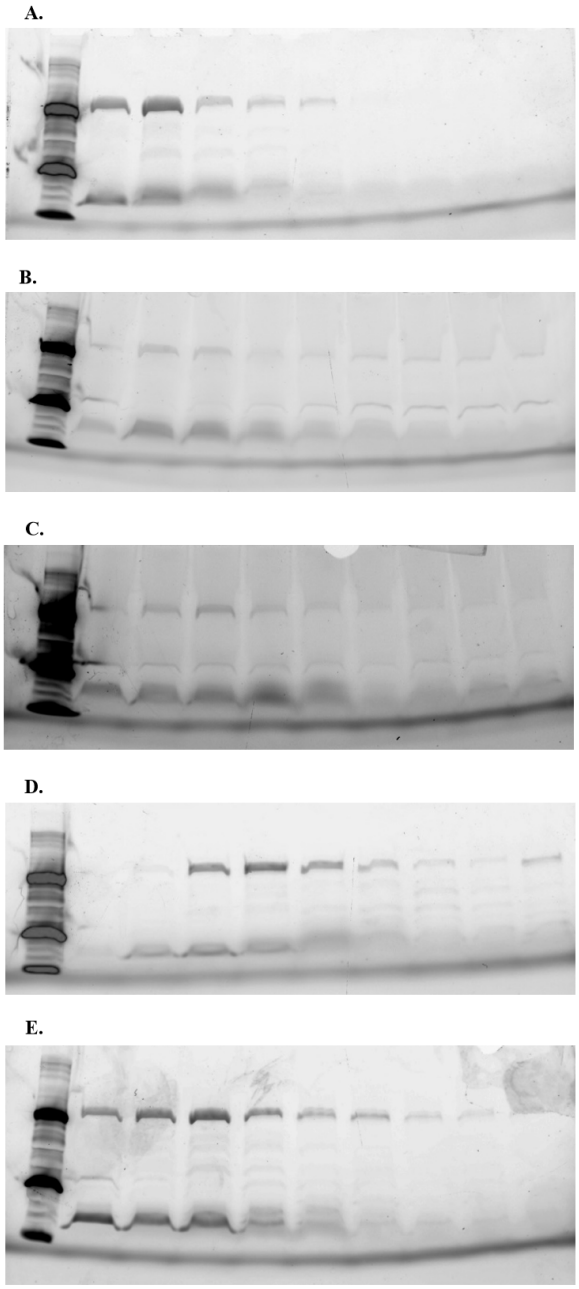
SDS-PAGE Gel Results

The SDS-PAGE gels were imaged using a Chemi-doc and results were saved and compiled into Figure 2 below. Going from top to bottom, each gel shows the results from one of the experimental trials. Figure 2A shows the results for lowering the flow rate applied to the column to 100 mL/hour while keeping the standard KCl salt concentration gradient of 25 mM to 700 mM. The next optimization result shown in Figure 2B was with a nominal flow rate of 200 mL/hour while still keeping the standard 25 mM to 700 mM KCl salt concentration gradient. Figure 2C shows the results for increasing the flow rate to 250 mL/hour while again maintaining the standard 25 mM to 700 mM KCl salt concentration gradient. These three gel results were then compared to determine the optimal flow rate. The final experimental trial conducted was at a flow rate of 250 mL/hour. Fractions 4-12 collected from the trial were then run on an SDS-PAGE gel and the results are shown in Figure 2C. (The gel results from the three flow rate trials were compared and analyzed to determine that the optimized flow rate was 100 mL/hour.)

The last two optimizations used the optimal flow rate, determined as 100 mL/hour, with varying KCl salt concentration gradients. Figure 2D shows the results for optimal flow rate of 100 mL/hour with a 25 mM to 500 mM KCl salt concentration gradient. Lastly, Figure 2E shows the results for optimal flow rate of 100 mL/hour with a 25 mM to 900 mM KCl salt concentration gradient.

For each trial in this study, fractions 4-12 were chosen as the predicted fractions with the most *Taq* polymerase. Elution of the fraction typically does not immediately occur, so it was determined that gel analysis would begin at fraction 4. Due to the limitation in the wells, only 9 samples could be run at once, which is why fraction 12 was the last one used. By maintaining the same 9 fractions, the gels could be easily compared for more accurate results.

Figure 2: SDS-PAGE gel results

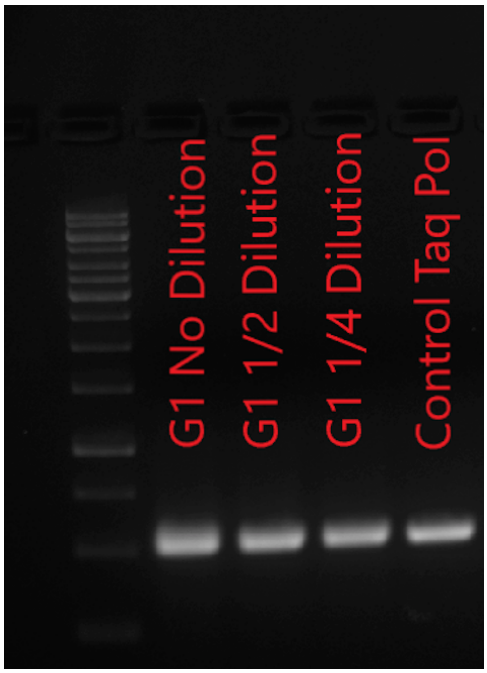


Note. **A.** Results for 100 mL/hour flow rate, 25-700mM KCl; **B.** Results for 200mL/hour flow rate, 25-700 mM KCl; **C.** Results for 250 mL/hour flow rate, 25-700 mM KCl; **D.** Results for 100 mL/hour flow rate, 25-500 mM KCl; **E.** Results for 100 mL/hour flow rate, 25-900 mM KCl

PCR Results

In order to test the activity of *Taq* polymerase collected, the fractions from the trial with the best results were dialyzed and then PCR analysis was performed. Fractions 4 and 5 from the experimental trial with flow rate of 100 mL/hour and concentration gradient of 25 mM to 700 mM were selected as the samples for these tests to be conducted on. These fractions were the most concentrated and pure out of all the trials. The results from PCR analysis are shown below in Figure 3.

Figure 3: PCR gel results



Discussion and Analysis

From the SDS-PAGE gel and PCR gel results, each was analyzed to determine which experiment trial produced the best results. Fractions that contained highly concentrated and pure amounts of the target protein, *Taq* polymerase, showed the most optimized protocol.

SDS-PAGE Gel Analysis

For each gel image, fractions in wells 2-10 were compared to the ladder in well 1. Specifically, analysis was done to look for bands that occurred at 94 kilodaltons, which represented the approximate weight of *Taq* polymerase. This band is represented by the first dark band on the ladder. More bold and prominent bands alluded to highly concentrated *Taq* polymerase present in the fraction, whereas lighter bands hinted at less *Taq* polymerase present in the fraction. Other eluted proteins were also seen as bands further down the gel, which suggested contaminated and less pure fractions.

In order to determine an optimal flow rate, the SDS-PAGE gel images from Figure 2A, YB, and YC, were compared and evaluated. Figure 2A shows the trial at a flow rate of 100 mL/hour, Figure 2B shows the trial at a flow rate of 200 mL/hour, and Figure 2C shows the trial at a flow rate of 250 mL/hour. All three figures show trials conducted at the standard KCl salt concentration gradient of 25mM to 700mM to allow for flow rate comparison. When analyzing Figure 2B and Figure 2C, all the wells show faint and low density bands of *Taq* polymerase -- meaning each fraction contained low amounts of the target protein. This suggests lower, less concentrated protein yield of *Taq* polymerase was collected during both trials. This low protein yield could potentially be a result of the flow rate being too high, as that both trials were at higher speeds. Rather than concentrating the *Taq* polymerase, it seems faster flow rates may

cause the target protein to be more spread out and potentially be more contaminated with other unwanted proteins. Additionally, it has been found that DEAE-cellulose columns may become impacted at high flow rates due to large amounts of pressure being applied by the flow pump ("Ion Exchange Chromatography", 2020). Therefore, the high flow rates in Figure 2B and Figure 2C may have created an impacted column, causing the DEAE-cellulose to be inefficient protein binding and elution. Incidentally, the gels run in Figure 2B and Figure 2C had longer sample preparation durations during electrophoresis. As a result, the *Taq* polymerase could have degraded over time; although *Taq* polymerase is typically thermostable at room temperature, the sample did not contain any stabilizing media to ensure activity was maintained.

The most concentrated and dense bands were found when the experiment was conducted at a flow rate of 100 mL/hour, as seen in Figure 2A. Bold and prominent dark bands were seen in the first two fractions with three subsequent fractions showing lighter bands at 94 kilodaltons -- similar to those seen in Figures YB and Figure 2C. Therefore, the dense bands found in only two fractions in Figure 2A suggest a large and highly concentrated protein yield of *Taq* polymerase was eluted and collected when using a flow rate of 100 mL/hour. It is hypothesized that slower flow rates allow column binding and elution to be more gradual, complete, and efficient which allows protein collection to be more concentrated and successful. Overall, the results of each protein gel conducted at different flow rates highlighted that flow rate does indeed affect the purity of protein elution. It was concluded that the trial conducted at a flow rate of 100 mL/hour (Figure 2A) yielded the most pure *Taq* polymerase and is the most optimized flow rate.

Determination of the optimal salt concentration was conducted by comparing Figure 2A, Figure 2D, and Figure 2E, as all were run at the same flow rate of 100 mL/hour with varying

concentration gradients. Figure 2D and Figure 2E represent results with lower and higher, respectively, concentration gradients and both show that dense fractions of *Taq* polymerase were obtained. However, these dense fractions were spread out over many fractions. Figure 2D shows dense bands of *Taq* polymerase appearing in later fractions. Figure 2D's trial was conducted at a lower salt concentration from the baseline and it is possible that the protein eluted off the column in later fractions because the smaller concentration gradient was not as efficient or quick in the protein elution. Likewise, it has been found that at low flow rates and small salt concentration gradients, stronger bound proteins may not elute ("3.4.3. Ion Exchange", 2019). In comparison, Figure 2E displays that the dense bands begin showing during the earlier fractions. A higher salt concentration from the baseline was conducted in Figure 2C. It is suggested that the higher concentration gradient caused many different proteins to elute out quickly due to the large change in salt concentration. This resulted in the protein being present in the earlier fractions and spread out over many fractions. It has also been found that if the salt concentration gradient is too high that protein purity is diminished (GE Healthcare, n.d.). Therefore, adjustments in the concentration gradient did not increase the protein yield of *Taq* polymerase eluted in comparison to the standard concentration shown in Figure 2A. It determines that the trial from Figure 2A provided the most optimal protocol: adjusting for a flow rate of 100 mL/hour and salt concentration gradient of 25 to 700 mM.

PCR Analysis

After the SDS-PAGE gels were run, dialysis occurred to dialyze and clean the best samples. The experimental trial that had the best results was from Figure 2A, which represents a flow rate of 100 mL/hour and salt concentration gradient of 25 to 700 mM. From this trial, it was

determined that fractions 4 and 5 produced the most concentrated and pure *Taq* polymerase since those fractions had the most prominent bands. Once these two fractions were dialyzed, they were used in a PCR to test that activity of *Taq* polymerase was maintained. Results from the PCR are shown in Figure P.

It can be concluded that the *Taq* polymerase collected in the fractions were active when compared to the control, which used commercially available *Taq* polymerase. The bands from the eluted *Taq* polymerase matched the control *Taq* polymerase band, demonstrating that the prior provided accurate results and maintained proper activity. Additionally, as the eluted *Taq* polymerase fractions were diluted, the bands became more focused with less blurring or fading of the result. This shows more accuracy as the sample concentration decreased, allowing for better results with fewer contaminants present. Overall, the optimizations implemented provided results with concentrated, pure, and active *Taq* polymerase and successfully improved the purification and elution process.

Conclusions

Overall, an optimized *Taq* polymerase extraction and purification protocol resulting in increased protein yield and maintained protein activity was obtained as a result of the experimental optimizations. Although the five trials were not run in triplicates due to time and lab constraints, data extrapolation from each trial still allowed for baseline data and trends to be analyzed. The experimental changes did lead to a new idealized flow rate, however the salt concentration found at this optimum flow rate stayed at the nominal range. The results of this study in relation to the flow rate optimization followed the model proposed by Knox, a modified version of the van Deemter equation. It could be seen that flow velocity through the

DEAE-cellulose column was found to affect separation efficiency and quality of protein isolation. Utilization of this enhanced protocol holds the potential to increase overall yield of mass production of recombinant proteins and to diminish production costs that surround the biopharmaceutical industry. As a result, these changes can trickle down to aid in decreasing drug costs that surround patients of recombinant protein therapies. However, future trials may need to be done in order to confirm findings and show reproducibility.

Recommendations

In addition to the optimizations conducted, a multitude of other potential optimizations to the protein purification protocol were researched and found. However, due to time constraints and laboratory resources they were not able to be conducted. The first example of potential optimization is changing the growing media. During the growth stage of the protocol LB Broth can be switched to Terrific Broth (TB). Key components of TB broth include glycerol, which provides the cells with a carbohydrate source, and phosphates, helping to act as a pH buffer (Kram & Finkel, 2015). This allows TB broth to be richer than LB broth and promote long-term cell survival. By using a broth media that provides better growing conditions, more bacteria proliferation may occur thus leading to higher protein yields (Lessard, 2013).

Another example of a future optimization that could be carried out is washing the raw lysate with Acetone. It has been found that the activity of collected *Taq* polymerase was maintained and as active as commercial *Taq* polymerase when a cold acetone wash was applied to crude supernatant (Kadir et al., 2018). This would therefore be a beneficial additional step to the protocol to potentially help maintain the enzymatic activity of *Taq* polymerase throughout the entire process, which is key for its utilization after purification.

An additional probable expansion would be to heat the fractions collected after chromatography. In comparison to other types of DNA polymerases, *Taq* polymerase is highly thermostable and can maintain its activity at high temperatures, typically around 75°C to 80°C (Din et al., 2020). For other proteins, denaturing can begin at temperatures higher than 41°C (Holme, n.d.). Due to *Taq* polymerase's thermal stability, heating the fractions of protein collected after chromatography has the potential to denature unwanted proteins and maintain the stability of *Taq* polymerase.

Lastly, a general recommendation would be to recreate this experiment and run the trials in triplicates to see if reproduction of results are attainable. As that *E. coli* grows slightly differently every trial, based on the culture chosen or the specific growing conditions, this may have altered some of the final outcomes. Additionally, the process of purification and extraction is quite lengthy. Initial trials had longer run times due to the inefficiency and inexperience of the process; these variances may have caused biased results. By running the experiment for each optimization three times, it would ensure better accuracy and precision of the results.

References

3.4.3. Ion Exchange Chromatography. (2019). LibreTexts https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/CHEM_4320//5320:_Biochemistry_1/03:_Methods_of_Protein_Purification_and_Characterization/3.4:_Chromatography/3.4.3._Ion_Exchange_Chromatography.

Archuleta, A. (2020, November 16). *SDS-PAGE Demystified*. PhosphoSolutions. https://www.phosphosolutions.com/sds-PAGE-demystified/.

Dialysis Methods for Protein Research. (n.d.) Thermo Fisher Scientific - US. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/dialysis-methods-protein-research.html.

Dimitrov D. S. (2012). Therapeutic proteins. Methods in molecular biology (Clifton, N.J.), 899, 1–26. https://doi.org/10.1007/978-1-61779-921-1_1

Din, R. U., Khan, M. I., Jan, A., Khan, S. A., & Ali, I. (2020). A novel approach for high-level expression and purification of GST-fused highly thermostable Taq DNA polymerase in Escherichia coli. *Archives of Microbiology*, *202*(6), 1449–1458. https://doi.org/10.1007/s00203-020-01860-9

Eskandani, M., Hamishehkar, H., & Ezzati Nazhad Dolatabadi, J. (2013). Cyto/Genotoxicity study of polyoxyethylene (20) sorbitan monolaurate (tween 20). *DNA and cell biology*, *32*(9), 498–503. https://doi.org/10.1089/dna.2013.2059

GE Healthcare. (n.d.) Protein purification troubleshooting guide. Retrieved December 1, 2020, from

http://wolfson.huji.ac.il/purification/PDF/Columns/GE_ProteinPurificTroubleshGuide.PDF

GE Healthcare Bio-Sciences AB. (2016). Appendix 5: Converting from flow velocity to volumetric flow rates. In Ion Exchange Chromatography Principles and Methods (p. 145).

Hartung D. M. (2017). Economics and Cost-Effectiveness of Multiple Sclerosis Therapies in the USA. Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics, 14(4), 1018–1026. https://doi.org/10.1007/s13311-017-0566-3

Holme, T. A. (n.d.) *Denaturation*. Chemistry Explained. http://www.chemistryexplained.com/Co-Di/Denaturation.html

Ion Exchange Chromatography (2020). Bio-Rad Laboratories. Retrieved October 13, 2020, from https://www.bio-rad.com/en-us/applications-technologies/ion-exchange-chromatography?ID=MWHAY9ESH

Jia, B., & Jeon, C. (2016). High-throughput recombinant protein expression in Escherichia coli : current status and future perspectives. *Open Biology*, *6*(8), 160196. https://doi.org/10.1098/rsob.160196

Johnson, M. (2020). Detergents: Triton X-100, Tween-20, and More. In *Labome*. Retrieved October 11, 2020, from https://doi.org/10.13070/mm.en.3.163

Kadır, Turan & Eken, B.. (2018). A simplified method for the extraction of recombinant Taq DNA polymerase from Escherichia coli. *Journal of Microbiology, Biotechnology and Food Sciences*. 7. 445-448. 0.15414/jmbfs.2018.7.5.445-448

Kram, K. E., & Finkel, S. E. (2015). Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Applied and environmental microbiology*, *81*(13), 4442–4450. https://doi.org/10.1128/AEM.00722-15

LB Media. (n.d.) Teknova. Retrieved December 2, 2020 from https://www.teknova.com/resource/lb-media/.

Lessard, J. C. (2013, October 31). *Growth Media for E. coli*. https://www.sciencedirect.com/science/article/pii/B9780124200678000118?via=ihub.

Libretexts. (2020). The Van Deemter equation. Retrieved October 19, 2020, from https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Chromedia/01Gas_Chromotography_(GC)/Gas_Chromotography:_Basic_Theory/13The_Van_Deemter_equation

López-Otín, C., & Bond, J. S. (2008). Proteases: multifunctional enzymes in life and disease. *The Journal of biological chemistry*, *283*(45), 30433–30437. https://doi.org/10.1074/jbc.R800035200

Martin, K. (n.d.). A Deep Dive Into Induction with IPTG. In *Gold Biotechnology*. Retrieved October 11, 2020, from https://www.goldbio.com/articles/article/a-deep-dive-into-iptg-induction

Mónico, A., Martínez-Senra, E., Cañada, F. J., Zorrilla, S., & Pérez-Sala, D. (2017). Drawbacks of Dialysis Procedures for Removal of EDTA. *PLOS One*, *12*(1). https://doi.org/10.1371/journal.pone.0169843

National Center for Biotechnology Information (2020). PubChem Compound Summary for CID 23831. Retrieved October 12, 2020 from https://pubchem.ncbi.nlm.nih.gov/compound/hepes

Protein Purification. (n.d.). In *European Medical Biology Laboratory*. Retrieved October 11, 2020, from https://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/lysis_buffer_additives/

Remove detergent from protein samples (n.d.). In *Thermo Scientific*. Retrieved December 2, 2020, from https://proteomicsresource.washington.edu/docs/protocols03/Thermo_TechTip19_Detergent_Removal.pdf

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. Protein science : a publication of the Protein Society, 28(8), 1412–1422. https://doi.org/10.1002/pro.3668

Sandlin, R. D., Carter, M. D., Lee, P. J., Auschwitz, J. M., Leed, S. E., Johnson, J. D., & Wright, D. W. (2011). Use of the NP-40 detergent-mediated assay in discovery of inhibitors of beta-hematin crystallization. *Antimicrobial agents and chemotherapy*, *55*(7), 3363–3369. https://doi.org/10.1128/AAC.00121-11

Slide-A-Lyzer Dialysis Cassettes (n.d.). In *Thermo Scientific*. Retrieved December 2, 2020, from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0011337_SlideALyzer_Dialy_Cass_UG.pdf

Smit, W., de Waart, D. R., Struijk, D. G., & Krediet, R. T. (2000). Peritoneal transport characteristics with glycerol-based dialysate in peritoneal dialysis. *Peritoneal dialysis international : journal of the International Society for Peritoneal Dialysis*, *20*(5), 557–565.

Tan, A. (2018). What Is the Function of a Tris Buffer in DNA Extraction?. In *Sciencing*. Retrieved from https://sciencing.com/function-tris-buffer-dna-extraction-6370973.html

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Frontiers in bioengineering and biotechnology, 7, 420. https://doi.org/10.3389/fbioe.2019.00420

Tris Buffer. (n.d.) Bio. Retrieved 2 December 2020 from https://www.bio-rad.com/featured/en/tris-buffer.html.

Tryptone. (2020). Retrieved 2 December 2020, from https://www.usbio.net/media/T8750

Use of sodium hydroxide for cleaning and sanitizing chromatography media and systems (2001). In *Amersham Bioprocesses*. Retrieved December 2, 2020, from http://www.proteinguru.com/protocols/NaOH%20and%20column%20cleaning.pdf

Wingfield P. T. (2001). Use of protein folding reagents. *Current protocols in protein science*, *Appendix 3*, Appendix–3A. https://doi.org/10.1002/0471140864.psa03as00

Yeast Extracts (n.d.) Sigma. (2020). https://www.sigmaaldrich.com/analytical-chromatography/microbiology/microbiology-products.html?TablePAGE=8657623.