THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

DESIGNING OPTIMIZED PROTEIN MOLECULAR WIGHT MARKERS APPLICABLE TO BOTH COOMASSIE STAINING AND WESTERN BLOTTING

YOSHITAKA SHIBATA
SPRING 2020

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Song Tan
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Wendy Hanna-Rose
Department Head of Biochemistry and Molecular Biology
Professor of Biochemistry and Molecular Biology
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Molecular weight markers are essential tools to measure protein sizes by gel electrophoresis. However, conventional markers are not detected in Western blots which use antibodies to recognize specific proteins. Together with undergraduate colleagues, I have designed and prepared 11 recombinant proteins of defined molecular weights from 10 to 250 kDa that contain *Staphylococcus aureus* Protein A IgG antibody binding domains through subcloning. These recombinant proteins can be visualized on polyacrylamide gels with Coomassie Blue stain and also on Western blots with non-specific secondary antibodies. In addition to migrating appropriately on SDS-PAGE gel, each recombinant protein was selected for high level expression in *E. coli* and contains a HIS tag for efficient metal affinity purification. In addition, two set of polycistronic expression vectors were created to simplify the expression and purification processes. The first vector co-expresses 10, 30, 50, and 100 kDa proteins, and the second vector co-expresses 20, 40, 60, and 80 kDa proteins at 21°C. The Penn State protein molecular weight markers should provide an inexpensive method to produce protein molecular weight markers for research laboratories.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have received a tremendous amount of support and guidance from many people during the past four years of my time in the Tan laboratory and at Penn State, and I wish to thank everyone in the lab who has helped me along the way. First and foremost, I wish to thank Dr. Song Tan for everything you have done for me. Thank you for believing in me and giving me an opportunity to be a part of your lab in the very beginning. It was never easy for me navigating through my research projects, and there were a number of obstacles along the way, but you have always been there to guide and assist me in finding solutions. During my time in the Tan lab, I not only mastered various fundamental techniques of molecular biology but also learned the skills of critical thinking; you taught me not only how to do research but also how to approach and solve a problem. This is something that will benefit me in the future as I plan to continue my education and career in research. I would also like to thank fellow undergraduate students Ryan Santilli and Jack Williamson for always having my back and working with me in completing our project. I am also very grateful for my upperclassmen mentor Rosie Sowers for your patience in training me as I started in the lab and your contribution to our research project. Additionally, I would like to thank all of the lab technicians, Lauren McCarl, Bryan Tornabene, Jeff Hall, Gabe Epstein, Owen Karsmarski, and Szu-yu Kuan, who were always there when I needed assistance and who provided many resources for our experiments. I appreciate their dedications in ensuring our project going smoothly. Finally, I would like to thank my honors adviser, Dr. Wendy Hanna-Rose, for your guidance since the summer before my freshman year. It was truly a privilege for me to get to know and work with everyone during my undergraduate years, and I am very grateful for everything you all have done to make me who I am today.

This work was supported by Penn State Eberly College of Science Undergraduate Research grants and NIH grants R01GM088236, R01GM111651 and R35 GM127034 to Dr. Song Tan. Any opinions, findings, and conclusions or recommendations expressed in this thesis are those of the author and do not necessarily reflect the views of Penn State Eberly College of Science and NIH.

Chapter 1

Introduction

1.1 SDS-PAGE and Western Blotting

Protein analysis is an important aspect of biochemistry and molecular biology research. One of the most common and economical way to measure protein molecular weight is sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). In SDS-PAGE, proteins are denatured by SDS resulting in a uniform charge to mass ratio. Proteins are then separated based solely on their sizes on a polyacrylamide gel. SDS is a dissociation agent that denatures secondary and non-disulfide-linked tertiary structures and adds negative charges to all proteins to differentiate proteins by their molecular weights only (Brunelle and Green 2014). While all proteins travel towards the anode because of the negative charges proportional to their lengths, proteins with smaller molecular weights travel faster towards the bottom due to sieving effects, in which smaller molecules can pass through pores faster (Fritsch and Krause 2003).

Although the protein migration patterns generally correspond to the molecular weights, proteins can migrate anomalously due to irregular detergent binding that disrupts the charge to mass ratio (Rath *et al.* 2009). One study showed that transmembrane proteins tend to interact with more SDS molecules because their hydrophobic helical structures attract hydrophobic tails of SDS, resulting in more negatively charged SDS head groups facing outside (Rath *et al.* 2009). There are many factors that can influence protein migration patterns, so it is important to understand the identity and characteristics of proteins that are being analyzed in SDS-PAGE.

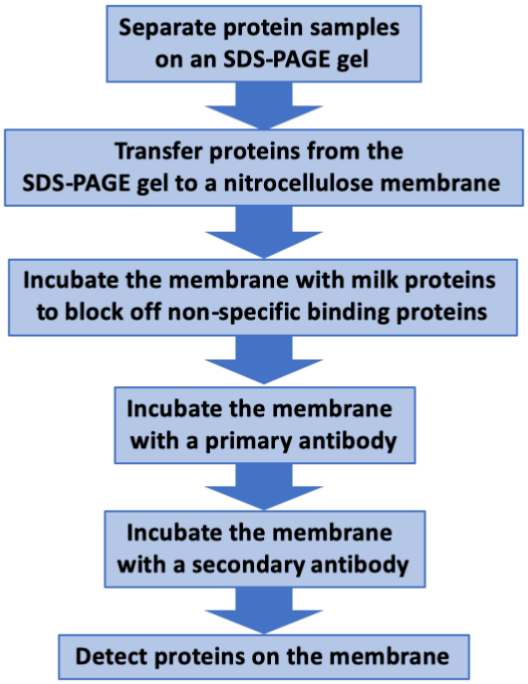


Figure 1.1 Flow chart of Western blotting

Proteins on an SDS-PAGE gel can be visualized through various staining methods, but most of these methods only allow non-specific staining of proteins. In other words, all proteins present in a sample will be visualized through staining. If one wishes to identify specific proteins based on their characteristics, one way to detect and quantify specific proteins is allowing proteins to interact with antibodies engineered to emit signals. This method, done in Western blotting, allows a variety of antigen-specific antibodies to interact with proteins, so only proteins that can bind specific antibodies can be visualized with special imaging techniques. During Western blotting, the proteins are transferred from an SDS-PAGE gel onto a nitrocellulose membrane because an SDS-PAGE gel is not an ideal support for protein-antibody interaction due to the lack of sensitivity for antibodies. It is the best to transfer protein samples to a more antibody-sensitive nitrocellulose membrane for more efficient antibody binding. Another benefit

of transferring proteins to a membrane is that proteins are immobilized on the membrane for better antibody interaction (Blancher and Jones 2001). Similar to that in SDS-PAGE, the rate of electrophoretic elution of proteins, or the migration of proteins, from the gel to the membrane depends on molecular weights, but the duration of electrophoresis can also impact whether proteins are transferred to the membrane completely (Burnette 1981). After the successful transfer of protein samples, the membrane is incubated with primary and secondary antibodies in order to identify proteins based on their characteristics in antibody binding.

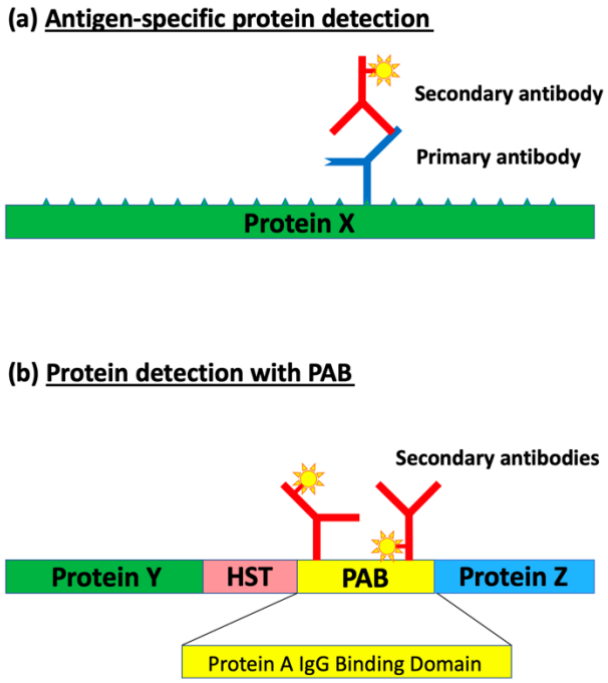


Figure 1.2 Antibody-protein recognition

Proteins X, Y Z = fusion proteins incorporated in recombinant constructs. HST = polyhistidine tag for metal affinity chromatography. PAB = Protein A IgG binding domain for antibody binding

Primary antibodies are usually specific to the target proteins because the binding occurs at the variable domain of an antibody, as shown in part (a) of Figure 1.2, so one specific primary antibody would theoretically only bind to one specific protein. In contrast, secondary antibodies

bind primary antibodies or antigens at the constant domain (Murphy and Weaver 2017). Therefore, secondary antibodies are ideal candidate as protein detection tools by conjugating with molecules that can emit signals for detection: some secondary antibodies are radiolabeled and can be detected through radioactivity, and others are conjugated with chemiluminescent molecules that emit light upon exposure (Burnette 1981). All antibodies only bind to specific binding domains or epitomes, so the proteins that lack the binding sites cannot be recognized or detected in Western blotting.

1.2 Protein Molecular Weight Markers

SDS-PAGE and Western blotting are commonly used techniques for biochemical studies of proteins (Brunelle and Green 2014, Burnette 1981). One main purpose of these two methods is isolation and identification of proteins based on their sizes and characteristic antigens, and protein molecular weight markers play an essential role in providing a molecular weight standard for protein size measurements (Blancher and Jones 2001). The conventional molecular weight markers currently available can be visualized on an SDS-PAGE gel through either a pre-staining or post-staining technique (Kumar 2018). Pre-stained markers are treated with dyes and can be visualized during SDS-PAGE. Markers that are not pre-stained can be stained with a dye after electrophoretic separation. However, a problem arises when the proteins are transferred from the SDS-PAGE gel to a nitrocellulose membrane in Western blotting because most of these marker proteins lack antibody binding domains for antibodies used to probe proteins. After blotting, proteins are incubated with primary and secondary antibodies, which are conjugated to enzymes or fluorophores. These conjugated antibodies bound to proteins can either produce colored

precipitates or light that can be detected. For chemiluminescence detection that is commonly used, a nitrocellulose membrane with antibody-bound proteins is treated with a chemiluminescent reagent, and the light emitted is detected on an X-ray film. Because almost none of the conventional protein marker bands are detected, scientists have to manually overlay the positions of protein marker bands on the film. However, this process is unreliable since the outcome depends on the proportion and fit of the SDS-PAGE gel, the nitrocellulose membrane, and the X-ray film (Schüchner *et al.* 2016). If the marker band positions were recorded incorrectly, then the sample proteins would not align accurately with the markers and cannot be analyzed correctly. There are a very few molecular weight marker products available that can be visualized in both SDS-PAGE and Western blotting, but they still have limitations with antibody recognition. Bio-Rad Laboratories, Inc. has a selection of Western blotting protein standards. However, the marker proteins are tagged with *Strep*-tag that can only be recognized by Strep Tactin conjugates instead of primary or secondary antibodies (Bio-Rad Laboratories, Inc.). These products still require additional steps for antibody interaction, and there is a potential difference in detection intensity between the Strep Tactin conjugates and secondary antibodies used. Using a different approach, Kao *et al.* developed a set of molecular weight markers that contains 14 commonly used epitope tags, and the study describes that both marker proteins and epitope-tagged sample proteins can be simultaneously detected in Western blotting by using anti-tag antibodies (Kao *et al.* 2012). These marker proteins also have the limitations in the applicable antibodies because the proteins being tested have to contain one of the 14 epitope-tags.

A solution to these problems that we proposed is the insertion of IgG antibody binding domains of the *Staphylococcus aureus* Protein A into each marker protein. The IgG binding

domains of Protein A are recognized by a wide range of secondary antibodies used in Western blotting (Nilsson *et al.* 1987). Because IgG antibody is the most abundant type of antibody in many organisms, the IgG binding domain essentially allows universal binding of antibodies (Murphy and Weaver 2017). The interaction between proteins and conjugated secondary antibodies allows detection of target proteins on a nitrocellulose membrane (Burnette 1980). Therefore, all proteins containing an IgG binding domain can theoretically be detected on a membrane in the Western blot assay. By incorporating IgG binding domains into the protein markers, these markers can be used in both SDS-PAGE and Western blotting and eliminate potential errors caused by manual positioning of the protein markers. In addition, because the main goal of this project was to provide an efficient and inexpensive means of producing universal protein molecular weight markers appropriate for both SDS-PAGE and Western blotting, I worked with two other undergraduate students, Ryan Santilli and Jack Williamson, and our research adviser Dr. Song Tan to optimize several features of the molecular weight markers: the size accuracy of each molecular weight marker, the migration pattern of each marker on SDS-PAGE gel, the detection intensity of each marker in western blotting, protein expression and purification conditions, and the polycistronic expression system that co-expresses multiple proteins simultaneously.

1.3 Protein A IgG Binding Domain

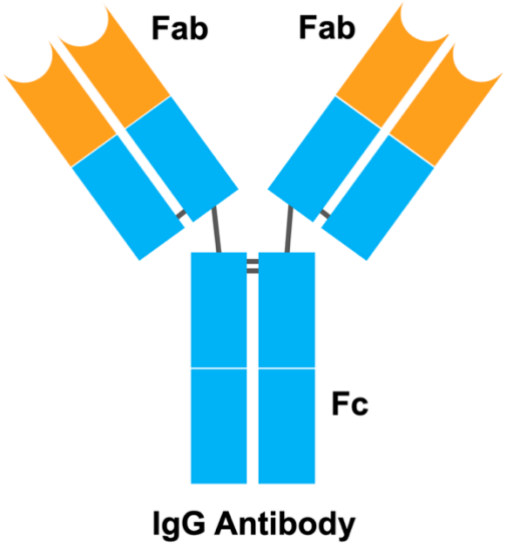


Figure 1.3 Structure of IgG antibody

Staphylococcus aureus Protein A is a commonly used immunological tool for studying immunoglobulin binding interaction with antigens, one of the most studied interactions between proteins (Jansson *et al.* 1998). There are five IgG binding domains of Protein A (domains E, D, A, B, C in order starting from the N-terminus), and each of them is sufficient to bind to the Fab or Fc fragment of an IgG antibody (Nilsson *et al.* 1987). Illustrated in Figure 1.3, the Fab fragment of an antibody contains the amino-terminal end of a heavy chain and the entire light chain of IgG antibody, and this region also forms an antigen-binding site (Murphy and Weaver 2017). Fc fragment, or Fc region, is the crystallizable carboxyl-terminal end of the IgG antibody heavy chains, and it contains very low to no antigen-binding activity (Murphy and Weaver 2017). This is partially due to the fact that the Fc region is highly conserved across IgG antibodies of most species (Murphy and Weaver 2017). Therefore, the ability of IgG binding domains of Protein A to bind both Fab and Fc fragments of IgG antibodies makes them the ideal

candidates to be incorporated into the universal molecular weight marker proteins (Figure 1.2, part b). The binding specificity of primary antibodies will not be a limiting factor in using the molecular weight markers with internal Protein A IgG binding domain because the markers directly interact with a wide range of secondary antibodies.

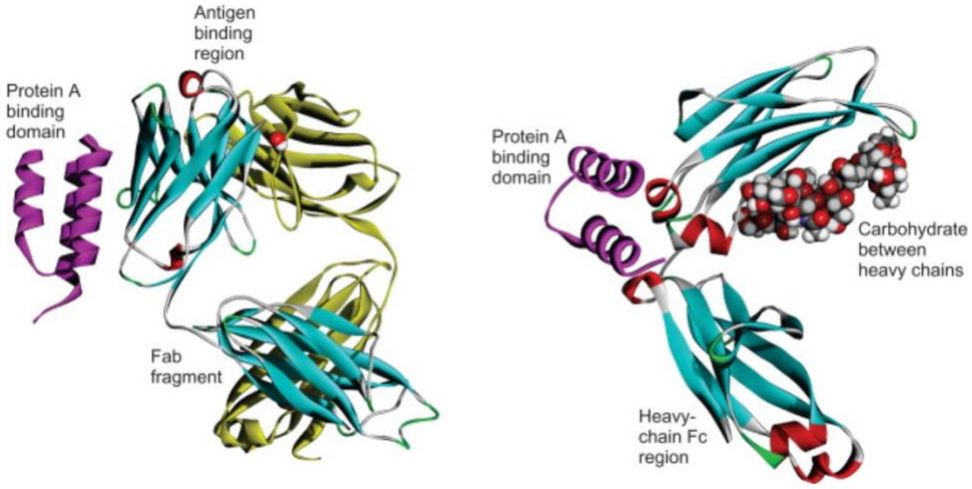


Figure 1.4 Structures of protein A binding domain and antibody fragments

The protein A binding domain is indicated by the purple double helix structure, which interacts with both Fab and Fc fragments of an antibody (Hermanson 2013, Figure 15.98).

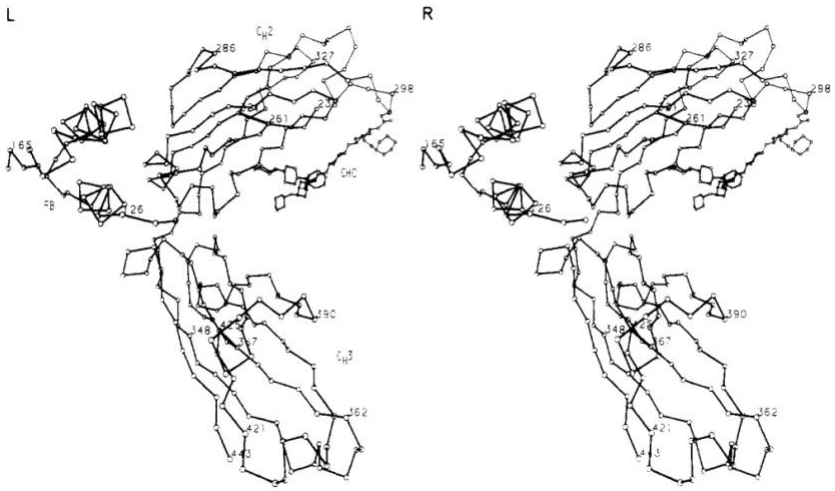


Figure 1.5 Structures of protein A binding domain B and antibody Fc fragments

A stereo drawing of a complex of Protein A IgG binding domain B and Fc fragments of antibodies (Deisenhofer 1981, Figure 6).

Among the five binding domains of Staphylococcal Protein A, domain B was the best candidate for the marker protein constructs because its sequence is closest to the hypothetical ancestral consensus sequence of the IgG binding domains (Moks *et al.* 1986). In addition, a study by Johann Deisenhofer reports the structure of protein A domain B-IgG antibody complex via crystallization (Figure 1.5), so the atomic details of interaction between the binding domain B and IgG is better understood compared to those of the other domains (Deisenhofer 1981). Because protein A binding domain B can bind both Fab and Fc regions of IgG antibodies, which is one of the most common types of immunoglobulins in many mammals, there are many commercially available secondary antibodies derived from different mammals that can bind to the binding domain B. Therefore, the molecular weight markers with the IgG binding domain can be used in various conditions that require different secondary antibodies. In addition, the binding domain C of protein A was also incorporated into several marker protein constructs in order to enhance the antibody binding efficiency and detection in Western blotting based on the experiment results we had (more details in Results and Discussion Section 3.3). The binding domain C is also ideal to be incorporated into the protein markers because its sequence is highly conserved just like that of the binding domain B; both domains have no changes in the α helices that are important for antibody binding (Moks *et al.* 1986).

1.4 Polycistronic Expression Vector

Polycistronic cloning allows co-expression of multiple proteins from a single expression vector. This technique has played an essential role in improving purification efficiency and reconstitution of protein complexes that have multiple subunits. Molecular cloning of

recombinant plasmids expressing various types of proteins is a commonly used method in biochemical studies of proteins. However, because many protein components actually form complexes with other components *in vivo*, the reconstitution of protein subunits is required after the purification of each subunit. There are other methods for protein complex reconstitution, including *in vitro* protein reconstitution, which requires the protein components to be individually expressed and combined together (Tan *et al.* 2005). In addition, co-infection of insect cells with multiple baculoviruses expressing single protein subunits is another way to prepare protein complexes, but this method may potentially yield incomplete complexes due to stoichiometric difference in subunit expressions (Tirode et al. 1999). The polycistronic expression system simplifies the process by enabling co-expression of multiple protein components in a cell to allow *in vivo* reconstitution (Tan 2001). This method also allows the protein complexes to be purified as a unit, which can then be used for biochemical studies directly. The pST44 expression system used in this project (Figure 1.6) was developed by Dr. Song Tan and his colleagues (Tan *et al.* 2005). The system allows four protein sequences to be inserted into the coding region of the vector for co-expression; each protein construct is first inserted into a translation cassette vector with a unique pair of restriction endonucleases. The four cassettes are then cleaved and transferred into the pST44 expression vector sequentially.

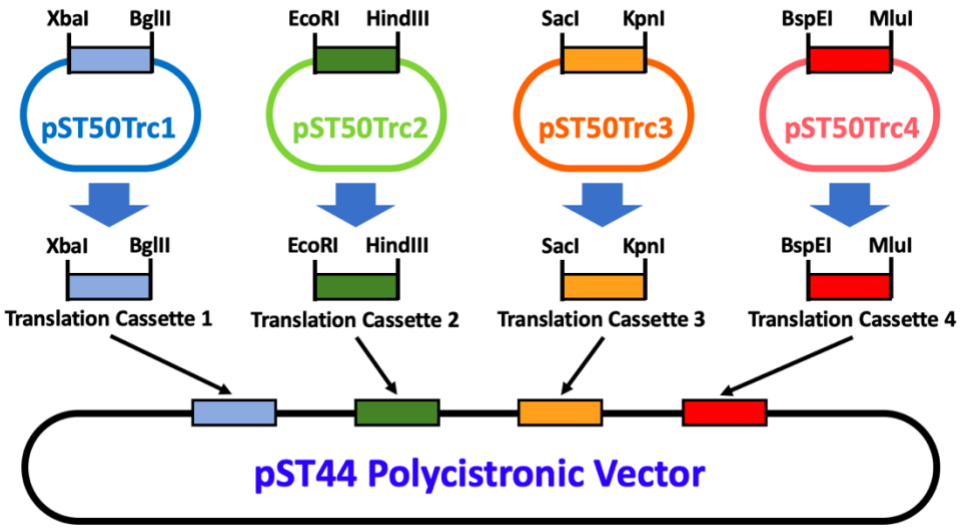


Figure 1.6 Structure of pST44 polycistronic vector

Because one of the goals for my project was to develop a more efficient protocol for preparing the molecular weight markers, polycistronic expression vector was a solution to reduce the number of plasmids needed to express all protein markers. By incorporating four protein constructs into one polycistronic vector, all four proteins were expressed simultaneously; it is also worth noting that the four co-expressed proteins do not form a protein complex in this case. Moreover, the time required to express and purify the set of eight molecular weight marker proteins was reduced greatly as a result.

1.5 Metal Affinity Chromatography

Affinity chromatography is commonly utilized to separate a desired protein from other proteins in a sample based on specific binding characteristics of the protein. In this procedure, the target protein binds to its affinity ligand cross-linked to a resin, while all other proteins that do not have affinity for the specific ligand do not bind to anything and will be discarded. The

type of ligand that the protein interacts with can be a variety of biological materials including inhibitor, substrate, antibody, antigen, and subunit (Urh *et al.* 2009). Affinity chromatography is widely used because it does not require any preceding processes to remove cellular materials other than protein for protein purification (Lichty 2005). Specifically, metal affinity chromatography results in purities of up to 95% with 90% recovery of the tagged proteins readily in a single purification process because of the great stability of metal ions attached to resins via four coordination sites and their ability to remain functional via two transition metal coordination sites under various conditions (Bornhorst and Falke 2000).

In my project, metal affinity chromatography, exploiting the interaction between a transition metal ion immobilized on the resin and a sequence of histidine residues, was used to separate the recombinant proteins from other materials. All of the marker proteins designed are tagged with a sequence expressing 10 histidine residues because a cobalt- or nickel-charged resin interacts strongly with the imidazole side chain of histidine residue (Clontech Laboratories, Inc. 2007). While a polyhistidine-tagged protein can bind to the resin, other materials will be washed off. In addition, the interaction between the resin and polyhistidine-tagged protein is reversible. In order to elute the target protein from the resin, a high concentration of free imidazole molecules is added to the resin; because imidazole has strong affinity to the metal ions and can compete with the histidine side chain, the polyhistidine-tagged protein is released from the resin. In order to develop efficient means of producing protein markers, the recombinant marker proteins were expressed on a 100 mL scale and purified with metal affinity chromatography. Along with two other undergraduate students, I prepared a set of universal protein molecular weight markers that can be separated on an SDS-PAGE gel and also detected in a Western blot.

Chapter 2

Materials and Methods

2.1 Nomenclature

11 molecular weight marker protein constructs were created. For the reason of simplification, all constructs and corresponding proteins are referred as their molecular weights in kilodalton (kDa).

Table 2.1 List of recombinant protein markers

Molecular Weight (kDa)	Protein Name
10	HSTPABHPCx1
20	STRHSTPABPACCBPx3
30	HSTPABPACS100BCBP
40	HSTPABPACGST
50	STRHSTPABMBPt1
60	HSTPABPACdRCC1t8
80	STRHSTPABPACIL1bt1dRCC1t8
100	STRHSTPABPACIL1bt1QRS
150	STRHSTPABMBPpepN
240	STRHSTPABPACIL1bt1QRSMBPpepN
250	STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN

Following are the definitions of the abbreviations of the fusion protein components:

- HST = 10x polyhistidine tag that binds to cobalt and nickel metal affinity resins
- STR = Strep II tag that binds to Strep Tactin
- PAB = Protein A IgG binding domain B
- PAC = Protein A IgG binding domain C
- HPC = Haemolysin co-regulated protein
- S100B = S100 calcium binding protein

- CBP = Calmodulin binding peptide that binds to calmodulin
- GST = Glutathione S-transferase
- MBP = Maltose binding protein that is also a chaperon to guide folding of fusion protein
- dRCC1 = Drosophila regulator of chromosome condensation 1
- IL1b = Interleukin 1 beta
- QRS = Glutaminyl-tRNA synthase
- pepN = Aminopeptidase N

2.2 Subcloning

In order to create recombinant protein expression plasmids that express the protein molecular weight markers with IgG binding domains, the technique of molecular cloning was utilized for incorporation of protein expression sequences into expression vectors.

2.2.1 IgG Binding Domain Insert PCR Amplification

PCR, or polymerase chain reaction, is a technique to amplify a specific fragment of DNA using compatible forward and reverse primers in order to obtain sufficient amounts of a gene to clone into an appropriate cloning vector. These primers were designed as complementary to the template DNA sequence, and each primer contained the sequence of a specific restriction endonuclease cutting site for digestion in subcloning process. PCR was utilized to amplify protein expressing DNA inserts from templates for subcloning of the recombinant expression vectors. The template sequences of Protein A IgG binding domains were obtained from a

synthetic sequence, and other fusion protein expression sequences were obtained from recombinant expression vectors previously made in the lab (work completed by Dr. Tan). The PCR reaction mixture was prepared by adding 58.5 μL water, 20 μL of 5x Q5 buffer, 10 μL of 2.5 mM dNTP, 1 μL of template DNA, 5 μL of 10 μM forward primer, 5 μL of 10 μM reverse primer, and 0.5 μL of 2 units/μL Q5 polymerase (see appendix A for reaction mixtures). The PCR mixture was placed in a thermal cycler for amplification. The reaction was set for the following durations and temperatures: 98°C for 30 seconds, 5 cycles of [98°C for 5 seconds, primer annealing temperature minus 5°C for 30 seconds, 72°C for 30 seconds], 25 cycles of [98°C for 5 seconds, 60°C for 30 seconds, 72°C for 30 seconds], 72°C for 2 minutes, and hold at 15°C until the mixture was taken out of the cycler. 5 μL of the PCR mixture was combined with 1 μL of 6x gel loading buffer (GLB) [0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol, 60 mM EDTA] and loaded on to a high gelling temperature (HGT) agarose gel for analysis (see detailed protocol for agarose gel electrophoresis in Section 2.2.4). The remaining of the PCR mixture was treated with one phenol/CIA extraction, one CIA extraction, and ethanol precipitation (see detailed protocols in Sections 2.2.2 and 2.2.3).

2.2.2 Phenol/CIA or CIA Extraction

Phenol/CIA or CIA extraction is a method to extract proteins and isolate nucleic acid in the aqueous phase. CIA is chloroform-isoamyl alcohol [24 volumes chloroform, 1 volume isoamyl alcohol], and phenol/CIA is a 1:1 mixture of TE-equilibrated phenol and CIA. The phenol/CIA extraction was prepared by adding equal volumes of DNA sample and phenol/CIA mixture. The mixture was vortexed at full speed for 15 seconds and centrifuged in a

microcentrifuge at 13000 rpm at room temperature for 1 minute. The aqueous phase at the top was transferred to a new Eppendorf tube. The CIA extraction was done in the same way but using CIA mixture instead of phenol/CIA mixture.

2.2.3 Ethanol Precipitation of DNA

Purification of DNA via phenol/CIA or CIA extraction was followed by ethanol precipitation of DNA. For 1 volume of DNA sample, 0.1 volumes of 3 M NaAc pH 5.2 and 2.5 volumes of 100% ethanol were added. The mixture was vortexed for 5 seconds and centrifuged in the microcentrifuge at 13000 rpm at room temperature for 10 minutes. The supernatant was aspirated, and the pellet was dried completely and resuspended in 30 μL of TE(10, 0.1) [10 mM Tris, 0.1 mM EDTA].

2.2.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis is a method for separating DNA fragments based on fragment sizes because shorter DNA fragments migrate faster on an agarose gel while longer DNA fragments migrate slower. The migration pattern of DNA samples on an agarose gel also depends on the concentration of agarose in the gel.

Table 2.2. Agarose Concentrations for Different DNA Fragment Lengths

DNA Fragment length	Required Agarose Concentration
< 300 bp	2%
300-600 bp	1.5%
600-1500 bp	1.0%
1500-3000 bp	0.7%

The desired concentration of high gelling temperature (HGT) agarose was combined with 30 mL of 0.5x TBE gel buffer [45 mM Tris, 45 mM boric acid, 1.5 mM EDTA] in a 125 mL Erlenmeyer flask (refer to Table 2.2 for the concentration of agarose required). 3 mL of deionized water was also added to compensate for water loss due to boiling in the microwave. The agarose was dissolved in 0.5x TBE gel buffer by heating in microwave for 90 seconds. The solution was mixed gently and heated for another 15 seconds. While the agarose solution was allowed to cool down, a gel-casting apparatus was set up and leveled. 1.5 μL of 10 mg/mL ethidium bromide was mixed into the agarose solution, and the agarose solution was poured into the gel-casting setup. A 10-well comb, holding 30 μL samples, was installed for preparative gels used for agarose gel purification; a 15-well comb, holding 10 μL samples, was installed for analytical gels used for PCR screening and restriction mapping. The agarose solution was allowed to cool and solidify for at least 30 minutes. Once the gel was solidified, the casting tray containing the agarose gel was transferred to the agarose gel electrophoresis box with the comb side aligned with the negative electrode (black). 300 mL of 0.5x TBE gel buffer mixed with 10 μL of 10mg/mL ethidium bromide was poured into the box until the gel is completely covered, and the comb was carefully removed from the gel. Each DNA sample was mixed with 6x gel loading buffer (GLB) [0.25% bromophenol blue, 0.255 xylene cyanol, 30% glycerol, 60 mM EDTA] with a 5:1 ratio of sample to GLB. Samples were loaded into the wells along with appropriate DNA ladders (Penn State 1 kb or 100 bp DNA ladders). The gel was run at 125 V for about 40 minutes or until the bromophenol blue dye reached the bottom of the gel. The gel was then photographed in a UV transilluminator to visualize the DNA bands in each well. Ethidium bromide was handled with care with gloves due to its carcinogenic property.

2.2.5 Restriction Endonuclease Digestion

The PCR product and DNA plasmids were digested with two appropriate restriction endonucleases, which cleave double-stranded DNA at specific recognition sites. Double restriction enzyme digestion was done to prepare an insert and vector for ligation by cutting DNA sample into fragments with desired length and specific ends. These ends are unique based on the enzymes used, so two different fragments with the compatible ends could attach together. The digestion mixture was prepared by adding appropriate amounts of water, DNA sample (PCR product or plasmid), 10x NEB restriction enzyme buffer (New England BioLabs), 100 mM DTT, and restriction enzymes (see appendix A for details). The digestion mixture was incubated at 37°C for 2 hours, and the reaction was stopped by addition of 0.2 volumes of 6x GLB. The sample was run on an agarose gel, and the desired DNA fragment was purified via agarose gel purification (see detailed protocol in Section 2.2.6).

2.2.5.1 Phosphatase Treatment

Vectors digested with single restriction endonuclease should be treated with phosphatase to prevent self-ligation, in which linearized vectors re-ligate to take the circular form. During the phosphatase treatment, the phosphate group at the 5’ ends of DNA fragments are removed by the enzyme. Because this 5’ phosphate is very important in forming interaction with the new nucleotide of the elongating DNA sequence, the absence of the 5’ phosphate prevent 3’ end from ligating with the phosphate-stripped 5’ end. Only an compatible end with phosphate attached can ligate with the phosphatase-treated 5’ end, so none of the fragments cleaved out in the vector sample can re-ligate after phosphatase treatment. 0.1 units of calf-intestinal phosphatase (CIP)

was added to the digested DNA sample and incubated at 50°C for 30 minutes. 0.2 volumes of 6x GLB was added to terminate the reaction. The sample was run on an agarose gel, and the desired DNA fragment was purified via agarose gel purification (see detailed protocol in Section 2.2.6).

2.2.5.2 Partial Digestion

If the desired DNA vector or insert contains an internal copy of the restriction enzyme recognition site that is also one of the two enzymes used in the double digestion, partial digestion can be used to separate the desired partial digest product from the complete digest products. This method can only be used when the fragment size of the partial digest product is unique enough to be isolated from those of the complete digest products via agarose gel electrophoresis. In this project, Protein MBP-coding fragment was digested from plasmid pWM529-MBP. MBP fragment (1110 bp) included an internal BglII recognition site. First, the DNA sample (pWM529-MBP plasmid) was digested with restriction enzyme BamHI. The digestion mixture was prepared by adding 44 μL water, 5 μL DNA, 6 μL 10x NEB restriction enzyme buffer (New England BioLabs), 2 μL 100 mM DTT, and 3 μL 20 units/μL BamHI (see appendix A for details). The digestion mixture was incubated at 37°C for 2 hours. Then, 1 μL of 10 units/μL BglII was added to the BamHI-digested DNA sample for partial digestion. The mixture was incubated at 37°C for 30 minutes total, and 20 μL aliquots of the sample were removed at 5 minutes, 15 minutes, and 30 minutes to be mixed with 4 μL 6x GLB each. The 3 time point samples were run in successive lanes on a preparative HGT agarose gel, and the desired DNA fragment (1110 bp) was purified via agarose gel purification (see detailed protocol in Section 2.2.6).

2.2.6 Agarose Gel Purification

The desired DNA fragment is present in a mixture of other DNA fragments, digestive enzymes, and reaction buffer. The desired DNA fragment can be purified from other contents through a method using agarose gel electrophoresis. A digestion DNA sample containing desired insert or vector fragment was electrophoresed on a preparative HGT agarose gel (10 wells) at 125 V for about 40 minutes. However, a longer running time was required for separation of DNA fragments of similar sizes. While the gel was running, a filter assembly was prepared with a 0.5 mL Eppendorf tube, 1.5 mL Eppendorf tube, and siliconized glass wool. A hole was pierced at the bottom of the 0.5 mL tube using a heated 25 gauge needle, and a small amount of siliconized glass wool was stuff into the bottom of the 0.5 mL tube. The 0.5 mL tube was then placed into the 1.5 mL Eppendorf tube. After the gel was photographed in a UV transilluminator, the gel was viewed under long wavelength UV light, and the band of desired DNA fragment was cut out with a razor blade. The gel slice was placed into the 0.5 mL tube of the filter assembly, and the assembly was centrifuged in a microcentrifuge at 7000 rpm for 3 minutes. The gel-purified DNA fragment collected in the 1.5 mL Eppendorf tube was ready for ligation reaction.

2.2.7 Sticky-End DNA Ligation

The sticky-end DNA ligation method was used to fuse an insert and expression vector with compatible overhang strands resulted from restriction endonuclease digestion. Two reaction conditions were set up: reaction A only contained the vector DNA and is prepared as a negative control, and reaction B contained both the vector DNA and insert DNA. The ligation reaction mixture was made by combining 4 μL of water, 1 μL of 10x T4 DNA ligase buffer, 0.5 μL of

100 mM DTT, 2 μL of gel-purified vector DNA [3 ng/μL], 1.5 μL of gel-purified insert DNA (only in the tube B), and 1 μL of 40 units/μL T4 ligase (see appendix A for details). The ligation mixes were incubated at room temperature for 1 hour and used for transformation that followed.

2.2.8 Transformation

In order to grow and amplify the recombinant DNA plasmid in bacterial cells, the ligated plasmid was incorporated into competent cells through transformation. The competent cells used varied depending on the procedures following the transformation. For cloning and DNA isolation processes, TG1 *E. coli* cells were used. For protein expression and purification, BL21(DE3)pLysS *E. coli* cells were used. All competent cells were stored in aliquots of 100 μL at -80°C before use, and a 100 μL aliquot of cells was thawed on ice for each DNA sample. 2 μL of a DNA ligation sample (1 μL of DNA plasmid sample) was added to an aliquot of competent cells and incubated on ice for 15 minutes. The competent cell tube was then heat shocked in 42°C water bath for 30 seconds and placed on ice for another 20 seconds. 0.5 mL of 2xTY media [1.6% (w/v) bacto tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl] was added to each aliquot and incubated in 37°C shaking incubator for 15 minutes. After the 37°C incubation, 0.3 mL of cell sample was plated on the appropriate plate. TYE plates with ampicillin [1.0% (w/v) bacto tryptone, 0.5% (w/v) yeast extract, 0.8% (w/v) NaCl, 1.5% (w/v) agar, 100 μg/mL ampicillin] were used for TG1 *E. coli* cells, and TYE plates with ampicillin and chloramphenicol [1.0% (w/v) bacto tryptone, 0.5% (w/v) yeast extract, 0.8% (w/v) NaCl, 1.5% (w/v) agar, 100 μg/mL ampicillin, 25 μg/mL chloramphenicol] were used for BL21(DE3)pLysS *E. coli* cells. Each plate was incubated at 37°C for 10 to 18 hours.

2.2.9 PCR Screening

PCR screening is a fast and convenient method to confirm the presence of the insert DNA in the ligated plasmid before moving forward to isolate the recombinant DNA plasmid. Colonies from the transformation plate were picked and applied to PCR for amplification of insert region of the vectors. The forward and reverse primers were selected to bind DNA flanking the insert region to check the correct insert DNA length. 6 colonies were picked from the transformation plate with sterile loops and transferred to separate aliquots of 100 μL water, and the sterile loops were immediately re-streaked on a new TYE plate with ampicillin that was separated into 6 sections. This re-streaked plate was incubated at 37°C for 10 to 18 hours for plasmid isolation. The cell suspensions were vortexed for 3 seconds. A PCR reaction master mix was prepared by combining 116.1 μL of water, 18 μL of 10x Thermo Pol buffer, 18 μL of 2.5 mM dNTP, 9 μL of 10 μM forward primer, 9 μL of 10 μM reverse primer, and 0.9 μL of 2 units/μL Pfu polymerase (see appendix A for reaction mixtures). 19 μL of the reaction master mix was combined with 1 μL of cell suspension, and the PCR samples were placed in a thermal cycler for amplification. The reaction was set for the following durations and temperatures: 95°C for 2 minutes, 25 cycles of [95°C for 30 seconds, primer annealing temperature minus 10°C for 30 seconds, 75°C for 45 seconds] and hold at 4°C until the mixture was taken out of the cycler. Each PCR sample was mixed with 3 μL of 6x GLB and loaded on to a HGT agarose gel for analysis (see detailed protocol for agarose gel electrophoresis in Section 2.2.4). If bands with the desired DNA fragment size were present in the colony samples, 2 positive clones were selected from the 6 samples for isolation of the plasmid DNA.

2.2.10 Plasmid Preparation

A single colony picked from each of the 2 selected clones was transferred to 100 mL 2xTY media [1.6% (w/v) bacto tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl] in a 500 mL Erlenmeyer flask for cell growth. 100 μL of 100 μg/mL ampicillin was also added to the flasks, and the cells were incubated in the shaking incubator at 37°C for 10-18 hours. After the incubation, the cell cultures were poured into separate 250 mL centrifuge bottles and centrifuged in a tabletop centrifuge at 4000 rpm at room temperature for 5 minutes. All supernatant of each clone sample was poured and aspirated off; the cell pellet was resuspended in 5 mL of Lysis buffer [50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM NaEDTA] and transferred to a 50 mL polypropylene Falcon tube. Each clone sample was mixed with 10 mL of NaOH/SDS [0.2 M NaOH, 1% (w/v) SDS] and was shaken rigorously 5 times until the mixture was clear without clumps. The mixture was incubated on ice for 5 minutes to break open the *E. coli* cells. 10 mL of chilled KAc/HAc solution [5 M KAc, 2.5 M HAc] was added to the mixture and shaken rigorously for 5 times. The mixture was incubated on ice for 5 minutes to precipitate chromosomal DNA. The sample was spun in the tabletop centrifuge at 4000 rpm at room temperature for 3 minutes, and the clear supernatant was transferred to a new 50 mL polypropylene Falcon tube. 12.5 mL isopropanol was added to the mixture, which was mixed and incubated at room temperature for 5 minutes. The sample was spun in the tabletop centrifuge at 4000 rpm at room temperature for 5 minutes. The supernatant was poured off, and the pellet was resuspended with 0.5 mL 70% ethanol and transferred to a 1.5 mL Eppendorf tube. The sample was spun in a microcentrifuge at 13000 rpm at room temperature for 1 minute. The supernatant was aspirated off and spun for addition 30 seconds in the microcentrifuge at 13000

rpm at room temperature before being aspirated off completely again. Then, 125 μL of TE(10, 50) [10 mM Tris, to 50 mM EDTA] was added to the sample to resuspend the pellet until no large clumps remained in the resuspension. 1 μL of 10 mg/mL RNase A (DNase free) was added to the sample to incubate in 37°C water bath for 15 minutes. While the sample was incubated with RNase A, a Sephacryl S400 HR spun column was assembled for each sample. The spun column was made by placing the top half of a 1.5 mL Eppendorf tube onto a clean 5 mL polypropylene tube and placing a Gilson blue pipette tip stuffed with siliconized glass wool on the Eppendorf tube/polypropylene tube stand. The Gilson blue pipette tip was then filled with Sephacryl S400 HR resin equilibrated in TE(10, 0.1). The spun column was put into the tabletop centrifuge at 2000 rpm at room temperature for 3 minutes, and the liquids collected in the tube were discarded. The RNase-treated sample was spun in the microcentrifuge at 13000 rpm at room temperature for 1 minute, and the resulting supernatant was extracted with 150 μL of phenol/CIA mixture twice, followed by one extraction with 300 μL of CIA (see detailed extraction protocol in Section 2.2.2). The extracted sample was transferred to the Sephacryl S400 HR spun column and spun in the tabletop centrifuge at 2000 rpm at room temperature for 3 minutes. The eluted TE(10,0.1) solution containing the plasmid was transferred to a labelled Eppendorf tube for storage at -20°C.

2.2.11 UV Quantitation of DNA

After the plasmid sample was isolated, a UV spectrophotometer (NanoDrop) was used to determine the concentration and purity of the isolated DNA sample. The absorbance of the DNA was taken from 220 nm to 320 nm. 1.5 μL of water was applied to the spectrophotometer as the

blank, and 1.5 μL of the DNA sample was applied to the device for measurement. The spectrophotometer measured the DNA concentration, 260/280 ratio, and 260/230 ratio to also test the purity of the sample.

2.2.12 Restriction Mapping and Sequencing

Restriction mapping is a digestion method to cleave a plasmid in certain patterns in order to provide strong evidence that the plasmid is the desired construct. The ideal restriction mapping experiment confirms the correct ligation of an insert with a vector in the correct orientation. The experiment was carried out in two separate sets of restriction enzymes yielding two different fragment patterns distinguishable from that of the original template construct. In general, one enzyme that cuts once in the insert region and another enzyme that cuts once in the vector region were paired up in a double digestion. Each digestion mixture was prepared by combining an appropriate amounts of water, isolated plasmid DNA, 10x NEB restriction enzyme buffer (New England BioLabs), 100 mM DTT, and restriction enzymes (see appendix A for details). The mixtures were incubated at 37°C for 2 hours, and they were mixed with 0.2 volumes of 6x GLB to end the reactions. The samples were analyzed on an analytical HGT agarose gel (see detailed protocol for agarose gel electrophoresis in Section 2.2.4). If the plasmid samples produced the predicted size fragments that are distinctive from the parent vector, the plasmid was confirmed to include the desired insert and was sent for sequencing.

Plasmid constructs that incorporated inserts created from PCR require sequencing because there is a chance that a mutation was introduced during PCR process that would affect

the protein expression. The primers generally used for sequencing were T7 and T7term primers that recognize sequences flanking the insert region. 5 μL of plasmid DNA was aliquoted for sequencing, and sequencing analysis was carried out by the Genomics Core Facility at the Huck Institutes of the Life Sciences at the Pennsylvania State University. The sequencing result was analyzed with 4Peaks software.

2.2.13 Polycistronic Vector

A single polycistronic vector was designed to insert 4 different protein expressing sequences into 4 separate cassettes. These inserts were first incorporated into separate transfer cassette vectors with unique restriction enzyme recognition sites flanking each insert region. The cassette 1 in transfer plasmid pST50Trc1 is flanked by enzymes XbaI and BglII. The cassette 2 in transfer plasmid pST50Trac2 is flanked by enzymes EcoRI and HindIII. The cassette 3 in transfer plasmid pST50Trac3 is flanked by enzymes SacI and KpnI. The cassette 4 in transfer plasmid pST50Trac4 is flanked by enzymes BspEI and MluI. Then each protein expressing insert was sequentially transferred into pST44 polycistronic vector via subcloning. The first polycistronic vector containing 10, 30, 50, and 100 kDa protein sequences was subcloned in the order of 50, 100, 30, 10 kDa sequences. The second polycistronic vector containing 20, 40, 60, and 80 kDa protein sequences was subcloned in the order of 20, 60, 80, 40 kDa sequences.

2.3 Small-Scale Expression

Small-scale expression was carried out to test the expression level of each recombinant protein and the optimal expression temperature. The appropriate plasmid was transformed into

BL21(DE3)pLysS *E coli.* competent cells, and the TYE + ampicillin + chloramphenicol plate was incubated at 37°C for 12 to 16 hours. A labelled 500 mL Erlenmeyer flask of 100 mL 2xTY media, 100 μL 50 μg/mL ampicillin, and 100 μL 25 μg/mL chloramphenicol was inoculated with 3 colonies taken from the transformation plate. The culture was allowed to grow in 37°C shaking incubator at 220 rpm. For 37°C expression, after the culture became cloudy (after about 3 hours), the optical density at 600 nm (OD_{600}) of the culture was measured periodically until the OD_{600} reached the range of 0.5 to 0.9 (the measurement was blanked against 2xTY media). Once the OD_{600} was between 0.5 and 0.9, 100 μL of 0.2 M Isopropyl β-D-thiogalactopyranoside (IPTG) was added to the flask to induce protein expression. The time of induction was considered to be the Hour 0 of the expression. Before the induction, 250 μL of uninduced culture was transferred to a 1.5 mL Eppendorf tube and spun in the microcentrifuge at 13000 rpm at room temperature for 1 minute. The supernatant was aspirated off, and the pellet was resuspended in 50 μL of protein gel loading buffer (PGLB) [0.5 M Bis-Tris pH 6.8, 20% (v/v) glycerol, 10% (w/v) SDS, 5 M 2-mercaptoethanol, 0.4 mg/mL bromophenol blue]. For 37°C expression, 125 μL of the induced culture was transferred to an Eppendorf tube every hour after induction until 3 hours. Each transferred sample was spun in the microcentrifuge at 13000 rpm at room temperature for 1 minute. The supernatant was aspirated off, and the pellet was resuspended in 50 μL PGLB. At 3 hours after induction, the cell culture was transferred to two 50 mL Falcon tube and spun in the tabletop centrifuge at 4000 rpm at room temperature for 10 minutes. The supernatant was poured off, and the cell pellet was resuspended in 10 mL of P300 – EDTA buffer [50 mM sodium phosphate pH 7.0, 300 mM NaCl, 1 mM benzamidine, 5 mM 2-mercaptoethanol]. The cell sample was flash frozen with liquid nitrogen and stored at -20 °C.

The expression time point samples were analyzed on an SDS-PAGE gel (see detailed protocol of SDS-PAGE in Section 2.5). For expression at a lower temperature (21°C), the cell growth was started in 37°C shaking incubator was well. However, when the OD_{600} reached the range between 0.05 and 0.15, the cell culture was transferred to a 21°C shaking incubator. Once the OD_{600} reached between 0.5 and 0.9, the culture was induced with 100 μL of IPTG. After the Hour 0 uninduced cell sample was collected, the culture was incubated overnight for 14 to 16 hours, and the last time point sample was collected before the cells were harvested. The sample grown at 21°C was harvest and stored in the same methods as the 37°C sample.

2.3.1 Solubility Test

The solubility of overexpressed protein was checked by examining the amounts of expressed protein in the whole cell extract, pellet, and supernatant of the cell sample. The frozen resuspended cells were thawed by immersing the tube in lukewarm water. The sample was transferred to an sonication cup (bottom half of Vivaspin 20 centrifuge concentration device) and sonicated for 14 seconds twice in Branson S-450D sonicator at 50% maximum power. The sample was placed on ice for 20 to 30 seconds between sonication steps. 25 μl of the sonicated sample was mixed with 25 μL PGLB and labelled as “WCE” (whole cell extract). Another 0.5 μL of the sonicated sample was transferred to an Eppendorf tube and spun in the microcentrifuge at 13000 rpm at room temperature for 5 minutes. The supernatant was transferred to a new Eppendorf tube, and the pellet was spun for additional 30 seconds, and the remaining supernatant was aspirated off. The pellet was resuspended with 0.5 mL of P300 – EDTA. 25 μL each of the supernatant and pellet samples was mixed with 25 μL PGLB and labelled as “Supernatant” and

"Pellet." The three samples were analyzed on an SDS-PAGE gel along with small-scale purification samples.

2.4 Small-Scale Batch Purification

The small-scale purification was utilized in purifying polypeptides fused with His tag because the Talon metal affinity resin used in this purification method binds the imidazole side chain of histidine with high affinity. This purification step was carried out immediately following the solubility test (Section 2.3.1) The sonicated cell sample from the solubility test was aliquoted to 4 Eppendorf tubes with 1.3 mL in each . The tubes were spun in the microcentrifuge at 13000 rpm at room temperature for 3 minutes. 25 μL of the centrifuged sample labelled as "Talon Fraction 0" was mixed with 25 μL PGLB. While the cell sample was being centrifuged, ABT cobalt Talon resin was prepared by transferring 1 mL of resuspended Talon resin (about 50% suspension) to a 15 mL Falcon tube. 10 mL of water was added to the tube and spun in a tabletop centrifuge at 1800 rpm at room temperature for 2 minutes to sediment resin. The supernatant was discarded, and 10 mL of P300 – EDTA was added to the resin. The tube was mixed several times and spun in the tabletop centrifuge at 1800 rpm at room temperature for 2 minutes, and the supernatant was discarded. At this point, the supernatant from the centrifuged sonicated cell sample was transferred from the 4 tubes to the 15 mL Falcon tube with washed Talon resin. The sample was incubated at room temperature on a rotator for 20 minutes. The incubated sample was then spun in the tabletop centrifuge at 1800 rpm at room temperature for 5 minutes, and the supernatant was transferred, with a P1000 Pipetman, to a 15 mL Falcon tube labelled "Talon FT." 10 mL of P300 – EDTA was added to the resin tube, mixed several times, and spun in the

tabletop centrifuge at 1800 rpm at room temperature for 5 minutes. The supernatant was transferred to a 15 mL Falcon tube labelled “P300 Wash A.” Another 10 mL of P300 – EDTA was added to the resin tube, mixed several times, and spun in the tabletop centrifuge at 1800 rpm at room temperature for 5 minutes. The supernatant was transferred to a 15 mL Falcon tube labelled “P300 Wash B.” While the sample was centrifuged, a BioRad BioSpin column stand was set up by clamping the column to a retort stand. The resin was resuspended in 3 mL P300 – EDTA and transferred to the BioSpin column. The flow through was collected in a 15 mL Falcon tube labelled “P300 Wash C.” After the last wash, the BioSpin column with the resin was positioned above a rack with 4 open Eppendorf tubes labelled “Fr 1” to “Fr 4.” An elution buffer was created by adding 200 mM imidazole to 10 mL P300 – EDTA. Total of 3 mL of the elution buffer was added to the column, and about 0.5 mL of the flow through was collected for each of the 4 fractions. The fractions were prepared for SDS-PAGE by mixing 25 μL fraction sample with 25 μL PGLB, and the remaining fraction samples were mixed with 25% (w/v) glycerol for long-term storage at -20°C.

2.5 SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was used to analyze the expression and purification levels of proteins.

2.5.1 Preparing SDS-PAGE Gels

SDS-PAGE gels were made in a block consisting about 10 gels. A Mini-Protean II gel-pouring block was prepared by stacking pairs of clean long glass plate and short glass plates with two spacers sandwiched in between, and acrylic blocks were used to fill up the remaining space.

60 μL of 18% acrylamide separating gel was prepared by adding 8 mL of water, 36 mL of 30%/0.5% acrylamide/Bis, 120 μL of 15 BPB in ethanol, 15 mL of 3 m Tris-Cl pH 8.8, 600 μL of 10% SDS, 60 μL of TEMED, and 240 μL of 25% ammonium persulfate. The mixture was deaerated before adding the last three ingredients, and the separating mixture was quickly injected into the gel assembly through the stopcock valve and inlet port using a 60 mL syringe. Injecting was alternated with sucking to achieve even filling without air bubbles all the way up to about 0.5 cm below the bottom of the comb when installed. Water-saturated 1-butanol was poured on the top to form flat surface, and the gel mix was allowed to polymerize until completely solidified. Once the separating gels are completely polymerized, the butanol was drained and washed off with ethanol and water. 20 mL of stacking gel mixture was prepared by adding 5 mL of water, 10 mL of 10%/0.5% acrylamide/Bis, 4.8 mL of 0.5 M Bis-Tris pH 6.8 (deaerated briefly), 0.2 mL 10% SDS, 15 μL of TEMED, 80 μL of 25% ammonium persulfate. The stacking gel mixture was quickly added to the top of gels, and combs (10 wells of 7 μL volume or 15 wells of 5 μL volume) were inserted between the long and short glass plates before the gels were polymerized for 10 minutes. The polymerized gels were carefully removed from the block, wrapped in paper towels dampened with deionized water, and stored in a box in refrigerator.

2.5.2 Running SDS-PAGE Gels

An SDS-PAGE gel was installed into the clamp assembly of the gel electrophoresis apparatus with the short glass plate facing the inside of the clamp. When one gel was run, a plastic block plate was placed on the other side, and when two gels were run, the second gel was

placed on the other side. The clamp assembly was placed into the outer buffer chamber in the correct orientation by matching the color-coded electrodes with corresponding sides. The 1x gel running buffer was prepared by diluting the 5x stock solution [50 mM Tris base, 0.38 M glycine, 0.1% (w/v) SDS, water to 2.5 L] with water. The 1x gel running buffer was poured into the inner and outer chambers to the appropriate level indicated on the box, and the comb was removed from the gel carefully. At the same time, protein samples with PGLB were boiled for 2 minutes. The samples were loaded onto the gel using a Hamilton syringe (7 μL sample per 10 well slot or 5 μL sample per 15 well slot). The gel was run at 10 W (about 350 V and 35 mA) for 40 minutes or until the bromophenol blue dye reached the bottom. The gel was removed from the glass plates carefully, placed in a box filled with FIX solution [1:1 water-ethanol solution, 0.1 volumes glacial acetic acid], and incubated on the rocker platform at room temperature for 5 minutes. Then FIX solution was removed, and STAIN solution [0.5% (w/v) Coomassie Blue R in FIX solution] was added to the box containing the gel. The gel was stained at room temperature for 5 minutes, and STAIN solution was removed. The gel was rinsed in deionized water to remove excess STAIN. DESTAIN solution [7% (w/v) ethanol, 5% (w/v) glacial acetic acid in water] was added with a paper towel to absorb the dye. The gel soaked in DESTAIN solution was incubated at 60°C in a shaking water bath for 30 minutes or until the bands on the gel were visible. DESTAIN solution was discarded and rinsed off, and the destained gel was soaked in deionized water for at least 2 hours.

2.6 Western Blotting

The purified protein samples were run on an SDS-PAGE gel. While the gel was running, the nitrocellulose membrane and 4 pieces of blotting paper were cut to the appropriate sizes and equilibrated in transfer buffer containing 20% methanol [25 mM Tris, 192 mM glycine, 20% methanol, water]. After the gel electrophoresis was complete, the gel was soaked in 30 mL of transfer buffer without methanol for 2 to 5 minutes. A specific corner of the gel was cut to identify the direction of the gel (the corresponding corner of the membrane was cut as well). The gel was then placed in the transfer assembly with the nitrocellulose membrane in a specific order: blotting papers were placed at the bottom facing the positive electrode, nitrocellulose membrane placed on next, SDS-PAGE gel placed on top of the membrane, and more blotting papers placed on top facing the negative electrode. The nitrocellulose membrane was blotted at 100 V and 350 mA for 30 minutes. The blotted membrane was removed from the assembly and placed in 30 mL of 1x TBS buffer [25 mM Tris, 150 mM NaCl, HCI (pH 8.0), water] to equilibrate for 2 to 5 minutes. The nitrocellulose membrane was transferred to 50 mL of preincubation buffer [2% nonfat dry milk, 1x TBS buffer] and incubated on a rocker for 30 minutes. The membrane was washed twice in 50 mL 1x TTBS buffer [0.05% Tween 20, 1x TBS buffer] for 5 minutes each to remove excess preincubation buffer. The membrane was transferred to a plastic pouch with 10 mL of 1x TTBS buffer containing a 1:5000 dilution of the appropriate secondary antibody and incubated for 1 hour. The membrane was washed 3 times for 5 minutes each in fresh 50 mL of 1x TTBS buffer. 1.5 mL each of ECL detection solution 1 and ECL detection solution 2 was mixed and added to the protein side of the membrane to allow exposure of the membrane to the film through detection of chemiluminescent reaction.

Chapter 3

Results and Discussion

3.1 Subcloning of MW Marker Protein Constructs

Table 3.1 List of recombinant protein constructs created

Protein Molecular Weight (kDa)	Protein Construct Name
10	pST50Trc4-HSTPABHPCx1
20	pST50Trc2-STRHSTPABPACCBPx3
30	pST50Trc3-HSTPABPACS100BCBP
40	pST50Trc2-HSTPABPACGST
50	pST50Tr-STRHSTPABMBPt1x2
60	pST50Trc4-HSTPABPACdRCC1t8x27
80	pST50Trc3-STRHSTPABPACIL1bt1dRCC1t8x27
100	pST50Trc2-STRHSTPABPACIL1bt1QRSx2
150	pST50Tr-STRHSTPABMBPpepNx1
240	pST50Tr-STRHSTPABPACIL1bt1QRSMBPpepN
250	pST50Tr-STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN

A total of 11 expression vectors were prepared in a group effort with fellow undergraduate students in the laboratory (Ryan Santilli and Jack Williamson) and my research adviser Dr. Song Tan. The original inserts of Protein A IgG binding domain B (PAB) and binding domain C (PAC) were amplified from synthetic DNA templates (Integrated DNA Technologies). Other inserts and vectors used to prepare the marker proteins were derived from plasmids previously created in the laboratory (work completed by Dr. Tan). The presence of an insert in a plasmid was confirmed by restriction mapping, and the sequence of the coding region of a plasmid was verified through sequencing. As shown in Figure 3.1, every recombinant protein contains a polyhistidine tag (HST) used for metal affinity protein purification and an IgG binding domain B (PAB) for antibody interaction. Most proteins except for three also contain an

IgG binding domain C (PAC) for enhanced antibody interaction. The unique protein components fused to the HSTPAB sequence were carefully selected based on the criteria that proteins can (a) be expressed under soluble conditions at high levels in *E. coli* cells, (b) be purified efficiently by metal affinity chromatography, (c) migrate appropriately on SDS-PAGE gel, and (d) be recognized by secondary antibodies directly in Western blotting. The last criterion was fulfilled by the incorporation of PAB and PAC. The candidate proteins that fulfilled the first three criteria were selected and tested by the undergraduate group with the assistance of Dr. Tan.

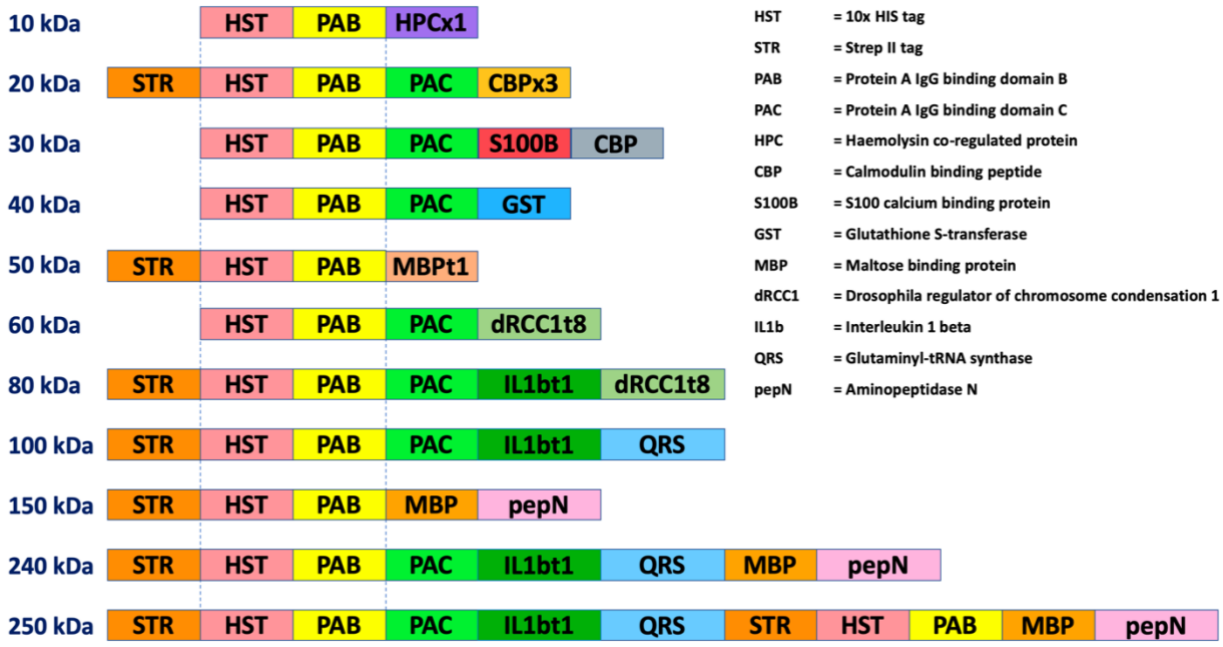


Figure 3.1 Designs of recombinant molecular weight markers

The abbreviated fusion protein labels are listed in the legend on the right.

Table 3.2 lists the finalized designs of the molecular weight marker proteins, and it shows the actual molecular weight of each recombinant proteins compared to the target molecular weight. As shown in Table 3.2, the molecular weights of most proteins are within 0.5 kDa of their target sizes (except for 100 kDa protein). Many of these protein constructs were modified

from their originally proposed designs in order to resolve various issues, such as low expression level, proteolysis, and low antibody interaction in Western blotting.

Table 3.2 Intended and actual molecular weights of recombinant markers

Protein	Target Molecular weight (kDa)	Molecular Weight (kDa)
HSTPABHPCx1	10	10.015
STRHSTPABPACCBPx3	20	19.877
HSTPABPACS100BCBP	30	29.773
HSTPABPACGST	40	40.362
STRHSTPABMBPt1	50	50.310
HSTPABPACdRCC1t8	60	60.029
STRHSTPABPACIL1bt1dRCC1t8	80	79.830
STRHSTPABPACIL1bt1QRS	100	98.269
STRHSTPABMBPpepN	150	149.946
STRHSTPABPACIL1bt1QRSMBPpepN	240	239.921
STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN	250	249.895

In the process of preparing the 50 kDa protein STRHSTPABMBP, a partial digestion of MBP insert with restriction enzymes BamHI and BglII was required due to the presence of an internal recognition site of BglII in MBP. The parent DNA plasmid containing MBP was first digested with BamHI for one hour at 37°C, and the linearized DNA was then treated with BglII at 37°C for a total of 30 minutes. Aliquots of the reaction mixture were taken at 5, 15 and 30 minute time points. The resulting three partial digestion samples were run on an agarose gel, and a DNA band with the desired length was isolated and purified via agarose gel purification. The resulting 50 kDa protein construct pST50Tr-STRHSTPABMBP was also modified through PCR site-directed mutagenesis twice to remove internal BglII and BspEI recognition sites because the 50 kDa protein construct STRHSTPABMBPx2 was also designed as an intermediate to prepare the 150 kDa construct STRHSTPABMBPpepN (Table 3.2). In addition, STRHSTPABMBPx2

construct was also truncated in order to bring down the its molecular weight from 50.542 kDa to 50.310 kDa (This experiment was completed by fellow undergraduate student Jack Williamson).

In order to simplify the protein expression and purification processes, eight different marker proteins were subcloned into two separate polycistronic expression vectors, each containing four protein sequences, using the polycistronic cloning system described previously in Section 2.2.13 (Polycistronic Vector). Constructs expressing the 10, 30, 50, and 100 kDa proteins were subcloned into one vector (Figure 3.2), and constructs expressing the 20, 40, 60, and 80 kDa proteins were subcloned into another vector (Figure 3.3) The polycistronic cloning was done sequentially with one protein expression sequence incorporated into the vector after another using distinctive pairs of restriction enzymes. In addition, restriction mapping was done after each round of plasmid preparation to confirm the correct insertion of a protein coding sequence.

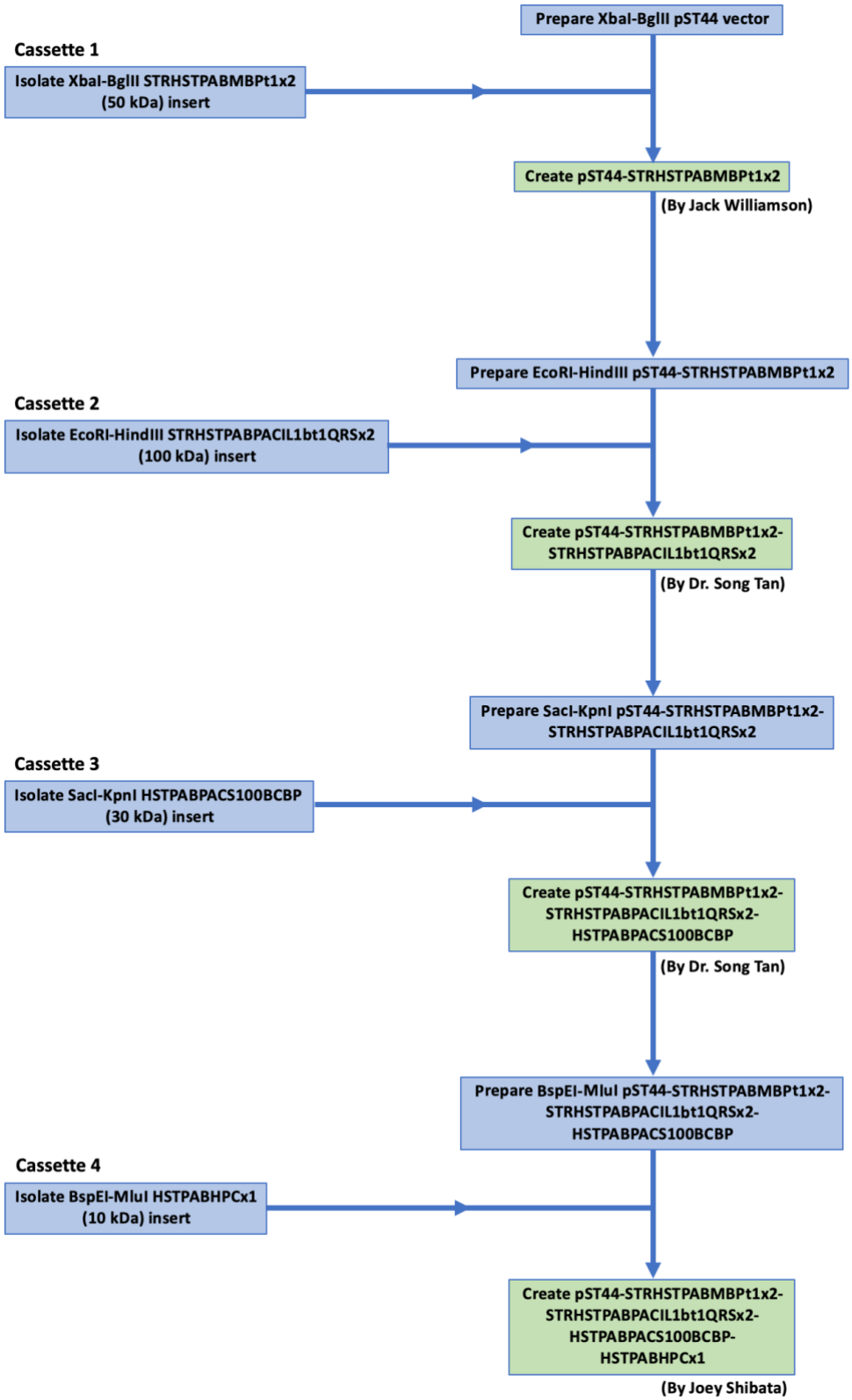


Figure 3.2 Flow chart for polycistronic cloning of 10-30-50-100 kDa protein vector

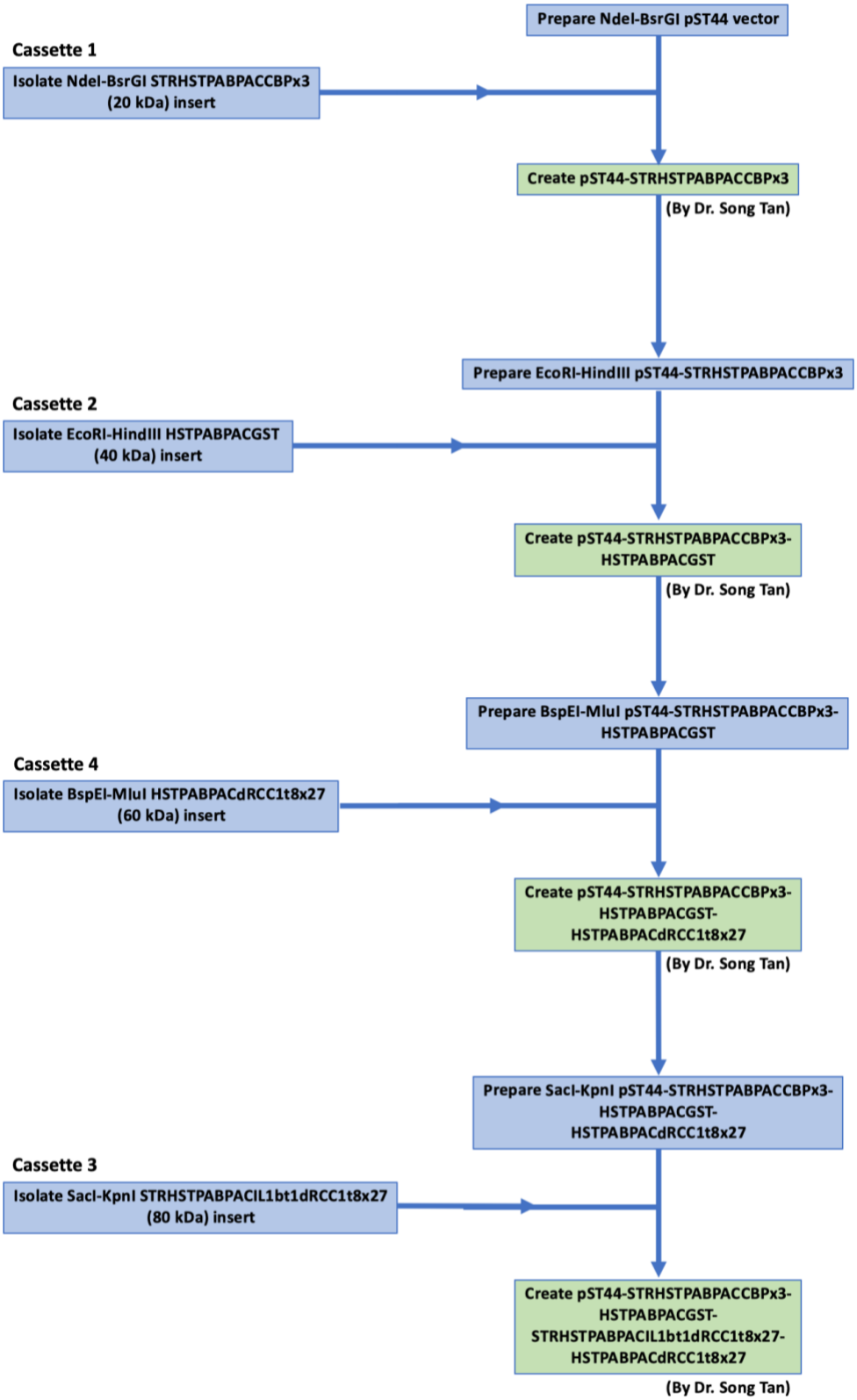


Figure 3.3 Flow chart for polycistronic cloning of 20-40-60-80 kDa protein vector

3.2 Small-Scale Expression and Batch Purification

Table 3.3 Summary of collaborators completed each experiment in the project

Protein Markers (kDa)	Construct Cloning	Expression	Purification	Western Blotting
10	Jack Williamson	Jack Williamson	Jack Williamson	Jack Williamson
20	Joey Shibata Song Tan Jack Williamson	Jack Williamson	Joey Shibata	Jack Williamson
30	Ryan Santilli	Ryan Santilli	Ryan Santilli	Jack Williamson
40	Ryan Santilli	Joey Shibata	Joey Shibata	Jack Williamson
50	Joey Shibata Jack Williamson	Jack Williamson	Jack Williamson	Jack Williamson
60	Jack Williamson	Jack Williamson	Jack Williamson	Jack Williamson
80	Jack Williamson	Jack Williamson	Ryan Santilli	Jack Williamson
100	Ryan Santilli Song Tan Jack Williamson	Joey Shibata	Joey Shibata	Jack Williamson
150	Ryan Santilli	Ryan Santilli	Ryan Santilli	Ryan Santilli Joey Shibata Jack Williamson
240	Song Tan	Joey Shibata	Joey Shibata	Ryan Santilli Joey Shibata Jack Williamson
250	Song Tan	Joey Shibata	Joey Shibata	Ryan Santilli Joey Shibata Jack Williamson
10-30-50-100 Polycistronic	Joey Shibata Song Tan	Joey Shibata	Joey Shibata	(not completed)
20-40-60-80 Polycistronic	Ryan Santilli Song Tan	Ryan Santilli	Joey Shibata	(not completed)

*This table only shows the records for the modified constructs that we prepared; there were many constructs cloned, expressed, and purified in the process of designing the ideal markers, and all members of the project contributed equally to the experiments.

All recombinant proteins were individually expressed at 37°C from corresponding expression vectors in a group effort (Table 3.3). Each protein was expressed at 37°C for three hours after induction, and each protein was purified from expression cell extracts using metal affinity chromatography. Purified protein samples were applied to SDS-PAGE to determine the

expression level, purification level, and size of marker proteins. However, recombinant proteins containing dRCC1 were found to have low expressions at 37°C (Figure 3.4), and because previous experiments done in the laboratory showed that dRCC1 expresses better at lower temperatures, proteins containing dRCC1 (60 kDa, 80 kDa) were expressed at 18°C. As shown in Figures 3.5, the 18°C expression resulted in high expression of proteins (experiments done by Jack Williamson). *E.coli* cells expressing the dRCC1 proteins at lower temperatures (18°C and 21°C) were incubated overnight after induction.

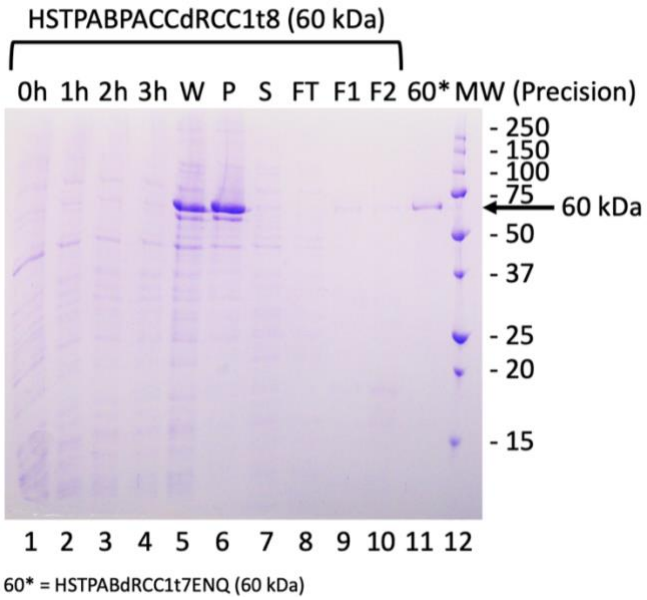


Figure 3.4 Small-scale expression of HSTPABPACdRCC1t8 (60 kDa) at 37°C

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract; where insoluble proteins are collected. S = supernatant of the whole cell extract; where soluble proteins are present. FT = flow through in metal affinity chromatography; non-specific proteins are discarded during this process in protein purification. "F1, F2" = fractions collected from the elution of purified proteins; this is where the desired proteins should be present after purification. 60* = a sample of a 60 kDa protein with a different design; tested for comparison between the two 60 kDa proteins. The proteins were expressed and purified by Jack Williamson.

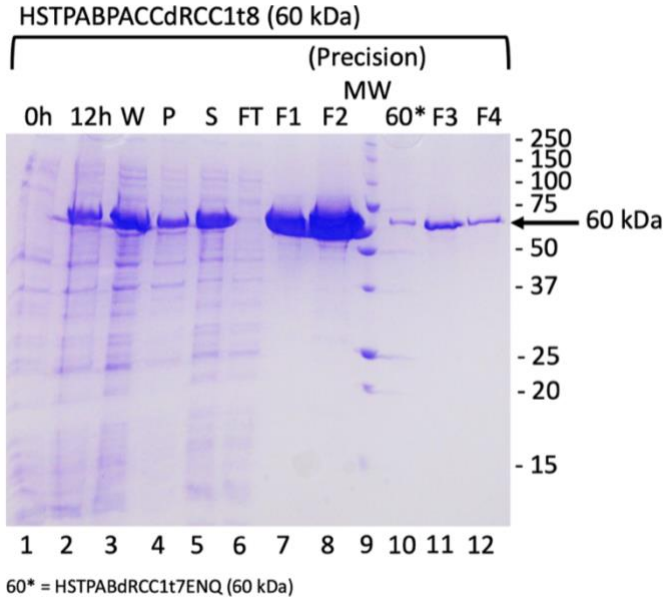


Figure 3.5 Small-scale expression of HSTPABPACdRCC1t8 (60 kDa) at 18°C

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. 60* = a sample of a 60 kDa protein with a different design; tested for comparison between the two 60 kDa proteins. The proteins were expressed and purified by Jack Williamson.

Because one of the goals for our project was to develop a protocol to efficiently prepare the protein markers with the most basic laboratory equipment, we aimed to prove that these proteins can be expressed very well even at room temperature without requiring a temperature-controlled incubator. Therefore, all marker proteins were expressed at 21°C, which is close to room temperature. Table 3.4 below describes the results for all marker proteins expressed at 21°C. Proteins were considered to be soluble if most protein contents were present in the supernatant instead of the pellet during the solubility test. The criteria for purification was based on purity and band intensity. Purification was considered to be successful if very few or no impurity was present and if the intensity of the desired protein band was at least three times the intensity of corresponding molecular weight marker bands.

Table 3.4. Results of small-scale expression and batch purification

Molecular Weight (kDa)	Temperature	Expression Level	Solubility	Purification
10	21°C	High	Soluble	Successful
20	21°C	High*	Soluble	Successful
30	21°C	High	Soluble	Successful
40	21°C	High	Soluble	Successful
50	21°C	High	Soluble	Successful
60	21°C	High	Soluble	Successful
80	21°C	High	Soluble	Successful
100	21°C	High	Soluble	Successful
150	21°C	High	Mostly soluble	Successful
240	21°C	Low	Mostly soluble	Successful
250	21°C	Low	Mostly soluble	Successful

*High individual expression but reduced expression when co-expressed from the polycistronic vector

All of the proteins from sizes 10 to 150 kDa were expressed at high levels and successfully purified in the end when expressed individually (Table 3.4). 240 and 250 kDa proteins had low levels of expression, but these proteins were migrating appropriately and identifiable on the corresponding SDS-PAGE gels. Moreover, because both proteins have very large sizes, it was predicted that these proteins were harder to be expressed and purified compared to other proteins with smaller sizes (Figures 3.6 and 3.7). An interesting fact to note is that both 240 and 250 kDa proteins yielded large amounts of by-products, which may have resulted from proteolysis or premature termination of expression since both proteins are very large in size. These by-products have estimated molecular weights of 40, 75, and 100 kDa, which are likely various N-terminal fragments of the 240 and 250 kDa proteins because these by-products also contain HIS tags. Moreover, because the 240 and 250 protein markers were intended to be used separately from the other lower molecular weight markers, the presence of protein by-products at the mentioned defined positions can be beneficial in that they can be included as a part of the molecular weight ladder.

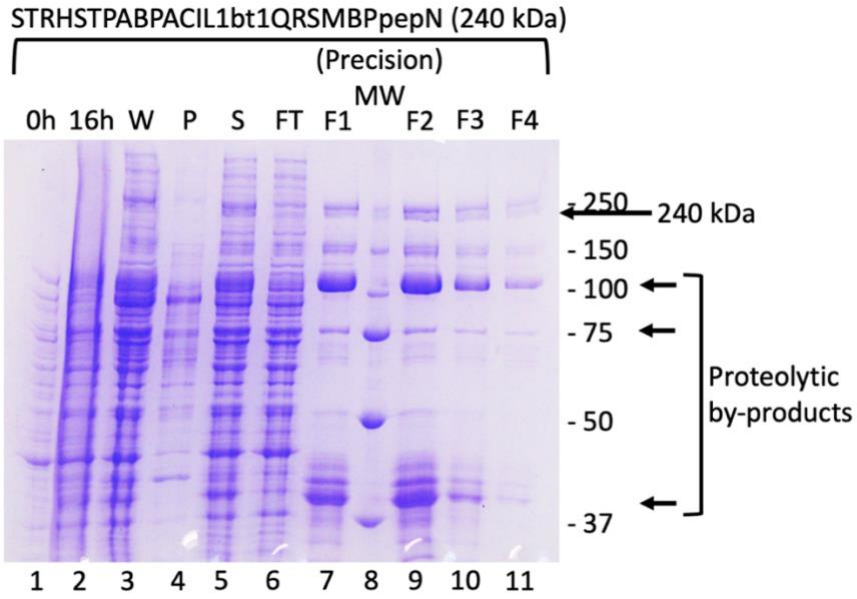


Figure 3.6 Small-scale purification of STRHSTPABPACIL1bt1QRSMBPpepN (240 kDa)

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Proteolytic by-products observed in purified proteins (lanes 7, 9, 10, and 11)

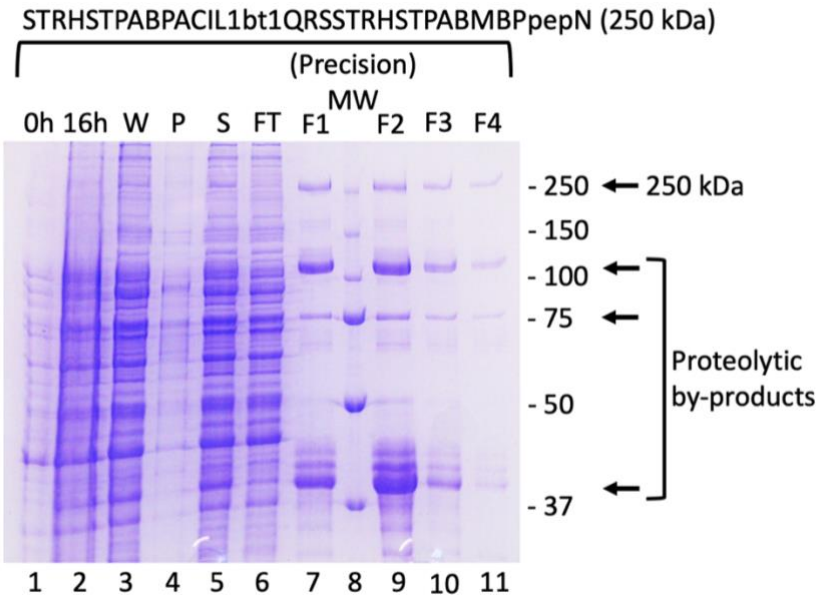


Figure 3.7 Small-scale purification of STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN (250 kDa)

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Proteolytic by-products observed in purified proteins (lanes 7, 9, 10, and 11)

When the first version of 20 kDa protein STRHSTPABPACCBP was expressed and purified, there was a sign of proteolysis observed on the SDS-PAGE gel (Figure 3.8). This was potentially caused by native proteases found in the expression cells that cleaved a part of the marker proteins.

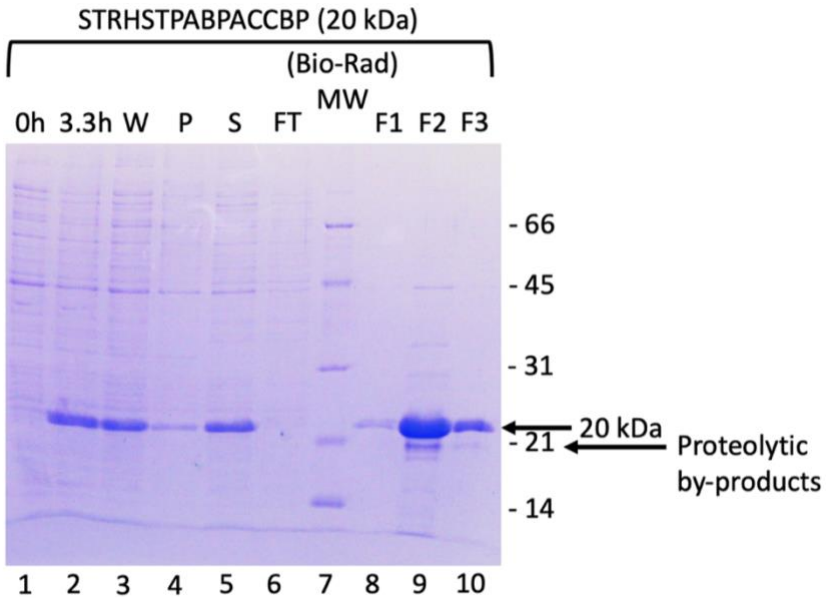


Figure 3.8 Small-scale purification of STRHSTPABPACCBP (20 kDa)

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Proteolytic by-products observed in purified proteins (lanes 9 and 10).

In order to solve this problem, fellow undergraduate student Jack Williamson, with the help from Dr. Tan, created two different versions of the 20 kDa protein by introducing a mutation in each to replace a phenylalanine residue with a leucine residue to prepare STRHSTPABPACCBPx2 and replace a tryptophan residue with a leucine residue to prepare STRHSTPABPACCBPx3. The two aromatic amino acid residues were replaced with leucine because aromatic amino acids can be cleavage targets for proteases due to their uniquely

recognizable side chains. The two new versions of the 20 kDa proteins, expressed by Jack Williamson and purified by myself, yielded results with reduced proteolytic by-products. While STRHSTPABPACCBPx2 still showed a small amount of proteolysis, STRHSTPABPACCBPx3 had minimal to no sign of proteolysis (Figures 3.9 and 3.10).

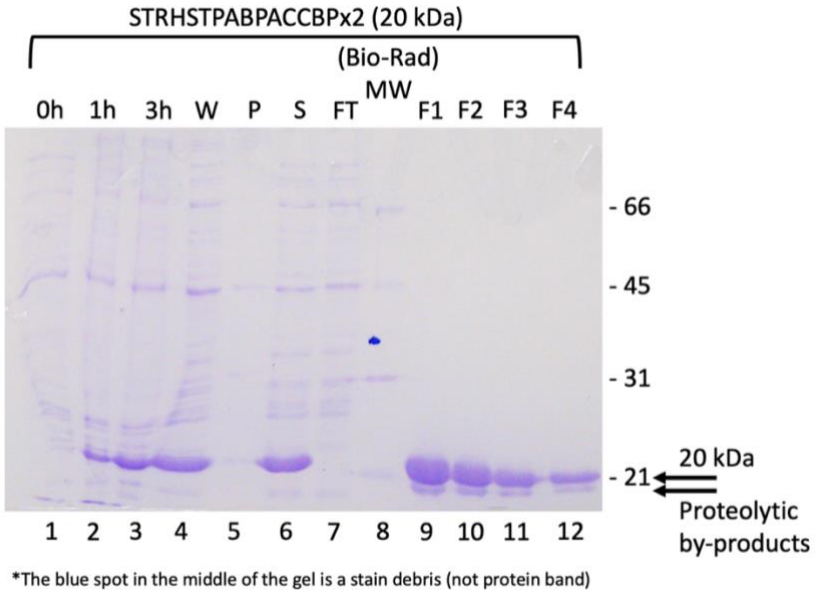


Figure 3.9 Small-scale purification of STRHSTPABPACCBPx2 (20 kDa)

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Proteolytic by-products observed in purified proteins (lanes 9, 10, 11, and 12). The protein was expressed by Jack Williamson and purified by Joey Shibata.

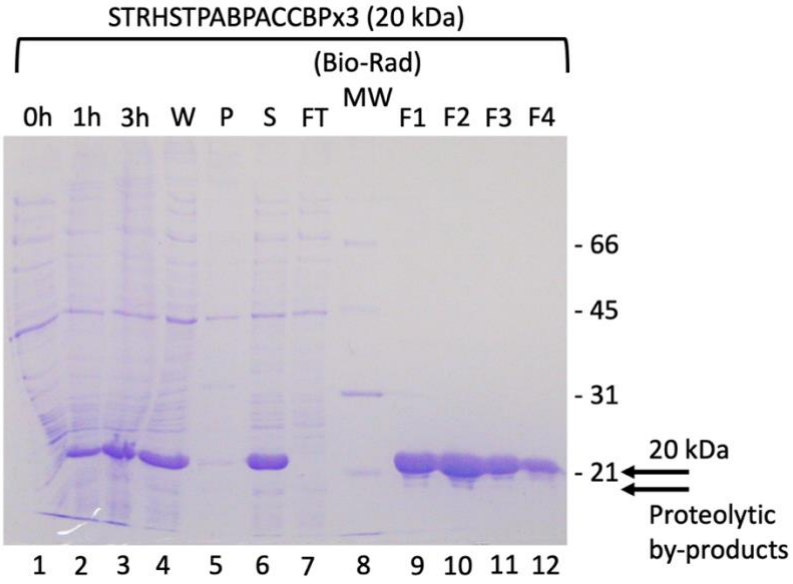


Figure 3.10 Small-scale purification of STRHSTPABPACCBPx3 (20 kDa)

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Minimal amounts of proteolytic by-products observed in purified proteins (lanes 9, 10, 11, and 12).The protein was expressed by Jack Williamson and purified by Joey Shibata.

The two polycistronic vectors were transformed into *E. coli* cells which were grown at 21°C to co-express the marker proteins. The first vector that expressed 10, 30, 50, and 100 kDa proteins yielded high levels of expression and purification for all four proteins (Figure 3.11) In addition, the expression levels of the four proteins were similar to each other, and each protein was clearly visible without significant sign of proteolysis. Therefore, the four proteins of 10, 30, 50, and 100 kDa co-expressed from the polycistronic vector are assumed to be ideal as the universal protein molecular weight markers.

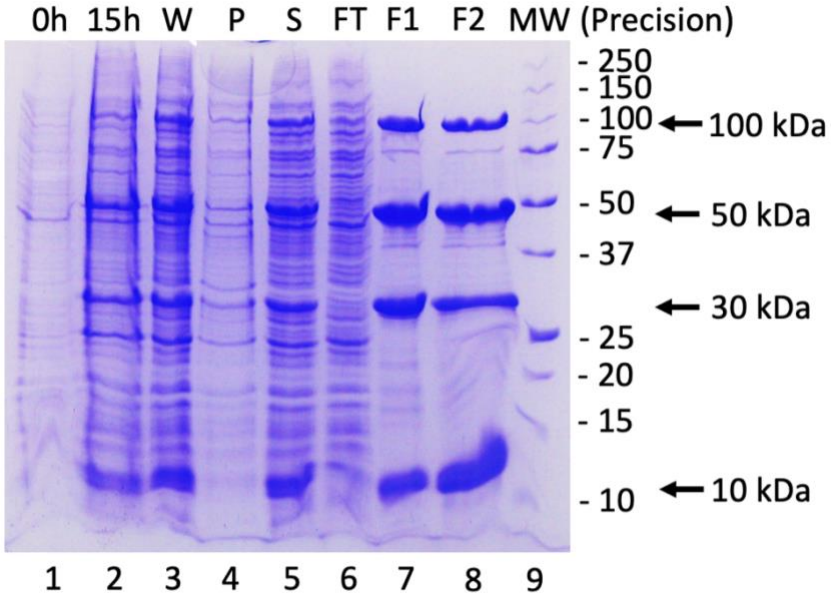


Figure 3.11 Co-expression of 10-30-50-100 kDa markers from the polycistronic vector

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins.

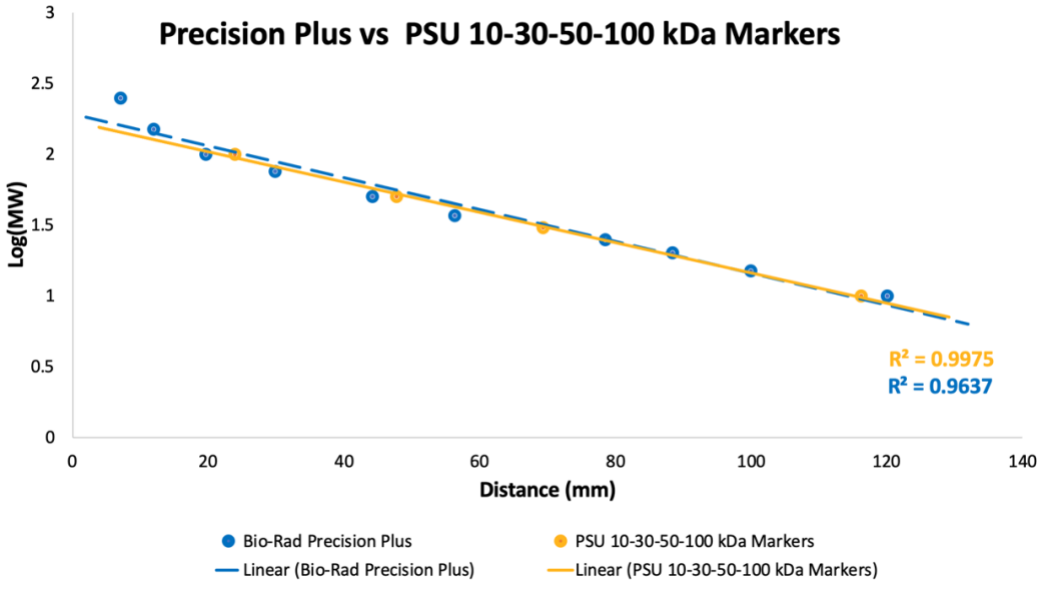


Figure 3.12 Semi-log migration plot of co-expressed 10-30-50-100 kDa markers

Blue dots and line represent the Bio-Rad Precision Plus markers. Orange dots and line represent PSU markers.

Based on the superimposed semi-log plot of our co-expressed proteins and Bio-Rad Precision Plus Molecular Weight Markers, it is noticeable that the migration patterns of the marker proteins developed (10, 30, 50, 100 kDa) are appropriate and very similar to those of the commercially available markers (Figure 3.12).

On the other hand, the co-expression of 20, 40, 60, and 80 kDa proteins was not optimal in the beginning. While 20 kDa protein was expressed at high level in the individual expression condition (indicated as "high*" in expression in Table 3.4), its expression level visibly decreased when the protein was expressed from the polycistronic vector containing sequences for 20, 40, 60, and 80 kDa proteins (Figure 3.13).

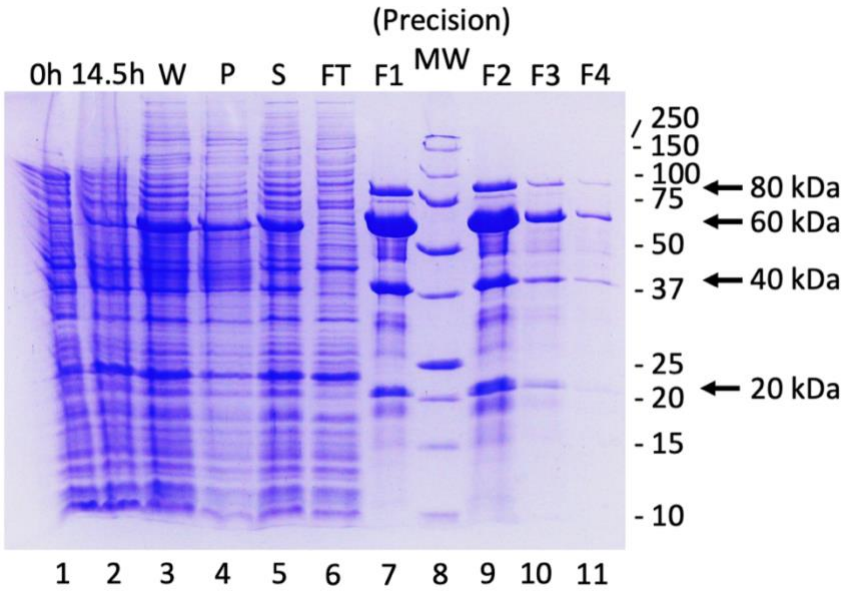


Figure 3.13 Co-expression of 20-40-60-80 kDa markers from the polycistronic vector

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins.

Despite the reduced expression level of 20 kDa proteins, the expression levels of the other three proteins were not affected greatly by the co-expression (Figure 3.13). According to a previous study on developing the polycistronic expression system, the order of expression of the proteins encoded in a polycistronic plasmid did not affect the expression level of each protein (Tan *et al.* 2005). However, because of the clear difference in expression levels of the 20 kDa proteins in individual expression and co-expression, it was suspected that the order of expression in the plasmid was the potential cause of the reduced 20 kDa protein expression because 20 kDa protein construct was inserted into the very last cassette, the farthest from the expression promoter, of the vector. The order of expression was originally decided in this pattern in order to avoid using a restriction enzyme that has an internal recognition site in one of the protein expression sequences. To test the potential cause of the reduced expression, additional expression and purification experiments were set up to determine the expression levels of 20 kDa protein in individual expression and in co-expression with two other proteins (20, 60, 80 kDa), and these expression levels were compared to that of the 20-40-60-80 kDa co-expression shown in Figure 3.13. While the 20 kDa protein was expressed very well in individual expression (Figure 3.14), the expression level of the 20 kDa protein significantly decreased when co-expressed with two other proteins from an intermediate vector (Figure 3.15). The results partially validated the hypothesis that the placement of 20 kDa protein expression sequence in the last cassette reduces the protein expression level.

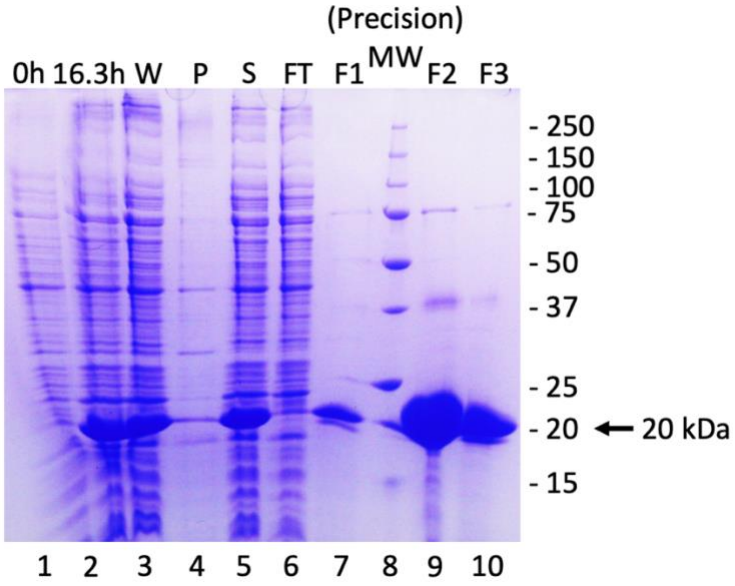


Figure 3.14 Single small-scale expression of the 20 kDa marker

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins.

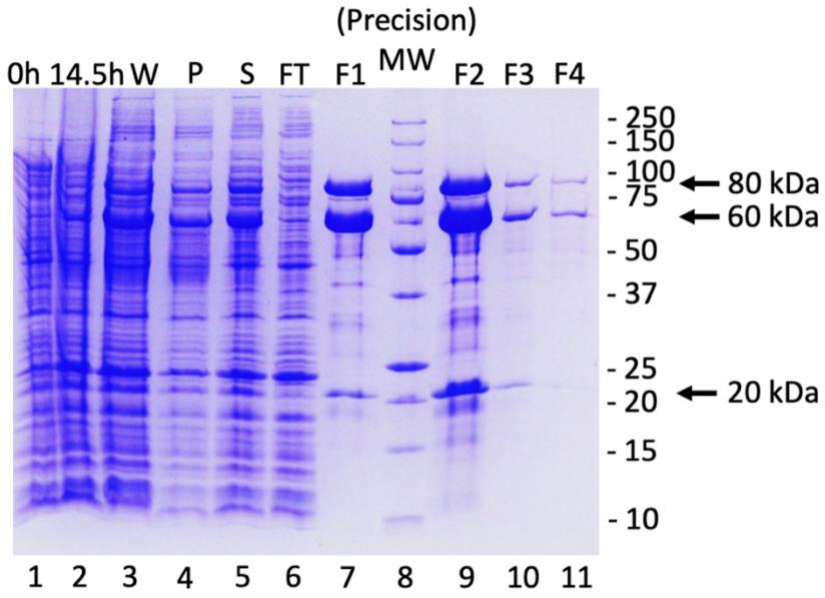


Figure 3.15 Co-expression of 20-60-80 kDa markers from an intermediate polycistronic vector

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins.

In order to clearly identify the problem and solve this issue, Dr. Tan assisted in the process to rearrange the order of protein expressions by transferring the 20 kDa protein construct to cassette 1. The other three protein constructs were also rearranged in the order of 40, 80, 60 kDa. Partial digestion of 60 kDa protein expression insert with enzymes BspEI and MluI had to be done because of the presence of an BspEI internal recognition site in the sequence (subcloning work was completed by Dr. Tan). Seen in Figure 3.16, the expression and purification results of the rearranged 20-40-60-80 kDa polycistronic vector showed a clear improvement in the expression level of the 20 kDa protein in co-expression compared to the expression of 20 kDa proteins from the old polycistronic vector (Figure 3.13). However, it was still noticeable that the expression level of 20 kDa protein was slightly lower than expression levels of other three proteins. Despite of this drawback, all four protein bands were clearly visible with minimal amounts of proteolytic by-products or other impurities. In addition, when compared to the protein expressions from an intermediate polycistronic vector containing 20 and 40 kDa protein, the expressions from the full polycistronic vector (20, 40, 60, 80 kDa proteins) was visibly lower overall (Figure 3.17). These results indicated that the number of protein constructs inserted into a polycistronic vector can negatively impact the expression levels of co-expressed proteins.

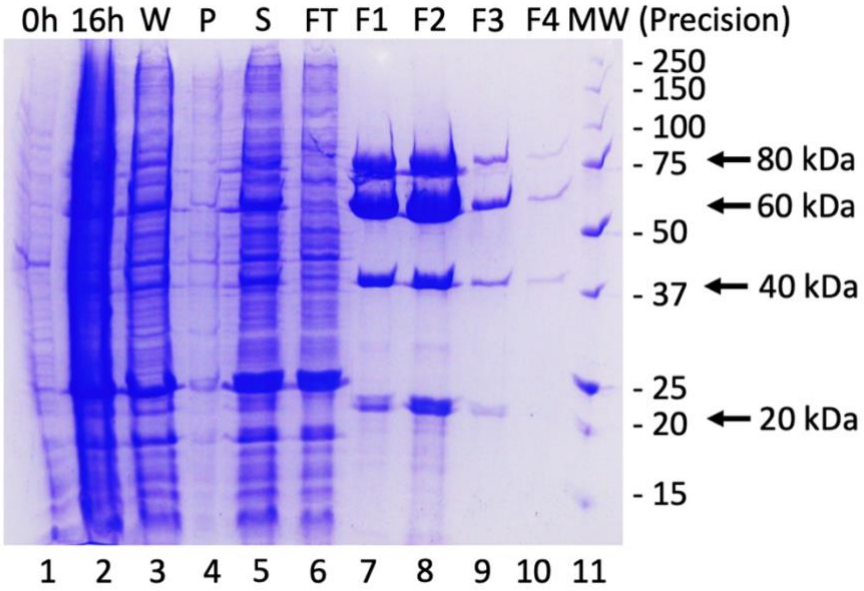


Figure 3.16 Co-expression of 20-40-60-80 kDa markers from the rearranged polycistronic vector

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. The proteins were expressed by Ryan Santilli and purified by Joey Shibata.

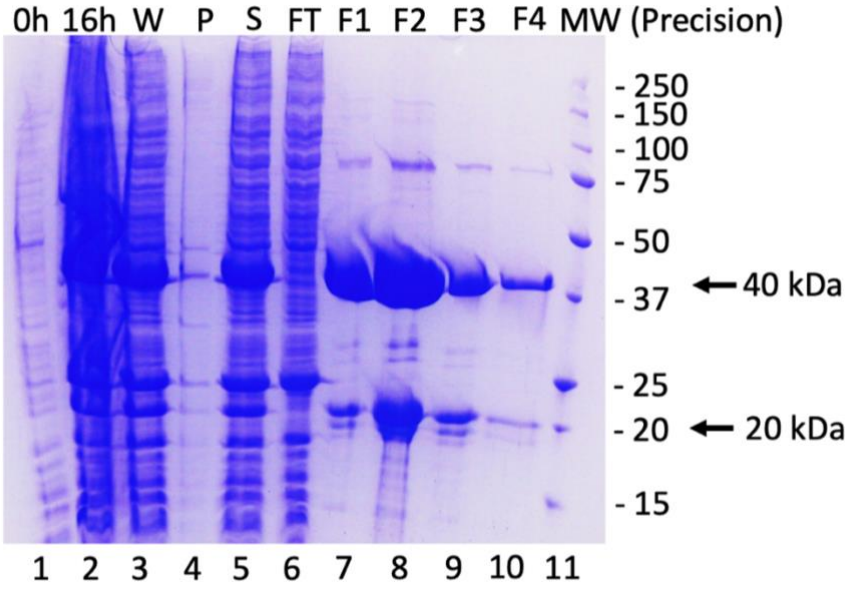


Figure 3.17 Co-expression of 20-40 kDa markers from an rearranged intermediate polycistronic vector

"h0, h1, ..." = hour time points after induction during expression; h0 is the uninduced cell culture sample. W = whole cell extract of protein-expressing *E. coli*. P = pellet collected from the whole cell extract. S = supernatant of the whole cell extract. FT = flow through in metal affinity chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. The proteins were expressed by Ryan Santilli and purified by Joey Shibata.

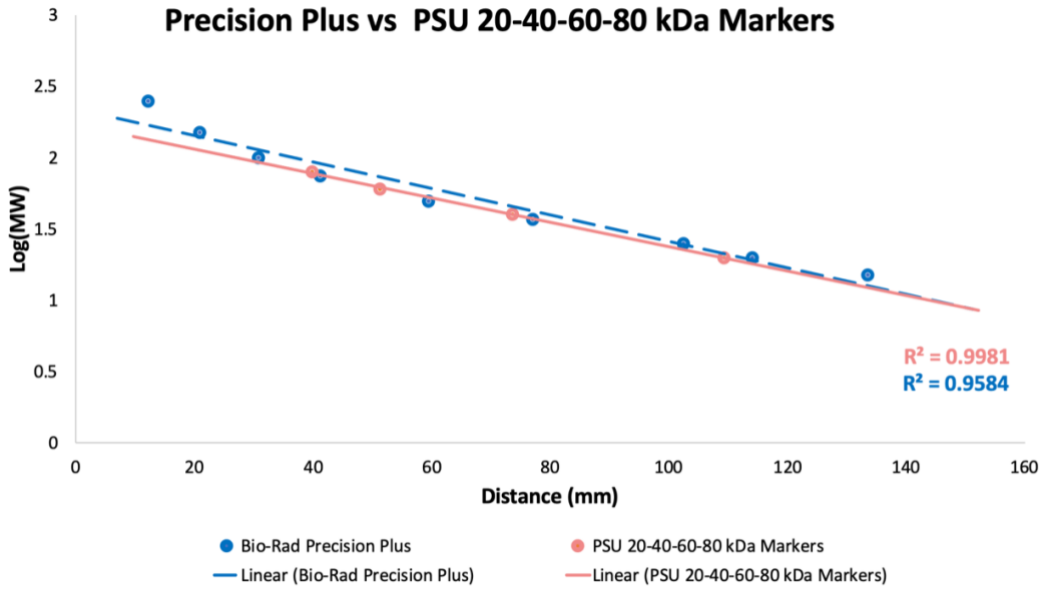


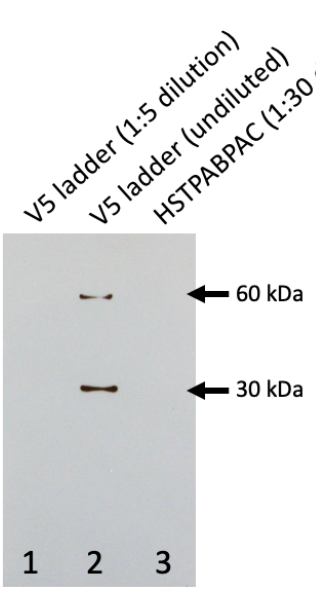
Figure 3.18 Semi-log migration plot of co-expressed 20-40-60-80 kDa markers

Blue dots and line represent the Bio-Rad Precision Plus markers. Pink dots and line represent PSU markers.

The semi-log plot for the 20, 40, 60, 80 kDa proteins also shows that these proteins migrated appropriately on the SDS-PAGE gel since the deviation in the trendlines of our proteins and Precision Plus markers is relatively small (Figure 3.18). Although optimization of the expression condition will be required to normalize the expression levels of all proteins, the bands corresponding to the proteins were clearly identifiable on the SDS-PAGE gels. The polycistronic vector encoding 20, 40, 60, and 80 kDa proteins and the other vector encoding 10, 30, 50, and 100 kDa proteins fulfill our criteria for the universal protein molecular weight markers. Based on the amount of protein purified from the small-scale polycistronic expression cultures, a 100 mL culture of polycistronic expression (10-30-50-100 or 20-40-60-80 co-expression) was estimated to provide enough proteins for 1000 lanes in SDS-PAGE.

3.3 Western Blotting

Western blotting is another important step in the process of developing the universal protein molecular weight markers because one of the key purposes of our markers is its application in western blotting that enables marker detection through IgG antibody interactions. These Western blotting experiments were completed in collaboration with the undergraduate group in the laboratory mentioned previously. In order to test our marker proteins in Western blotting, eight different purified proteins were combined together with equivalent concentrations, including GKTHSTPAB (10 kDa), HSTPABS100B (20 kDa), HSTPABPACS100BCBP (30 kDa), CBPSTRHSTPABGST (40 kDa), STRHSTPABMBP (50 kDa), HSTPABPACdRCC1t8 (60 kDa), HSTPABGSTdRCC1t8 (80 kDa), and STRHSTPABGSTQRS (100 kDa). The ladder was also diluted in water with a ratio of 1:5. A 1:30 dilution of protein HSTPABPAC (15 kDa) was also tested alongside the combined ladder (the sample was diluted because of the high concentration of purified protein). All three samples were run on an SDS-PAGE gel first, and these proteins were transferred onto a nitrocellulose membrane through a 30-minute transferring process with 25 V of voltage and maximum 1 mA of current. The membrane was treated with 1:5000 dilution of donkey HRP-linked anti-rabbit whole antibodies (Pharmacia) for detection of the marker proteins, and a film was exposed to the chemiluminescence from the bound secondary antibodies (Figure 3.19).



*MW of HSTPABPAC = 15 kDa

*MW of V5 ladder proteins

Proteins	MW (kDa)
GKTHSTPAB	10
HSTPABS100B	20
HSTPABPACS100BCBP	30
CBPSTRHSTPABGST	40
STRHSTPABMBP	50
HSTPABPACdRCC1t8	60
HSTPABGSTdRCC1t8	80
STRHSTPABGSTQRS	100

*film exposure time = 10 min

Figure 3.19 Western blot of recombinant protein markers containing the Protein A IgG binding domains

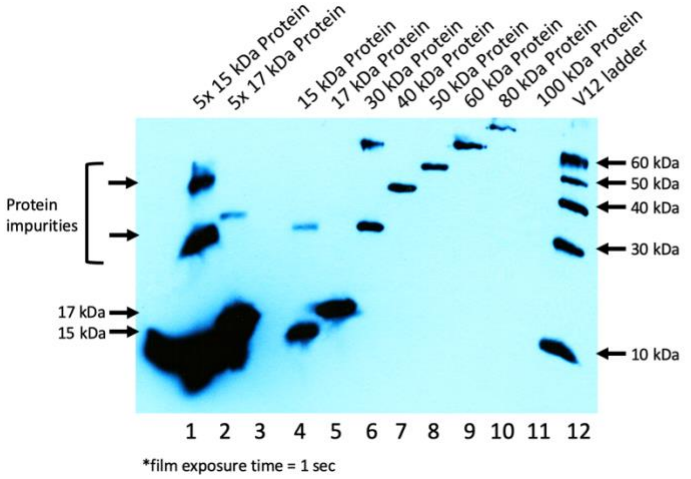
Western blotting was completed by Jack Williamson.

It was interesting to note that only two proteins (30 and 60 kDa) in the undiluted lane (lane 2) were visible on the exposure film, and both proteins had one thing in common in that the two proteins contained two IgG binding domains, PAB and PAC, while most of the other proteins contained only PAB (HSTPABPAC was an exception). Another experiment was done with a different type of secondary antibodies, goat HRP-linked anti-rabbit whole antibodies at 1:2000 dilution, and the similar result was shown with the 30 and 60 kDa proteins visible while none of the other proteins were detected (the experiment was done by Jack Williamson). These results suggested that the incorporation of an additional IgG binding domain in a marker protein sequence can enhance the antibody interactions and lead to stronger protein detection in Western blotting. Therefore, many of the protein constructs were redesigned to include PAC (modified proteins listed in Table 3.2). Specifically, 20, 40, 80, and 100 kDa protein constructs were rearranged to have both PAB and PAC. All experiments presented in Section 3.2 (Small-Scale

Expression and Batch Purification) were done after the incorporation of PAC, so the proteins described in the previous section already included PAC in their constructs. The final 10 kDa protein does not include PAC because the modified core component of the marker protein HSTPABPAC alone is 15 kDa. 50 kDa protein was not modified to include PAC due to difficulties in preparing recombinant proteins with molecular weights that are exactly their corresponding defined sizes.

In another experiment, a new secondary antibody, mouse m-IgGkappa BP-HRP (Santa Cruz), was used for protein detection in order to test the antibody compatibility of the universal IgG binding domain PAB and PAC. The marker proteins were again combined into a ladder, which included HSTPABHPCx1 (10 kDa), HSTPABPAC (15 kDa), STRHSTPABPACCBP (20 kDa), HSTPABPACS100BCBP (30 kDa), HSTPABPACGST (40 kDa), STRHSTPABMBPt1 (50 kDa), HSTPABPACdRCC1t8 (60 kDa), STRHSTPABPACIL1bt1dRCC1t8 (80 kDa), and STRHSTPABPACIL1bt1QRS (100 kDa); the protein ladder was tested in Western blotting (Figure 3.20). The figure shows that most proteins, except the 100 kDa protein, were detected in their individual lanes, but some proteins were undetected when combined with others in the ladder lane, specifically the 15, 20, 80, and 100 kDa proteins. It is interesting to note that while all four that were undetected contained both PAB and PAC, the two proteins without PAC, 10 and 50 kDa, were clearly visualized on the film. Therefore, more testing will be required to investigate the detection issues and optimize the protein detection in Western blotting. In addition, undefined proteins were detected in lanes containing the 15 and 17 kDa proteins; it is unlikely that proteolysis occurred since the unknown proteins are larger than the marker proteins applied. They may be some form of contaminants of by-products from expression, but

investigation is needed to identify and eliminate the impurities. Fortunately, the 15 and 17 kDa proteins tested in this experiments were not included in the finalized protein markers, and those markers that contain the HSTPABPAC component (15 kDa) as an intermediate did not have problems with impurity in purification. In the aspect of secondary antibodies in this experiment, the mouse m-IgGkappa BP-HRP antibody yielded stronger detection signals for the 30 and 60 kDa proteins compared to donkey HRP-linked anti-rabbit and goat HRP-linked anti-rabbit whole antibodies used previously. Another thing worth noting was that X-ray film was only exposed for one second in the experiment using the mouse m-IgGkappa BP-HRP antibody compared to the 10-minute exposure time required for the first experiment using the donkey HRP-linked anti-rabbit whole antibody. However, it is also necessary to acknowledge that the donkey HRP-linked anti-rabbit whole antibody used previously was old and had a minor issue with reduced sensitivity. Overall, all three types of secondary antibodies tested showed binding with certain proteins (30 and 60 kDa proteins especially), and this means that the PAB and PAC in the protein markers are capable of binding a variety of secondary antibodies.



*5x HSTPABPAC (15 kDa) and STRHSTPABPAC (17 kDa) as reference of protein positions

*MW of V12 ladder proteins

Proteins	MW (kDa)
HSTPABHPCx1	10
HSTPABPAC	15
STRHSTPABPACCBP	20
HSTPABPACS100BCBP	30
HSTPABPACGST	40
STRHSTPABMBPt1	50
HSTPABPACdRCC1t8	60
STRHSTPABPACIL1bt1dRCC1t8	80
STRHSTPABPACIL1bt1QRS	100

Figure 3.20 Western blot of modified protein markers containing the double IgG binding domains (PABPAC)

The two large bands in lane 1 were impurities. Large molecular weight impurities were also found in lanes 2, 4, and 6. Western blotting was completed by Jack Williamson.

Finally, the higher molecular weight markers, 100, 150, 240, and 250 kDa proteins, were applied to an SDS-PAGE gel and transferred to a Western blot in order to check their abilities to interact with antibodies (Figure 3.21). All transferring conditions were kept constant, and mouse m-IgGkappa BP-HRP (Santa Cruz) was used as the secondary antibody for detection.

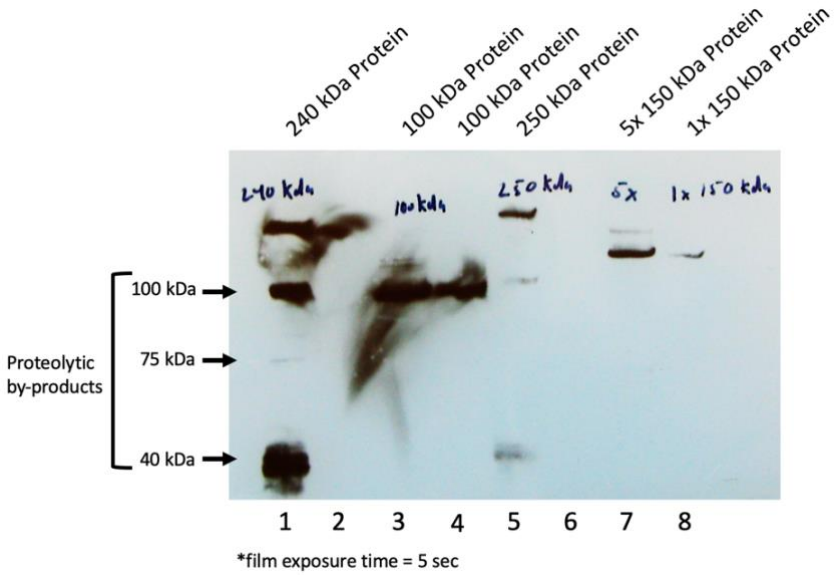


Figure 3.21 Western blot of higher molecular weight markers (100, 150, 240, 250 kDa)

Proteolytic by-products were present in 240 and 250 kDa samples (lanes 1 and 5). 5x 150 kDa protein was run next to 1x 150 kDa protein to test the efficiency of antibody binding. Western blotting was completed by Jack Williamson, Ryan Santilli, Joey Shibata.

All proteins were successfully detected and visualized on the exposure film. 100 and 150 kDa proteins showed single sharp-edged bands indicating their high purity and high concentration of the proteins. For 240 and 250 kDa proteins, the proteolytic by-products positioned at 100 and 40 kDa were also visualized along with the large marker proteins. This also proved that the proteolytic by-products all contain the antibody binding domains that allow antibody interaction. Although optimization is still necessary to normalize the intensities of the marker protein bands on Western blots, it is promising that the recombinant marker proteins were able to interact with antibodies and to be visualized on Western blots.

Chapter 4

Conclusion

4.1 Summary

Expression vectors for 11 recombinant protein markers were prepared, and all of the proteins were successfully expressed at 21°C and purified using metal affinity chromatography. Analysis of these proteins via SDS-PAGE revealed that every protein fulfilled the criteria of (a) high expression under soluble conditions in *E. coli* cells, (b) efficient purification by metal affinity chromatography, and (c) appropriate migration pattern on SDS-PAGE gel. In addition, the expression sequences of protein between 10 kDa and 100 kDa were arranged into two separate polycistronic vectors that are able to co-express four proteins at 21°C. Even though there was a slight reduction in the expression level of 20 kDa proteins in the co-expression, all eight proteins were expressed at sufficient levels and purified to a degree in which these proteins can be combined together and directly applied to an SDS-PAGE gel as molecular weight markers. The three large molecular weight markers of 150, 240, and 250 kDa were developed individually and can be prepared in an efficient process as well. There is a significant issue of by-products remaining after purification of 240 and 250 kDa, and further testing will be required to eliminate impurity in the preparation of these markers.

In addition to meeting the three important criteria for protein molecular weight markers, these recombinant proteins also fulfilled another criterion of direct interaction with secondary IgG antibodies in Western blotting. The experimental results showed that the incorporation of protein A IgG binding domains B and C in the protein markers allowed efficient antibody

interaction and the detection of these markers in Western blot. The use of double IgG binding domains enhanced the detection of protein markers by secondary antibodies, and the experiments testing numerous types of secondary antibodies proved that these markers can be recognized by IgG antibodies of various mammal species. The combination of the markers' abilities to bind a wide range of IgG antibodies in western blotting, to migrate properly on SDS-PAGE gel according to their precisely designed molecular weights, and to be prepared through an efficient expression and purification process allows these markers to be the frontrunner among the currently available molecular weight markers.

4.2 Future Directions

Due to the COVID-19 pandemic that caused the closure of all Penn State campuses in Spring 2020, we were unable to fully complete our project. The next step following the co-expression of the protein markers will be testing the co-expressed proteins in Western blotting for antibody binding and detection efficiencies. This is the final phase of the project that will validate the functionalities of our molecular weight markers.

In addition, another future goal will be expressing and purifying the markers in large scale because this improvement will allow whoever is using the molecular weight markers to prepare a high quantity of stock protein markers for future uses. During the development process of the molecular weight markers, proteins were mainly expressed and purified in the small scale for convenience. However, the preparation of these marker proteins should be something that is

done only once in a long time because a large-scale preparation can yield a large amount of proteins that can then be easily stored.

Another plan for the markers that has been explored by my adviser Dr. Tan is the pre-staining of the protein markers. Pre-staining will allow direct visualization of the markers while they are still running down on an SDS-PAGE gel, and this application can benefit researchers greatly in estimating the duration of gel electrophoresis. The conventional process of running a gel is usually controlled by checking the position of a staining dye that runs ahead of all proteins. However, this only gives one an idea of how far the smallest protein is running down, and the positions of other proteins on the gel are still unknown until the gel is stained. Pre-stained molecular weight markers will allow one to visualize how far the markers are running down on the gel during electrophoresis to estimate when to stop the process for optimal separation of proteins on the gel.

Appendix A

Preparation of Reaction Mixtures

*Adapted from the protocols written by Dr. Song Tan

PCR Amplification
PCR reaction mixture was prepared in a 0.2 mL PCR tube

Water	58.5 μL
5x Q5 Buffer	20 μL
2.5 mM dNTP	10 μL
200 ng/μL DNA	1 μL
10 μM forward primer	5 μL
10 μM reverse primer	5 μL
2 units/μL Q5 polymerase	0.5 μL
	100 μL

Vector DNA Preparation via Digestion
Appropriate restriction endonucleases were selected to yield desired vector DNA.

Water	21 μL
10x NEB Buffer	3 μL
100 mM DTT	1 μL
200 ng/μL plasmid DNA	3 μL
10-20 units/μL restriction endonuclease	1 μL
10-20 units/μL restriction endonuclease	1 μL
	30 μL

Insert DNA Preparation via Digestion
Appropriate restriction endonucleases were selected to yield desired insert DNA.

Plasmid DNA as the template:

Water	21 μL
10x NEB Buffer	3 μL
100 mM DTT	1 μL
200 ng/μL plasmid DNA	3 μL
10-20 units/μL restriction endonuclease	1 μL
10-20 units/μL restriction endonuclease	1 μL
	30 μL

PCR product as the template:

Water	14 μL
10x NEB Buffer	3 μL
100 mM DTT	1 μL
PCR product	10 μL
10-20 units/μL restriction endonuclease	1 μL
10-20 units/μL restriction endonuclease	1 μL
	30 μL

Partial Digestion

Restriction endonuclease 1 (without internal recognition site) was selected to digest one end of the DNA fragment.

Plasmid DNA as the template:

Water	44 μL
10x NEB Buffer	6 μL
100 mM DTT	2 μL
200 ng/μL plasmid DNA	6 μL
10-20 units/μL restriction endonuclease 1	2 μL
	60 μL

Restriction endonuclease 2 (with internal recognition sites) was added to the mixture for specific amount of time.

Reaction mixture with restriction endonuclease 1	60 μL
10-20 units/μL restriction endonuclease 2	1 μL

Ligation

Two conditions of reaction were set up: (A) vector only and (B) vector + insert

	A	B
Water	5.5 μL	4.0 μL
10x T4 ligase Buffer	1 μL	1 μL
100 mM DTT	0.5 μL	0.5 μL
Vector DNA	2 μL	2 μL
Insert DNA	--	1.5 μL
750 units/μL T4 DNA ligase	1 μL	1 μL
	10 μL	10 μL

Colony PCR screening

Master PCR mixture was prepared based on the number of colony samples

	4 samples	6 samples	8 samples
Water	58.5 μL	90.3 μL	116.1 μL
10x Thermo Pol Buffer	9 μL	14 μL	18 μL
2.5 mM dNTP	9 μL	14 μL	18 μL
10 μM forward primer	4.5 μL	7 μL	9 μL
10 μM reverse primer	4.5 μL	7 μL	9 μL
2 units/μL Pfu polymerase	0.5 μL	0.7 μL	0.9 μL
	86 μL	133 μL	171 μL

Add 19 μL PCR reaction mix to a 0.5 mL PCR tube
Add 1 μL of cell suspension to each PCR tube

Restriction Mapping

Appropriate combination of restriction endonucleases was selected to uniquely digest a plasmid DNA in order to distinguish the recombinant plasmid from its parent plasmid.

Water	6.5 μL
10x NEB Buffer	1 μL
100 mM DTT	0.5 μL
200 ng/μL plasmid DNA	1 μL
10-20 units/μL restriction endonuclease	0.5 μL
10-20 units/μL restriction endonuclease	0.5 μL
	10 μL

Appendix B

Solutions and Buffers

2xTY Media

1.6%	bacto tryptone
1.0%	yeast extract
0.5%	NaCl

TYE Agar Plates

1.0%	bacto tryptone	
0.5%	yeast extract	
0.8%	NaCl	
1.5%	agar	
(100 μg/mL	ampicillin)	depending on the type of plates preparing
(25 μg/mL	chloramphenicol)	

LYSIS Buffer

50 mM	glucose
25 mM	Tris-Cl pH 8.0
10 mM	Na, EDTA

NaOH/SDS

0.2 M	NaOH
1%	SDS

5 M KAc/2.5 M HAc

5 M	KAc
2.5 M	HAc

TE (10, 0.1)

10 mM	Tris-Cl pH 8.0
0.1 mM	EDTA

TE (10, 50)

10 mM	Tris-Cl pH 8.0
50 mM	EDTA

P300 – EDTA

50 mM	sodium phosphate pH 7.0
300 mM	NaCl
1 mM	benzamidine
5 mM	2-mercaptoethanol

0.5x TBE

45 mM	Tris base
45 mM	boric acid
1.5 mM	EDTA

6x Gel Loading Buffer (6x GLB)

0.25%	bromophenol blue
0.25%	xylene cyanol
30%	glycerol
60 mM	EDTA

Protein Gel Loading Buffer (PGLB)

0.5 M	Bis-Tris pH 6.8
20%	glycerol
10%	SDS
5 M	2-mercaptoethanol
0.4 mg/mL	bromophenol blue

Protein Gel Running Buffer

10 mM	Tris
76 mM	glycine
0.02%	SDS

FIX Solution

50%	ethanol
0.1 volumes	glacial acetic acid

STAIN Solution

0.5%	Coomassie Blue R
	FIX solution

DESTAIN Solution

7%	ethanol
5%	glacial acetic acid

Transfer Buffer

25 mM	Tris
192 mM	glycine

1x TBS Buffer

25 mM	Tris
150 mM	NaCl
HCI	adjust to pH 8.0

Preincubation Buffer

2%	nonfat dry milk
	1x TBS buffer

1x TTBS Buffer

0.05%	Tween 20
	1x TBS buffer

Appendix C

Car9 Tag Purification

Prior to the current project developing the universal protein molecular weight markers, I worked on another project involving evaluation of Car9 affinity purification tag in the hope of improving purification of recombinant proteins and protein complexes. The main purification method used in the laboratory is the metal affinity chromatography using Talon resin that binds polyhistidine tags. However, such methods, along with many other affinity purification methods, require specialized and costly chromatography resin. One potential solution to improve this matter was an affinity purification that uses silica gel as resin to bind a 12-amino acid peptide tag called Car9. A group of scientists developed this system and exhibited success in purification using this method in a publication (Coyle and Baneyx 2014). The amino acid sequence of Car9 tag is DSARGFKKPGKR, and the basic lysine (K) and arginine (R) are theorized to interact with hydroxyl groups present on silica (SiO_2) surfaces (Coyle and Baneyx 2014). Car9-tagged proteins would bind to silica gel, and these proteins can be eluted with 1 M L-lysine.

I prepared two protein constructs, CR9HSTNDHFR and STRHSTNDHFRCR9, through subcloning. DHFR (dihydrofolate reductase) was chosen as the protein to express because it exhibited high expression and purification levels in previous experiments done in the lab. The two recombinant proteins differed by the location of Car9 tag on the N-terminal or C-terminal end. After small-scale expression at 37°C, the proteins were purified using silica gel and Talon resins in parallel to compare the efficiency of the two purification methods. Unlike the positive results reported by the Coyle and Baneyx paper, the initial purification with Car9 tag was not

successful in that only a very low amount of proteins was purified from the silica gel set-up using pH 7.5 Tris-Cl + EDTA buffer. Some of the adjustments made to the buffer were buffer pH, sodium chloride concentration, and EDTA concentration. However, none of the modified purification conditions significantly improved the protein yield in the end. Based on the small-scale purification of STRHSTNDHFRCR9 (Figure C.1), the desired protein with Car9 tag was actually still bound to the silica gel resin after elution with lysine. This indicated that the interaction between silica gel and Car9 tag was functional as reported, but the elution buffer tested was not effective enough to replace the bound proteins for elution.

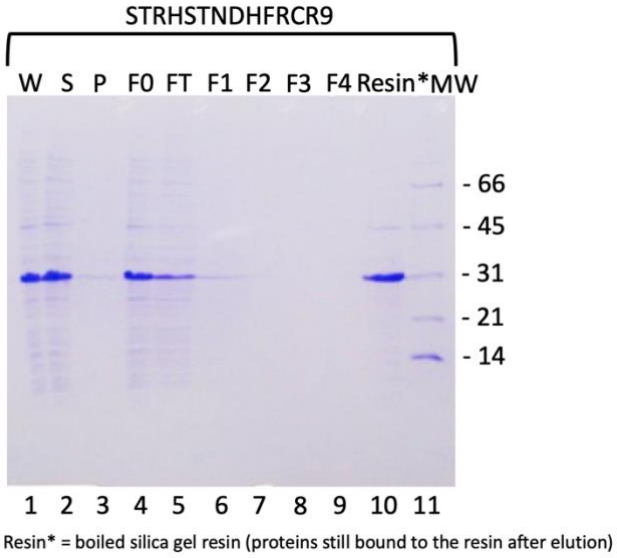


Figure C.1 Silica gel purification of STRHSTNDHFRCR9

W = whole cell extract of protein-expressing *E. coli*. S = supernatant of the whole cell extract. P = pellet collected from the whole cell extract. F0 = protein sample before resin binding. FT = flow through in silica gel chromatography. "F1, F2,..." = fractions collected from the elution of purified proteins. Resin* = silica gel resin after elution was boiled to check the presence of proteins still bound to the resin.

BIBLIOGRAPHY

Bio-Rad Laboratories, Inc. Instruction Manual for Precision StrepTactin HRP-Conjugate and AP-Conjugate.

Blancher C, Jones A. 2001. SDS-PAGE and Western blotting techniques. Methods in Molecular Medicine. 57:145-162.

Bornhorst JA, Falke JJ. 2000. [16] Purification of Proteins Using Polyhistidine Affinity Tags. Methods in Enzymology. 326:245-254.

Brunelle JL, Green R. 2014. Chapter Twelve - One-dimensional SDS-Polyacrylamide Gel Electrophoresis (1D SDS-PAGE). Methods in Enzymology. 541:151-159.

Burnette WN. 1981. "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Analytical Biochemistry. 112(2):195-203.

Clontech Laboratories, Inc. 2007. TALON Metal Affinity Resins User Manual.

Coyle BL, Baneyx F. 2014. A cleavable silica-binding affinity tag for rapid and inexpensive protein purification. Biotechnology and Bioengineering. 111(10):2019-2026.

Deisenhofer J. 1981. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. Biochemistry. 20(9):2361-2370.

Fleischman A. 2010. Western Blot Marker with Universal Immunoglobulin Binding Domain. University Park (PA): The Pennsylvania State University.

Fritsch and Krause. 2003. Electrophoresis. Encyclopedia of Food Sciences and Nutrition 2nd Edition. Cambridge (MA): Academic Press.

Hermanson GT. 2013. Chapter 15 - Immobilization of Ligands on Chromatography Supports. In: Bioconjugate Techniques, Third Edition. Cambridge (MA): Academic Press. 589-740.

Jansson B, Uhlén M, Nygren P. 1998. All individual domains of staphylococcal protein A show Fab binding. FEMS Immunology and Medical Microbiology. 20(1):69–78.

Kumar G. 2018. Multicolored Prestained Standard Protein Marker Generation Using a Variety of Remazol Dyes for Easy Visualization of Protein Bands During SDS-PAGE. Methods in Molecular Biology Protein Gel Detection and Imaging. 1853:19-25.

Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. 2005. Comparison of affinity tags for protein purification. Protein Expression and Purification. 41(1):98-105.

Moks T, Abrahmsén L, Nilsson B, Hellman U, Sjöquist J, Uhlén M. 1986. Staphylococcal protein A consists of five IgG-binding domains. European Journal of Biochemistry. 156:637-643.

Murphy K, Weaver C. 2017. Chapter 4 – Antigen Recognition by B-cell and T-cell Receptors. In: Janeway's Immunobiology 9th Edition. New York (NY): Garland Science. 139-172.

Nilsson B, Moks T, Jansson B, Abrahmsén L, Elmblad A, Holmgren E, Henrichson C, Jones TA, Uhlén M. 1987. A synthetic IgG-binding domain based on staphylococcal protein A. Protein Engineering. 1(2): 107-113.

Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of the National Academy of Sciences. 106(6):1760-1765.

Schüchner S, Andorfer P, Mudrak I, Ogris E. 2016. Anti-RAINBOW dye-specific antibodies as universal tools for the visualization of prestained protein molecular weight markers in Western blot analysis. Scientific Reports. 6:31363.

Tan S, Kern RC, Selleck W. 2005. The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. Protein Expression and Purification. 40(2):385-395.

Tan, S. 2001. A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. Protein Expression and Purification. 21(1):224-34.

Tan S. 2020. Tan Lab Protocols. University Park (PA): The Pennsylvania State University.

Tirode F, Busso D, Coin F, Egly JM. 1999. Reconstitution of the Transcription Factor TFIIH: Assignment of Functions for the Three Enzymatic Subunits, XPB, XPD, and cdk7. Molecular Cell. 3(1):87-95.

Urh M, Simpson D, Zhao K. 2009. Affinity Chromatography: General Methods. Methods Enzymology. 463:417–38.

ACADEMIC VITA

Yoshitaka (Joey) Shibata

EDUCATION

Bachelor of Science in Biochemistry and Molecular Biology
Minor in Political Science

Class of 2020

The Pennsylvania State University – Schreyer Honors College | *University Park, PA*

RESEARCH EXPERIENCE

Undergraduate Research – Dr. Song Tan Laboratory

December 2016 – May 2020

The Pennsylvania State University – Department of Biochemistry and Molecular Biology | *University Park, PA*

- Developed universal protein molecular weight markers applicable to SDS-PAGE and Western Blotting
- Worked to improve protein purification using protein affinity tag Car9

Schreyer MD/PhD Summer Exposure Program – Dr. Laura Carrel Laboratory

May 2018 – August 2018
May 2019 – August 2019

The Penn State College of Medicine | *Hershey, PA*

- Studied genetic correlations between X chromosome inactivation and systemic lupus erythematosus
- Tested expression activity of specific genes within X chromosome that contribute to X chromosome inactivation
- Conducted clinical shadowing with physicians in different areas for total of 103 hours (pathology, gastroenterology, hematology, vascular surgery)

LEADERSHIP EXPERIENCE

Schreyer Honors College Medical Career Development Team

January 2019 – May 2020

The Pennsylvania State University – Schreyer Honors College | *University Park, PA*

- Organizing career development panels and seminars for scholars interested in pursuing medical professions
- Assisting Schreyer Honors College Administrative Team in planning college-wide career development events

Resident Assistant – Schreyer Honors College Special Living Option

August 2018 – May 2020

The Pennsylvania State University – Penn State Residence Life | *University Park, PA*

- Ensured safety and well-being of residents living in residence hall by responding to various situations including roommate conflict, suicidal ideation, sexual assault, and alcohol/drug abuse
- Provided guidance, support, policy enforcement, and campus resources for 60 residents on assigned floor and 700 residents in the residence hall

President

April 2017 – April 2018

Penn State Japanese Friendship Association | *University Park, PA*

- Oversaw the Executive Board of the organization and conducted weekly meeting with officers
- Communicated with leaders from other organizations to improve the international community at Penn State and promote cultural diversity

VOLUNTEERING EXPERIENCE

Mentor

January 2019 – May 2020

LifeLink PSU | *University Park, PA*

- Assisted college-level students with disabilities or special needs in classes and on-campus activities
- Provided academic supports for LifeLink students such as class assignments and homework

Clinic Volunteer

September 2018 – May 2020

Remote Area Medical – Penn State Chapter | *University Park, PA*

- Assisted professionals in medical, dental, and vision cares during Remote Area Medical clinics to provide free healthcare services to people in rural areas (total of 40+ hours of volunteering)

AWARDS & HONORS

- Eberly College of Science Charles R. Gerth Scholarship Award
- Eberly College of Science Sinha Family Trustee Scholarship
- J. W. Van Dyke Memorial Scholarship
- Penn State Undergraduate Exhibition University Fellowship & Phi Kappa Phi Peter T. Luckie Award
- Schreyer Honors College Academic Excellence Scholar Award
- Dean's List (Fall 2016, Spring 2017, Fall 2017, Spring 2018, Fall 2018, Spring 2019)

POSTERS & PRESENTATIONS

- 2020 Penn State Undergraduate Exhibition
- 2019 Penn State College of Medicine Summer Undergraduate Research Symposium
- 2019 Penn State Undergraduate Exhibition
- 2018 Penn State College of Medicine Summer Undergraduate Research Symposium
- 2018 Penn State Undergraduate Exhibition