© Copyright 2019

Pei-Rung Wu

Design and Construction of His-tagged Protein Purification Workflow in Aquarium

Pei-Rung Wu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2019

Reading Committee:

Eric Klavins, Chair

Barry Lutz

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Design and Construction of His-tagged Protein Purification Workflow in Aquarium

Pei-Rung Wu

Chair of the Supervisory Committee:

Professor Eric Klavins

Department of Electrical and Computer Engineering

Protein purification is a long process starting from cloning, protein induction, extraction, isolation to protein verification. Carrying out this experiment can be very time-consuming and pricey, especially when some high-end instruments are required. Proper laboratory experiences and background knowledge of protein properties are also crucial for people who are conducting protein purification. As a result, the aforementioned considerations lead to the motivation of designing and building a robust protein purification workflow in Aquarium. The laboratory operating system, Aquarium, will compute required information such as protein concentration, sample amount, and buffer volume to yield step-by-step instructions for laboratory technicians for generating reproducible outcomes. The protein purification using poly-histidine-tag (His-tag) is implemented to provide an inexpensive purifying method. With Aquarium, the protocols are represented as

executable code with negative controls. Moreover, the system automates multiple purifications in batches, which not only significantly saves time and budget, but also elevates the success rate. The success rate is verified by 10 identical purifications which are performed in 2 batches; each batch includes 5 purifications in parallel. This demonstrates that this workflow is capable of processing multiple batches of purification. The results of the test run also demonstrate this workflow is robust. The isopropyl-β-D-thio-galacto-pyranoside (IPTG) induced samples have a significant enrichment on the protein expression by comparing with uninduced samples. The high purity of purified proteins is verified by gel images of SDS-PAGE analysis as well as abundant protein yield is determined by NanoDrop. This workflow in Aquarium provides a robust, straightforward, economic and highly applicable tool to perform protein purification for further biomedical research.

TABLE OF CONTENTS

List of Figures iii

List of Tables v

Chapter 1. Introduction 1

1.1 Motivation 1

1.2 Overview of Chapters 1

1.3 Overview of Protein Purification Workflow in Aquarium 2

Chapter 2. Cloning 4

2.1 Characteristics of Vector 4

2.2 Cloning of Insert Containing Gene of Interest into Vector 5

Chapter 3. Protein Induction 7

3.1 Characteristics of Host Cell 7

3.2 Protein Induction by IPTG 8

3.3 Optimal Scheme for Protein Induction 8

3.3.1 Optimal Condition for Induction 8

3.3.2 IPTG Concentration 9

Chapter 4. His-tagged Protein Purification 11

4.1 Cell Lysis 11

4.2 Affinity Column 11

4.2.1 His-tags 11

4.2.2 NI-NTA 12
4.2.3 Buffer System 13
4.2.4 Purification Process 15
4.3 Concentrate and Buffer Exchange 17
Chapter 5. Protein Verification 19
5.1 Sample Preparation 19
5.2 Gel Preparation 20
5.3 Scan Gel and Verify Protein in Aquarium 22
Chapter 6. Guidelines for Designing A Plan in Aquarium 24
6.1 Start A New Plan 24
6.1.1 Start with A DNA Fragment of Interest 24
6.1.2 Start with A DNA of Interest That Are Already Cloned to pET Vector 26
6.2 Launch A Plan 27
Chapter 7. Test Runs 30
7.1 Results 30
7.2 100% Success Rate 32
Chapter 8. Conclusions 35
Bibliography 37

LIST OF FIGURES

Figure 1 Operations of the protein purification workflow in Aquarium. The operations are divided into four categories according to specific purposes. 2
Figure 2 Schematic diagram of the protein purification workflow. 3
Figure 3 The DNA sequence of pET-29 vector [1]. 4
Figure 4 The DNA sequence of pET-22b vector [2]. 5
Figure 5 Ligation of a vector and DNA of interest with Gibson Assembly®. 6
Figure 6 The expression of recombinant proteins encoded by the pET vector which can be rapidly induced by the addition of IPTG. 9
Figure 7 The affinity resin, when charged with nickel ions, has a selectivity for a target protein with a His-tag [13]..13
Figure 8 A flow chart of His-tagged protein purification using Ni-NTA agarose beads. 16
Figure 9 Protein verification with SDS-PAGE analysis. 23
Figure 10 Enter the molecular weight of a target protein and a desired concentration of imidazole in the wash buffer. 24
Figure 11 An example screenshot of Aquarium planner illustrates a plan of cloning the DNA of interest into pET vectors...25
Figure 12 An example screenshot of Aquarium planner illustrates the His-tagged protein purification workflow. 26
Figure 13 Select BL21DE3 (indicated with an orange arrow) and a plasmid DNA of interest (indicated with a blue arrow) as inputs for transformation. 27
Figure 14 Select a protein sample as an output (indicated with a blue arrow). 28
Figure 15 Select Qiagen Ni-NTA for column packing (indicated with a blue arrow), then save and launch the plan. 29
Figure 16 Samples for test runs. 30
Figure 17 Concentrations of purified protein samples (from the first row to the third row): Spycatcher (16.354 mg/mL), SnoopTag-Spycatcher (37.001 mg/mL), and SpyTag-Snoopcatcher (31.378 mg/mL). 31

Figure 18 Gel image of the test run with three protein sequences. The protein sequences are (from left to right) SnoopTag-Spycatcher, Spycatcher, and SpyTag-Snoopcatcher, respectively. The three wells in each data set are (from left to right) uninduced, induced and purified protein sample, respectively. 32
Figure 19 Success rate tested by 10 purifications of SnoopTag-Spycatcher. 33
Figure 20 Concentrations of purified SnoopTag-Spycatcher (mg/mL). 33
Figure 21 The statistical results of the 10 purifications: (a) the concentrations of each purification in bar chart, and (b) the histogram of the distribution in probability where the mean is 48.76 mg/mL and the standard deviation is 7.3 mg/mL (15%). 34

LIST OF TABLES

Table 1 Timeline and Experiment Duration..3
Table 2 Buffer System..15

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Professor Klavins, for the continuous encouragement to my research, for his patience, motivation, and knowledge. His guidance helped me throughout the research and writing of the thesis. I really appreciate having the opportunity to join the team.

Besides my advisor, I would like to thank my thesis committee: Professor Barry Lutz, for his insightful comments and encouragement which widen my research from various perspectives.

My sincere thanks also go to lab managers and lab mates who provided me valuable information, discussions, and all the fun we have had in the last two years. Without their kindness and help, this project could not have been successfully developed.

Finally, I must express my very profound gratitude to my parents and to my husband for providing me with unfailing support and continuous encouragement through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you from the bottom of my heart.

Chapter 1. INTRODUCTION

1.1 MOTIVATION

Protein purification is a series of processes to yield a protein of interest, isolated from a complex mixture. A purified protein can be applied to numerous biological studies such as drug discovery, structure determination, and function prediction. Therefore, obtaining a protein of interest is a primary and highly desirable step before conducting further studies. In the procedure of protein purification, the main challenge is to extract target proteins under the native condition without interfering biological functions as well as to obtain protein samples with high purity and yield. Moreover, performing protein purification highly depends on the experience and relevant knowledge of protein properties. Even with high-quality commercial kits or well-described protocols, researchers might still experience unsatisfying results, especially if they try to perform multiple protein purifications at the same time, due to the possibility of cross contaminations and being overwhelmed by the batch sizes. Furthermore, some methods can provide proteins with high purity, like fast protein liquid chromatography (FPLC), but it requires high-end instruments which might not be accessible to everyone. To minimize the possibilities of aforementioned nonidealities, such as low yield and low purity as well as the requirements for high-end instruments, we designed and built a His-tagged protein purification workflow in our laboratory operating system, Aquarium, to facilitate researchers to obtain reliable protein samples for their further studies.

1.2 OVERVIEW OF CHAPTERS

In this thesis, we present the principles, construction, and results of His-tagged protein purification workflow as well as software approaches to improve the protein yield and purity. This

thesis is divided into eight chapters, starting with an overview of plasmid DNA preparation in Chapter 2, followed by three chapters that focus on the strategies of performing protein induction, purification, and verification. Chapter 6 describes the guidelines for designing a plan for this protein purification workflow in Aquarium. The test run results are provided in Chapter 7. Lastly, Chapter 8 concludes this thesis.

1.3 Overview of Protein Purification Workflow in Aquarium

Operations of this workflow in Aquarium [1] are divided into four categories, including cloning, protein induction, His-tagged protein purification, and protein verification. Each category is comprised of several operations, all built in Aquarium, see Figure 1.

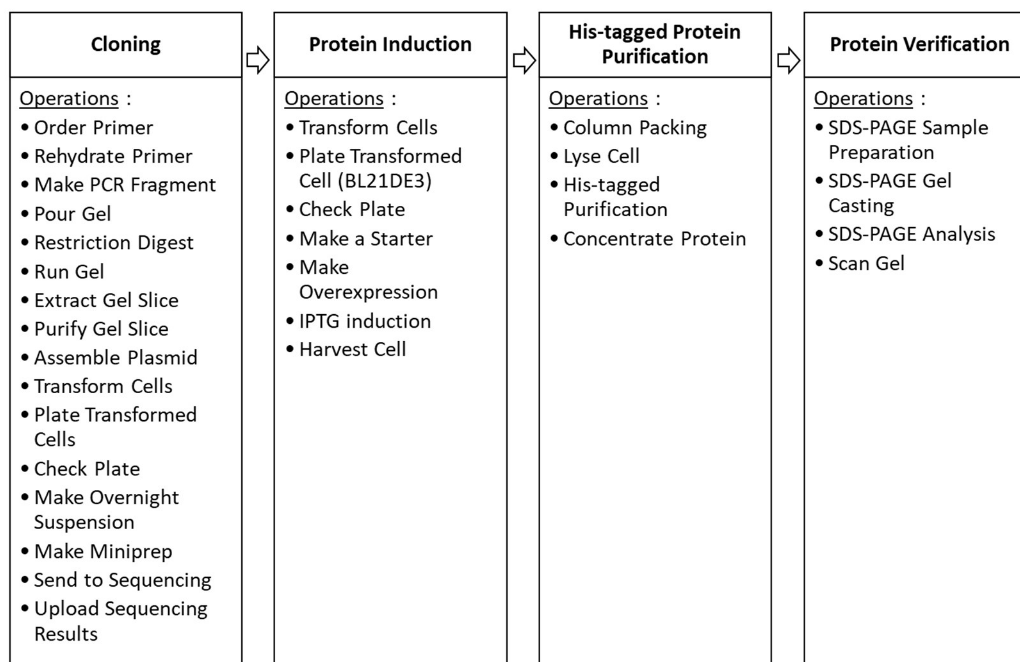

Figure 1. Operations of the protein purification workflow in Aquarium. The operations are divided into four categories according to specific purposes.

The schematic diagram of the protein purification workflow in Figure 2 depicts an example timeframe and steps to perform this protein purification workflow. More detailed time duration of

executing each operation is listed in Table 1. It takes totally 5 days to complete this His-tagged protein purification workflow, starting from protein induction to protein verification.

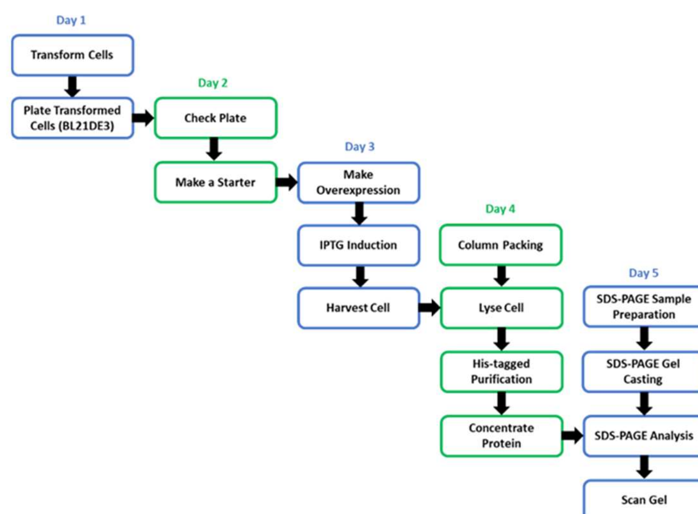

Figure 2. Schematic diagram of the protein purification workflow.

Table 1. Timeline and Experiment Duration

Timeline	Operation	Duration (minutes)
Day 1	Transform Cells	20
	Plate Transformed Cell (BL21DE3)	5
Day 2	Check Plate	3
	Make a Starter	10
Day 3	Make Overexpression	15
	IPTG Induction	10
	Harvest Cell	35
Can be paused here. Store samples at -80°C.		
Day 4	Column Packing	20
	Lyse Cell	25
	His-tagged Purification	60
	Concentrate Protein	20
Can be paused here. Store samples at -20°C.		
Day 5	SDS-PAGE Sample Preparation	30
	SDS-PAGE Gel Casting	50
	SDS-PAGE Analysis	40
	Scan Gel	10

Chapter 2. CLONING

2.1 CHARACTERISTICS OF VECTOR

To be compatible with this workflow, the selected vectors have to meet two requirements for protein expression. First, the vectors should be inducible to up-boost a target protein expression in *E. coli* (Escherichia coli) in order to produce great abundance of desired proteins. Second, the vectors need to be capable of expressing His-tag that can be chelated with Ni^{2+} (Nickel ion) in complex with nitrilotriacetic acid (NTA), providing an approach for capturing His-tagged protein. Considering these two required characteristics, pET vectors such as pET-22b (+) and pET-29b (+) are qualified to serve this workflow well. As shown in Figure 3 and Figure 4, both pET-22b (+) and pET-29 (+) are controlled by T7 promoter and contain a hexa-histidine as a His-tag on the C-terminus.

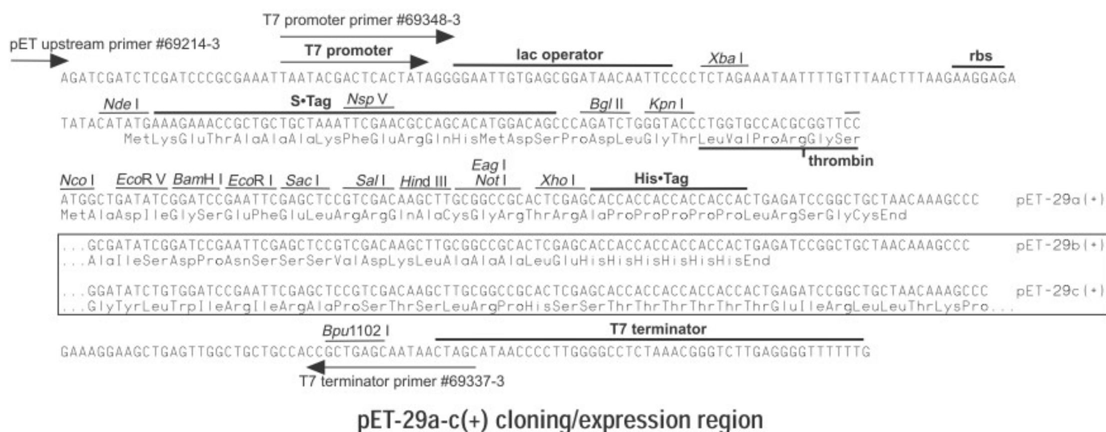

Figure 3. The DNA sequence of pET-29 vector [2].

The T7 promoter in pET vectors is not recognized by *E. coli* RNA polymerase, indicating it will give no or very little expression unless T7 RNA polymerase is provided. Moreover, the coding sequence for T7 RNA polymerase is presented in the chromosome under the control of the inducible *lac*UV5 promoter in the host cell, BL21(DE3).

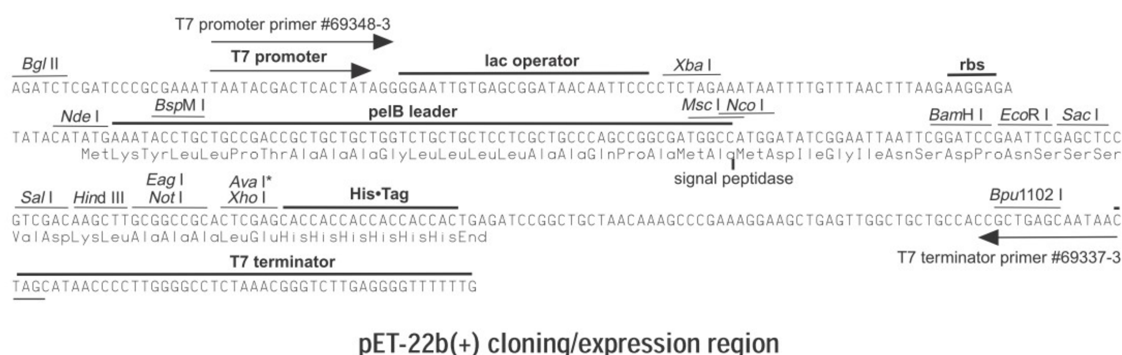

Figure 4. The DNA sequence of pET-22b vector [3].

In BL21DE3 host cells, the *lacI* repressor binds to the *lac* operator which prevents transcription initiated in the *lac*UV5 promoter and then blocks the production of T7 RNA polymerase for protein expression in pET vectors. Upon addition of a chemical inducer, isopropyl-β-D-thio-galacto-pyranoside (IPTG), to the host cell, the repression by *lacI* repressor is released. T7 RNA polymerase and this in turn starts transcription from T7 promoter and protein production [4].

2.2 CLONING OF INSERT CONTAINING GENE OF INTEREST INTO VECTOR

This section addresses cloning steps including designs of constructions, preparation of inserts and vectors, ligations, transformations, and sequence identifications. As illustrated in Figure 5, the first step is to design the primers with overlapping sequences between the adjacent DNA fragments for them to be assembled into a cloning vector. With the primer sets, the desired inserts will be amplified by polymerase chain reaction (PCR), while a linearized vector can be prepared by either PCR amplification or by restriction digestion. For plasmid assembly, the inserts and linearized vectors are mixed with the Gibson Assembly® Master Mix which contains three

enzymes - exonuclease, DNA polymerase, and DNA ligase - in one buffer mix. This mixture will then be incubated at 50°C for 1 hour.

In this cloning part, constructs are expressed in DH5α by taking advantages of its high transformation efficiency. Once a bacterial clone that expresses a His-tagged protein is established, the experiment proceeds to the next steps in series, which are induction, purification, and verification of His-tagged proteins described in next few chapters.

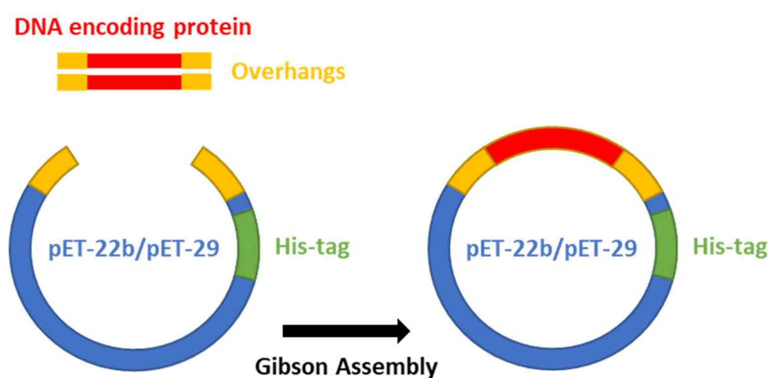

Figure 5. Ligation of a vector and DNA of interest with Gibson Assembly®.

Chapter 3. PROTEIN INDUCTION

The Aquarium operations in this category include *Transform Cells*, *Plate Transformed Cells (BL21DE3), Check Plate*, *Make a Starter*, *Make Overexpression*, *IPTG Induction*, and *Harvest Cell*, see Figure 1 in Chapter 1.

3.1 CHARACTERISTICS OF HOST CELL

Once target genes have been cloned into a pET vector, this established plasmid will be transformed into the competent cell, BL21DE3. There are several advantages of using BL21DE3 as a host cell for protein purification.

First, we can tell that BL21DE3 is classified as B strain directly from its name. First, this strain is deficient in both *lon* protease in the cytoplasm and *OmpT* protease in the outer membrane, which decreases the possibility of protein degradation during the expression process. Second, the name, DE3, means this strain contains the λDE3 lysogen, a lambda derivative that has the immunity region of phage 21 and carries a DNA fragment containing *lacI* gene, the *lac*UV5 promoter, and the gene for T7 RNA polymerase [5]. As mentioned in Section 2.1, the gene of interest cloned to pET vector could be inducible because it contains T7 promoter, and T7 promoter is controlled by T7 RNA polymerase instead of *E. coli* RNA polymerase. Protein expression is induced while T7 RNA polymerase is provided. Since BL21DE3 carries the gene for T7 RNA polymerase, it would be a suitable strain for expressing target proteins. Also, the production of T7 RNA polymerase is regulated by a *lac*UV5 promoter in BL21DE3. The *lac* operon has two mechanisms to regulate transcription. One is to down-regulate transcription by *lac* repressor and the other is to stimulate transcription by inducers, such as IPTG. Once IPTG is added, the *lac*UV5

promoter directs transcription of T7 RNA polymerase gene which in turn transcribes the target DNA in the plasmid.

3.2 Protein Induction by IPTG

In the absence of inducers or lactose in BL21DE3, *lac* repressor binds tightly to the *lac* operon and represses gene transcription by *E. coli* RNA polymerase [6]. However, in the presence of inducers such as IPTG, the inducer will bind to the *lac* repressor and cause the repressor to disassociate from the *lac* operon, leading to the transcription of genes which can proceed under *lac* operon control. Recombinant proteins can then be largely expressed. This mechanism is widely applied to generate recombinant proteins with T7 polymerase-based pET systems. pET vectors carry *lacI* genes which provide *lac* repressors to repress the chromosomal gene for initiating the transcription of T7 RNA polymerase which is controlled by the *lac*UV5 promoter. However, upon the addition of inducers, IPTG, the T7 RNA polymerase will be generated which will then transcribe DNA sequences controlled by the T7 promoter in pET vectors.

Under some conditions, although IPTG is absent, very little T7 RNA polymerase or target protein can still be produced by a leak of *lac* promoter that leads to basal expression of uninduced samples. Therefore, with SDS-PAGE analysis, we expect a very low or no basal expression of target genes is present before induction, but high levels of expression are obtained upon induction.

3.3 Optimal Scheme for Protein Induction

3.3.1 *Optimal Condition for Induction*

The induction in *E.coli* can be performed using conditions as follows: either slow induction at low temperatures such as 15-20 °C overnight or fast induction at high temperatures such as 37 °C for 3-4 hours. Fast induction can produce a large amount of target protein within a short period

of time. However, protein aggregates may also accumulate to insoluble inclusion bodies in cells when expressed at high levels [7]. These protein aggregates are accumulations of misfolded or partially folded intermediates while proteins are rapidly expressed in a short period of time. Therefore, slow induction with lower temperature is a more preferable condition to produce recombinant proteins which prevents protein aggregation and improves the protein solubility [8]. In this workflow, we optimized the conditions for the induction at 30 °C for 5 hours with an IPTG concentration of 0.5 mM in cell culture (Figure 6).

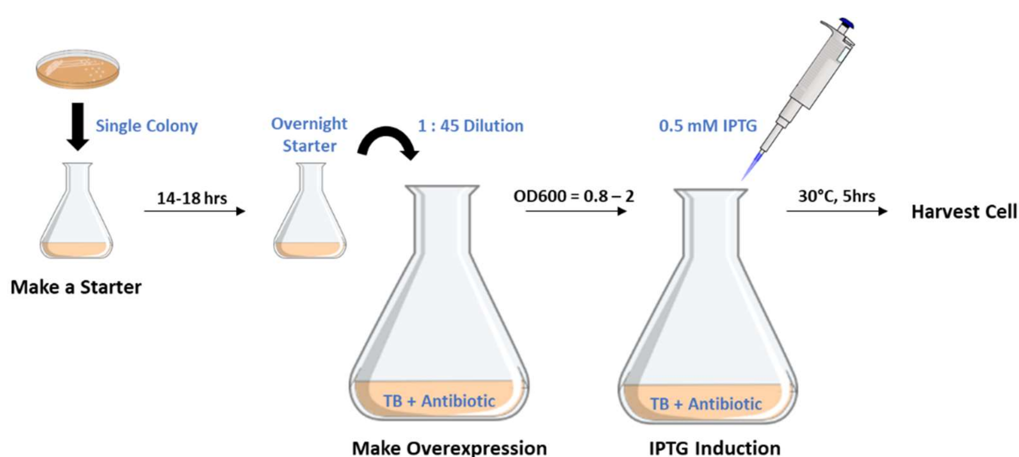

Figure 6. The expression of recombinant proteins encoded by the pET vector which can be rapidly induced by adding IPTG in a cell culture using Terrific Broth (TB).

3.3.2 *IPTG Concentration*

Though the optimal concentration of IPTG might vary in different given cases [9], a commonly used IPTG concentration is in the range from 0.01 to 1 mM. A previous study reported that the IPTG-induced protein expression imposes physiological burdens on the BL21DE3 host cell [10]. The induction generates an overexpression of exogenous genes that results in a harsh physiological condition for host cells with possible stresses originated from the toxicity of substrates, protein products, or intermediates. Therefore, a proper IPTG concentration is crucial

for maintaining the stability of host cells. The optimal IPTG concentration found for this workflow is 0.5 mM in a cell culture, which is within the range of commonly used concentration for induction.

Chapter 4. HIS-TAGGED PROTEIN PURIFICATION

The Aquarium operations in this category includes *Column Packing, Lyse Cell*, *His-tagged Purification,* and *Concentrate Protein*, see Figure 1 in Chapter 1.

4.1 CELL LYSIS

The protein induction is followed by harvesting and lysing cells, to release expressed proteins from cells. Sonication is a widely used method for lysing cells by generating mechanical forces to disrupt bacterial cell walls that ensure the expressed proteins are not affected by any solubilizing lysis agents, like detergents, that can affect solubility or stability [11].

In the operations of *Harvest Cell* and *Lyse Cell*, cell pellets are collected by a centrifuge and then soaked in an ice-cold PBS with EDTA-free protease inhibitors (Pierce™) to prevent protein degradation before performing sonication. The setting of sonication is using 95% power delivered to the probe with 10 seconds ON and 30 seconds OFF, repeating 30 rounds that results in total 20-minute processing time for each sample. Bacteria cell walls are lysed by sonication to release expressed proteins from cells. This whole cell lysate is then separated into cell pellet and supernatant after high-speed centrifugation at 24,000 g for 30 minutes. Expressed His-tagged proteins are soluble in the supernatant and will be captured by Ni-NTA agarose beads after applying the supernatant to an affinity column later on.

4.2 AFFINITY COLUMN

4.2.1 *His-tags*

Recombinant proteins are commonly expressed with an affinity tag fused to the N- or C-terminus to facilitate purification and detection. One of the most commonly used fusion tags for

recombinant protein purification is His-tag, which contains six consecutive histidine residues [8]. His-tags have gained great popularity over the last decade as a purification tool for recombinant proteins. Because of its small size, 0.84 kDa, and uncharged at physiological pH, His-tags do not affect folding or interfere with the functions of fusion proteins [12]. A previous study demonstrated that His-tags are not hazardous to the protein structures by a comparison of protein crystal structures with and without His-tags [13]. His-tagged protein purification is based on the affinity of histidine residues with nickel ions (Ni^{2+}) which are immobilized on chromatographic matrices, nitrilotriacetic acid (NTA), by a chelating ligand. Ni-NTA agarose beads are widely used for His-tagged protein purification since it has a stable binding affinity with His-tagged proteins [14].

4.2.2 *NI-NTA*

Ni-NTA affinity purification for His-tagged proteins is a procedure of bind-wash-elute that can be performed under the native condition of proteins. This type of purification procedure that separates proteins according to their affinity for metal ions is also called immobilized metal ion affinity chromatography (IMAC). The concept of IMAC was first formulated in 1975 [15]. It is based on the known affinity of transition metal ions such as Ni^{2+}, and Cu^{2+} to histidine in liquid solutions [16]. The concept has been extended to use metal ions as supports to capture target proteins with His-tags.

Nitrilotriacetic acid (NTA) was introduced in the 1980s as a chelating ligand used to fix the metal ion to agarose [15]. In the meantime, a genetic approach to facilitate this purification model was developed in 1988 [17]. The poly-histidine extended tag was generated to be applied to Ni-NTA resin. The model of the interaction between His-tag and Ni^{2+} is presented in Figure 7 [14].

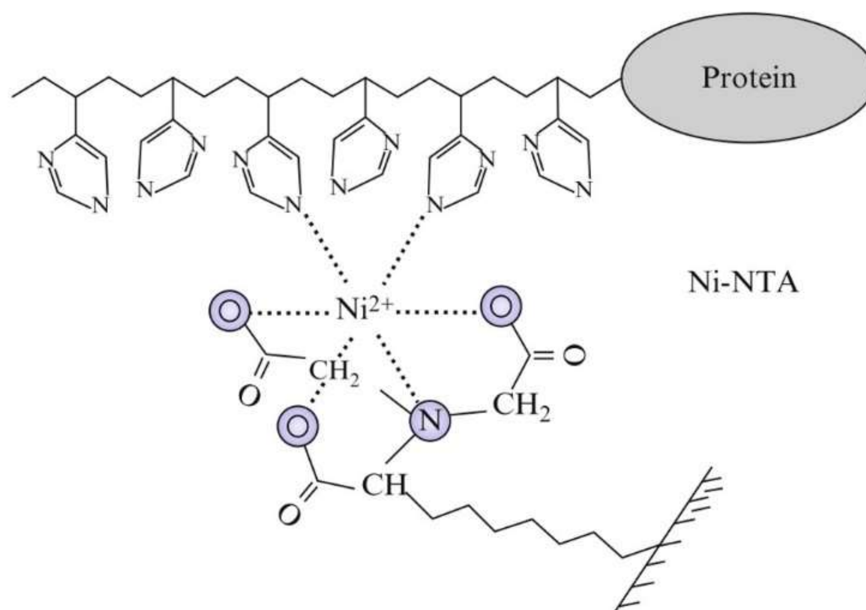

Figure 7. The affinity resin, when charged with nickel ions, has a selectivity for a target protein with a His-tag [14].

4.2.3 *Buffer System*

To avoid protein loss and degradation affected by buffering compounds and pH value, the buffer system in protein purification needs to be well-established as it plays an important role in facilitating protein purity, stability, and yield. The buffer system used in this workflow is listed in Table 2.

The following characteristics are carefully considered while establishing the buffer system in the workflow.

4.2.3.1 Salt and pH

Phosphate buffer is one of the commonly used biological buffers with a near neutral pKa of 6.86. As a result, it provides effective buffering in the pH range of 6.4 to 7.4 which is similar to a physiological pH value. Moreover, the phosphate buffer contributes to keeping protein stability by avoiding dramatic changes in pH value irreversibly affects protein stability. A previous study

presents that the pH value of phosphate buffer is not sensitive to temperature, therefore, using phosphate buffer for purification and protein storage prevents from a dramatic pH variation [18]. We also optimized the phosphate buffer concentration of 50 mM to ensure adequate buffering capacity.

4.2.3.2 Protease Inhibitors and Stabilizing Elements

The addition of EDTA[*1]-free protease inhibitors (Pierce™) in lysis buffer prevents protein degradation from endogenous proteases (Table 2). Chelators such as EDTA, EGTA[*2] and citrate should be excluded from lysis buffer, as they will strip the metal ions from agarose beads which affect affinity in the IMAC procedure. Triton X-100 is one type of osmolyte that serves as a stabilizing element, which stabilizes protein structure and enhances solubility.

4.2.3.3 Imidazole

Ni-NTA ligands are covalently attached to cross-linked agarose matrices for the selective purification of His-tagged proteins. Histidine residues can coordinate with the Ni^{2+} ion by replacing the bound water molecules and then retain His-tagged protein on column matrices. Imidazole has a structural similarity to histidine that can be used to elute the protein, through their ability to coordinate with the Ni^{2+} ion and to displace the bound protein. The addition of imidazole in low concentration can interfere with the weak bindings of other proteins and wash out unwanted proteins weakly bound to Ni-NTA. This wash step removes non-specific protein binding and contributes to elevating the purity of target proteins. In our workflow, users will be asked to select a proper imidazole concentration of the wash buffer for their target proteins. Nevertheless, a high imidazole concentration of 250 mM is used in elution buffer for eluting target proteins.

[*1] EDTA: Ethylenediaminetetraacetic-acid
[*2] EGTA: Ethylene glycol-bis(β -aminoethyl ether)-N,N,N’,N’-tetraacetic acid

Table 2 Buffer System for His-tagged Protein Purification

Buffers (All kept in 4 °C)	Composition
Lysis Buffer, pH 7.4	▪ Phosphate Buffered Saline (PBS), pH 7.4 ▪ 1% Triton X-100 ▪ Protease Inhibitor EDTA-free (Pierce™): Aprotinin (Serine Proteases Inhibitor) Bestatin (Aminopeptidases Inhibitor) E-64 (Cysteine Proteases Inhibitor) Leupeptin (Serine and Cysteine Proteases Inhibitor) AEBSF (Serine Proteases Inhibitor) Pepstatin A (Aspartic Acid Proteases Inhibitor)
Purification Buffer, pH 7.7	▪ Phosphate Buffered Saline (PBS), pH 7.4 ▪ 300mM Sodium Chloride, NaCl ▪ 0.01 % Trition X-100 ▪ 50mM Sodium Phosphate Buffer: Sodium Phosphate Dibasic, 7-Hydrate ($Na_2HPO_4 \cdot 7H2O$) Sodium Phosphate Monobasic, Monohydrate ($NaH_2PO_4 \cdot H2O$)
Wash Buffer, pH 7.7	▪ Purification Buffer ▪ 0 / 5 / 10 / 15 / 30 mM Imidazole, pH 8.0
Elution Buffer, pH 7.7	▪ Purification Buffer ▪ 250 mM Imidazole, pH 8.0
Protein Storage Buffer, pH 7.4	▪ Phosphate Buffered Saline, pH 7.4

4.2.4 *Purification Process*

In Aquarium, the purification process follows *Harvest Cell* that includes *Column Packing*, *Lyse Cell,* and *His-tagged Purification*. As shown in Figure 8, the purification process starts from a column is packed with Ni-NTA agarose beads. The Ni-NTA agarose beads (Qiagen) have a binding capacity of 5-10 mg protein per mL of the matrix resin and a high binding affinity (K_d $=10^{-13}$M) for His-tag at pH value of 8.0 [19]. Pre-equilibrium agarose beads with purification

buffer. Pre-equilibrium agarose beads with purification buffer ensure that agarose beads are soaked in the same buffer condition as target proteins, decreasing dramatic changes in pH values and components in the background. Described in chapter 4.1, bacteria cell walls are lysed by sonication to release proteins from cells into cell lysates. The cell lysate is then separated to supernatant and pellet by high-speed centrifuge. This supernatant, which target proteins are soluble in, is then applied to a column packed with Ni-NTA agarose beads. In this binding process, depending on the affinity of His-tags toward the Ni-NTA, desired proteins interact effectively with the ligand and are then captured by Ni-NTA agarose beads. Wash out non-specific proteins with the wash buffer that contains a low concentration of imidazole. Target proteins are retained on the column and not affected by the wash buffer because of the low concentration of imidazole. (e) His-tagged proteins are eluted from agarose beads with elution buffer. In this step, the high imidazole concentration of 250 mM generates intense competition between imidazole and His-tag for binding to the Ni-NTA column which is utilized for the elution of the desired protein.

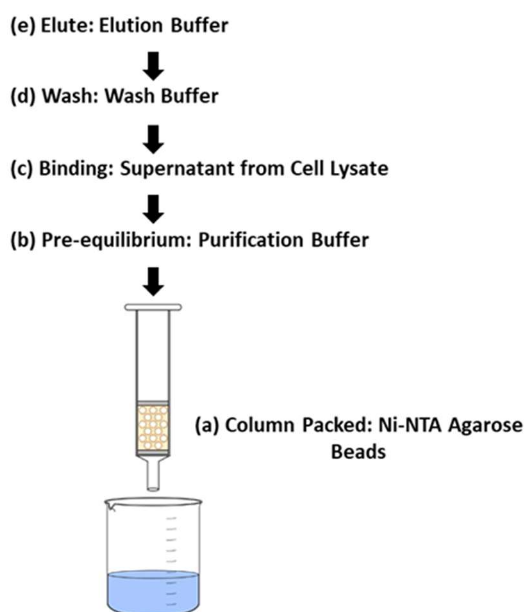

Figure 8. A flow chart of His-tagged protein purification using Ni-NTA agarose beads.

4.3 Concentrate and Buffer Exchange

In this workflow, proteins are purified under their native state and thus maintain their biological functions. Therefore, the preparation of concentrated protein samples is necessary for long-term storage which prevents spontaneous proteolytic degradation. Moreover, it is advisable to closely mimic the *in vivo* environment of proteins and replace the protein storage buffer with compatible buffers for further biological studies. In the operation of *Concentrate Protein* in Aquarium, concentrating is carried out by centrifugation at the temperature of 4 °C to maintain protein stability. A protein elution is placed into a Centriprep® tube (MilliporeSigma) for concentrating in a centrifuge until the protein solution, which will still remain under the filter, is at the desired volume (500 µL-1 mL).

Centriprep® tubes are an efficient tool for sample concentrating, desalting and buffer exchange. The concept of Centriprep® tubes is based on ultrafiltration which involves forcing a protein solution through a cellulose membrane with a defined molecular weight cutoff (10 or 30 kDa). A selection of the Centriprep® tube with an appropriate cutoff membrane for a target protein is according to protein molecular weight. The cutoff size of the membrane should be smaller than the molecular weight of a target protein, allowing smaller background proteins to pass through the membrane and be removed from the protein solution.

Centriprep® tubes also perform desalting and imidazole removal. Target proteins will not be allowed to pass through the membrane, whereas the other buffer components will. Some downstream applications such as nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, enzyme activity assay, and mass spectrometry are sensitive to salts or detergents. Chemical compounds, salts, and detergents such as imidazole and Triton X-100 can be removed

by performing this step. This method is usually relatively rapid and does not adversely affect the protein sample.

Chapter 5. PROTEIN VERIFICATION

To verify the presence of target proteins in purified protein samples, we identify purified proteins by using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as the analytical tool. Protein separation by SDS-PAGE can be used to estimate relative molecular weight, to determine the relative abundance of major proteins in samples, and to determine the distribution of proteins among fractions. This category contains Aquarium operations of *SDS-PAGE Sample Preparation, SDS-Gel Casting, SDS-PAGE Analysis,* and *Scan Gel*, see Figure 1 in Chapter 1.

5.1 SAMPLE PREPARATION

1 mL of cell cultures are reserved from the cell culture before and after IPTG induction, respectively. OD_{600} values of the cell culture will be measured by NanoDrop (Thermo) and recorded in Aquarium. Cell pellets are then collected by centrifuge at 8000 g for 3 minutes and then dissolved in the lysis buffer. Aquarium calculates the required amount for the lysis buffer according to their OD_{600} values to normalize cell numbers of induced and uninduced samples. Equal cell lysate loadings are required to make accurate comparisons of target protein expression data. Normalized expression is then compared across samples to verify that changes in the target protein expression represent the differences between uninduced and IPTG-induced samples. 40 µL of each sample, induced and uninduced, are enough to provide proper intensities of protein bands in gel lanes of an SDS-PAGE gel after Coomassie blue staining.

Purified protein samples are measured by NanoDrop under protein mode. A_{280} values will also be recorded in Aquarium. Since 20-µg total protein is needed for a gel lane on a gel apparatus (1.5-mm thickness, 10-well comb), Aquarium calculates the sample amounts for each sample that

are needed for gel lanes as well as normalizes loading amounts for each sample with PBS. The next step is to suspend uninduced, induced, and purified protein samples in a 2X sample buffer (Sigma-Aldrich®) and then heat samples at a temperature of 98 °C for 10 minutes. The sample buffer contains 0.125-M Tris HCl, 0.004 % bromophenol blue, 20 % glycerol, 4 % SDS (sodium dodecyl sulfate), and 10 % 2-mercaptoethanol. The ingredients of the sample buffer are explained as follows: (a) Tris acts as a buffer, which provides a specific pH value since the stacking process in the discontinuous electrophoresis requires a specific pH value; (b) Glycerol is thick which makes samples become dense so the samples will remain in the bottom of the well rather than float out; (c) bromophenol blue allows researchers to track the progress of the electrophoresis. (d) SDS, 2-mercaptoethanol, and the heat are responsible for the actual denaturation of the sample. SDS breaks up the two- and three-dimensional structures of the proteins by adding negative charges to the amino acids. Proteins are turned into linear shapes for rendering them functionless immediately. Some quaternary structures may remain due to disulfide bonding [20]. Many proteins have significant hydrophobic properties and may be tightly associated with other molecules, such as lipids, throughout hydrophobic interaction. Heating the samples at 98 °C and mix samples well, allowing SDS to be bound in the hydrophobic regions and complete the denaturation.

5.2 Gel Preparation

SDS-PAGE with a discontinuous buffer system is the most-widely used electrophoresis technique that separates proteins by molecular weight. This system uses a discontinuous buffer system where the pH value and ionic strength of the buffer used for running the gel (Tris/Glycine/SDS, pH 8.3) is different from the buffers used for the stacking gel (Tris-HCl, pH 6.8) and resolving gel (Tris-HCl, pH 8.8). With this discontinuous system, high gel resolution can be easily obtained even with large volumes of diluted protein samples applying to a gel. This high

resolution is a direct result of the way that proteins stack into narrow zones during migration through the large-pore stacking gel as well as resolve under the given conditions in the small pore resolving gel.

SDS-PAGE separates proteins primarily by the molecular weight because the ionic detergent SDS denatures and binds to proteins to make them uniformly-charged negatively. Thus, when a current is applied, all SDS-bound proteins in a sample will migrate through the gel toward the positively-charged electrode. Proteins with less mass travel more quickly through the gel than those with greater mass. Consequently, when these samples are electrophoresed, proteins separate according to the molecular weight.

In the operation of *SDS-PAGE Gel Casting*, the prepared gel solution is poured into the thin space between two glass or plastic plates that form a gel cassette. To obtain an optimal resolution of proteins, a stacking gel is cast over the top of the resolving gel. The stacking gel has a lower concentration of acrylamide (10%), a lower pH value (6.8), and a different ionic content than a resolving gel (12%, pH 8.8). This allows proteins in a loaded sample to be concentrated into one tight band during the first 20 minutes of electrophoresis before entering the resolving gel. The components in gels, acrylamide mixed with bis-acrylamide, form a crosslinked polymer network when the polymerizing agent, ammonium persulfate (APS), is added. TEMED (N, N, N, N'-tetramethylenediamine) catalyzes the polymerization reaction by promoting the production of free radicals by APS. Once the gel polymerizes, the cassette is mounted into an apparatus so that the top and bottom edges are placed in contact with buffer chambers containing a cathode and an anode. The buffer chambers then filled with running buffer which contains ions that conduct current through the gel.

In the operation of *SDS-PAGE Analysis*, protein samples are added in wells at the top edge. Once a current is applied, the proteins are drawn by the current through the gel matrix. Electrophoresis gels are formulated in buffers comprising a stacking gel component that helps focus proteins into sharp bands at the beginning of the electrophoretic run and the resolving gel component that separates the proteins based on size. After electrophoresis, gels are stained in gel staining reagent (EZBlue™) which is composed of coomassie brilliant blue G-250 and formulated as an acidic solution in methanol. In acidic conditions, the dye binds to proteins primarily through basic amino acids (arginine, lysine, and histidine), and the number of coomassie dye ligands bound to each protein molecule is approximately proportional to the number of positive charges found on the protein. Protein-binding causes the dye to change from reddish-brown to bright blue. After staining for 1 hour, gels are washed with water until the gel background matrix is clear and blue protein bands are distinct.

5.3 Scan Gel and Verify Protein in Aquarium

Distinct protein bands in a gel with clear background matrix are used to verify protein expression in this workflow. As shown in Figure 9, every three consecutive wells are viewed as a data set for a batch of protein which are labeled in black, orange, and purple, respectively. The intensities of uninduced protein bands are expected to be very little or nearly no expression, whereas the intensity of induced protein bands will be increased because of the addition of IPTG for protein induction in cell culture. Purified protein samples are expected to be in a correct molecular weight with high purity. Gel images will be uploaded, reviewed, and saved in Aquarium.

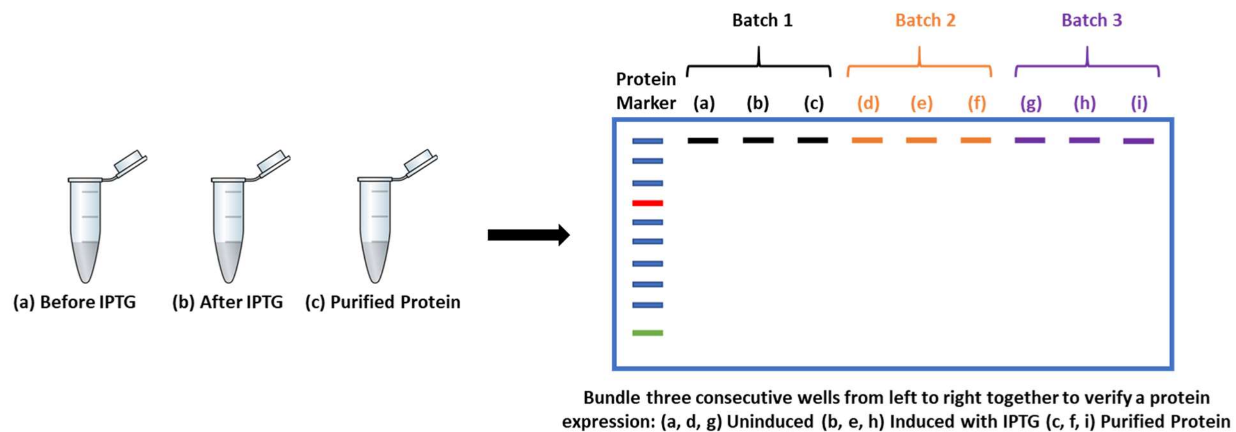

Figure 9. Protein verification with SDS-PAGE analysis.

Chapter 6. GUIDELINES FOR DESIGNING A PLAN IN AQUARIUM

6.1 Start A New Plan

A new protein sample must be created in Aquarium inventory before starting a new plan. To create a new sample, click SAMPLES in the Aquarium main menu, and then click the NEW button to select “Protein” in the sample type. Users will be asked to enter the protein size (in molecular weight, kDa) which will be used for protein identification in SDS-PAGE analysis. Then select a desired imidazole concentration (mM) in the wash buffer that will be used for washing out non-specific binding (Figure 10).

Figure 10. Enter the molecular weight of a target protein and a desired concentration of imidazole in the wash buffer.

6.1.1 *Start with A DNA Fragment of Interest*

Start from *Make PCR Fragment* to add overhangs on DNA fragments with designed primers. Run PCR product and digested pET vector on an agarose gel to check size and yield, and then apply them in Gibson Assembly® reaction. Next, sequence the plasmid DNA to verify the clone (Figure 11). Once a plasmid DNA is established, the plasmid DNA can be transformed into the BL21DE3 competent cell to start the purification workflow (Figure 12).

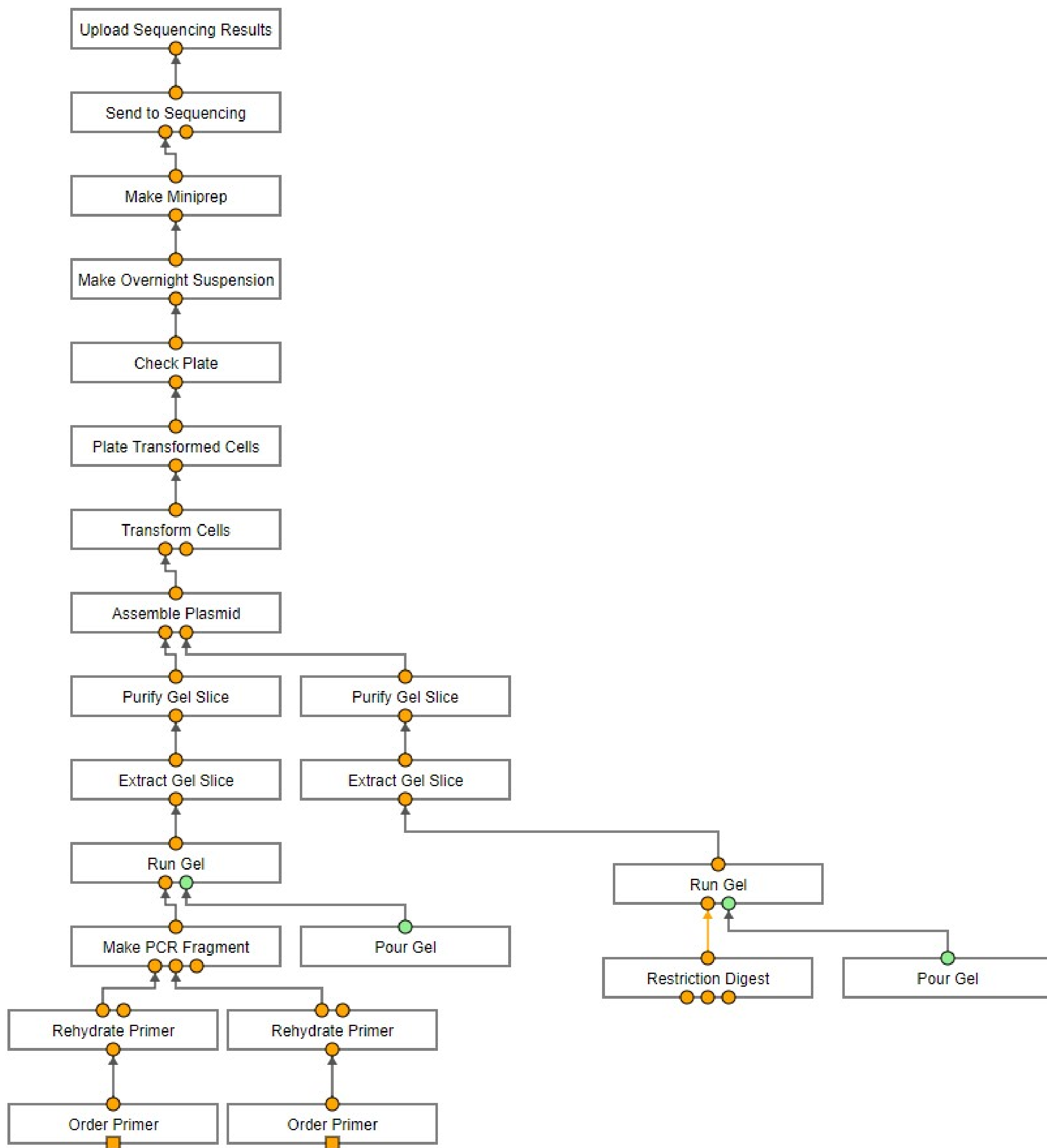

Figure 11. An example screenshot of Aquarium planner illustrates a plan of cloning the DNA of interest into pET vectors.

6.1.2 *Start with A DNA of Interest That Are Already Cloned to pET Vector*

Directly start from the operation of *Transform Cells*, and then proceed to the purification workflow (Figure 12).

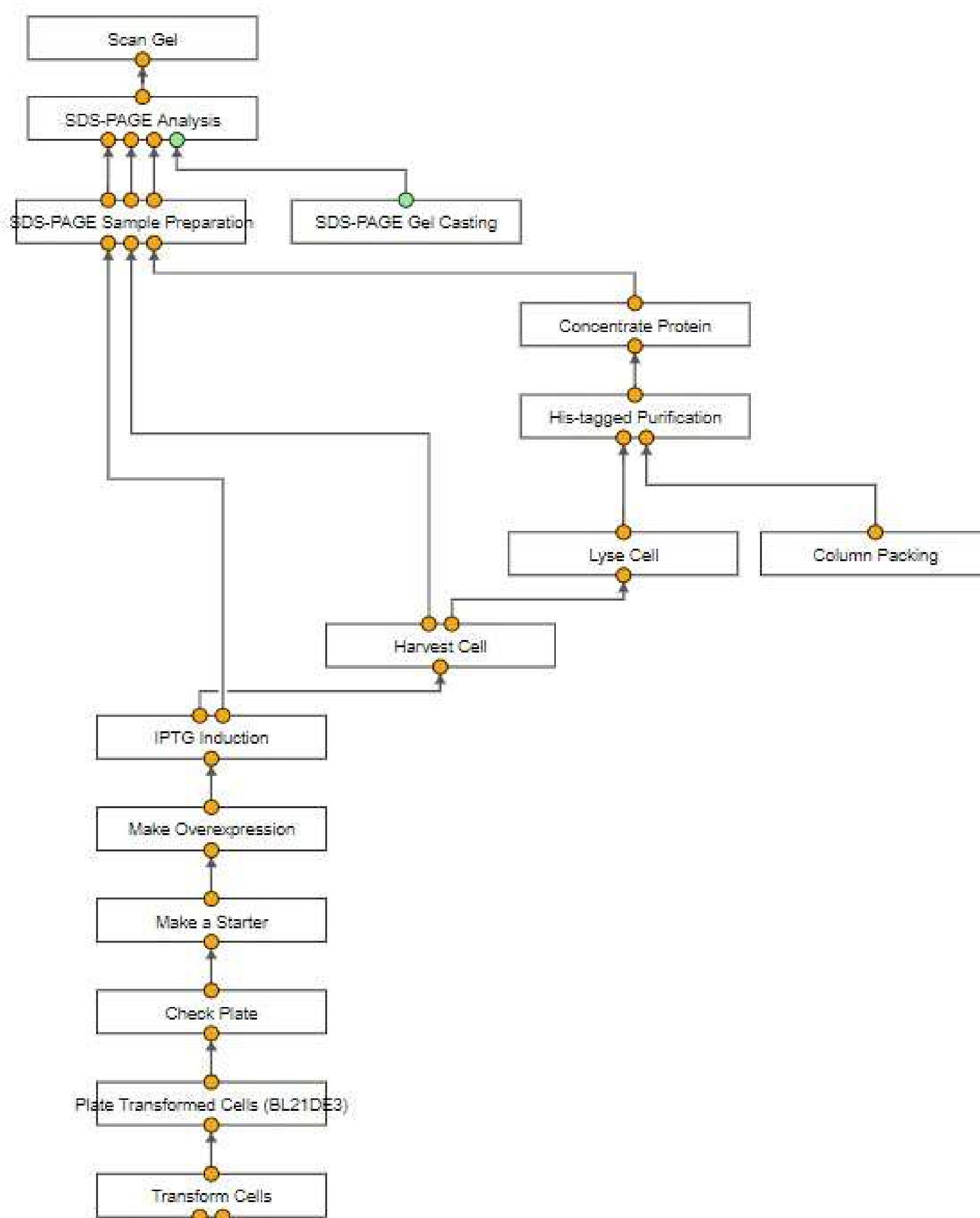

Figure 12. An example screenshot of Aquarium planner illustrates the His-tagged protein purification workflow.

6.2 LAUNCH A PLAN

Before launching a new plan, there are three steps to ensure the plan is submitted correctly. Firstly, as shown in Figure 13, select the BL21DE3 competent cell from Aquarium inventory. Any plasmid DNA of interest that uses pET vectors with His-tag and T7 promotor would be compatible with this protein purification workflow. Secondly, assign a target protein that will be purified in this plan as an output (Figure 14). The final step before launching is to select Qiagen Ni-NTA as the matrix for column packing (Figure 15).

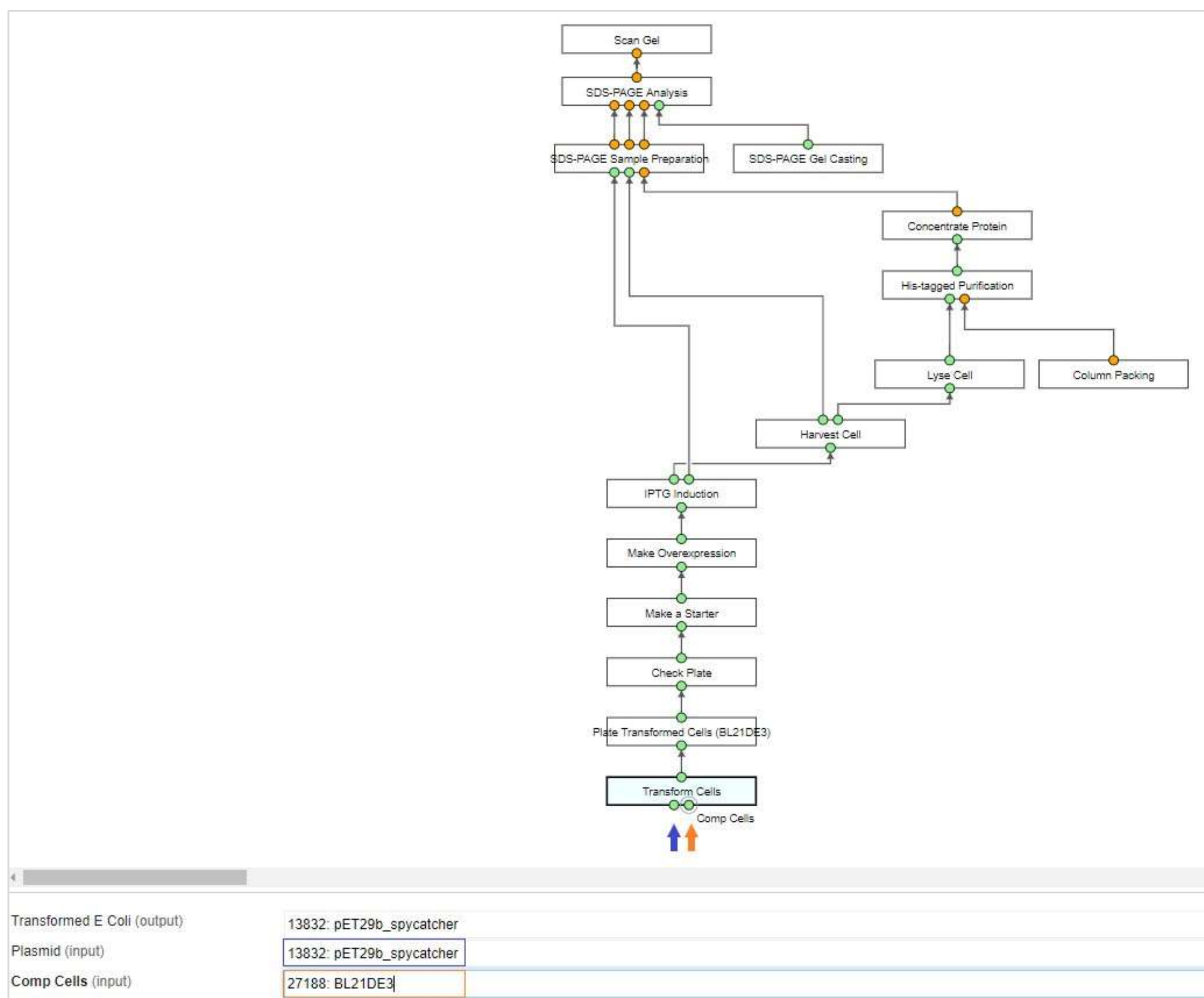

Figure 13. Select BL21DE3 (indicated with an orange arrow) and a plasmid DNA of interest (indicated with a blue arrow) as inputs for transformation.

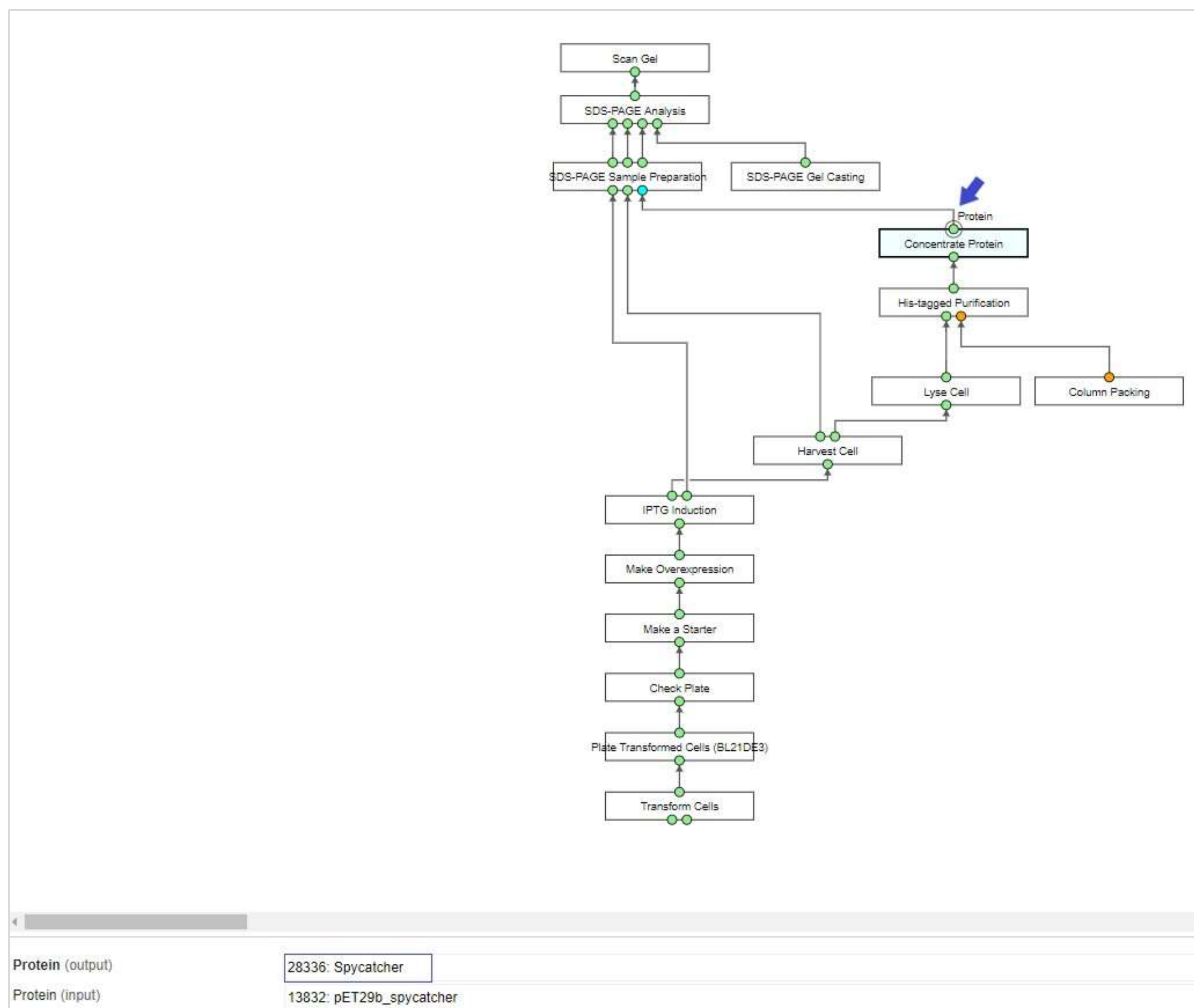

Figure 14. Select a protein sample as an output (indicated with a blue arrow).

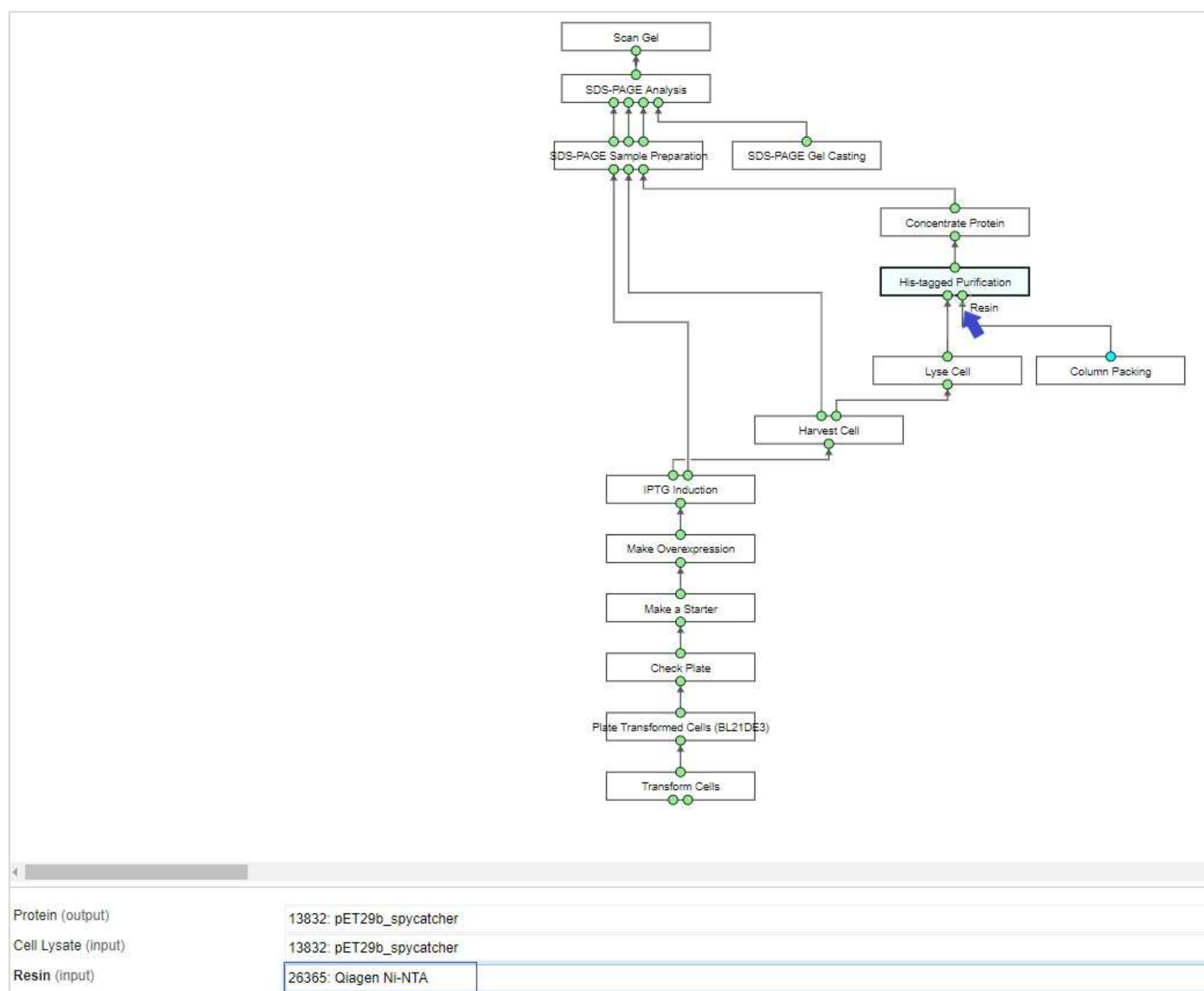

Figure 15. Select Qiagen Ni-NTA for column packing (indicated with a blue arrow), then save and launch the plan.

Chapter 7. TEST RUNS

7.1 RESULTS

To test this workflow, we used three plasmid DNAs (SpyTag-Snoopcatcher, SnoopTag-Spycatcher, and Spycatcher). DNA fragments of interest had been cloned to the pET-29b vector (Figure 16).

▼ **28338:** SpyTag-Snoopcatcher

Description:	DNA fragment cloned to pET29b with His-tag
Sample Type:	Protein
Created:	Feb 18, 2019
Size:	21.5
Wash_Buffer:	5

▼ **28337:** SnoopTag-Spycatcher

Description:	DNA fragment cloned to pET29b with His-tag
Sample Type:	Protein
Created:	Feb 18, 2019
Size:	20
Wash_Buffer:	5

▼ **28336:** Spycatcher

Description:	DNA fragment cloned to pET29b with His-tag
Sample Type:	Protein
Created:	Feb 18, 2019
Size:	18
Wash_Buffer:	5

Figure 16. Samples for test runs.

The plan starts from the transformation in BL21DE3 competent cell and then through the protein purification workflow. We did buffer change for protein elution with PBS and concentrated protein sample volume to 1 mL that prevents the protein from decaying by keeping it in high concentration. A_{280} value is measured by Nanodrop to determine protein concentrations (Figure 17).

Plots | Report

Test type: Protein A-280

2/22/2019 4:17 PM

Exit

Report Name

Report Full Mode: Ignore

Sample ID	User ID	Date	Time	mg/ml	260/280	A280 10mm	E 1%	Ext. Coeff x10e3	Mol. Wt. kDa	Cursor Pos.	Cursor abs.
	Default	2/22/2019	4:15 PM	16.35	0.73	16.354	10.00	NaN	NaN	280	16.354
	Default	2/22/2019	4:16 PM	37.00	0.61	37.001	10.00	NaN	NaN	280	37.001
	Default	2/22/2019	4:16 PM	31.38	0.66	31.378	10.00	NaN	NaN	280	31.378

Figure 17. Concentrations of purified protein samples (from the first row to the third row): Spycatcher (16.35 mg/mL), SnoopTag-Spycatcher (37.00 mg/mL), and SpyTag-Snoopcatcher (31.38 mg/mL).

Protein expression and purity were verified by SDS-gel image. In the gel, the order of well is from left to right, where every three wells form a data set for one batch (Figure 9). The three wells are uninduced, induced and purified protein sample, respectively. In all three purifications (Figure 18), very little protein was observed at the expected size without induction as compared to a significant amount of protein after induction and a significant enrichment of the desired protein after purification. The first two purified proteins have high purity while the last sample, SpyTag-Snoopcatcher, has some non-specific proteins in the purified sample. In some cases, purified protein samples may express more than one single bands on a gel lane, indicating the purity is lower than expected. To solve this problem, one of the possible solutions is adding a higher concentration of imidazole in wash buffer such as 15 or 30 mM to wash out the non-specific binding more completely. Users have the control of selecting the parameter of wash buffer concentration to get a better result (Figure 10). If a very high purity is desired, purified protein

samples could be applied to a second purification method for further enhancement such as fast protein liquid chromatography (FPLC) or biotinylation to obtain a final product with high purity.

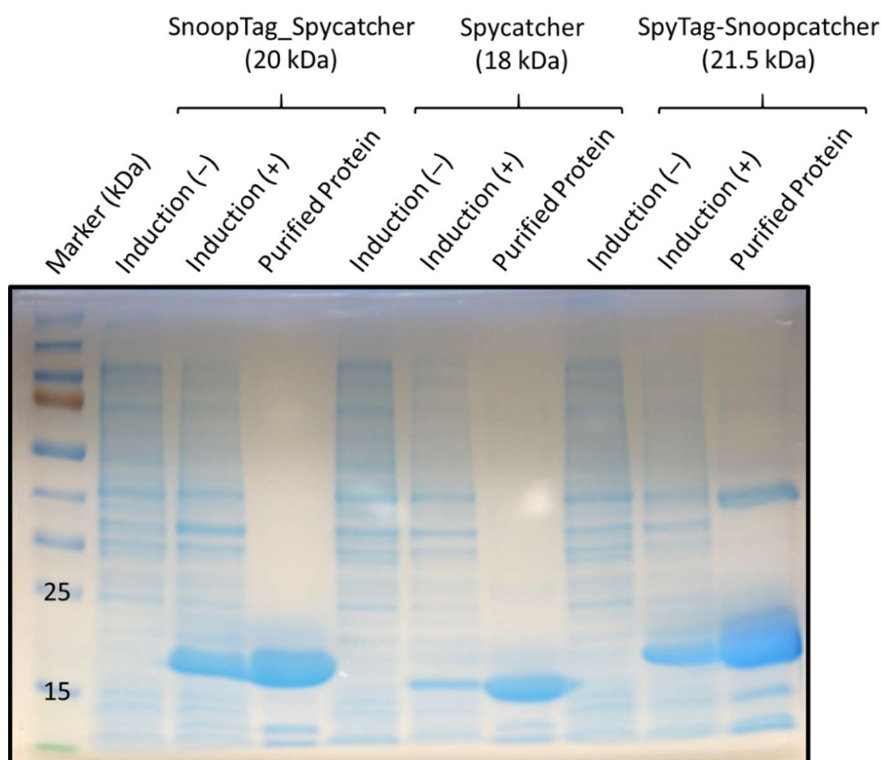

Figure 18. Gel image of the test run with three protein sequences. The protein sequences are (from left to right) SnoopTag-Spycatcher, Spycatcher, and SpyTag-Snoopcatcher, respectively. The three wells in each data set are (from left to right) uninduced, induced and purified protein sample, respectively.

7.2 100% Success Rate

The high purity and great protein abundance of purified SnoopTag-Spycatcher have been proved in the test run. Here, in order to test the success rate of this workflow, we submitted 10 identical purification plans for SnoopTag-Spycatcher and performed them in 2 batches, each batch includes 5 purifications in parallel. The gel image in Figure 19 shows protein bands of each batch have the same pattern repeated in all of uninduced, induced and purified protein samples, indicating high reproducibility and reliability of this workflow. In Figure 20, the concentrations of

SnoopTag-Spycatcher samples in 10 purifications are measured by NanoDrop. The statistical result of these 10 purifications is shown in Figure 21. The mean concentration is 48.76 mg/mL and the standard deviation is 7.3 mg/mL (15%), indicating the robustness of this workflow.

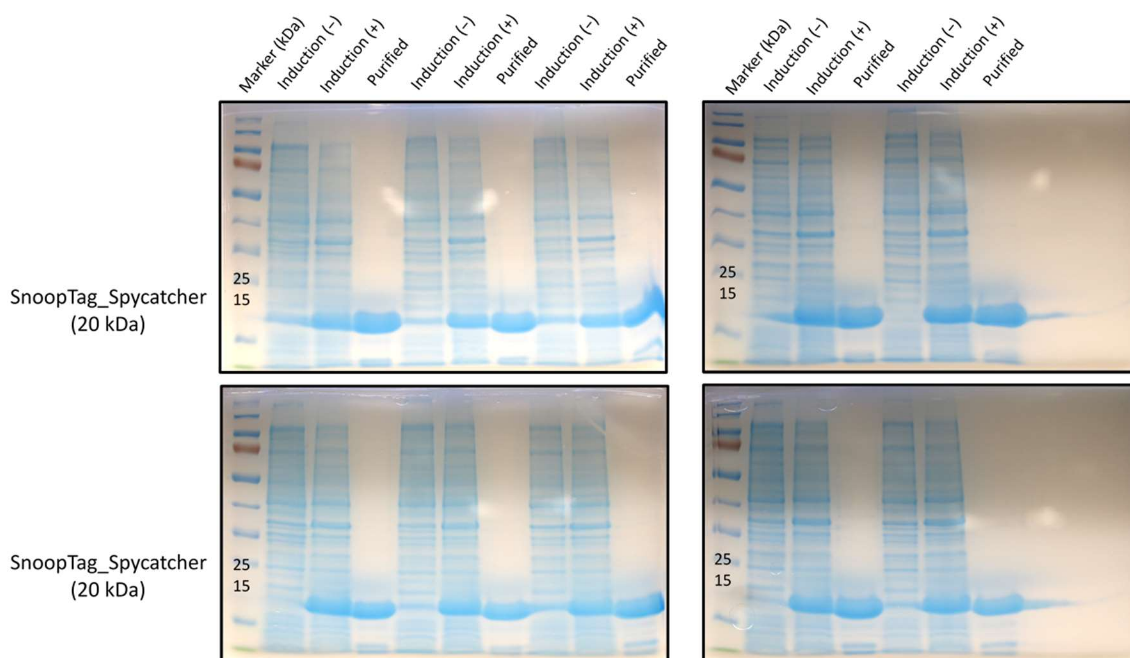

Figure 19. Success rate tested by 10 purifications of SnoopTag-Spycatcher.

Plots | Report — Test type: Protein A-280 — 3/14/2019 6:24 PM — Exit

Report Name — Report Full Mode: Ignore

Sample ID	User ID	Date	Time	mg/ml	260/280	A280 10mm	E 1%	Ext. Coeff x10e3	Mol. Wt. kDa	Cursor Pos.	Cursor abs.
	Default	3/14/2019	6:21 PM	37.08	0.62	37.077	10.00	NaN	NaN	280	37.077
	Default	3/14/2019	6:22 PM	49.24	0.65	49.236	10.00	NaN	NaN	280	49.236
	Default	3/14/2019	6:22 PM	43.01	0.66	43.011	10.00	NaN	NaN	280	43.011
	Default	3/14/2019	6:23 PM	49.45	0.67	49.454	10.00	NaN	NaN	280	49.454
	Default	3/14/2019	6:23 PM	49.97	0.66	49.968	10.00	NaN	NaN	280	49.968

Plots | Report — Test type: Protein A-280 — 3/15/2019 6:21 PM — Exit

Report Name — Report Full Mode: Ignore

Sample ID	User ID	Date	Time	mg/ml	260/280	A280 10mm	E 1%	Ext. Coeff x10e3	Mol. Wt. kDa	Cursor Pos.	Cursor abs.
	Default	3/15/2019	6:18 PM	38.01	0.68	38.010	10.00	NaN	NaN	280	38.010
	Default	3/15/2019	6:19 PM	51.30	0.68	51.301	10.00	NaN	NaN	280	51.301
	Default	3/15/2019	6:19 PM	54.68	0.66	54.682	10.00	NaN	NaN	280	54.682
	Default	3/15/2019	6:20 PM	57.70	0.67	57.699	10.00	NaN	NaN	280	57.699
	Default	3/15/2019	6:20 PM	57.20	0.71	57.198	10.00	NaN	NaN	280	57.198

Figure 20. Concentrations of purified SnoopTag-Spycatcher (mg/mL).

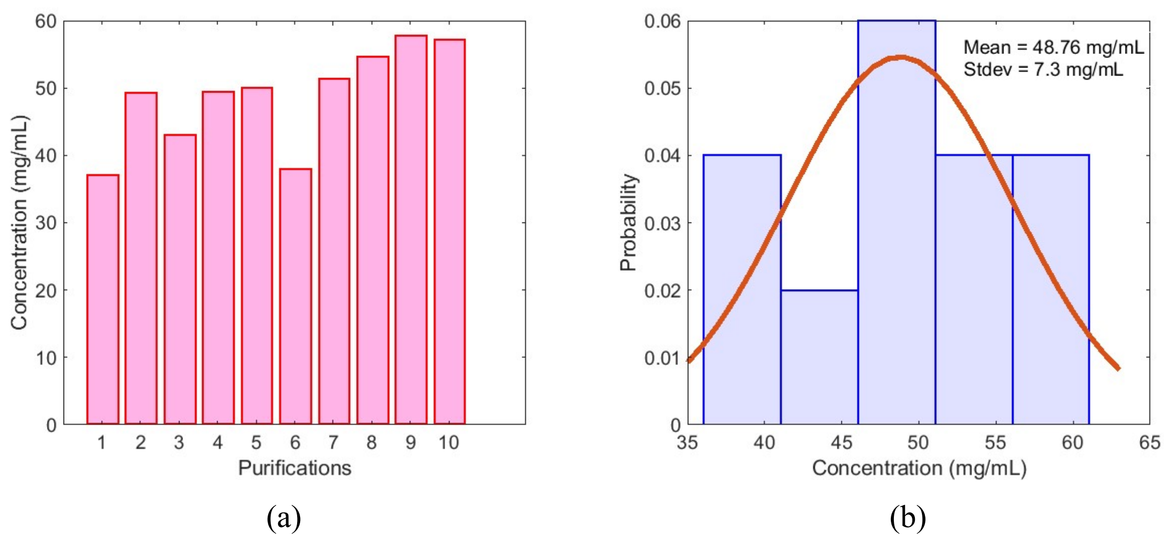

Figure 21. The statistical results of the 10 purifications: (a) the concentrations of each purification in bar chart, and (b) the histogram of the distribution in probability where the mean is 48.76 mg/mL and the standard deviation is 7.3 mg/mL (15%).

Chapter 8. CONCLUSIONS

This thesis presented a reproducible workflow of the His-tagged protein purification which plays an essential role in the success of downstream biochemical studies such as enzyme activity assays and structural research. Poly-histidine exhibits the strongest interaction with immobilized metal ion matrices which is applied as an electron donor group on the imidazole ring of histidine that forms coordination bonds with the immobilized transition metal. Also, because of its small size (0.84 kDa), His-tag does not interfere with the structure or function of desired proteins. Following by washing of the matrix material, the proteins containing poly-histidine sequences can be easily eluted by adding free imidazole into the elution buffer. SDS-PAGE electrophoresis is used to analyze and verify the purity and size of target proteins. Gel images are used as the final results to demonstrate the excellent outcome of executing this proposed workflow. After thorough demonstrations by test runs, this workflow showed promise of providing protein purity and reproducibility with a very high success rate (100% in 10 identical purifications).

The His-tagged protein purification workflow described in this thesis is implemented in the computer-assisted platform, Aquarium, to minimize the possibility of getting unsatisfying results that the users or laboratory technicians might experience due to deficient laboratory experiences or insufficient background knowledge of protein purification. With Aquarium, the protocols are represented as executable code. Aquarium records every detail such as concentration, molecular weight, and OD value that is needed for the process, formula or inventory interaction. Aquarium computes every required information and generates step-by-step instructions for laboratory technicians. All these features assist not only to elevate the success rate of the purification but also provide data saved in Aquarium inventory for tracking down errors. Users can evaluate the data in Aquarium and re-submit a purification plan with modifications to get a

better result. Moreover, Aquarium automates multiple purifications in batches which can efficiently save time and reduce the cost of conducting experiments, as well as increase the success rate. Overall, the use of this workflow to purify His-tagged proteins provides a rapid, robust, and inexpensive purification method for obtaining a protein of interest with high purity and yield for further biomedical research.

BIBLIOGRAPHY

[1] Pei-Rung Wu, "His-tagged Protein Purification: An Aquarium Workflow: peirung/Protein_Purification," 14-May-2019. [Online]. Available: https://github.com/peirung/Protein_Purification. [Accessed: 13-May-2019].

[2] "pET-29b(+) DNA - Novagen | 69872." [Online]. Available: http://www.emdmillipore.com/US/en/product/pET-29b+-DNA-Novagen,EMD_BIO-69872#anchor_CITATION. [Accessed: 09-Apr-2019].

[3] "pET-22b(+) DNA - Novagen | 69744." [Online]. Available: http://www.emdmillipore.com/US/en/product/pET-22b+-DNA-Novagen,EMD_BIO-69744. [Accessed: 09-Apr-2019].

[4] P. G. Blommel, K. J. Becker, P. Duvnjak, and B. G. Fox, "Enhanced Bacterial Protein Expression During Auto-Induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition," *Biotechnol. Prog.*, vol. 23, no. 3, pp. 585–598, 2007.

[5] F. W. Studier and B. A. Moffatt, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes," *J. Mol. Biol.*, vol. 189, no. 1, pp. 113–130, May 1986.

[6] J. W. Dubendorfft, "Controlling Basal Expression in an Inducible T7 Expression System by Blocking the Target T7 Promoter with Zac Repressor," p. 15.

[7] R. R. Kopito, "Aggresomes, inclusion bodies and protein aggregation," *Trends Cell Biol.*, vol. 10, no. 12, pp. 524–530, Dec. 2000.

[8] D. S. Waugh, "Making the most of affinity tags," *Trends Biotechnol.*, vol. 23, no. 6, pp. 316–320, Jun. 2005.

[9] R. S. Donovan, C. W. Robinson, and B. R. Glick, "Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter," *J. Ind. Microbiol.*, vol. 16, no. 3, pp. 145–154, Mar. 1996.

[10] P. Dvorak *et al.*, "Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway," *Microb. Cell Factories*, vol. 14, no. 1, Dec. 2015.

[11] C. W. Ho, T. K. Chew, T. C. Ling, S. Kamaruddin, W. S. Tan, and B. T. Tey, "Efficient mechanical cell disruption of Escherichia coli by an ultrasonicator and recovery of intracellular hepatitis B core antigen," *Process Biochem.*, vol. 41, no. 8, pp. 1829–1834, Aug. 2006.

[12] A. Newconceptinaffinitychromatographyforrecombinantproteins, "Chelating Peptide-immobilized Metal Ion Affinity Chromatography," p. 5.

[13] M. Carson, D. H. Johnson, H. McDonald, C. Brouillette, and L. J. DeLucas, "His-tag impact on structure," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 63, no. 3, pp. 295–301, Mar. 2007.

[14] H. Block *et al.*, "Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC): A Review," in *Methods in Enzymology*, vol. 463, R. R. Burgess and M. P. Deutscher, Eds. Academic Press, 2009, pp. 439–473.

[15] J. Porath, J. Carlsson, I. Olsson, and G. Belfrage, "Metal chelate affinity chromatography, a new approach to protein fractionation," *Nature*, vol. 258, no. 5536, pp. 598–599, Dec. 1975.

[16] J. Z. Hearon, "The Configuration of Cobaltodihistidine and Oxy-Bis(Cobaltodihistidine)2," *JNCI J. Natl. Cancer Inst.*, vol. 9, no. 1, pp. 1–11, Aug. 1948.

[17] R. Gentz and D. Stuber, “GENRIC APPROACH TO FACILITATE PURIFICATION OF RECOMBINANT PROTEINS WITH A NOVEL MEIAL CHELATE ADSORBENT,” p. 5, 1988.

[18] C. Mohan, “A guide for the preparation and use of buffers in biological systems,” p. 37.

[19] J. Schmitt, H. Hess, and H. G. Stunnenberg, “Affinity purification of histidine-tagged proteins,” *Mol. Biol. Rep.*, vol. 18, no. 3, pp. 223–230, Oct. 1993.

[20] U. K. Laemmli, “Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4,” *Nature*, vol. 227, no. 5259, pp. 680–685, Aug. 1970.