THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

THE PENN STATE PROTEIN LADDERS: AN INEXPENSIVE SOURCE OF PROTEIN MOLECULAR WEIGHT MARKERS APPROPRIATE FOR BOTH COOMASSIE STAINING AND WESTERN BLOTTING

RYAN SANTILLI
SPRING 2020

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biology- Vertebrate Physiology
with honors in Biology

Reviewed and approved* by the following:

Dr. Song Tan
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Dr. Bernhard Luscher
Professor of Biology, Biochemistry and Molecular Biology, and Psychiatry
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Although protein molecular weight markers are among the most commonly used reagents in molecular biology laboratories, simple and inexpensive means to produce such markers are lacking. We have therefore designed and prepared recombinant proteins of defined molecular weights from 10 to 150 kDa that migrate appropriately on SDS-PAGE gel. Each protein was selected for high level expression in E. coli and contains a HIS-tag for efficient metal-affinity purification. In addition to detection by Coomassie or silver staining, each protein contains an IgG binding domain and is therefore visible on Western blots without specific second antibodies. The Penn State protein ladder system provides an inexpensive method to produce protein molecular weight markers for the research laboratory.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to acknowledge everyone that has crossed my path in my four years here at The Pennsylvania State University for shaping me into the young scholar that I am today: To those that came and went, to my mentors, to my friends, to my family and those whom I do not directly mention in my thoughts that follow.

When I reflect on the values of education and the power that it entrusts to those who wield it, two very special people come to mind, my Grandmom and Grandpop. I want to thank them endlessly for instilling an integral component of my core beliefs that fuels my determination to learn and broaden my horizons. Grandpop, I want to commemorate your accomplishments in the medical field, and I can only strive to be half the physician that you were. Thank you for providing me with a direction and a path that you have carved merely with intelligence and drive. Grandmom, thank you for showing me the products of resilience and hard work. Your qualities have imprinted onto me and taught me when the "going gets rough, suck it up and push through." You both have been pillars of wisdom, generosity, and honesty throughout my entire life.

Next, to my family, Mom, Dad, and Zach. Thank you for your continual and everlasting support. Mom and Dad, you both have made many sacrifices to ensure that I was able to follow my dreams and have the world at my fingertips. You both have created an environment full of love, guidance, and the right amount of push to mold me into the young man that I am today. Zach, thank you for always regarding my needs above your own and ensuring that I am equipped with the necessary knowledge to carve out my own path.

Finally, to the members of the Tan Laboratory. Song, I simply cannot express the amount of gratitude I have towards you in words, but here I must try. I firmly believe that I have learned at least one piece of valuable information, whatever it may have concerned, in every interaction we had; from my first interview, in which I quickly learned to never chew gum again, to the present, providing your insights on possible reasons we saw reduced expression in our polycistronic plasmid. Your eagerness to share your knowledge is admirable and the breadth of that very knowledge is commendable. Joining your laboratory has changed my entire college career. If it was not for you, I would not even be sitting here writing this, since it was your suggestion that persuaded me to apply to the Gateway Program. You saw something in me that I did not even see in myself and pushed me to lengths, that at the time, I considered unreachable. And even despite your sometimes-chaotic schedules you never failed to assist me. For that, I sincerely thank you as you are the best mentor and role model that I know.

To Jack and Joey, my fellow undergraduates that assisted me on this project, I could not have finished this without you. To Rosie who mentored me for my first semester in the lab, to the techs, Owen, Szu-Yu, Jeff, and Gabe, thank you all for being so patient and guiding. To the post-docs and grad students, thank you for always answering my repetitive, pestering questions that I know must have got on your nerves. Thank you, Tan lab, for providing an amazing intellectual experience and notable examples of accomplished scientists.

P.S. There is a very special software known as FileMaker Pro (AKA Song's Database) that I would like to give many thanks to. It helped tremendously in the organization of my laboratory experience and the writing of this thesis.

Chapter 1

Introduction

1.1 Proteins & Molecular Weight

Determining the structural, chemical, and functional properties of novel proteins is of vital importance in the field of biochemistry. Initial characterization of a protein's biochemical activity and function is central to many research projects around the world. A key parameter in identifying an unknown protein is determining its molecular weight.

Proteins are classified as one of the four macromolecules that constitute all living matter. They are made up of hundreds to thousands of amino acids connected by peptide bonds between the amino nitrogen and the carboxyl group to form a polypeptide. Proteins have a wide variety of functions, including but not limited to binding foreign particles to help protect the body (immunoglobulins), catalyzing biochemical reactions (enzymes), providing structural support to cells, and transporting molecules throughout the body.[1]

Differences in their chemical makeup allow them to carry out the wide variety of functions listed above. There are four different levels to protein composition: primary, secondary, tertiary, and quaternary. Primary structure is the unique sequence of amino acids in each polypeptide chain that makes up the protein. There are a total of 20 amino acids that are naturally produced in living organisms. The four classifications of these amino acids based on their side chains (non-polar and neutral, polar and neutral, acidic and polar, basic and polar) and their ability to be arranged in many combinations allows for the vast array of proteins. Secondary structure results from the interactions of the side chains to form -helices and -sheets because of

hydrogen bonding between carbonyl and amino groups in the peptide backbone. Tertiary structure is the completely folded 3-dimensional structure that is determined by hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages. Quaternary structure describes how proteins with multiple polypeptide chains, creating more than one subunit, are oriented in space.[2]

Protein molecular weight is determined by the primary structure of the protein. Each amino acid has a specific weight based on its chemical composition. In 1803, John Dalton proposed to use the atomic mass of hydrogen as the natural unit of atomic mass. This became the basis of the atomic weight scale. The Dalton, or unified atomic mass unit (symbols: Da or u) is currently used as a unit of mass in physics and chemistry and is defined as 1/12 of the mass of an unbound neutral atom of carbon-12 in its nuclear and electronic ground state. The atomic mass constant denoted, m_u is defined identically, giving:[3]

$$M_u = m\,(^{12}\mathrm{C})/12 = 1\ \mathrm{Da}$$

The Dalton is related to the molar mass of proteins. The mass of one mole of substance, measured in grams, is numerically equal to the average mass of one if its constituent particles, measured in Daltons.[3] Proteins are usually measured in kilodaltons (1 kDa = 1 kiloDalton or 1000 times the weight of a hydrogen atom) due to their massive size. This is calculated by adding up all of the weights of the amino acids in its primary structure. The average weight of an amino acid is 110 Da. The molecular weight of all 20 amino acids are listed below:

Table 1. Molecular Weight of Amino Acids

Amino Acid	Three-Letter Abbreviation	One-Letter Symbol	Molecular Weight
Alanine	Ala	A	89 Da
Arginine	Arg	R	174 Da
Asparagine	Asn	N	132 Da
Aspartic Acid	Asp	D	133 Da
Cysteine	Cys	C	121 Da
Glutamine	Gln	Q	146 Da
Glutamic Acid	Glu	E	147 Da
Glycine	Gly	G	75 Da
Histidine	His	H	155 Da
Isoleucine	Ile	I	131 Da
Leucine	Leu	L	131 Da
Lysine	Lys	K	146 Da
Methionine	Met	M	149 Da
Phenylalanine	Phe	F	165 Da
Proline	Pro	P	115 Da
Serine	Ser	S	105 Da
Threonine	Thr	T	119 Da
Tryptophan	Trp	W	204 Da
Tyrosine	Tyr	Y	181 Da
Valine	Val	V	117 Da

1.2 Polyacrylamide Gel Electrophoresis

Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis (SDS-PAGE) is the most common and reliable technique to obtain the molecular weight of proteins. While other methods

of determining molecular weight by sequencing the primary structure exist; such as N-terminus Edman Degradation (coupled with High Performance Liquid Chromatography), C-terminus Hydrazinolysis (coupled with High Performance Liquid Chromatography), and Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry (MALDI-TOF Mass Spec), they are often costly and time consuming. Examples of more methods include analytical ultracentrifugation and light scattering. Similarly, these methods are impractical due to their requirement of highly purified proteins and expensive equipment.[4]

SDS-PAGE is a technique for separating and analyzing proteins based on mass. The physical and chemical properties of proteins can affect the way that they migrate during gel electrophoresis. For example, a protein can have a net positive, net negative, or neutral charge depending on the pH of solution and its amino acid composition. A protein also exhibits a three-dimensional structure. The shape of this structure, whether it be more rod-like or more spherical, can influence how the protein migrates through the pores of the gel. To standardize these properties, sodium dodecyl sulfate (SDS) is used to denature the protein. SDS is a detergent that is comprised of a hydrocarbon chain bonded to a negative sulfate group. It binds to proteins in a constant ratio of 1.4 grams of SDS per gram of protein,[4] preserving the charge to mass ratio. The negative charge of SDS disrupts proper protein folding without breaking any covalent bonds. This causes the protein to unfold to a rod-like shape while retaining the primary sequence. However, some proteins contain disulfide linkages between cysteine residues resulting in reduced SDS binding. Decreased detergent binding can cause anomalous migration by disturbing the charge to mass ratio. Reducing agents such as ß-mercaptoethanol are added to break the disulfide bonds allowing the SDS to completely dissociate. The amount of negative charge that

the SDS confers greatly exceeds the intrinsic charges of the individual amino acids, effectively denaturing the protein.[4]

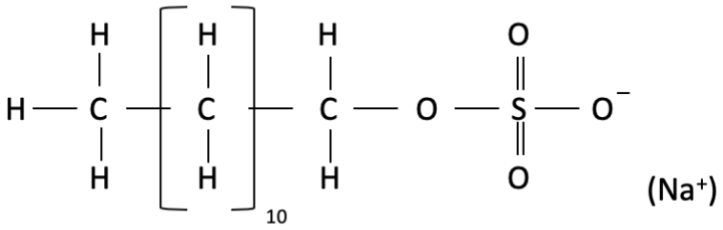


Figure 1. Chemical Structure of Sodium Dodecyl Sulfate (SDS)

The polyacrylamide gel that the protein samples migrate through is a three-dimensional network of the monomer, acrylamide, reacted with the cross-linking agent, methylenebisacrylamide (bis-acrylamide). A free radical generator, often ammonium persulfate, aids in polymerization. The pore size is controlled by the ratio of acrylamide to bis-acrylamide. A higher concentration of acrylamide, resulting in a smaller pore sized, is used for smaller proteins while a lower concentration is used for bigger proteins. Gels can either be prepared homogenously or on a gradient typically ranging from 4% (at the top) to 20% (at the bottom). The homogenous composition is typically used in the experimental setting due to its ability to create a wider separation of proteins that occupy narrow ranges of molecular weights.[4] When the samples are loaded in the gel, application of an electrical current through an electrophoretic chamber cause the proteins to migrate toward the positively charged anode (after being denatured by SDS). The smaller proteins pass through the gel matrix much easier, allowing them to migrate faster (closer to the bottom of the gel) while the bigger proteins migrate much slower (closer to the top of the gel). After the conclusion of the electrophoresis, the proteins can be

visualized by various staining methods such as Coomassie Blue, the most common and easy to use, or silver staining if a higher sensitivity is needed. Before running an unknown protein sample, a lane containing a protein ladder of known molecular weights must be present to provide a standard reference.

1.3 Commercially Available Protein Molecular Weight Markers & Their Limitations

Protein molecular weight markers are essential tools used in biochemistry and molecular biology research to estimate the molecular weight of proteins of interest. They are among the most ubiquitous reagents in molecular biology designed as a set of references for electrophoretic separation.

Commercially available markers are currently produced such as the BioRad Low Molecular Weight Markers (LMW); however, they have many limitations. The BioRad LMW set is based on non-recombinant proteins. Many of them are used predominantly for Coomassie Blue Staining because they do not possess a secondary antibody binding domain. The lack of this region impedes the ability for the markers to be visualized on another widely used visualization technique known as immunoblotting. Indeed, researchers have produced markers that contain linear epitopes that are derived from the heavy chain constant regions of mouse and rabbit immunoglobulin G (IgG Fc LE). However, they do not provide evidence of high expression and detailed purification to be visualized on SDS-PAGE.[5]

Other available protein molecular weight markers do not contain the polyhistidine affinity tag for easy protein expression by metal affinity chromatography. Protein molecular weight markers have been made to incorporate a “mega-tag” which contains 14 of the most

commonly used epitope tags, however, their findings do not incorporate the amount of protein they can purify from a single 100ml preparation.[6] If the amount of protein purified is low, many preps have to be conducted reducing the efficiency of the purification scheme.

Many protein molecular weight ladders do not include proteins with an appropriate size range or a regular fixed interval. This can cause confusion or inaccurate assessment of the protein of interest. A published ladder made of disulfide bridges carries the limitation of introducing any reducing agents that could compromise its ladder function during electrophoresis.[7] Also, the molecular weight kDa numbers are not based on the actual molecular weight of the protein amino acid sequence. This can cause anomalous migration on gels of varying acrylamide composition.

Lastly, commercial markers are expensive, costing approximately $1/lane.

1.4 Creation of Universal Protein Molecular Weight Ladder

Designing a universal protein molecular weight marker must address the aforementioned limitations while also providing a simple means of producing an inexpensive marker. For this project, we combined the use of a universal antibody binding domain utilizing recombination with proteins that were selected using the following criteria: high expression level in *E. coli*, solubility, high purification yield, and appropriate migration on SDS-PAGE.

1.4.1 Recombinant DNA & Proteins

The technology for propagating and expressing recombinant genes was invented by Stanley Cohen and Herbert Boyer in 1973.[8] Recombinant DNA allows for the combination of

genetic material that would not otherwise be found in the genome. For this project, the gene encoding a desired protein was isolated via restriction endonucleases and subcloned into *E.coli* expression vectors to make recombinant proteins. A theoretical process can be viewed in Figure 2. This technology allows for the assemblage and expression of multiple protein fragments to yield a desired molecular weight.

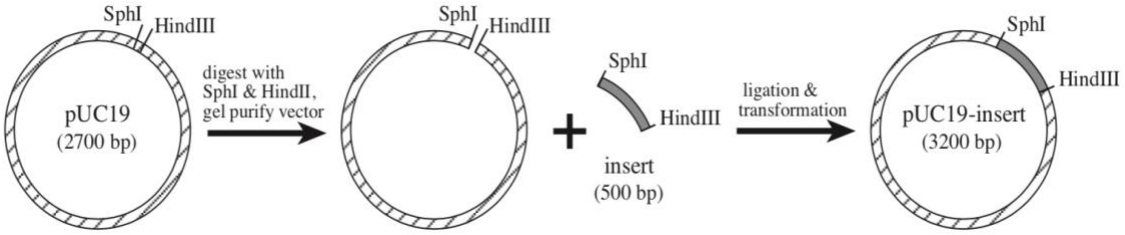


Figure 2. Subcloning of SphI-HindIII fragment into pUC19 (theoretical). From Dr. Song Tan's Guide to Subcloning protocol.

1.4.2 IgG Binding Domain and Immunoblotting

Protein A is a 42 kDa surface protein originally found in the cell walls of the bacteria, *Staphylococcus aureus*.[9] It contains five immunoglobulin (IgG) binding domains (E, D, A, B, C in order starting from the N-terminus), each with a similar but not identical sequence.[10] Immunoglobulins, also known as antibodies, are glycoprotein molecules that are comprised of two copies of each heavy chains (Fc fragments) and light chains (Fab fragments). Each of the five IgG binding domains of Protein A is sufficient to bind to the Fab or Fc fragment of an IgG antibody. The components of an immunoglobulin molecule is pictured in Figure 3.

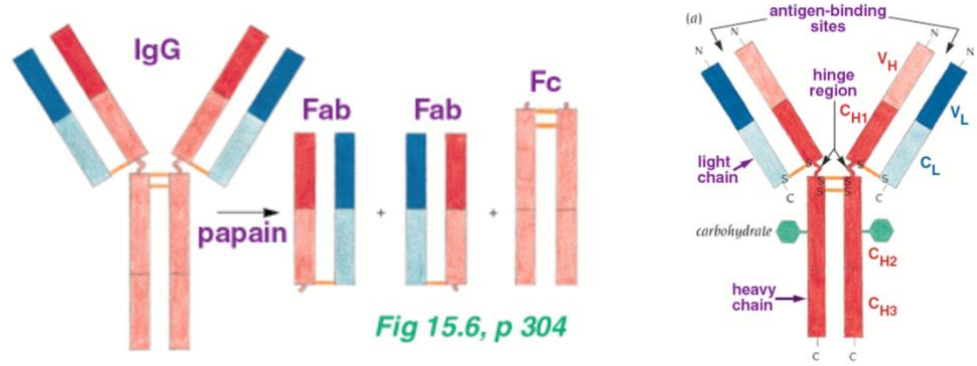


Figure 3. IgG Immunoglobulin Molecule. Light chain folds into variable and constant domains. Heavy chains folds into 4 domains: one variable and three constant domains. Each domain contains about 110 amino acids. Figure used by permission from Dr. Song Tan's BMMB531 Spring 2018 lecture slides, based on a figure in Branden and Tooze, 1999.

Immunoblotting, also known as Western blotting, utilizes antigen-antibody recognition for the detection and characterization of proteins. The technique consists of three major steps: separating by size on an electrophoresis gel, transferring the separated proteins from the gel to a membrane, and detecting of the target protein using antibodies.[11] One typically uses a primary antibody which specifically recognizes the protein of interest, and a secondary antibody which both recognizes the primary antibody and may be linked to an enzyme for enzymatic detection.

Previous work in the Tan lab demonstrated that the immunoglobulin binding domain of Protein A can be incorporated into a molecular weight marker to be visualized on a western blot. The Protein A ZZ domain, where the Z domain is a modified consensus of IgG binding domain and ZZ is a tandem repeat of that single domain, was shown to bind antibody linked for chemiluminescent detection.[10,12] Based on this result, Protein A domain B, abbreviated as PAB, was used for Western Blotting application.

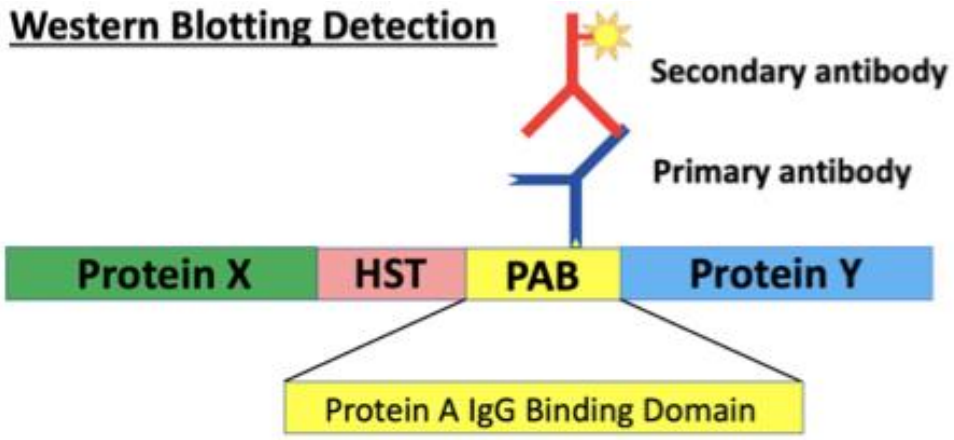


Figure 4. Protein A Domain B Western Blot Detection

1.4.3 Polycistronic Expression System

The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli (E. coli)* designed by Dr. Song Tan enables up to 4 genes, encoding different proteins, to be co-expressed from a single mRNA transcript.[13] The unique, but compatible origin of replication, specific restriction enzyme sites designed to correspond to four different transfer cassettes, and the addition of N- and C- terminal tags allows for a simple and flexible modular design system. Although this system was developed for coexpressing protein complex subunits to reconstitute a protein complex in vivo, we have utilized the system to express multiple different proteins at the same time. This development is important for the utility of our protein ladder system because it allows for an 8-protein molecular weight ladder to be expressed in two pST44 plasmids, highlighting the simplicity of expressing and purifying our markers.

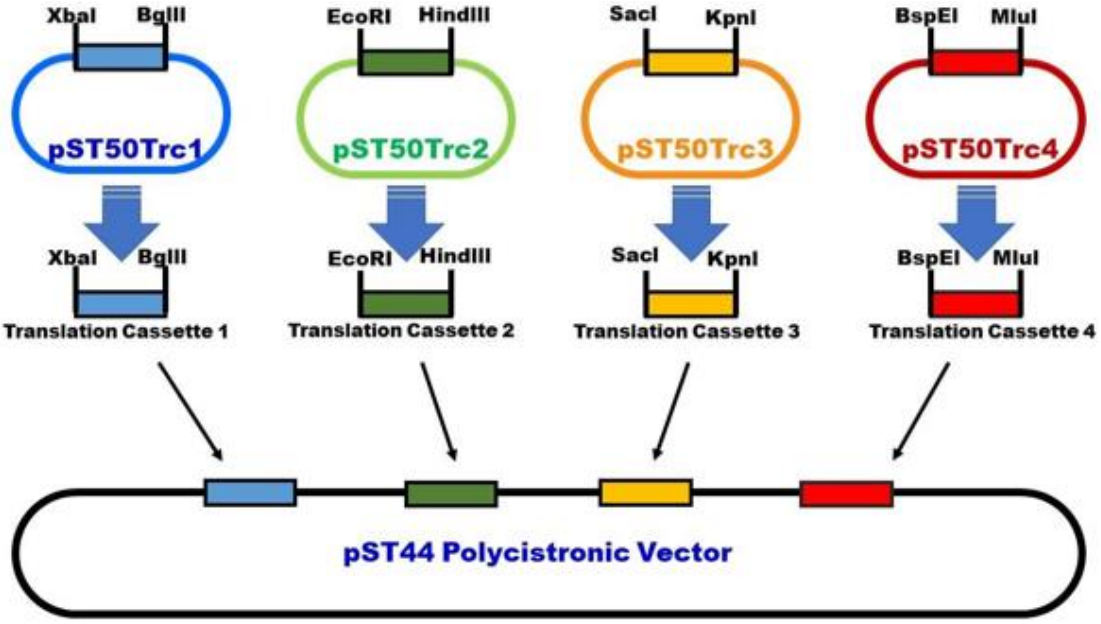


Figure 5. pST44 Polycistronic Co-expression Vector System. Adapted to Creation of Protein Ladder.

1.4.4 Polyhistidine Metal Affinity Chromatography

Immobilized Metal Affinity Chromatography (IMAC) is the most widely used technique for single-step purification of recombinant proteins.[15] It works by separating proteins or peptides according to their affinity for metal ions that have been immobilized by chelation to an insoluble matrix. The 10x polyhistidine tag that has been incorporated into all of our recombinant proteins readily forms coordination bonds with the immobilized transition metal, cobalt, that is used in the stationary phase (carboxymethyl aspartate agarose) of TALON resin. The resin is unaffected by nuclease or protease activity and are appropriate for fusion proteins from crude cell lysates.[16] The chelation allows for the desired protein to stick to the resin while the other unwanted protein fragments are washed away with a buffer. Imidazole, a substituent of histidine, is used to elute

the polyhistidine tagged proteins from the stationary phase of the TALON resin so they can be collected and analyzed.

1.5 Experimental Objectives

The goal of this work was to provide a simple and inexpensive means of producing protein molecular weight markers appropriate for both SDS-PAGE and western blotting. The markers were designed on a regular 10 kDa increment ranging from 10 kDa to 100 kDa (10, 20, 30, 40, 50, 60, 80, 100) across two plasmids using the polycistronic vector system. Single expression vectors of 150 kDa and 250 kDa were also made to accommodate researchers who work on bigger proteins. Each protein incorporates a commonly used affinity tag (HST, i.e. 10x repeat of histidine, His-Ten) and an IgG binding domain for simple metal affinity chromatography purification and western blot application, respectively. The markers were optimized for practical use and we plan to distribute the expression plasmids without any licensing fees, similar to what our laboratory has done for the equivalent DNA molecular weight markers. The Penn State Protein Molecular Weight Markers are expected to have worldwide impact on the scientific community as an economical source of robust biochemical reagents.

Chapter 2

Materials and Methods

2.1 Bacteriological Methods

2.1.1 Bacterial Strains Used

Escherichia coli (E.coli) cells were used for all experimental work. TG1 cells were used for DNA methods to design plasmids encoding the genes to express the proteins of interest. BL21(DE3)pLysS cells were used for the expression of proteins. In one experiment where the TG1 cells were ineffective at harboring the correct plasmid, HB101 cells were used.

Table 2. Bacterial Strains

Strain	Source	Genotype	Antibiotic Resistance
TG1	Toby Gibson	(lac-pro), supE, thi, hsdD5/F', traD36, $proA^{+}B^{+}$, $lacl^{q}$, lacZ, M15	Ampicillin
BL(DE3)pLysS	Stratagene	B F-, ompT, gal, dcm, lon, $hsdS_{B}$ (r_{B}^{-} m_{B}^{-}), (DE3), pLysS(cm^{R})	Ampicillin and Chloramphenicol
HB101	Stratagene	F- mcrB mrr hsdS20(r_{B}^{-} m_{B}^{-}) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl- rpsL20 (Sm^{R}) glnV44 -	Ampicillin

2.1.2 Bacterial Media

2xTY media was used for plasmid transformation, 100 mL plasmid preparations and expressing proteins using the T7 system. 1.6% (w/v) bacto-tryptone, 1.0% (w/v) yeast extract, and 0.5% (w/v) NaCl in deionized water was mixed and sterilized by autoclaving for 20 minutes at 121°C and 15psi. The appropriate antibiotic was added before use.

Solid TYE media was prepared by mixing 1.0% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 0.8% (w/v) NaCl, and 1.5% (w/v) agar in deionized water. The solution was then autoclaved, cooled to 60°C, combined with the appropriate antibiotic(s) and poured into sterile petri dishes.

Table 3. Antibiotic Concentrations

Antibiotic	Liquid Media Concentration	Solid Media Concentration
Ampicillin	50 µg/mL	100 µg/mL
Chloramphenicol	25 µg/mL	25 µg/mL

2.2 DNA Methods

2.2.1 Linear Chain Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used to amplify the genes of the desired proteins. The DNA sequence of the gene to clone was obtained. The sequence was then examined for any internal restriction sites. Forward and reverse primers were designed to anneal to the gene with a T_m (melting temperature of the primers) of 50°C or higher. Sequencing primers were then

designed if necessary. Before preparing PCR reaction mixture, the template DNA was diluted to 10 ng/µl in TE (10, 0.1) [10 mM Tris, 0.1 mM EDTA]. The PCR reaction mixture was then prepared with water, 10x Thermo Pol Buffer, 2.5 mM dNTP, 10 ng/µl DNA template, 10 µM forward primer, 10 µM reverse primer, and 2 units/ul Pfu polymerase. The mixture was reacted in the thermocycler as followed: 5 x (30 sec 95°C → 30 sec T_m -5°C → 30 sec 75°C) → 25 x (30 sec 95°C → 30 sec 60°C → 30 sec 75°C) → 3 min 75°C → 15°C. The PCR product was analyzed: added 5 µl of the PCR reaction to 1 µl 6xGLB and loaded 5 µl onto a 10% acrylamide (40:1), 1xTBE minigel if the expected product is less than 1000bp. Electrophorese in 1x TBE at 10W for 20-25 minutes. If the expected product was greater than 1000bp, 10 µl of the PCR reaction was added to 2 µl of 6xGLB and 10 µl was loaded onto a 1% agarose gel.

2.2.2 Phenol/Chloroform Extraction and Ethanol Precipitation of PCR Product

100 µl of the PCR product was extracted with 100 µl phenol/ 24:1chloroform: amyl alcohol, followed by one extraction with 400 µl 24:1chloroform: amyl alcohol. The sample was ethanol precipitated by adding 10 µl 3 M NaAc pH 5.2, 250 µl 100% ethanol, vortexing for 5 seconds and centrifuging in the microcentrifuge for 10 min at room temperature (RT). The supernatant was removed by aspiration and the pellet was allowed to dry for 5 min. The pellet was then resuspended in 30 µl TE (10, 0.1).

2.2.3 Preparation of Vector DNA

The correct DNA plasmid was digested with the appropriate restriction enzymes. The digestion mixture was comprised of water, 0.2 µg/µl plasmid DNA, 10x NEB Buffer, 100 mM

DTT, 10-20 units/µl restriction enzyme 1, and 10-20 units/µl restriction enzyme 2. A third enzyme was commonly used to separate the vector DNA for increased accuracy on agarose gel purification. The digestion was usually performed at 37°C for 2 hours, though the temperature may vary depending on the restriction enzymes used. A double digest was also used when digesting enzymes of different salt conditions. Here, one enzyme digested the DNA for 1 hour, and the second enzyme was added for 1 hour after the appropriate salt conditions were created. Singly cut vectors or vectors with compatible ends were treated with calf-intestinal phosphatase (CIP) to prevent self-ligation. Vectors with protruding 5' ends were treated with 0.1 units CIP and incubated at 50°C for 30 minutes. Vectors with blunt or recessive 5' ends were treated with 1-unit CIP and incubated at 50°C for 45 minutes. 2 µl of 6xGLB was added to the reaction mixture to stop the reaction. The DNA was agarose gel purified and isolated by centrifugation.

2.2.4 Preparation of Insert DNA

Insert DNA was prepared form digesting a plasmid or a PCR product. If a PCR product was used, it was first phenol/ClA extracted and ethanol precipitated. The reaction mixture contained water, 0.2 µg/µl plasmid DNA or PCR product in TE, 10x NEB Buffer, 100mM DTT, 10-20 units/µl restriction enzyme 1, and 10-20 units/µl restriction enzyme 2. The digestion was performed at 37°C unless specific enzymes required other temperatures. 2 µl of 6xGLB was added to the reaction mixture after 2 hours to stop the reaction. The DNA was agarose gel purified and isolated by centrifugation.

2.2.5 Partial Digest

Partial Digestion was used if the desired DNA fragment contained an internal restriction enzyme site. If a PCR product was used, it was first phenol/chloroform extracted followed by ethanol precipitation. The DNA was digested with restriction enzyme 1 for 1.5-2 hours at the appropriate temperature. Enzyme 2 was then added to the digested DNA using short incubation times. For enzymes other than BamHI, 60 µl of restriction enzyme 1 digested DNA, additional NaCl (if needed), and 1 µl of 10-20 units/u of restriction enzyme 2 were incubated at the appropriate temperature. 20 µl aliquots were removed at 1, 2, and 5 minutes and transferred to Eppendorf tubes with 4 µl of 6xGLB. For BamHI restriction enzyme 2, 60 µl of restriction enzyme 1 digested DNA, additional NaCl (if necessary), and 0.2 ul of 20 units/ul BamHI were incubated at 37°C. 20 µl aliquots were removed at 15, 30, and 60 seconds and transferred to Eppendorf tubes with 4 µl of 6xGLB. The time points were loaded on a preparative agarose gel and isolated by centrifugation.

2.2.6 Agarose Gel Electrophoresis of DNA

Agarose Gel Electrophoresis was used for analytical purposes when assessing the size of DNA fragments and was also used to purify vector/insert DNA. The agarose concentration was first determined based on the size of the desired fragment (1.0% agarose being the default). 0.30g of agarose was added to 30 ml of 0.5x TBE buffer in a 125 ml Erlenmeyer flask. 3 to 5 ml of deionized water was added, and the mixture was allowed to swell for 5 minutes at RT. The mixture was heated in the microwave for 90 to 120 seconds until all of the agarose was dissolved. 1.5 µl of 10 mg/mL ethidium bromide was added and the solution was allowed to cool

for 5 min. The gel casting block was set up using analytical combs (15 well; 10 µl) for analytical digests and preparative combs (8 well; 30 ul) for preparative digests. The solution was poured into the casting tray and allowed to polymerize for 30 minutes. Samples were loaded into the gel along with the appropriate molecular weight markers. The gel was run at 125 V until the bromophenol blue was near the bottom of the gel and imaged over UV transillumination.

2.2.7 Isolation of DNA from Agarose by Centrifugation

The DNA sample was electrophoresed on the 8 well preparative agarose gel for roughly 40 min at 125 V. A filter assembly was prepared by heating a 25-gauge needle in a flame and piercing the bottom of a 0.5mL Eppendorf tube. Siliconized wool was stuffed into the bottom of the Eppendorf tube and the 0.5 ml tube was placed in a 1.5mL Eppendorf tube. The gel was imaged over the UV transilluminator and moved to a long wavelength UV where the desired band was excised with a clean razor blade. The gel slice was transferred to the filter assembly and microcentrifuged at 7000 rpm for 3 min. The 0.5 mL Eppendorf tube was discarded to yield the purified DNA fragment in the 1.5 mL tube.

2.2.8 Ligation of Sticky Ended DNA

Two ligation mixtures were prepared to sub-clone the insert DNA into the vector DNA: one with only vector DNA and the other with vector and insert DNA. The vector only mixture consisted of 5.5 µl water, 1 µl10x T4 DNA ligase buffer, 0.5 µl 100mM DTT, 2 µl gel purified vector DNA (~2 ng/µl), and 40 units/µl T4 DNA ligase buffer (750 units/µl was used if the DNA was digested with NdeI due to the 2 bp overhang rather than 3). The vector plus insert mixture

contained 4 µl water, 1 µl T4 DNA ligase buffer, 0.5 µl 100mM DTT, 2 ul gel purified vector (~2 ng/µl), 1.5 ul gel purified insert DNA and 1 µl 40 units/ul T4 DNA ligase. Both mixes were incubated at RT for 20-40 minutes.

2.2.9 Plasmid Transformation into TG1 Competent Cells

The plasmids were then transformed into a bacterial host where the DNA was able to replicate. Frozen competent cells were thawed on ice. 1 µl of ligation mix was added to 50 µl aliquots of competent cells and incubated in ice for 20-40 min. The cells were then heat shocked at 42°C for 30 seconds and placed back on ice for 10 to 20 seconds. 0.3 mL of 2xTY media [1.6% (w/v) bactotryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl], was added to each aliquot and transferred to a 37°C shaking incubator for 20-40 min. 0.3 mL of cells were then plated onto TYE media plate [1.0% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.8% (w/v) NaCl, 1.5% (w/v) agar] and 100 ug/ml ampicillin. The plates were incubated at 37°C for 10-18 hours.

2.2.10 PCR Screening of Colonies

PCR screening was used to determine if the selected colonies contained the desired plasmid. Specific PCR primers were chosen to amplify a unique region of plasmid of interest. Each colony (usually 4-8) was transferred with a sterile loop to aliquots of 100 µl water. After the loop (containing the colony) was submerged, it was re-streaked on a TYE+amp plate to be incubated for 10 hours at 37°C. The cell suspensions were vortexed for 3 seconds to be used within 10 min. PCR reaction mix was prepared with varying volumes (depending on total

number of colonies to be screened) of water, 10x Thermo Pol Buffer, 2.5 mM dNTP, 10 µM forward primer, 10 µM reverse primer, and 2 units/µl Pfu polymerase. 19 µl of the PCR reaction mix was added to 1 µl of vortexed cell suspension in PCR tubes. The samples were amplified with 25 cycles in the thermocycler. An annealing temperature of T_m -10°C and extension time of 1 min per kilobase was used: 95°C, 2min → 25x [95°C, 30 sec → 50°C, 30 sec → 75°C, 45 sec] → 4°C. After the termination of the PCR reaction, 3 µl of 6xGLB were added to each of the samples and analyzed on an analytical agarose. The colonies that produced the desired fragment were selected for plasmid preparation.

2.2.11 100 ml Alkaline Lysis Plasmid Preparation

A single colony grown on the re-streak plate from two positive clones (selected from PCR screening) were inoculated in separate 500mL Erlenmeyer flasks of 100 mL 2xTY media and 100 µg/ml ampicillin. The cells were grown in a 37°C shaking incubator for 12 hours. After the allotted time, the cells were spun down at 4000 rpm for 5 minutes in a 250 ml centrifuge bottle in a tabletop centrifuge (Heraeus #7570G). The supernatant was discarded, and the cell pellet was resuspended in 5 ml LYSIS solution [50 mM glucose, 25 mM Tris-HCL pH 8.0, 10 mM NaEDTA]. The cells were transferred to a 50 mL polypropylene Falcon tube where 10 mL NaOH/SDS solution [0.2M NaOH, 1.0% (w/v) SDS] was added. The tubes were shaken vigorously and incubated on ice for 5 min. 10 mL of cold KaC/HAc [5 M KaC, 2.5 M HAc] was added to the solution, mixed 5 times, and incubated on ice for 5 min. The sample was spun down at 4000 rpm for 3 minutes at RT in the Haraeus tabletop centrifuge. The clear supernatant was transferred to a 50 mL polypropylene Falcon tube and 12.5 mL isopropanol was added. The

solution was incubated at RT for 5 min and then spun in the Haraeus tabletop centrifuge at 4000rpm for 5 minutes at 20°C. The supernatant was poured off and 0.5 mL 70% ethanol was added to the pellet. The pellet was transferred to a 1.5 mL Eppendorf tube and spun in the microcentrifuge at full speed for 1 min. The supernatant was aspirated off and the pellet was spun for an additional 30 seconds followed by another aspiration step to eliminate residual supernatant. The pellet was resuspended in 125 µl TE (10, 50) [10mM tris, 50 mM EDTA]. 1 µl of 10 mg/mL RNase A was added to the sample and incubated at 37°C for 20 minutes. While the sample was incubating, one Sephacryl S400 HR spun column was prepared for every 100 mL plasmid prep. The top half of a 1.5 mL Eppendorf tube was placed onto a 5 mL polypropylene tube. Siliconized glass wool was stuffed into the bottom of a 1 ml Gilson blue pipette tip and placed into the Eppendorf tube/polypropylene stand. The Gilson blue tip was filled with Sephacryl S400 resin equilibrated in TE (10, 0.1) and the spun column was centrifuged at 2000 rpm for 3 minutes. The resulting liquid in the 5 mL polypropylene tube was discarded. The RNase treated sample was spun in the microcentrifuge and the resulting supernatant was extracted with 150 µl phenol/24:1 chloroform: amyl alcohol two times, followed by one extraction with 300 µl 24:1 chloroform: amyl alcohol. The extracted sample was pipetted onto the prepared Sephacryl S400 HR spun column and centrifuged at 2000 rpm for 3 minutes at RT. The eluted plasmid, in TE (10, 0.1) was transferred to a labelled Eppendorf tube. The plasmid was then analyzed by restriction mapping and Sanger sequencing (if necessary).

2.2.12 Restriction Mapping and Sequencing

The purified plasmid DNA was restriction mapped using two endonucleases to confirm that the correct plasmid has been selected by comparing it to the parent plasmid DNA. The digest mixture typically was prepared with 4.5 µl water, 1 µl 10x NEBuffer (specific to the enzyme(s) used), 0.5 µl 100mM DTT, 2 µl 0.2 ng/µl DNA, and 1 µl of enzyme. After the samples were analyzed by agarose gel electrophoresis, the plasmid was submitted to the Penn State Huck Sciences Nucleic Acid Core Facility for Sanger sequencing. Sequencing data was aligned via a Smith-Walterman algorithm and was analyzed for mutations.

2.2.13 Site-Directed Mutagenesis Using Linear Amplification and DpnI Selection

Site-directed mutagenesis was used for the modification of individual base pairs in a DNA sequence to manipulate the amino acids present in protein primary structure. A 25 µl PCR mutagenesis reaction was prepared by adding 15.5 µl water, 5.0 µl 5x Q5 reaction buffer, 2.0 µl 2.5 mM dNTP, 1.0 µl 10ng/µl DNA template, 0.5 µl 10µM forward primer, 0.5 µl 10 µM reverse primer, and 0.5 µl 2 units/µl Q5 polymerase in a thin-walled PCR tube. A linear amplification was then performed in the thermocycler: 30 sec 98°C → 15x (5 sec 98°C → 15 sec 60°C → 1min/1Kbp 72°C) 5 min 72°C → 15°C.

Afterwards, 2 µl of the reaction mixture was stored in an Eppendorf tube. 0.5 µl of 10 units/µl DpnI was added to the remainder of the PCR mix and incubated at 37°C for 1 hour. During this time DpnI digests the contaminant methylated DNA while leaving the PCR product untouched (due to the absence of methylation). The digested sample, alongside the undigested sample, were transformed into competent TG1 cells (see 2.2.9).

The clones selected to be screened were attained from the DpnI treated mixture in expectation that all of the parent DNA was eliminated. PCR primers were strategically selected to incorporate the mutated site with a 500bp or less fragment. After the sample was thermocycled (see 2.2.10), 10 µl of reaction mixture was combined with 2 µl of 6xGLB and run on an agarose gel (see 2.2.6).

The remaining 10 µl of the PCR mix, along with the parent, were digested with endonucleases for the site that was either introduced or removed. Colonies that produced the desired fragments, indicating successful mutagenesis, were selected for plasmid preparation, restriction mapping, and sequencing.

2.2.14 Subcloning Into Polycistronic Expression Vector System

DNA methods 2.2.1 – 2.2.12 were used to subclone coding regions of desired proteins into the modular pST50(1-4) transfer cassettes. The cassettes were then placed into the pST44 expression vector where coexpression of up to four proteins can occur. Fig. 6 and 7 illustrate this scheme.

2.3 Protein Methods

2.3.1 Small Scale Expression Using the T7 System

Proteins were expressed using the T7 expression system. The plasmid that contained the desired DNA sequence was transformed into BL(DE3)pLysS competent cells (see 2.2.9) and plated on TYE + ampicillin + chloramphenicol media. 3 colonies of the expression clone from

the freshly transformed plate were inoculated into a flask containing 100 mL 2xTY + 50 µg/mL ampicillin (amp) + 25 µg/mL chloramphenicol (cm). Once the cultures reached an OD_{600} of 0.5-0.9 (blanked against 2xTY), protein expression was induced with 100 µl 0.2 M IPTG. If the cultures were to be grown at different temperatures, they were first inoculated at 37°C and transferred to the desired temperature when the OD_{600} was ~0.15. The cells were typically

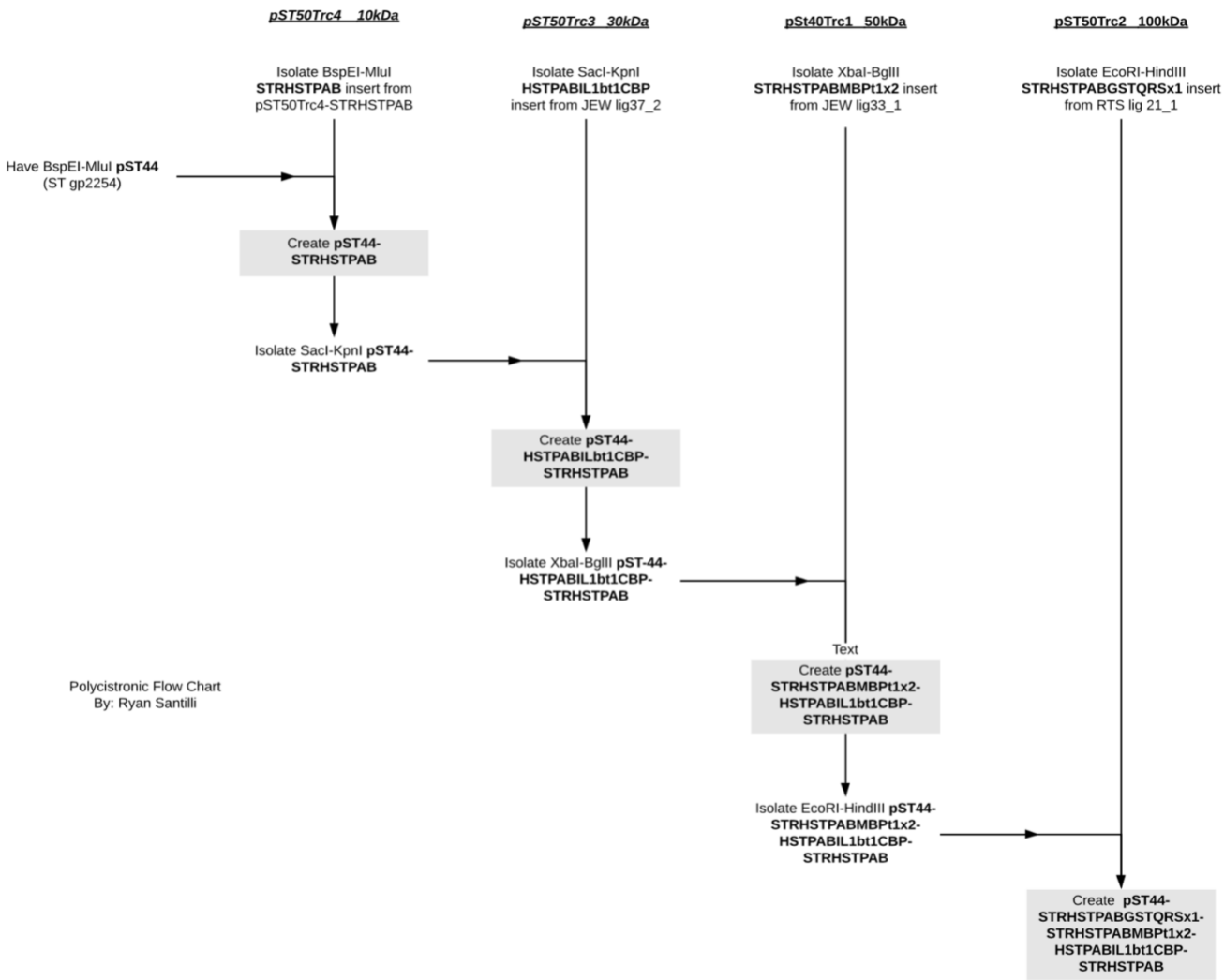


Figure 6. Subcloning 10kDa, 30kDa, 50kDa, and 100kDa proteins into pST44 Polycistronic Expression Vector

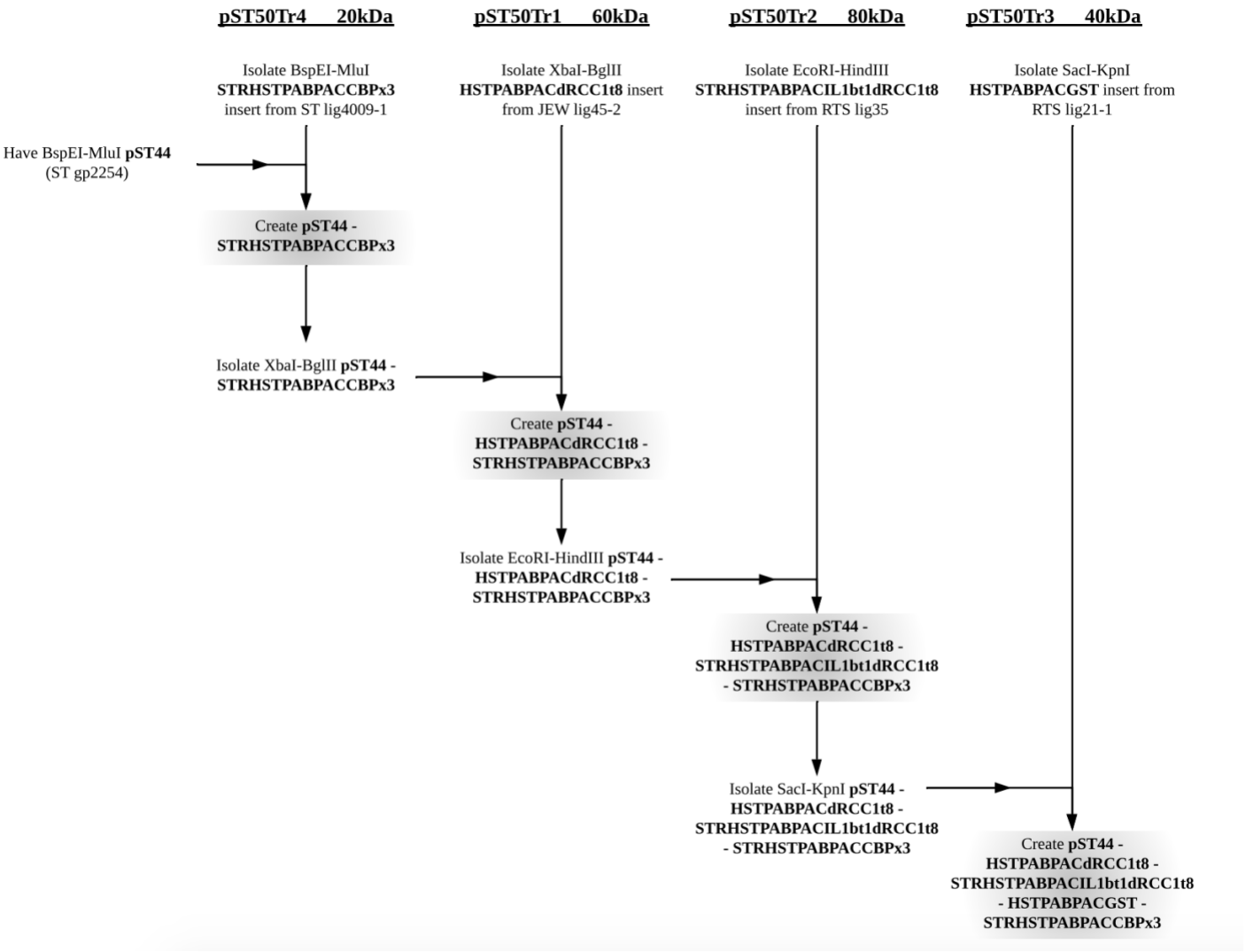


Figure 7. Subcloning 20kDa, 40kDa, 60kDa, and 80kDa proteins into pST44 Polycistronic Expression Vector

induced after two doubling times at an OD_{600} of 0.6-0.9. A 250 µl sample of uninduced culture, and hourly time points of 125 µl induced culture of was spun down for 1 min in the microcentrifuge and the remaining supernatant was aspirated off. The pellet was resuspended in 50 µl Protein Gel Loading Buffer (PGLB) (125 mM Bis-Tris, pH 6.8, 20% (w/v) glycerol, 4%

SDS, 15% (w/v) 2-mercaptoethanol). At 3 hours after induction, 50 mL of cell culture was harvested by spinning a 50 mL polypropylene falcon tube at 4000 rpm for 10 min at RT in a tabletop centrifuge. The supernatant was poured off and the cell pellet was resuspended in 10 mL of P300-EDTA (50 mM Na_3PO_4, 300 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM benzamidine), flash frozen in liquid nitrogen and stored at -20°C. Time points and harvest times were varied to optimize expression levels.

2.3.2 Small Scale Purification of HST-Tagged Polypeptides

All of the proteins used were N-terminally fused with a deca-histidine (His_{10}) affinity tag. This allowed for one-step purification with ABT cobalt resin or Clontech Talon Superflow.

Prepared Resin. The Talon resin was resuspended in the storage liquid. 1 mL of resuspended resin (~50% suspension) was transferred to a 15 mL Falcon tube. 10 mL of MilliQ water was added to the suspension, mixed, and centrifuged at 1800 rpm for 2 min at RT. The supernatant was poured off and 10 mL P300-EDTA was added, mixed, and spun in the centrifuge at the same parameters. The supernatant was poured off and the tube was capped to prevent the resin from drying out

Prepared Soluble Extract and Bind sample to Resin. The cells frozen in P300-EDTA from the small-scale expression (see 2.3.1) were thawed at RT. Once thawed, the cells were sonicated for 2 x 10 sec in a Branson S-450D sonicator at 40% maximum power, 50% cycle. The sample was stored on ice for 20-30 sec in between sonication steps. 25 µl of sonicated sample (whole cell extract) was mixed with 25 µl of PGLB and saved for SDS-PAGE electrophoresis. 1.3 mL of sonicated sample was then aliquoted into 4 1.5mL Eppendorf tubes. The tubes were

spun in the microcentrifuge for 3 min at RT. 25 µl of the supernatant from one tube was mixed with 25 µl of PGLB and saved for SDS-PAGE electrophoresis. The resulting supernatant from all 4 tubes was transferred to the prepared resin and incubated for 20 minutes. A pellet sample was manufactured by resuspending the pellet from one of the tubes with 1.3 mL P300-EDTA and combining 25 µl with 25 µl of PGLB.

Washed Resin. After the 20 min incubation period, the 15 mL Falcon tube was spun in the tabletop centrifuge at 1800 rpm for 5 min. A P1000 Pipetman was used to transfer the supernatant (flow through) to a 15 mL Falcon tube. 25 µl of the flow through was combined with 25 µl PGLB for SDS-PAGE analysis. Two series of washes with 10 mL of P300-EDTA was used to wash the resin and the supernatant was discarded after the same centrifugation parameters were applied. The resin was then resuspended in 3 mL of P300-EDTA and transferred to a disposable BioRad BioSpin column where the flow through was discarded. The proteins were eluted in 4, 500 µl fractions, with 2 mL P300-EDTA + 200 mM imidazole. 25 µl of each fraction was mixed with 25 µl of PGLB and analyzed via SDS-PAGE.

2.3.3 SDS-PAGE Electrophoresis

SDS-PAGE was used to analyze the proteins samples acquired from the expression and purification methods. 18% (w/v) gels were prepared by default however, on some occasions, 12% gels were made to accommodate bigger proteins. The gel block was assembled by first stacking glass plates into a BioRad Mini-Protean gel-casting block. The separating gel solution consisted of 18% (w/v/) acrylamide (30:0.5 acrylamide:bisacrylamide), water, and 3 M Tris-Cl pH 8.8. The solution was deaerated briefly before 10% SDS, TEMED, and 25% Ammonium

Persulfate in deionized water was used to polymerize the gel. Water saturated butanol was poured over the separating layer to level out the top of the gel while it incubated for 1 hour. The water saturated butanol was then poured off and the top layer was washed with ethanol, water, and air-dried with an air hose. A 5% (w/v) stacking gel was then prepared with water, 5% (w/v) acrylamide (10:0.5 acrylamide:bisacrylamide), 0.5 M Bis-Tris pH 7.0 and deaerated briefly. The same polymerizing agents of the separating gel were added, and the solution was poured onto the block. 10-well and 15-well combs were inserted, and the gel incubated at RT for 1 hour. The gel block was disassembled, and the gels were wrapped in wet paper towels and stored in the refrigerator.

Samples were prepared by adding equal volumes of PGLB and boiled for 1-3 minutes. While the samples were boiling, a pre-cast gel was fitted into a Mini-Protean II Electrophoresis cell and filled with Protein Gel Running Buffer (PGRB) (10 mM Tris, 76 mM glycine, 0.02% (w/v) SDS). Once the chamber was set up, 8 µl of each sample was loaded into corresponding lanes in the gel. BioRad Protein Molecular Weight Marker was loaded as standards for weight comparison. The gel was electrophoresed at 12 W for 30 min.

2.3.4 Immunoblotting

Immunoblotting, also known as Western Blotting, is another analytical technique to detect proteins. Proteins were first separated by SDS-PAGE electrophoresis. The gel was incubated in transfer buffer (25mM Tris, 192mM glycine, 20% (v/v), pH 8.3) and the proteins were transferred to a 0.2 µm nitrocellulose membrane (Amersham Hybond ECL) via trans

immunoblot electrophoresis at 0.9 A for 30 minutes. The gel was then stained with Coomassie blue to verify successful transfer.

After the transfer process, the membrane was incubated in Ponceau S Staining Solution (0.1% (w/v) Ponceau S in 5% (v/v) acetic acid) to demonstrate that the proteins were indeed on the membrane. The membrane was then equilibrated in 1xTBS (25mM Tris, pH 8.0, 0.15 M NaCl) for 5 minutes. A blocking solution containing 2% (w/v) nonfat dry milk and 1xTBS buffer was incubated with the membrane for 30 minutes. The membrane was then subjected to a solution containing 1:5000 dilution of anti-rabbit/mouse antibody in 1xTBS for 30 minutes to 1 hour. Three consecutive 5 min incubation washes with TTBS (25mM Tris, pH 8.0, 0.15 M NaCl, 0.05% Tween 20) were administered to the membrane to remove any unbound antibody. Enzymatic detection to visualize the proteins utilized chemiluminescence by covering the membrane with a 1:1 solution of ECL detection solutions 1 and 2. The membrane was wrapped in Saran plastic and exposed to X-Ray film.

Chapter 3

Results

3.1 Background and Design of Expression Plasmids

3.1.1 Components

Protein Data Bank and various literature were consulted to select proteins based on migration patterns, molecular weight, solubility, high expression yields, and efficient purification by metal affinity chromatography. The chosen array of proteins is illustrated in Table 4.

Table 4. List of Proteins Used

Protein (Abbreviation)	Protein (Full Name)	Molecular Weight
HST	10xHis tag	1.66 kDa
STR	Strep peptide	19.339 kDa
PAB	Protein A IgG binding domain B	6.598 kDa
PAC	Protein A IgG binding domain C	6.598 kDa
CBP	Calmodulin binding peptide	21.287 kDa
GST	Glutathione S-transferase	25.499 kDa
HPC	Heavy chain of protein C	19.594 kDa
MBP	Maltose binding protein	40.698 kDa
S100B	S100 protein B	11.591 kDa
dRCC1	Drosophila RCC1	58.806 kDa
IL1b	Interleukin 1 beta	30.748 kDa
QRS	Glutaminyl-tRNA	63.478 kDa
pepN	*E. coli* aminopeptidase	98.919 kDa

3.1.2 Construct Design

HST affinity tag and a dual IgG binding domain (PABPAC) were incorporated into all protein constructs except the 10 kDa due to its small size. Each construct was made using similar methods. First, with the help of Dr. Song Tan, the constructs were designed to possess the appropriate molecular weights at regular intervals. Various truncations and mutations have been made to accommodate problems that arose, which are discussed **3.2 Creation of Recombinant Expression Plasmids**. Many of the parent DNA templates were already available in the lab and were attained by endonuclease digestion from previously created constructs. For DNA templates that were not available in the lab, synthetic gBlock template DNA were ordered from Integrated DNA Technologies (Coralville, Iowa). The same 8 steps (with minor variations) were carried out to create all constructs mentioned in this thesis:

1. PCR amplification of template DNA with site-specific primers.
2. Digestion of PCR product with restriction endonucleases.
3. Agarose gel purification of PCR product.
4. Ligation of pST50Trc(1-4) expression vector with desired insert.
5. Transformation of plasmid into bacterial competent TG1 cells
6. PCR screen resulting colonies
7. 100 mL plasmid preparation of two colonies
8. Verification of plasmid with restriction mapping and sequencing.

3.1.3 Overexpression of Protein Constructs

pST50Trc-type expression vectors allowed expression our constructs. The pST50Trc1 was created by Dr. Song Tan by first by subcloning the blunt ended NgoMIV-EcoRV insert of pET3aTr2 into the blunt ended vector pST39. The pST50Trc(2-4) plasmids were constructed by amplifying regions from pST50Trc1 to exclude particular nested restriction sites.[13] The created constructs utilized these expression vectors to be overexpressed in *E. coli* (BL21(DE3)pLysS) and analyzed via SDS-PAGE. Expression times and temperatures were varied to yield optimal results. 3-6 consecutive 500 µl talon eluent fractions were collected into 1.5 mL Eppendorf tubes from purification by metal affinity chromatography. Usually, the proteins samples would elute into fraction 2 and these samples would be stored in 20% glycerol in the -20°C freezer. SDS-PAGE gels were loaded with a 1:1 ratio of sample to PGLB until appropriate dilutions were made for analyses. Fig. 8 shows the commercially available markers used for comparisons.

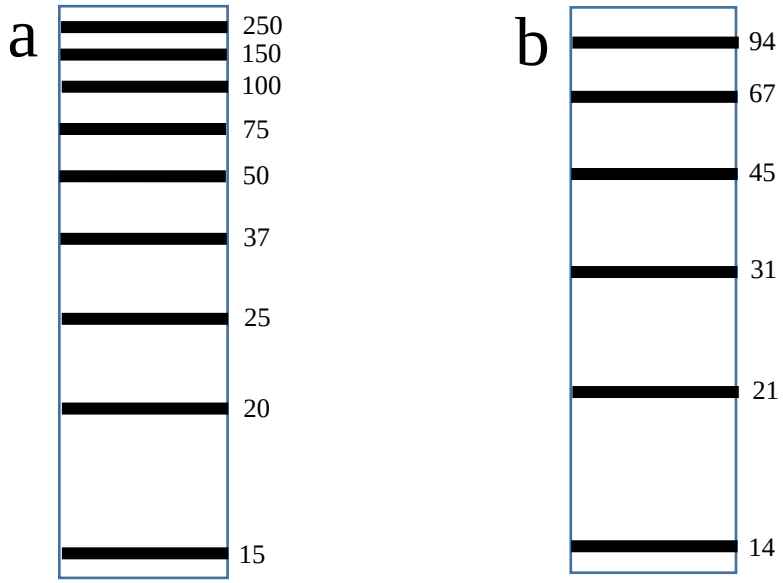


Figure 8. a.) BioRad Precision Plus Molecular Weight Markers. b.) BioRad Low Molecular Weight Markers

3.1.4 Immunoblot Testing of Individual Constructs

The BioRad Trans-Blot Turbo transfer system was used to transfer the protein markers from the SDS-PAGE gels to a 0.2 µm nitrocellulose membrane. All protocols were conducted under the "General" setting with a max current of 1 Amp (however, the current was consistently 0.8A). In order to optimize Western blotting, several parameters were manipulated. Two secondary antibodies were used for detection, anti-rabbit and anti-mouse. Also, the incubation times with the antibodies ranged from 30 minutes to an hour. Exposure time, the time the membrane is in contact with the X-ray film, was also adjusted several times. Overall, all other parameters were consistent. The next few sections highlight both these methods (**3.1.3 and 3.1.4**) in more detail.

3.2 Creation of Recombinant Expression Plasmids

3.2.1 10 kDa Protein Construct

The production of a 10 kDa marker protein construct (STRHSTPAB) was completed by my Schreyer Honors colleague, Yoshitaka Shibata.

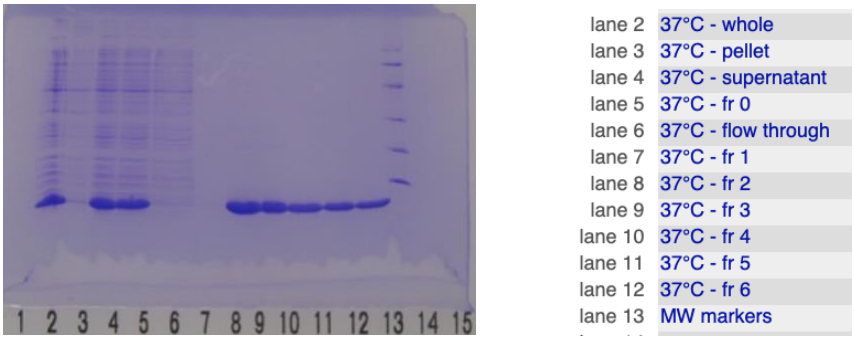


Figure 9. SDS-PAGE of small-scale metal affinity chromatography purification of STRHSTPAB performed by Yoshitaka (Joey) Shibata at 37°C.

In the early stages of our analysis, we compared the migration pattern of our created constructs to the migration pattern of known molecular weight markers (BioRad LMW). Although it seems as if the protein is migrating around 10 kDa based on Fig. 9, later semi-log plot analysis concluded that this protein was migrating anomalously slow.

We speculated that the STR tag, a strep peptide, was responsible for the anomalous migration. A new construct was designed, HSTPABHPC, which eliminated the STR tag and instead included the HPC affinity tag.[19] Despite its high purification and expression yield, it was still migrating too slow, representing a 13 kDa construct. The results are illustrated in Fig. 10.

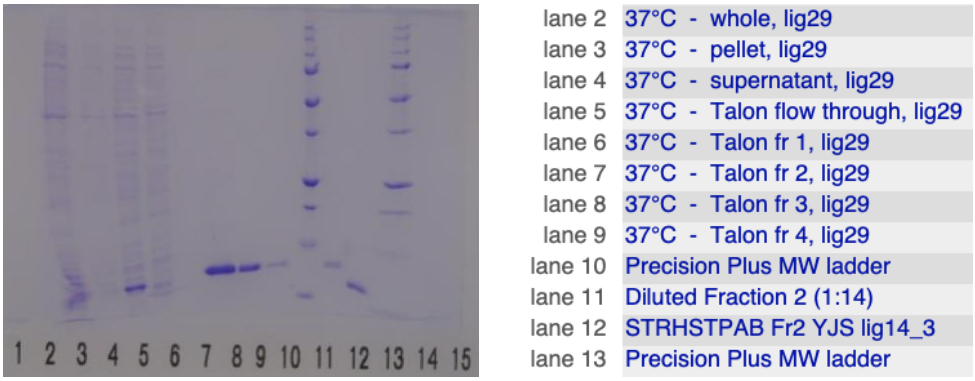


Figure 10. SDS-PAGE of small-scale metal affinity purification of HSTPABHPC at 37°C.

In order to combat the issue of slow migration, we planned to make a new construct with a lysine-rich tag at the N-terminus. However, a protein with a lysine residue after the START codon might be degraded according to the N-end rule.[20] The N-end rule governs the rate of protein degradation by determining its half-life based on its N-terminal amino acid. Howard Salis, a professor at The Pennsylvania State University, suggested using a poly-Lysine or Lysine-Threonine repeat at the N-terminus to reduce the likelihood of degradation. Therefore, three different variants were created in parallel to address the possible issue of reduced expression:

MKKKKKKKKK (LYSHSTPAB)

MKTKTKTKTKT (MKTHSTPAB)

MGKTKTKTKTKT (MGKHSTPAB)

Fig. 11 shows the results of the three constructs.

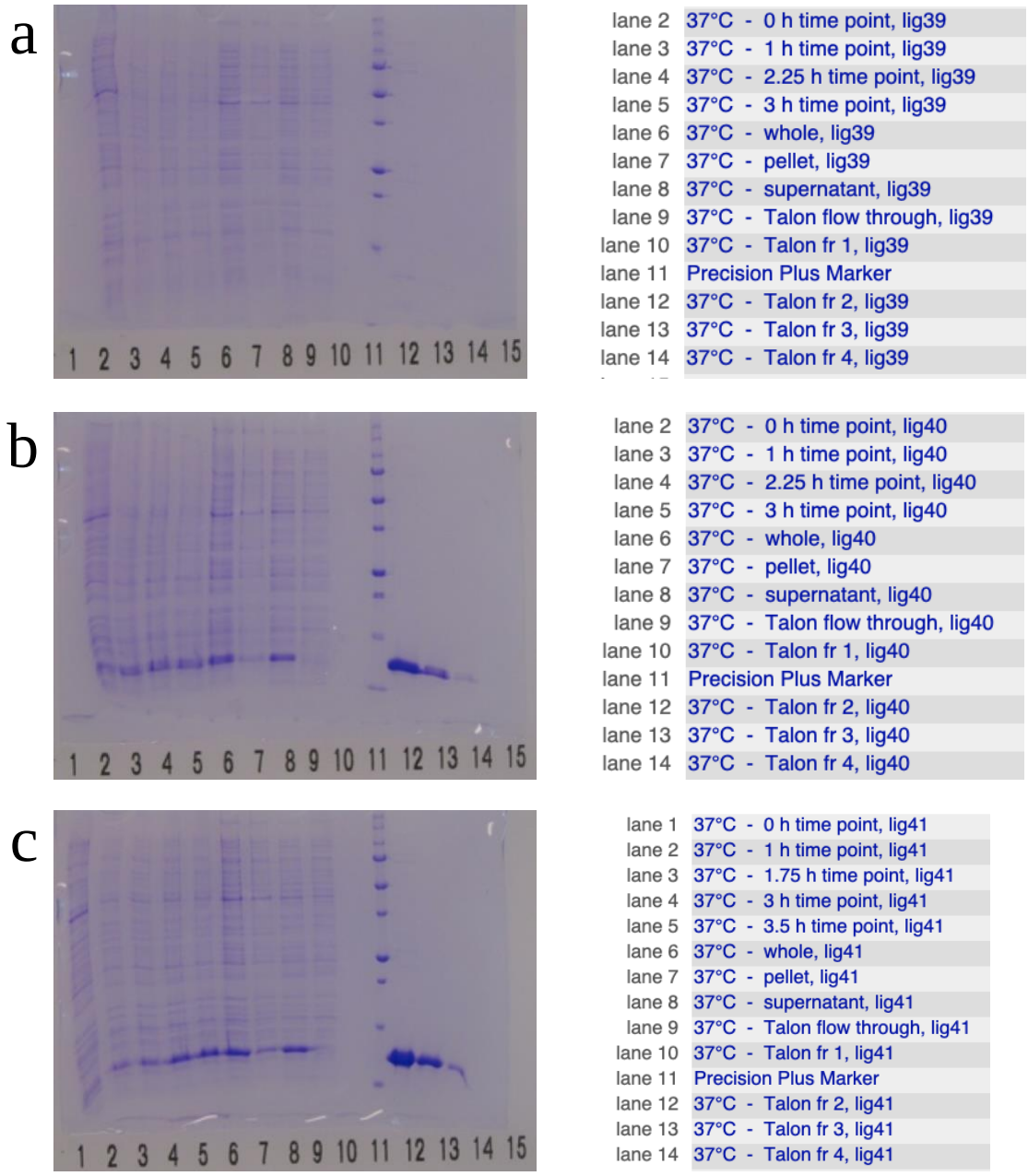
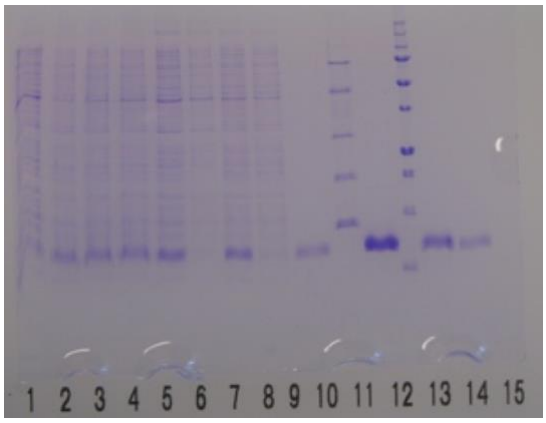


Figure 11. SDS-PAGE of small-scale expression and metal affinity purification of a. LYSHSTPAB b. MKTHSTPAB c. MGKTHSTPAB at 37°C.

The absence of protein expression of LYSHSTPAB could have been caused by either degradation due to the N-end rule, a stability issue of the mRNA, and/or a translation issue of the mRNA. The official reason is unknown. MKTHSTPAB showed significant expression and purification, however a doublet was visualized on the SDS-PAGE gel from conditions that were unknown. MGKHSTPAB showed even better purity and quantity of protein but migrated slower compared to MKTHSTPAB. Although we felt we had possible candidates, an experiment to modify the migration of HSTPABHPC was designed.

A mutagenesis experiment was designed to mutate a Proline residue to a Lysine residue in HSTPABHPC to facilitate faster migration. Proline has a cyclic structure that introduces natural "kinks" in the polypeptide chain, usually in alpha helices after the first turn (4^{th} residue). The kink is caused by proline being unable to complete the hydrogen bonding chain of the helix and steric effects keep it from adapting to its preferred helical geometry.[21] This change increases the Stokes radius, the radius of a hard sphere that diffuses at the same rate as that solute, causing slower migration. Because of this, proteins with a high proline count are notorious for running slower than their actual molecular weight. A PCR mutagenesis was performed to change a Proline residue to a Lysine residue at the 82^{nd} position in the HSTPABHPC construct. Two complementary primers, one with AAA encoding Lysine and one with TTT (the complement), flanked by 20 other bases were created and used in the reaction. The DNA with the mutated residue was then digested with appropriate restriction endonucleases and agarose gel purified. The macromolecule was ligated with pST50Tr expression vector and transformed into TG1 cells. The new construct, HSTPABHPCx1, was expressed and affinity purified. The result can be seen in Fig. 12.



lane 1	37°C - 0 h time point, lig35
lane 2	37°C - 1 h time point, lig35
lane 3	37°C - 2 h time point, lig35
lane 4	37°C - 3 h time point, lig35
lane 5	37°C - whole, lig35
lane 6	37°C - pellet, lig35
lane 7	37°C - supernatant, lig35
lane 8	37°C - Talon flow through, lig35
lane 9	37°C - Talon fr 1, lig35
lane 10	Biorad Molecular Weight Ladder
lane 11	37°C - Talon fr 2, lig35
lane 12	Precision Plus Molecular Weight Ladder
lane 13	37°C - Talon fr 3, lig35
lane 14	37°C - Talon fr 4, lig35

Figure 12. SDS-PAGE of small-scale expression and metal affinity chromatography purification of 10 kDa Protein Construct (HSTPABPACx1) at 37°C.

HSTPABHPCx1 is completely soluble, expresses at high levels, and shows excellent purity after small scale metal affinity purification. The faster migration pattern fit accurately into the expected migration pattern when it was analyzed via semi-log plots (distance vs. log(molecular weight). Refer to Fig. 29 (**3.3.1**) for comparison between MKTHSTPAB and HSTPABHPCx1. Based on these analytical techniques, HSTPABHPCx1 was chosen for the 10 kDa marker protein.

3.2.2 15 kDa Protein Construct

The 15 kDa protein construct originated from a gBlock template DNA. pST50Tr-type expression plasmids were used to express the cells because they contain the T7 promoter gene and antibiotic resistance. The HSTPABPAC G-block was subcloned into pST50Tr-type expression plasmid via the 8 steps mentioned in **3.1.2.** The protein was expressed at 37°C for 3 hours in BL21(DE3)pLysS bacterial strain (Fig. 13).

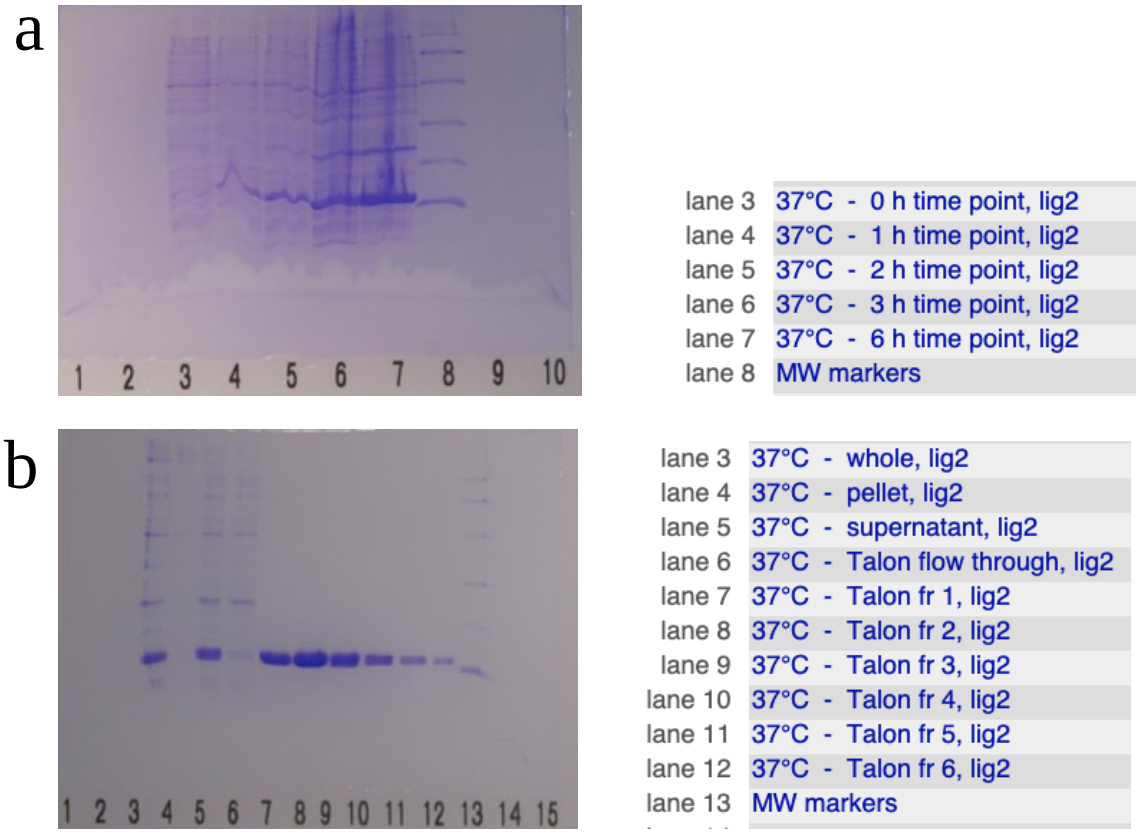


Figure 13. SDS-PAGE of small-scale a.) expression b.) metal affinity chromatography purification of HSTPABPAC at 37°C.

The 15 kDa construct demonstrated high solubility, expression and purification. The migration pattern was slightly slower than the commercial ladders.

Western Blot Analysis of the 15 kDa marker protein. The 15 kDa protein marker was chosen to validate the binding of secondary antibody to the IgG binding domain. It so happened that this 15 kDa construct contained two IgG binding domains. We did not realize at the time of this experiment that the tandem binding domain (PABPAC) influenced the Western blot results. Fig. 14 illustrates Western blot detection of the 15 kDa protein molecular weight marker at varying dilutions with 1:5000 dilution of anti-rabbit secondary antibody.

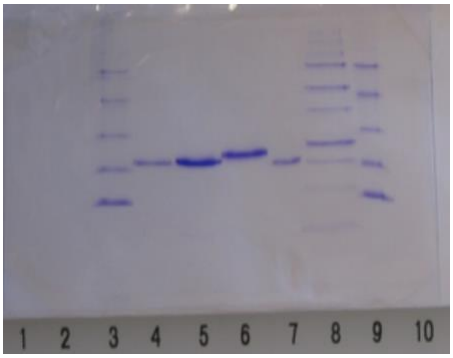
lane 4	HSTPABPAC test batch fr 2 1:10 dilution
lane 5	
lane 6	HSTPABPAC test batch fr 2 1:100 dilution
lane 7	
lane 8	HSTPABPAC test batch fr 2 1:500 dilution

Figure 14. Western blot of HSTPABPAC with 1:5000 dilution of anti-rabbit secondary antibody. Despite the numbers on the bottom of the photograph, the three samples correspond to lanes 4, 6, and 8.

The results show a strong signal with anti-rabbit secondary antibody. There are degradation products present, however the goal of this experiment was solely to confirm the success of our proteins in a Western blot.

3.2.3 20 kDa Protein Construct

The creation of the 20 kDa protein molecular weight construct proved to be challenging. A total of five different constructs were made to try and tackle anomalous migration issues. Before we experienced Western Blot detection problems (see **3.5.1**), four different constructs were subcloned, expressed, and analyzed via SDS-PAGE (Fig. 15).



lane 3	BioRad MW Markers
lane 4	HSTPABTrx
lane 5	STRHSTPABCAP
lane 6	HSTPABhTrx2
lane 7	HSTPABS100B
lane 8	Precision Plus MW Markers
lane 9	BioRad MW Markers

Figure 15. SDS-PAGE of small-scale metal affinity chromatography purification of "failed" 20 kDa protein markers.

The first 20 kDa protein construct, depicted in lane 4 utilizes *E. coli* thioredoxin (Trx) protein. Although the amino acid sequence sums to 20 kDa, the protein construct migrates anomalously slow. The second construct (lane 5) incorporates the STR tag along with a new CAP protein. The addition of these two elements enhanced the amount of material present but did not solve the migration issue. The third attempt in lane 6 uses human thioredoxin 2 (hTrx2) protein, however it caused an even slower migration pattern. Since none of the previous attempts resulted in a satisfactory migration pattern, the literature was consulted to find another protein that could be used with the appropriate amino acid sequence to confer a molecular weight of 20 kDa. S100B was successfully subcloned via recombination. The marker protein demonstrated an accurate migration pattern; however, it was later eliminated as a possible candidate due to the lack of the dual binding domain.

The new construct was re-engineered to include the PABPAC sequence. It also include the STR tag and CBP protein. When the construct was expressed at 37°C, a high purification yield was attained. When the 20 kDa was further manipulated to address Western blot issues (see **3.5.1** *Transfer of the 20 kDa protein*) expression experiments yielded a doublet band visualized in Fig. 16.

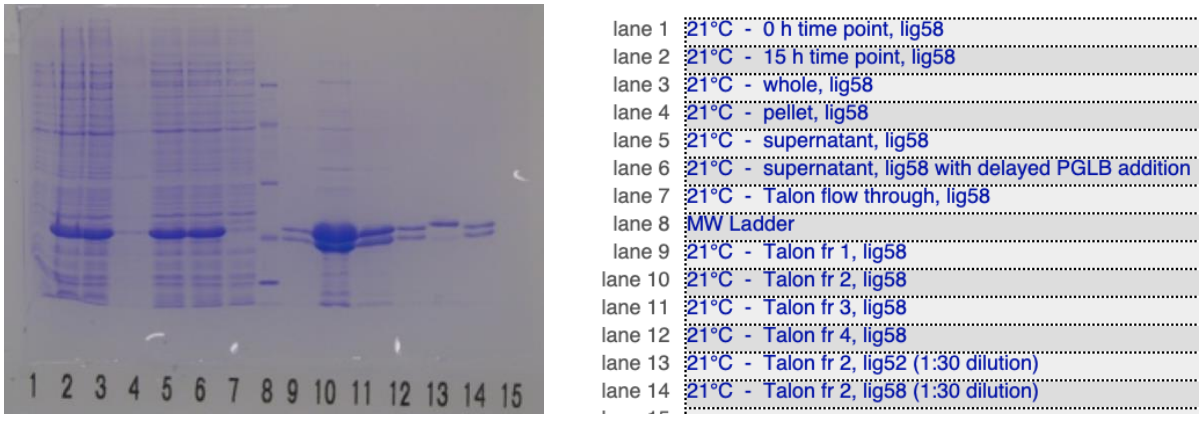


Figure 16. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABPACCBPx1 (optimized for western blotting) at 21°C.

Before attempting any mutations, small-scale expression and metal affinity chromatography parameters were manipulated to resolve the degradation issue. An experiment was conducted to grow the cells at 21°C and add proteolysis inhibitors to the Talon buffer (Fig. 17):

P300-EDTA + 1mM benzamidine + 0.5 mM PMSF + 5 mM 2-mercaptoethanol

0.5 mL of 0.1M PMSF was prepared by dissolving 8.7 mg of PMSF in 0.5 mL 100% ethanol.

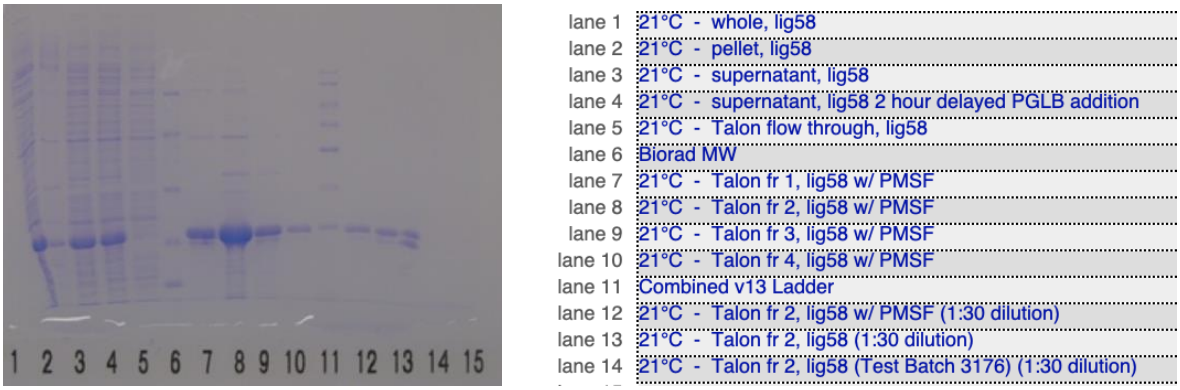


Figure 17. SDS-PAGE of small-scale expression and metal affinity chromatography purification of HSTPABPACCBPx1 with PMSF at 21°C.

Reduced temperature and protease inhibitor were ineffective at removing the degradation product. Based on the size of the contaminant, we hypothesized that the band might correspond to a degradation product of the CBP tag at the C-terminus of the protein. Visual inspection of the protein sequence revealed a tryptophan residue at position 157 and a phenylamine residue at position 161. Both amino acids contain an aromatic ring favored by chymotrypsin, and we speculated that chymotrypsin or a related protease might be responsible for the degradation

product. Therefore, two mutagenesis experiments were designed: mutation of the phenylalanine to a lysine and tryptophan to a lysine.

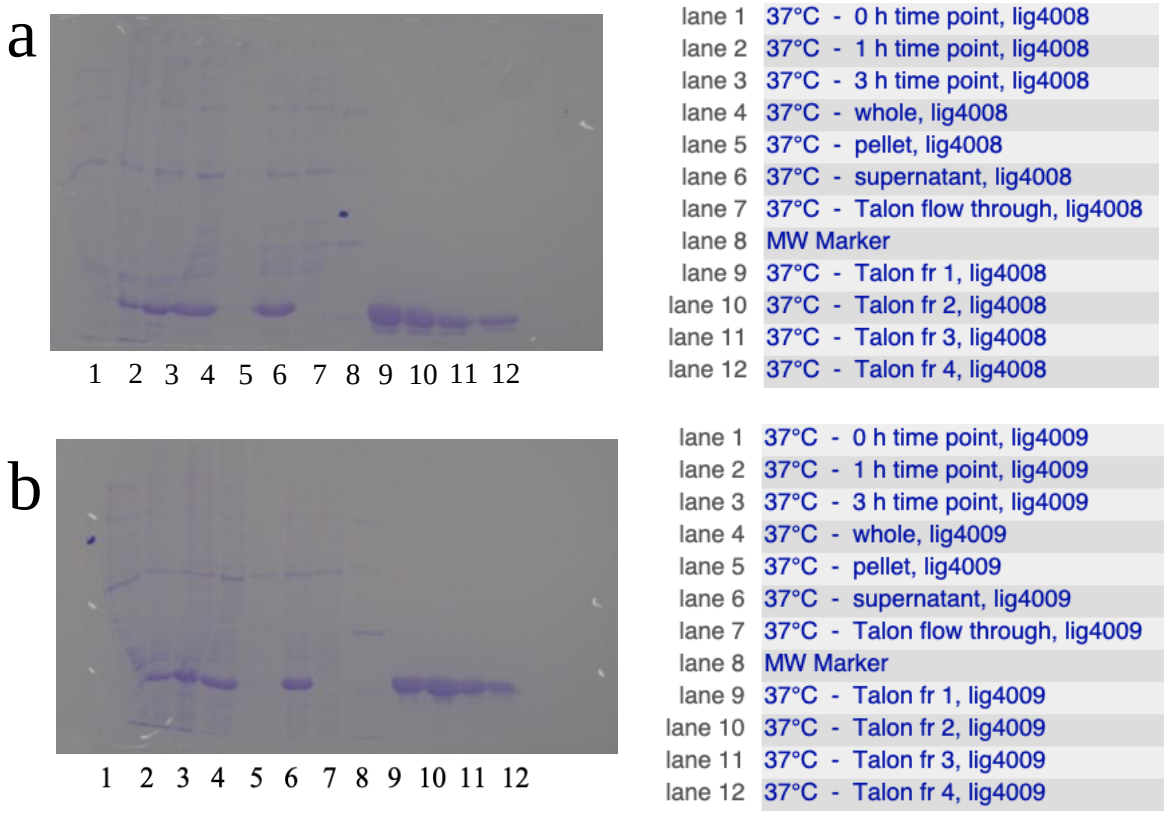
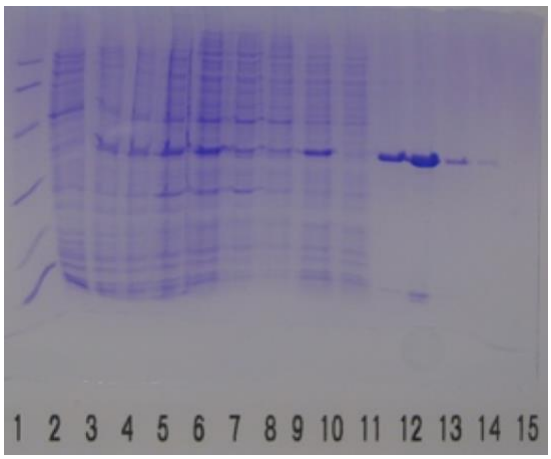


Figure 18. SDS-PAGE of small-scale expression and metal affinity chromatography purification of a.) STRHSTPABPACCBPx2 (F to L) and b.) STRHSTPABPACCBPx3 (W to L) at 37°C.

The modification of mutating the phenylalanine (Fig. 18a) to a lysine did not reduce the degradation product. The tryptophan mutation at position 157 (Fig.18b) significantly reduced the degradation product, so STRHSTPABCBPx3 was chosen as the new 20 kDa protein ladder construct.

3.2.4 30 kDa Protein Construct

Interleukin 1 beta (IL1b) is a member of the interleukin 1 family of cytokines. The literature confirmed that this protein could express well in *E. coli* with a molecular weight of 18.215 kDa. We decided to incorporate IL1bS1 (S1 indicating it contained a lanthanide binding loop) into the 30 kDa construct to make STRHSTPABIL1bS1. When we tried to express this construct at 37°C for 4 hours, the expression level of this construct was subpar (Fig. 19).



lane 1	MW markers
lane 2	37°C - 0 h time point, lig22
lane 3	37°C - 1 h time point, lig22
lane 4	37°C - 2 h time point, lig22
lane 5	37°C - 4 h time point, lig22
lane 6	37°C - whole, lig22
lane 7	37°C - pellet, lig 22
lane 8	37°C - pellet with 5x P300 buffer, lig22
lane 9	37°C - supernatant, lig22
lane 10	37°C - Talon flow through, lig22
lane 11	37°C - Talon fr 1, lig22
lane 12	37°C - Talon fr 2, lig22
lane 13	37°C - Talon fr 3, lig22
lane 14	37°C - Talon fr 4, lig22

Figure 19. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABIL1bS1 at 37°C. Lanes 11 and 12, indicating talon fractions 1 and 2 show medium level expression.

Due to a GC-rich 5' end of the mRNA transcript in IL1bS1, we suspected that a property of the lanthanide binding loop was affecting the expression of the protein. To test our hypothesis, we conducted two experiments: subclone and expression of IL1bS1 on its own, and site-directed deletion mutagenesis to remove the lanthanide-binding loop (deletion between residues 63 to 79).

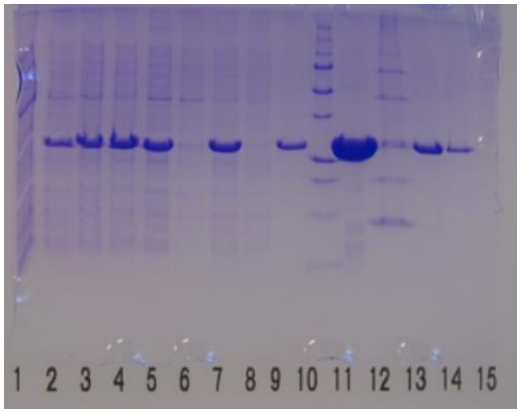
IL1b = IL1b (1-269). i.e. full length IL1b

IL1bt1 = IL1b (117-269) without S1 lanthanide binding loop insertion

IL1bS1 = IL1b (117-269) with lanthanide binding loop insertion

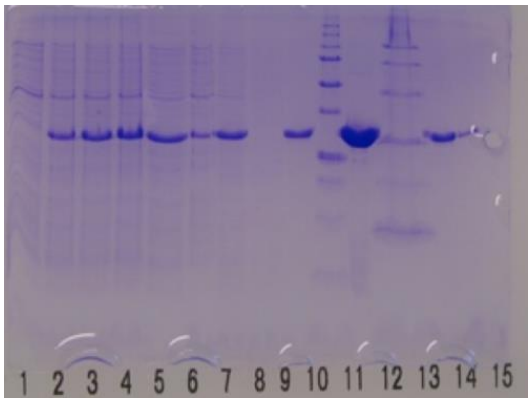
The expression plasmid containing solely IL1bS1 did not show any expression. This supported our hypothesis that the GC rich 5' end of the mRNA transcript impeded expression. To solve this issue, another experiment introduced a STRaHSTN to the coding region to eradicate the GC rich by using a cleavable tag. Although this proved to be a possible resolution to the expression problem, the results of the expression of the IL1bt1 construct (without the lanthanide binding loop) awarded significantly increased amounts of expression and purified product and was therefore chosen for the protein marker. The intermediate protein construct (26kDa), HSTPABPACIL1bt1(Fig. 20a), was also expressed at 37°C for 3 hours.

a



lane 1	37°C - 0 Hr
lane 2	37°C - 1 Hr
lane 3	37°C - 2 Hr
lane 4	37°C - 3 Hr
lane 5	37°C - whole
lane 6	37°C - pellet
lane 7	37°C - supernatant
lane 8	37°C - flow through
lane 9	37°C - fr 1
lane 10	MW markers
lane 11	37°C - fr 2
lane 12	MW markers
lane 13	37°C - fr 3
lane 14	37°C - fr 4

b

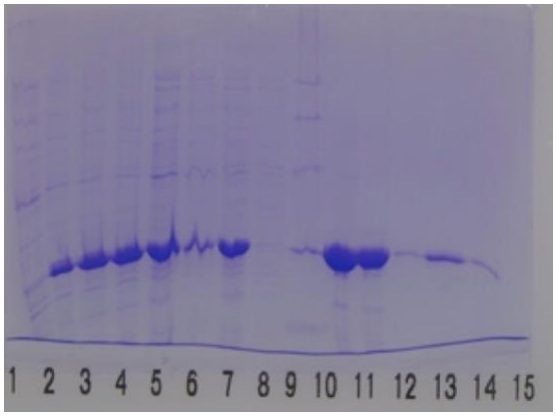


lane 1	37°C - 0 Hr
lane 2	37°C - 1 Hr
lane 3	37°C - 2 Hr
lane 4	37°C - 3 Hr
lane 5	37°C - whole
lane 6	37°C - pellet
lane 7	37°C - supernatant
lane 8	37°C - flow through
lane 9	37°C - fr 1
lane 10	MW markers (Precision Plus)
lane 11	37°C - fr 2
lane 12	MW markers (BioRad)
lane 13	37°C - fr 3
lane 14	37°C - fr 4

Figure 20. SDS-PAGE of small-scale expression and metal affinity chromatography purification of a.) HSTPABPACIL1bt1 (26 kDa intermediate) and b.) HSTPABPACIL1bt1CBP (30kDa) at 37°C.

Although the migration pattern was presumably accurate based on the commercially available ladders in lanes 10 and 12, the semi-log plots (Fig. 29 **3.3.1**) proved that this construct was migrating too slow.

We consulted the literature for another protein that was similar to Interleukin 1 beta because we wanted to keep the same construct backbone (HSTPABPAC[X]CBP). We found a protein of the S100 family which are localized in the nucleus and cytoplasm in wide range of cells. Specifically, S100B was used in place of IL1bt1 to speed up the migration pattern of the 30 kDa construct. The expression plasmid was expressed at 37°C and high amount of soluble protein was purified (Fig. 21).



lane 1	37°C - 0 h time point, lig24
lane 2	37°C - 1 h time point, lig24
lane 3	37°C - 2 h time point, lig24
lane 4	37°C - 3 h time point, lig24
lane 5	37°C - whole, lig24
lane 6	37°C - pellet, lig24
lane 7	37°C - supernatant, lig24
lane 8	37°C - Talon flow through, lig24
lane 9	Bioad MW Ladder
lane 10	37°C - Talon fr 1, lig24
lane 11	37°C - Talon fr 2, lig24
lane 12	37°C - Talon fr 2, lig24 (1:60 dilute)
lane 13	37°C - Talon fr 3, lig24
lane 14	37°C - Talon fr 4, lig24

Figure 21. SDS-PAGE of small-scale expression and metal affinity chromatography purification of HSTPABPACS100BCBP (30kDa) at 37°C.

Compared to Fig. 20b, the metal affinity chromatography purification fraction 2 of our desired construct is slightly below the 31 kDa marker (BioRad MW Ladder, lane 9). This means we have successfully increased the migration rate to produce an accurate 30 kDa marker protein. The final 30 kDa marker used in this project was named HSTPABPACS100BCBP.

3.2.5 40 kDa Protein Construct

From this point forward, the original constructs without the dual binding domain will not be discussed (see **3.5.1 Western Blot** *Single vs. Dual Binding Domain* for further details). Glutathione S-transferase (GST) is a naturally occurring 26 kDa protein found in eukaryotic cells.[18] It is known to express in high quantities and purify well due to its affinity tag properties. The construct, HSTPABPACGST was expressed at 37°C and proved to be soluble, have a high purification yield, and migrate accurately (Fig. 22).

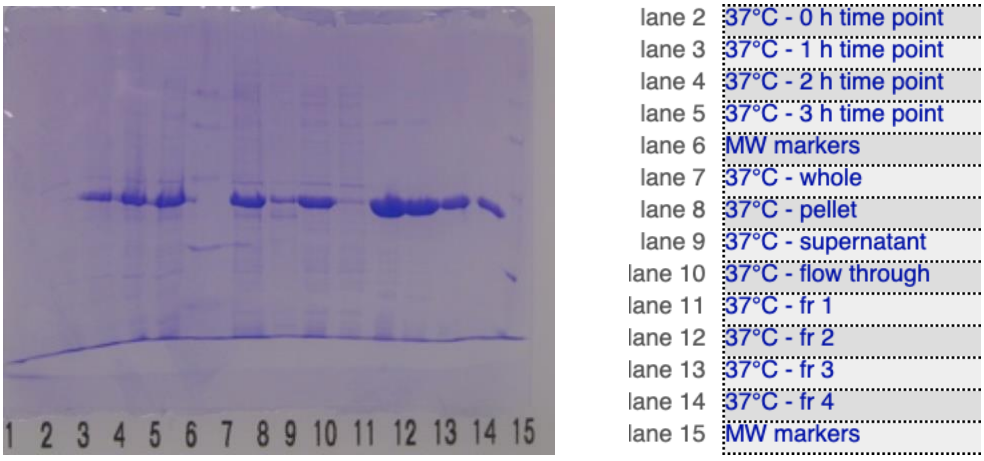
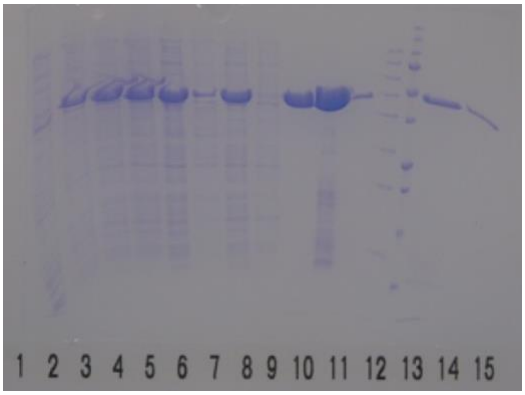


Figure 22. SDS-PAGE of small-scale expression and metal affinity chromatography purification of HSTPABPACGST (40 kDa) at 37°C.

3.2.6 50 kDa Protein Construct

Maltose Binding Protein (MBP) is a part of the maltose/maltodextrin system of *E. coli*, which is responsible for the uptake and efficient catabolism of maltodextrins.[22] Due to its size of roughly 40.5 kDa, it was used in the creation of the 50 kDa construct. My Schreyer Honors colleague, Joey Shibata, worked on this construct. It was expressed at 37°C and yielded high quantities of purified product. STRHSTPABPACMPBt1 is shown in Fig. 23.

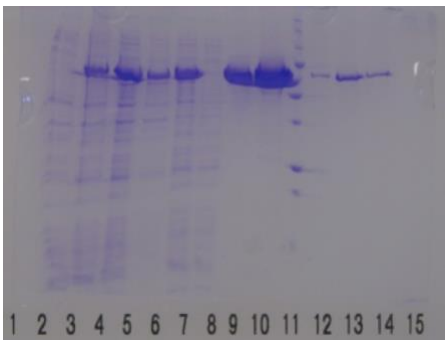


lane 1	37°C - 0 h time point, lig33
lane 2	37°C - 1 h time point, lig33
lane 3	37°C - 2 h time point, lig33
lane 4	37°C - 3 h time point, lig33
lane 5	37°C - whole, lig33
lane 6	37°C - pellet, lig33
lane 7	37°C - supernatant, lig33
lane 8	37°C - Talon flow through, lig33
lane 9	37°C - Talon fr 1, lig33
lane 10	37°C - Talon fr 2, lig33
lane 11	37°C - Talon fr 2, lig33 (1:29 dilution)
lane 12	JEW Combined Ladder v5
lane 13	Precision Plus
lane 14	37°C - Talon fr 3, lig33
lane 15	37°C - Talon fr 4, lig33

Figure 23. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABPACMBPt1 (50 kDa) at 37°C.

3.2.7 60 kDa Protein Construct

Drosophila melanogaster (fruit fly) Regulator of Chromosome Condensation 1 (dRCC1) is a protein that promotes the exchange of Ran-bound GDP by GTP.[23] This protein was utilized in the design of the 60 kDa marker, HSTPABPACdRCC1t8x27. My other colleague, Jack Williamson, facilitated the creation of this construct. When this construct was first expressed at 37°C, we noticed significantly reduced expression and only a small amount of product was able to be purified from metal affinity chromatography (see **3.5.2**). However, when we re-expressed this construct at 18°C, we noticed an increase in the amount of product that was comparable to our other marker proteins (Fig. 24).

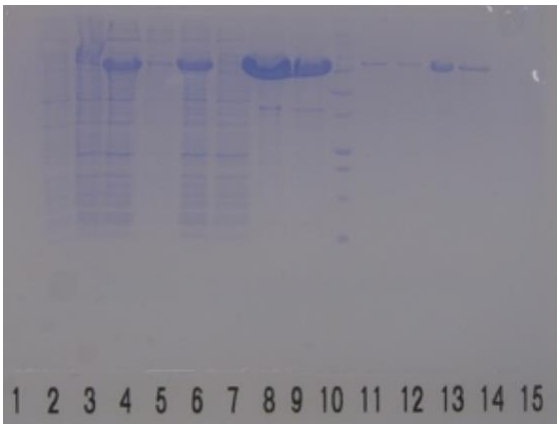


lane 2	18°C - 0 h time point, lig42
lane 3	18°C - 12 h time point, lig42
lane 4	18°C - whole, lig42
lane 5	18°C - pellet, lig42
lane 6	18°C - supernatant, lig42
lane 7	18°C - Talon flow through, lig42
lane 8	18°C - Talon fr 1, lig42
lane 9	18°C - Talon fr 2, lig42
lane 10	Precision Plus Ladder
lane 11	HSTPABdRCC1t7x27ENQ
lane 12	18°C - Talon fr 3, lig42
lane 13	18°C - Talon fr 4, lig42

Figure 24. SDS-PAGE of small-scale expression and metal affinity chromatography purification of HSTPABPACdRCC1t8 (60kDa) at 18°C.

3.2.8 80 kDa Protein Construct

The 80 kDa protein was also mainly constructed by Jack Williamson. It utilizes the combination of interleukin 1 beta (IL1bt1) and drosophila Regulator of Chromosome Condensation 1(dRCC1). Since it contained dRCC1, the small-scale expression was performed at 18°C to yield high expression and purity by metal affinity chromatography. The 80 kDa construct, STRHSTPABPACIL1bt1dRCC1t8, can be visualized in Fig. 25. Correct migration patterns were verified by semi-log plots (see **3.3.2**).



lane 1	18°C - 0 Hr
lane 2	18°C - 15 Hr
lane 3	18°C - whole, lig54
lane 4	18°C - pellet, lig54
lane 5	18°C - supernatant, lig54
lane 6	18°C - Talon flow through, lig54
lane 7	18°C - Talon fr 1, lig54
lane 8	18°C - Talon fr 2, lig54
lane 9	Precision Plus
lane 10	18°C - Talon fr 2, lig54 (1:30) dilute
lane 11	18°C - Talon fr 2, lig54 (1:60) dilute
lane 12	18°C - Talon fr 3, lig54
lane 13	18°C - Talon fr 4, lig54

Figure 25. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABOACIL1bt1dRCC1t8 (80 kDa) at 18°C.

3.2.9 100 kDa Protein Construct

The 100 kDa construct utilizes glutaminyl-tRNA synthetase (QRS), which is one of several aminoacyl-tRNA synthetase (ARSs) that form a macromolecular protein complex.[24] The original expression experiment was conducted at 37°C. The experiment demonstrated that the 100 kDa construct, STRHSTPABPACIL1bt1QRSx1, had high expression and purity yield via metal affinity chromatography. However, since our main objective was to have every protein express at 21°C (see **3.5.2**), we decided to use this construct as a test. Fig. 26 shows that we

achieved our desired results of high expression and excellent purification by metal affinity chromatography. Migration patterns were validated by semi-log plots (see **3.3**).

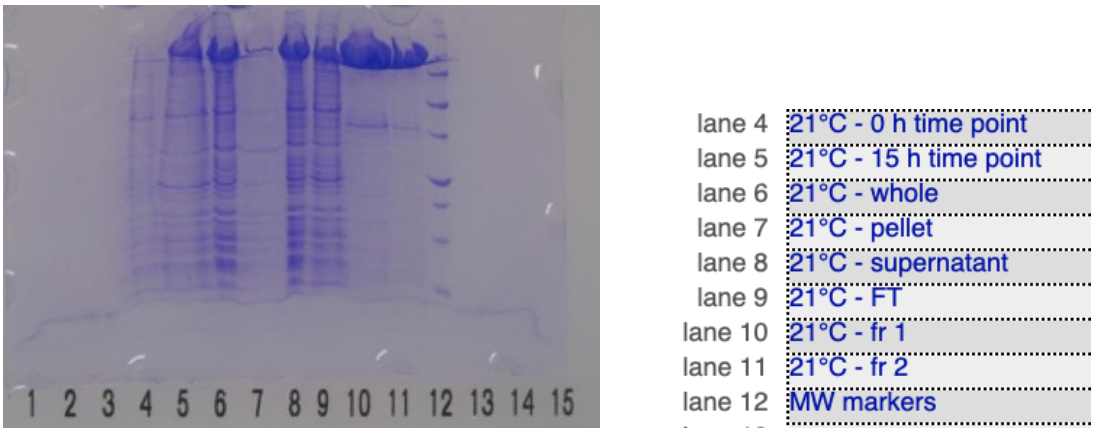


Figure 26. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABPACIL1bt1QRSx1 (100 kDa) at 21°C.

3.2.10 150 kDa Protein Construct

E. coli aminopeptidase (pepN) is involved in the ATP-independent steps during cytosolic protein degradation.[25] The subcloning process to make this construct was slightly different than the rest. *E. coli* genomic DNA was used as the template in the PCR amplification of the gene. The pepN gene was then cloned into an expression plasmid with a cleavable N term so it could be subcloned into the desired 100 kDa construct. Since this construct was not planned to be incorporated into our distributed ladder, it was not re-engineered to the incorporate the dual binding domain. It was also never tested on a Western blot. Once the first cloning step to transform the pepN gene into *E. coli* was finished, the following subcloning, expression and purification was carried out with ease. Although the first expression of this construct, STRHSTPABMBPpepNx1, was done at 37°C, we re-expressed it at 21°C to determine if it fit our criteria. Based on the results in Fig. 27, the construct yielded high quantities of protein, but

metal affinity chromatography was not efficient in purifying the protein. We concluded that this construct was viable because the desired protein was large in quantity compared to the degradation products to be easily recognizable as a protein marker. Further purification using HPLC by laboratory technician, Szu-yu Kuan, showed that this protein can be completely isolated from its degradation products (not shown).

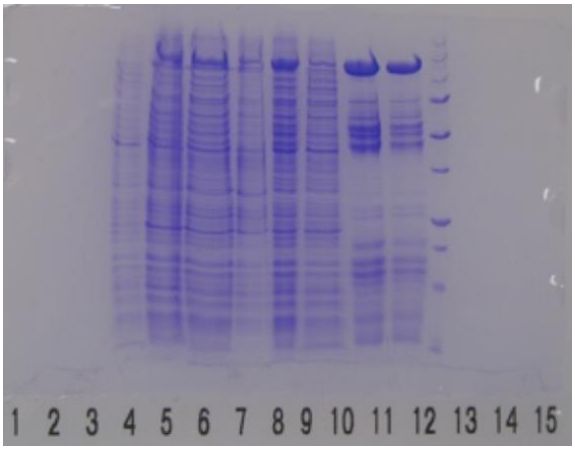
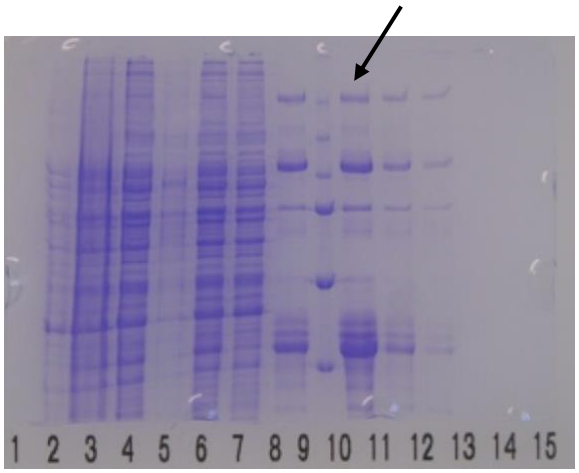


Figure 27. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABMBPpepNx1 (150 kDa) at 21°C.

3.2.11 250 kDa Protein Construct

The 250 kDa protein construct was a combination of the 100 kDa and 150 kDa protein markers to yield STRHSTPABPACIL1bt1QRSSTRHHSTPABMBPpepN. The subcloning was completed by Dr. Song Tan to help speed up the project. This protein was expressed at 21°C and purified by metal affinity chromatography. We were pleased to see that we were able to express such a large protein in *E. coli* despite the numerous degradation products that are seen in Fig. 28. Lanes 8 and 10 contain degradation products with a higher quantity than the desired protein. Ms. Szu-yu Kuan was able to purify this protein on a larger scale combining metal affinity chromatography with ion exchange chromatography (not shown).



lane 2	21°C - 0 Hr
lane 3	21°C - 15 Hr
lane 4	21°C - whole
lane 5	21°C - pellet
lane 6	21°C - supernatant
lane 7	21°C - flow through
lane 8	21°C - fr 1
lane 9	MW markers
lane 10	21°C - fr 2
lane 11	21°C - fr 3
lane 12	21°C - fr 4

Figure 28. SDS-PAGE of small-scale expression and metal affinity chromatography purification of STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN (250 kDa) at 21°C.

3.3 Combined Ladder Analysis (Individual Proteins)

A semi-log plot is a useful tool when analyzing the migration pattern of a protein. It has one axis on a logarithmic scale and the other on a linear scale. For this project, semi-log plots were generated by imaging gels and measuring the distance, x, based on pixels/distance in inches based off the ruler in Photoshop. The y-axis consisted of long (molecular weight) of the proteins. Since our markers were on a regular 10-interval scale, the distance between bands should decrease as molecular weight increases. Note, when analyzing the migration pattern of these proteins, it was important to dilute the fraction 2 samples to get thin bands on the SDS-PAGE gels to determine the protein's actual migration pattern. For example, in Fig. 26, the band is so thick, it is not clear if measurements should be taken from the bottom of the band or the top.

Several protein ladders were constructed by mixing dilutions of individual Talon Fraction 2 samples. This allowed us to analyze how each protein was migrating based on a line of best fit.

3.3.1 Comparison of 10 and 30 kDa Semi-log Plot.

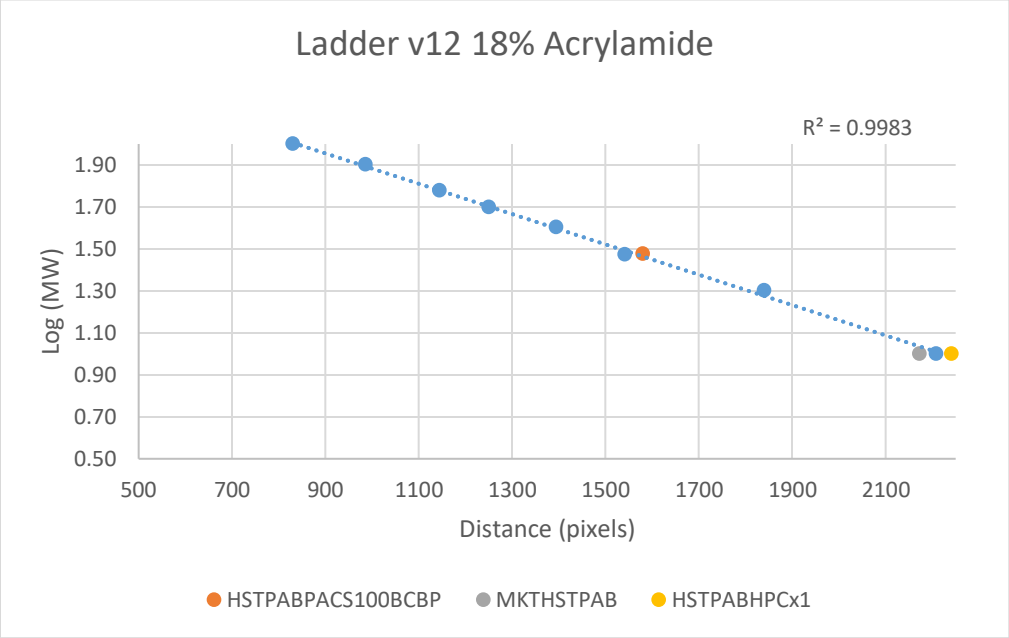


Figure 29. Semi-Log Plot of ladder of 18% SDS-PAGE including STRHSTPAB (10 kDa), HSTPABHPCx1 (10 kDa), MKTHSTPAB (10 kDa), STRHSTPABPACCBPx1 (20 kDa), HSTPABPACS100BCBP (30 kDa), HSTPABPABIL1bt1CBP (30 kDa), HSTPABPACGST (40 kDa), STRHSTPABMBPt1 (50 kDa), HSTPABPACdRCC1t8 (60 kDa), STRHSTPABPACIL1bt1dRCC1t8 (80 kDa), and STRHHSTPABPACIL1bt1QRSx1 (100 kDa). Note, the 15 kDa was not included in this plot because we planned to exclude this marker from the finished ladder.

Data points that are above the line represent proteins that are migrating too fast, whereas data points below line represent proteins that are migrating too slow. The original 10 kDa protein construct that Joey Shibata made was used in the calculation of the trend lane. However, the other candidate 10 kDa proteins, MKTHSTPAB and HSTPABHPCx1, were also plotted. Based on the graph, it was determined that HSTPABHPCx1 migrates the fastest (farthest to the right). Similarly, the faster migration of HSTPABPACS100BCBP compared to HSTPABIL1bt1CBP, would better fit the line.

3.3.2 Semi-Log Plots of Final Protein Markers

Once we figured out the correct constructs to yield an accurate semi-log plot, we reconstructed a semi-log plot with the correct proteins. Each individual sample was diluted in a final volume of 30 µl (Table 5). 10 µl of each dilution was mixed into a 1.5 mL Eppendorf to create a ladder master mix. 10 µl of the ladder master mix was loaded into an 18% acrylamide and BioRad 4-20% gradient acrylamide SDS-PAGE gel. We also added BioRad's commercially available markers next to ours for comparison in Fig. 31, 32 and 33.

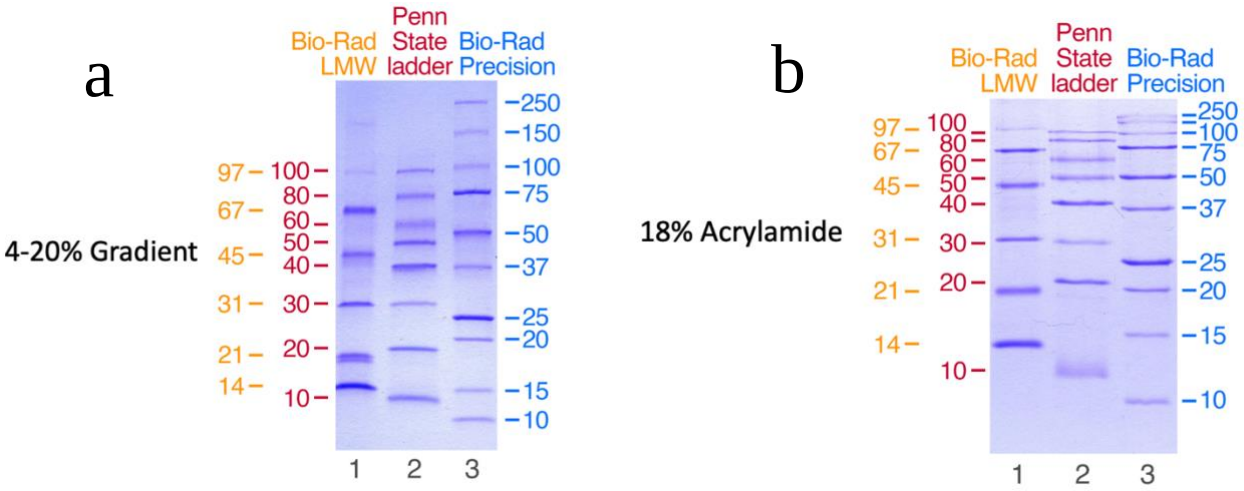
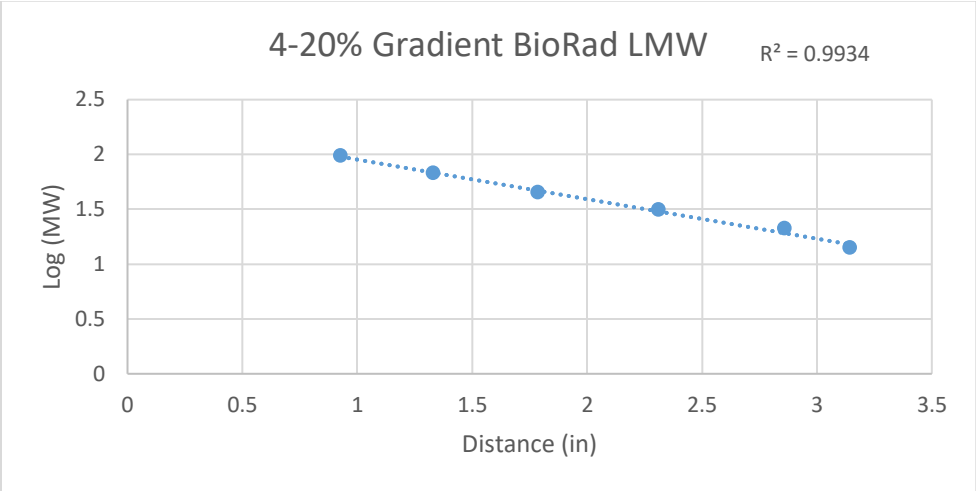
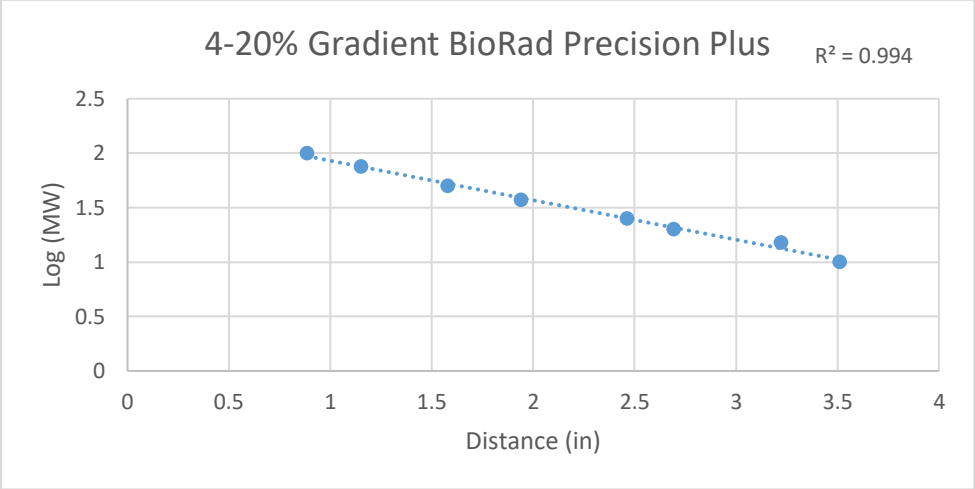


Figure 30. a.) 4-20% Gradient SDS-PAGE of 1- BioRad LMW, 2- Penn State Ladder Table 6), and 3- BioRad Precision Plus Marker. b.) 18% Acrylamide SDS-PAGE of Gradient gel of 1- BioRad LMW, 2-Penn State Ladder, and 3- BioRad Precision Plus Marker.

a



b



c

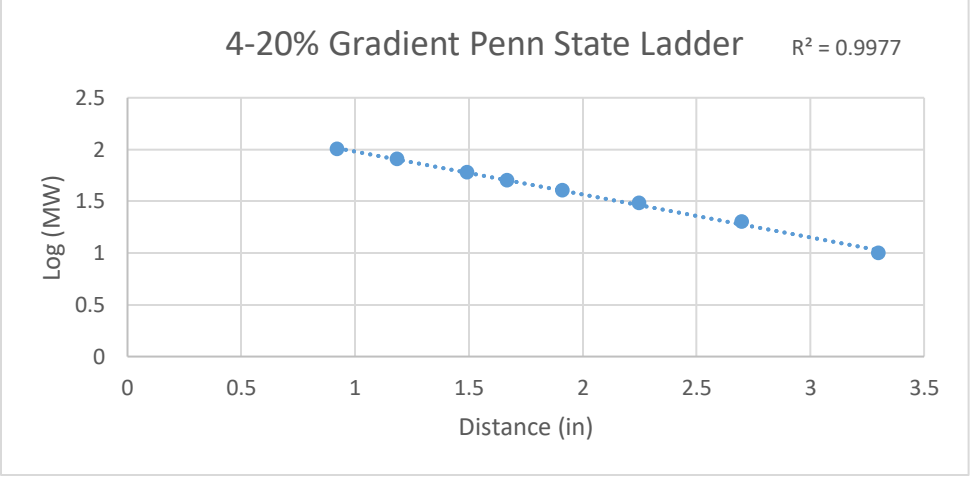
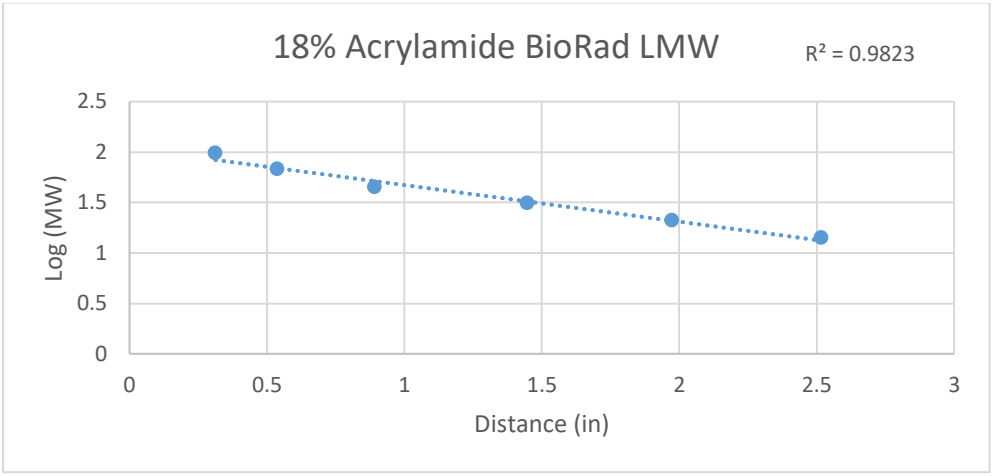
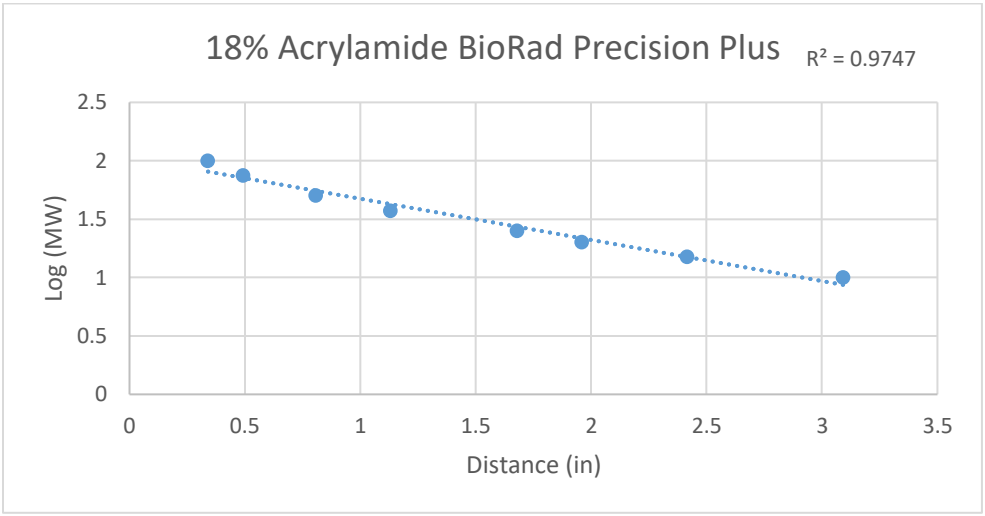


Figure 31. Semi-Log Plots of a.) BioRad LMW. b.) BioRad Precision Plus Ladder. c.) Penn State Ladder on BioRad 4-20% Gradient SDS-PAGE gels. Note, the 250 kDa and 150 kDa marker proteins in the BioRad Precision Plus Marker were excluded in semi-log plots for comparison purposes.

a



b



c

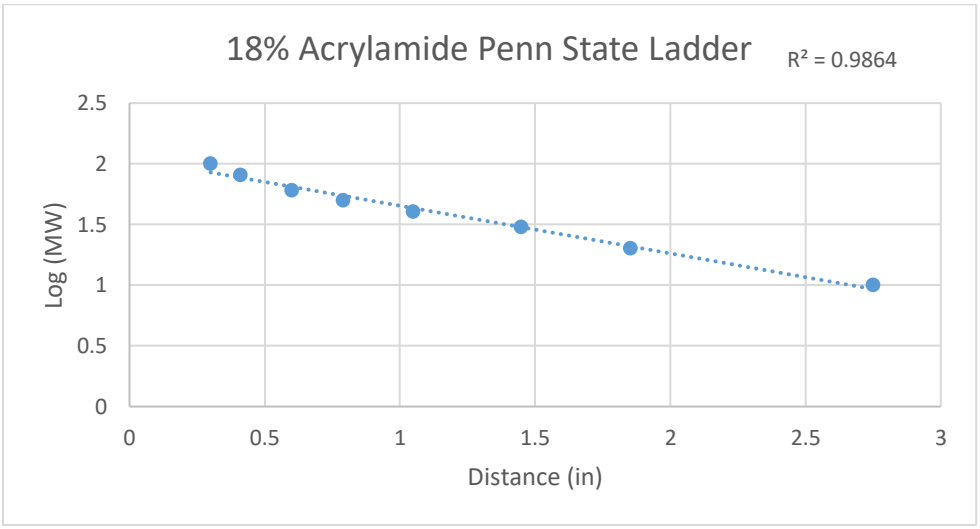


Figure 32. Semi Log Plots of a.) BioRad LMW. b.) BioRad Precision Plus Marker. c.) Penn State Ladder on 18% Acrylamide SDS-PAGE Gels. Note, the 250 kDa and 150 kDa marker proteins in the BioRad Precision Plus Marker were excluded in semi-log plots for comparison purposes.

4-20% Gradient SDS-PAGE gels are better suited to handle markers that span a broad range of molecular weights. The gradient allows the smaller proteins to migrate slower as they near the bottom of the gel creating an appreciable distance between proteins markers. Due to this accommodation, the R_2 of the trend lines for every plot in Fig. 31 are higher than their respective counter parts in Fig. 32. The Penn State ladder shows the highest R_2 value for both types of gels (Fig. 31c and Fig. 32c) denoting the superior accuracy that our ladders demonstrate over the commercially available products.

3.3.3 Summary of Individual Protein Molecular Weight Markers

The following figures and tables represent the results of our individual protein constructs.

Table 5. Individual Protein Marker Final Dilutions

Molecular Weight (kDa)	Protein Construct	Dilution (Fraction 2 sample: PGLB) total of 30µl
10	HSTPABHPCx1(P82K)	1:6
15	HSTPABPAC	1:2
20	STRHSTPABPACCBPx1(W157L)	1:2
30	HSTPABPACS100BCBP	1:15
40	HSTPABPACGST	1:15
50	STRHSTPABMBPt1	1:15
60	HSTPABPACdRCC1t8(4-422)	1:15
80	STRHSTPABPACIL1bt1dRCC1t8x27	1:2
100	STRHSTPABPACIL1bt1QRSx1	1:15
150	STRHSTPABMBPpepNx1	1:2

Table 6. Protein Constructs

component	component identity	MW
HSTPABHPCx1(P82K)	HSTPABHPCx1(P82K)	10014
HSTPABPAC	HSTPABPAC	15012
STRHSTPABPACCBPx1(W157L)	STRHSTPABPACCBPx1(W157L)	19876
HSTPABPACS100BCBP	HSTPABPACS100BCBP	29772
HSTPABPACGST	HSTPABPACGST	40361
STRHSTPABMBPt1	STRHSTPABMBPt1	50310
HSTPABPACdRCC1t8(4-422)	HSTPABPACdRCC1t8(4-422)	60029
STRHSTPABPACIL1bt1dRCC1t8x27	STRHSTPABPACIL1bt1dRCC1t8x27	79829
STRHSTPABPACIL1bt1QRSx1	STRHSTPABPACIL1bt1QRSx1	98269
STRHSTPABMBPpepNx1	STRHSTPABMBPpepNx1	149945
STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN	STRHSTPABPACIL1bt1QRSSTRHSTPABMBPpepN	249895

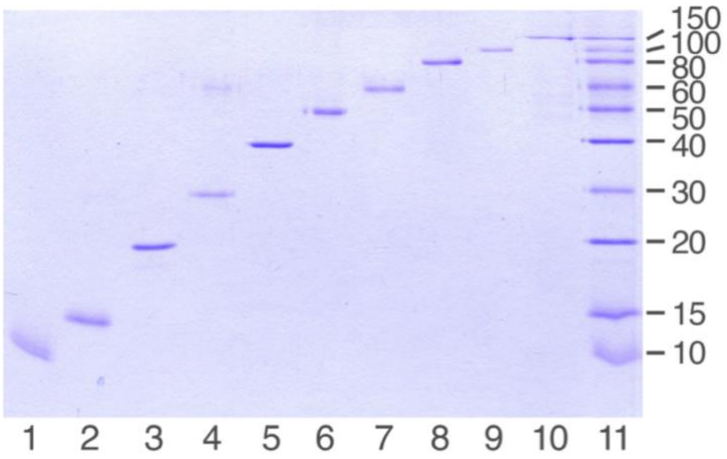


Figure 33. Individual marker proteins. Lane 1- 10 kDa. Lane 2- 15 kDa. Lane 3- 20 kDa. Lane 4- 30 kDa. Lane 5- 40 kDa. Lane 6- 50 kDa. Lane 7- 60 kDa. Lane 8- 80 kDa. Lane 9- 100 kDa. Lane 10- 150 kDa. Lane 11- Combined individual marker proteins. Note, the 15 kDa and 150 kDa proteins were not included in our semi-log plot analysis because they are excluded from the final ladder that is expressed via the polycistronic system

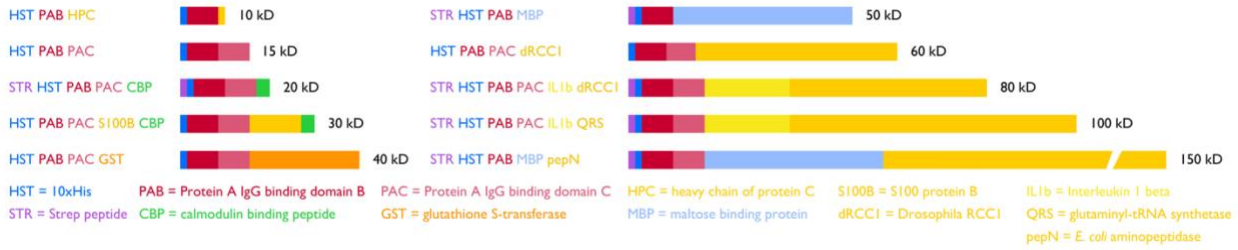


Figure 34. Relative proteins sizes and make-up of individual marker constructs. Used with permission by Dr. Song Tan.

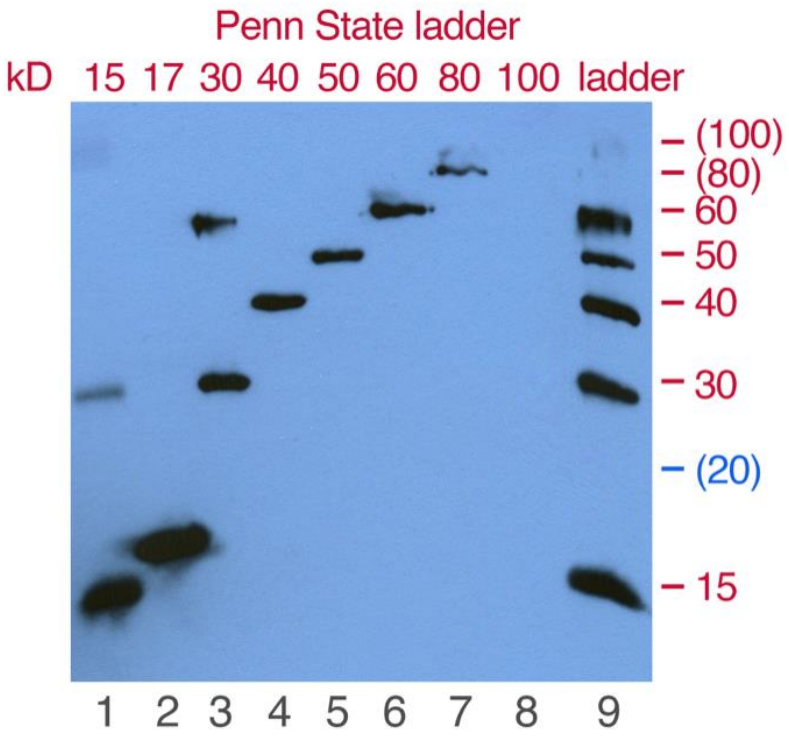


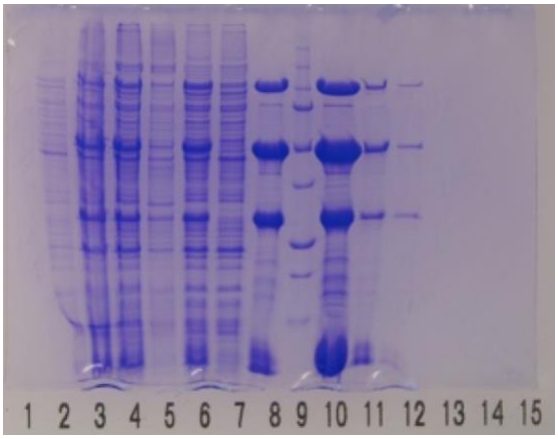
Figure 35. Western blot using HRP-linked anti-mouse IgG. Note, this was before the 20 kDa was optimized for successful transfer.

3.4 Polycistronic Analysis

Once the individual marker proteins were optimized for accuracy and precision, they were subcloned into pST44 expression plasmids. The pST44 polycistronic expression system is based on the pET3a expression vector where the use of a donor or transfer pET3aTr plasmid is

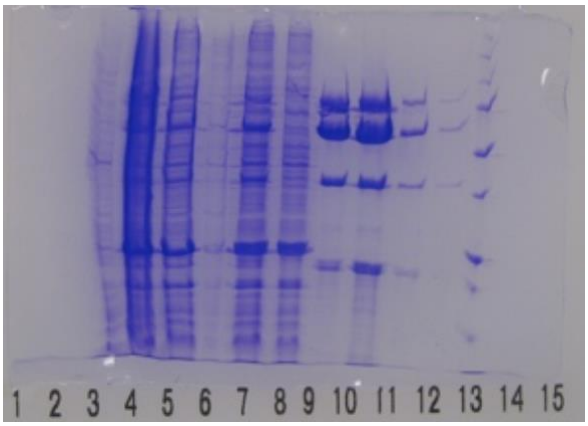
the source of a translation cassette (ribosomal binding site and coding region).[13] Although this system was originally designed to express a multicomponent eukaryotic protein complexes from *E. coli*, it was adapted to express 4 of our marker proteins simultaneously. We designed two plasmids to limit the total number of bases to be less than 10,000 in order to increase stability of the plasmid. Therefore, one plasmid was constructed to express the 10, 30, 50, and 100 kDa proteins while the other would express the 20, 40, 60, and 80 kDa proteins. Experiments were first conducted to subclone the protein markers into their respective transfer cassettes based on the appropriate restriction endonucleases to be inserted into the pST44 expression vector. Further subcloning was administered to equip the pST44 expression plasmids with the correct protein markers. The creation of these coexpression vectors was a group effort by me, Jack Williamson, Joey Shibata and Dr. Song Tan. Flow charts illustrating the scheme of this subcloning can be viewed in Fig. 6 and 7. The overexpression of this system was performed at 21°C (room temperature) for 16 hours. The results can be seen in Fig. 36.

a



lane 2	21°C - 0 Hr
lane 3	21°C - 16 Hr
lane 4	21°C - whole
lane 5	21°C - pellet
lane 6	21°C - supernatant
lane 7	21°C - flow through
lane 8	21°C - fr 1
lane 9	MW markers
lane 10	21°C - fr 2
lane 11	21°C - fr 3
lane 12	21°C - fr 4

b



lane 3	21°C - 0 hr
lane 4	21°C - 16 hr
lane 5	21°C - whole
lane 6	21°C - pellet
lane 7	21°C - supernatant
lane 8	21°C - flow through
lane 9	21°C - fr 1
lane 10	21°C - fr 2
lane 11	21°C - fr 3
lane 12	21°C - fr 4
lane 13	MW markers

Figure 36. SDS-PAGE of small-scale expression and metal affinity chromatography purification of a.) 10-30-50-100 kDa protein ladder and b.) 20-40-60-80 kDa protein ladder at 21°C.

Based on these results, we have successfully produced a protein marker ranging from 10 kDa to 100 kDa. The expression and purification of the 10-30-50-100 kDa protein complex (Fig. 36a) was slightly better than the 20-40-60-80 kDa complex (Fig. 36b) however, they both exhibit high expression and excellent purity via metal affinity chromatography.

3.5 Troubleshooting Details

3.5.1 Western Blot

Single vs Dual Binding Domain. According to a study on the interactions between an IgG-Binding domain based on the B domain of Staphylococcal Protein A and Rabbit IgG, a single-domain construct binds IgG fourfold weaker than a two-domain protein.[17] Therefore, in the initial design of our 11 total protein constructs, 8 had single domain IgG binding sites (PAB), while two had dual domains because we thought a fourfold decrease would not result in a significant reduction of signal. Verification of detection of the 15 kDa protein marker provided

us with a positive result to continue creating the previously planned markers. It was not until after a Western experiment was performed on markers ranging from 10-100 kDa that we noticed a reduction in signal in the single domain markers (Fig. 37).

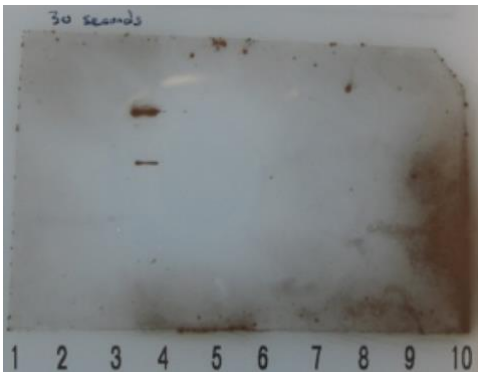


Figure 37. Western Blot of GKTHSTPAB, HSTPABS100B, HSTPABPACS100B, CBPSTRHSTPABGST, STRHSTPABMBP,HSTPABPACdRCC1t8, HSTPABGSTdRCC1t8, STRHSTPABGSTQRS incubated with anti-rabbit secondary antibody. The bottom signal in lane 4 represents HSTPABPACS100B (30 kDa) and the top signal represents HSTPABPACdRCC1t8 (60 kDa) after 30 second exposure time.

Only the constructs containing the dual IgG binding domains (HSTPABPACS100B and HSTPABPACdRCC1t8) are detected via chemiluminescence on the X-ray film. We realized that the single Protein A immunoglobulin binding domain was insufficient for detection on a Western blot. Therefore, every construct except the 10 kDa (due to size restraints) was redesigned in order to incorporate the dual IgG binding domain.

We had problems detecting our molecular weight marker proteins even after the new constructs were made. A Western blot analysis of the dual binding domain constructs with anti-rabbit secondary antibody still produced very low signals (Fig. 38).

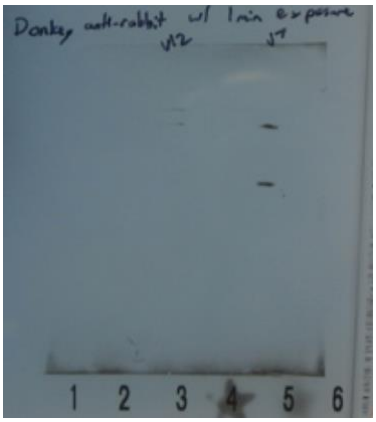


Figure 38. Western Blot of Ladder proteins. Lane 3. Ladder V12: HSTPABHPCx1, HSTPABPAC, HSTPABPACS100B, STRHSTPABMBP, HSTPABPACdRCC1t8, HSTPABPACGST, STRHSTPABPACCBP, STRHSTPABPACILb1t1 Lane 5. Ladder V7: HSTPABHPCx1, HSTPABS100B, HSTPABPACS100BCBP, CBPSTRHSTPABGST, STRHSTPABMBPt1, HSTPABPACdRCC1t8, STRHSTPABGSTQRS with anti-rabbit secondary antibody at minute exposure.

Lane 3 was loaded with ladder v12, containing the new protein ladder markers with the dual binding domain. Lane 5 was loaded with ladder V7, including the original constructs with only single IgG binding domains (except the 30 and 60 kDa), as a positive control. Based on the results, the incorporation of the dual binding domain seemingly did not appear to improve binding of the secondary antibody.

Table 7. Ladder v7 with original concentrations

V7							
10	20	30	40	50	60	80	100
HSTPABHPCx1	HSTPABS100B	HSTPABPACS100BCBP	CBPSTRHSTPABGST	STRHSTPABMBPt1	HSTPABPACdRCC1t8	HSTPABGSTdRCC1t8	STRHSTPABGSTQRS
6	2	0.5	15	0.5	0.5	0.75	1

Table 8. Ladder v12 with original concentrations

V12								
10	15	20	30	40	50	60	80	100
HSTPABHPCx1	HSTPABPAC	STRHSTPABPACCBP	HSTPABPACS100BCBP	HSTPABPACGST	STRHSTPABMBPt1	HSTPABPACdRCC1t8	STRHSTPABPACIL11bt1dRCC1	STRHSTPABPACIL11bt1QRS
6	1	1	0.5	0.5	0.5	0.5	1	0.5

Repeating the western blot experiment of ladder v12 and v7 with a three-fold increase in the individual protein concentrations and a longer exposure time (the time the membrane is in contact with the X-Ray film) increased signal intensity and quantity shown in Fig. 39.

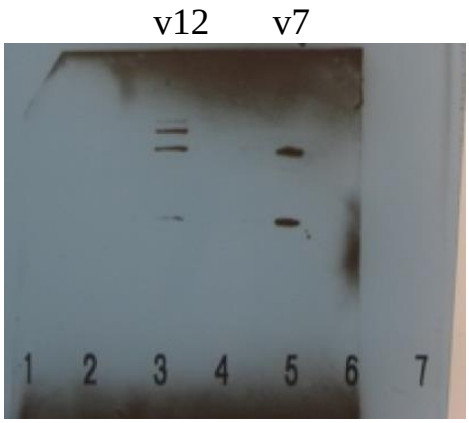


Figure 39. Western Blot of ladder v12 and v7 with three-fold increase in concentration and 5 min exposure time.

The v12 ladder in lane 3 shows the appearance of two more bands, presumably the 80 kDa and 100kDa protein markers. This differs from the v7 ladder in lane 3 because, because like before, only the 30 kDa and 60 kDa proteins are visualized but with a stronger intensity. This result verifies our hypothesis that the dual IgG binding domain fosters significantly increased secondary antibody binding. This positive result led to the conclusion that the proteins in the v12 ladder will be used in our final production molecular weight marker, however we sought to further optimize western blot conditions.

We were able to resolve the problem of weak Western blot signals when we realized the anti-rabbit second antibody we were using was several years old. An experiment utilizing a fresher source of anti-mouse antibody resulted in significantly increased detection. Instead of

using a 1:5000 dilution of anti-rabbit secondary antibody, a 1:5000 dilution of anti-mouse secondary antibody was used. Positive control samples were also included in Fig. 40.

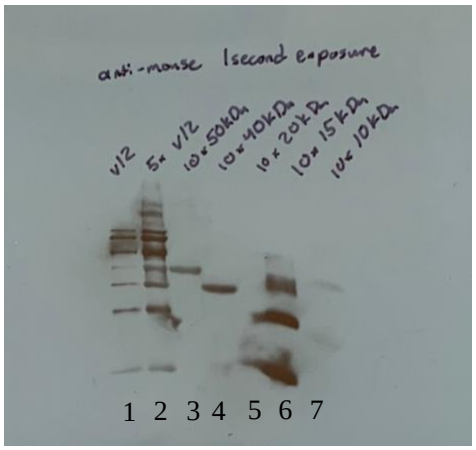
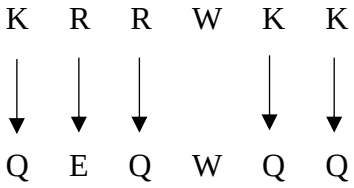


Figure 40. Western Blot analysis of Ladder V12 with positive controls (anti-mouse secondary antibody with 1 sec exposure time). Lane1. V12 ladder, Lane 2. 5x conc. V12 ladder, Lane 3. 10x conc of 50 kDa, Lane 4. 10x conc of 40 kDa, Lane 5. 10x conc of 20 kDa, Lane 6. 10x conc of 15 kDa, Lane 7. 10x conc of 10 kDa.

All bands were successfully transferred and visualized on the X-Ray film except the 10 kDa and 20 kDa protein as depicted from the lack of signal in lanes 5 and 7. Attempting to transfer a range of proteins ranging from 10 kDa to 100 kDa can be inefficient because the rate at which a protein transfers is directly correlated to its molecular weight. Small-sized proteins tend to transfer faster during the blotting process. Since the transblot process is set to a constant amperage (0.8 A), we concluded that the 10 kDa protein was over transferring, surpassing the membrane. However, since the two proteins bracketing the 20 kDa protein (15 and 30kDa) both transferred, we suspected that the issue lay in the 20 kDa protein itself.

Transfer of the 20 kDa Protein Marker. A protein's isoelectric point (pI) is defined as the pH at which the molecule confers a net electrical charge equal to zero. Therefore, at a pH below their PI, proteins carry a net positive charge and at a pH higher than the PI, the net charge is

negative. The PI of the original 20 kDa protein construct with the dual binding domain (STRHSTPABPACCBP) was 9.93. For comparison, the PI of all other marker proteins were roughly 7.0. The transfer buffer had a pH of ~8.3 conferring a positive charge to only the 20 kDa protein. This may have impeded the protein's ability to be transferred to the membrane during the transblot process. To fix this problem, 5 basic amino acids were mutated to acidic amino acids to decrease the PI to 7.23:



The resulting construct was coined STRHSTPABPACCBPx1.

3.5.2 Small Scale Expression

One of our main objectives of this project was too make these protein markers simple and inexpensive. In doing so, we wanted to make sure that our ladder would express highly at 21°C (room temperature) so no special laboratory equipment was required. In the initial small-scale expression experimental stages, we were expressing all of our protein at 37°C in the essence of time. When we tried expressing the 60 kDa construct at 37°C, we noticed that the construct did not express well, and the protein was insoluble. We suspected that dRCC1 was temperature sensitive since HSTPABPAC did not show any expression issues. In order to tackle this problem, our first troubleshooting solution was to re-express the construct at a lower temperature. Lowering the temperature usually improves the solubility of recombinantly expresses proteins because at lower temperatures, cell processes such as transcription, translation, and cell division

slow down. This reduces the probability that the protein will aggregate. We performed an expression of the 60 kDa construct, HSTPABPACdRCC1t8, at 18°C and noticed significantly increased expression yield. The comparison of expression yields can be seen in Fig. 41.

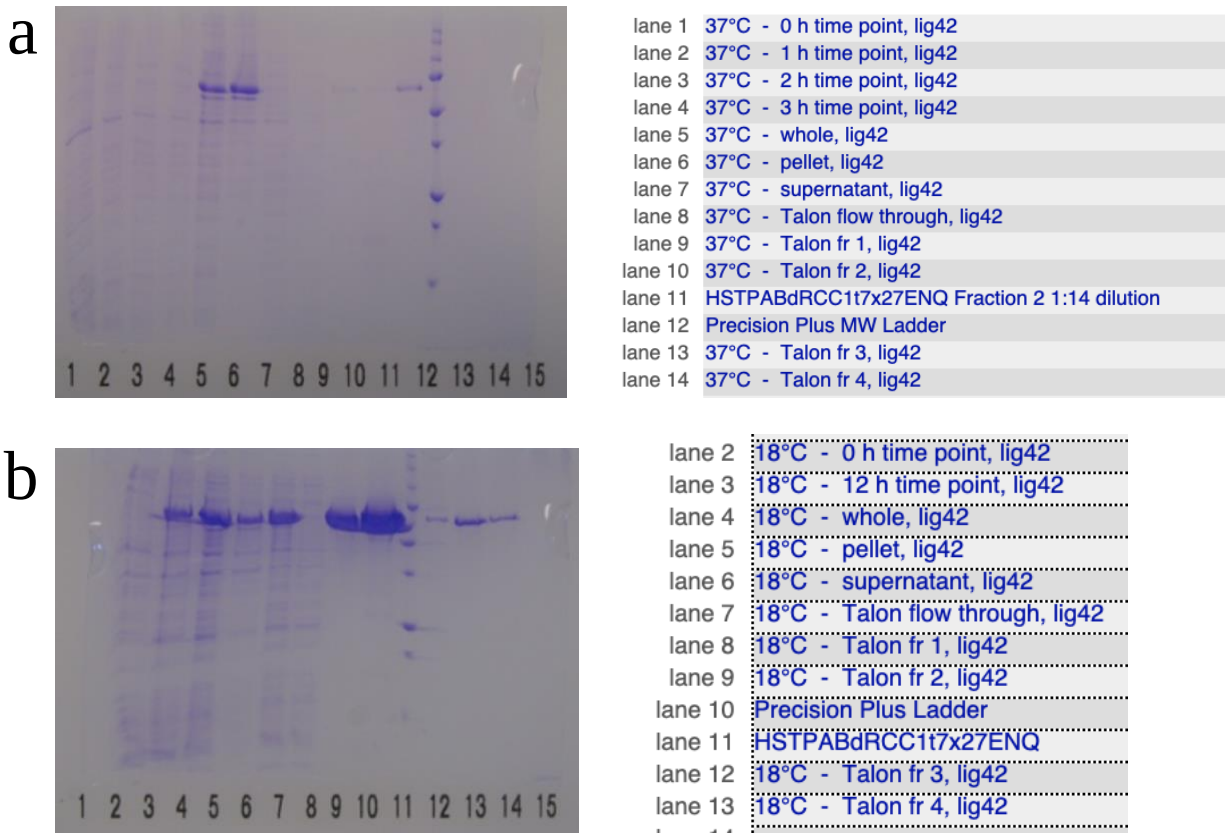


Figure 41. SDS-PAGE of small-scale expression and metal affinity chromatography purification of HSTPABPACdRCC1t8 at a.) 37°C and b.) 18°C.

Lowering the temperature significantly increased the solubility of HSTPABPACdRCC1t8. Although there is still some protein residing in the pellet, we were able to purify large amounts of the construct. Since increasing the temperature reduces expression, we were worried this protein might not express well at 21°C. However, we were pleased to see that

the there was ample expression of the 60 kDa marker in our polycistronic vector plasmid (Fig. 36b).

3.5.3 Polycistronic Expression Plasmid

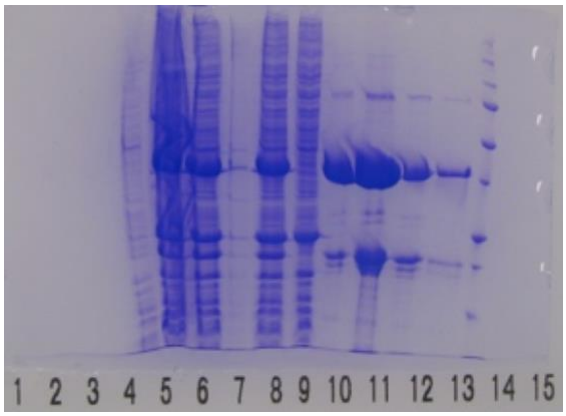
Our first attempt at expressing and purifying the 20-40-60-80 kDa protein plasmid was unsuccessful. When we expressed the construct at 21°C, we noticed reduced expression, specifically in the 20 kDa protein, and an overall poor purification by metal affinity chromatography (Fig. 42).

Figure 42. SDS-PAGE of small-scale expression and metal affinity chromatography purification of 1st attempt at 20-40-60-80 polycistronic expression plasmid at 21°C.

The only explanation we could think of that was causing this problem was the position of the proteins in the polycistronic vector system. Instead of the original subclone scheme of 20-60-80-40 (Fig. 7), we decided to move the position of the proteins and modify the subclone scheme by 20-40-60-80. In order to effectively troubleshoot this problem, we also expressed the precursor constructs (20-40 and 20-40-60) at 21°C. We noticed once the 60 kDa protein was incorporated, the expression of the 20 kDa and 40 kDa proteins were greatly reduced (Fig. 43b).

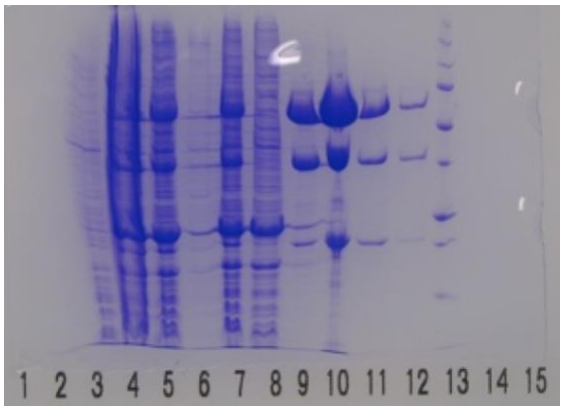
We currently do not have an explanation for this; however, it would be beneficial to go back and investigate this phenomenon. After the inclusion of the 80 kDa construct, the expression of all 4 proteins seemed to balance out and the final result showed significantly better purification by metal affinity chromatography (Fig. 43c).

a



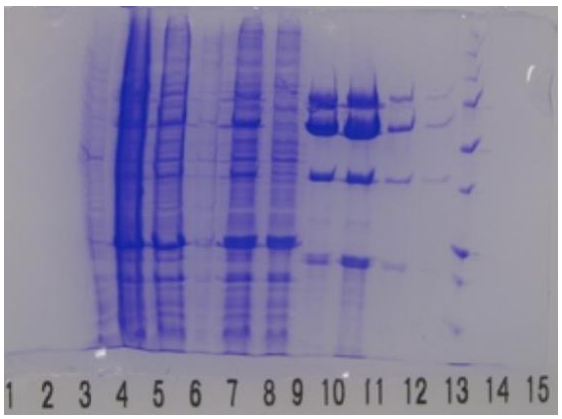
lane 4	21°C - 0 hr
lane 5	21°C - 16 hr
lane 6	21°C - whole
lane 7	21°C - pellet
lane 8	21°C - supernatant
lane 9	21°C - flow through
lane 10	21°C - fr 1
lane 11	21°C - fr 2
lane 12	21°C - fr 3
lane 13	21°C - fr 4
lane 14	MW markers

b



lane 3	21°C - 0 hr
lane 4	21°C - 16 hr
lane 5	21°C - whole
lane 6	21°C - pellet
lane 7	21°C - supernatant
lane 8	21°C - flow through
lane 9	21°C - fr 1
lane 10	21°C - fr 2
lane 11	21°C - fr 3
lane 12	21°C - fr 4
lane 13	MW markers

c



lane 3	21°C - 0 hr
lane 4	21°C - 16 hr
lane 5	21°C - whole
lane 6	21°C - pellet
lane 7	21°C - supernatant
lane 8	21°C - flow through
lane 9	21°C - fr 1
lane 10	21°C - fr 2
lane 11	21°C - fr 3
lane 12	21°C - fr 4
lane 13	MW markers

Figure 43. SDS-PAGE of small-scale expression and metal affinity chromatography purification of a.) 20-40 kDa construct. b.) 20-40-60 kDa construct. c.) 20-40-60-80 kDa construct at 21°C.

Chapter 4

Discussion

This study was successful in creating a simple and inexpensive recombinant protein ladder. The four criteria for each marker protein; high expression level in *E. coli*, solubility, high purification yield, and appropriate migration on SDS-PAGE have been met. We have also successfully incorporated a 10x HIS tag (HST) and a dual IgG binding domain for easy metal affinity chromatography purification and detection on Western blot, respectively. By expressing the two polycistronic expression plasmids at room temperature (21°C), we have demonstrated that the means of production was effectively simplified. Furthermore, the high purification yield of our ladder allows for the use of 1000 lanes in SDS-PAGE from 100 mL of *E. coli* culture.

4.1 Possible Explanations

During our troubleshooting process, we experienced issues with our polycistronic expression plasmid that did not have a clear explanation. When the 60 kDa protein was incorporated (Fig. 43B), the expression of both the 20 and 40 kDa proteins were reduced. Although we believe it to not be the case, MBP and GST stain unusually well when using Coomassie so we have to be careful when equating band intensity with expression level. The 60 kDa does not contain any of these protein components, but the 40 kDa construct does. This can explain why the 40 kDa protein marker appears to have a slightly higher expression than the 20 kDa marker but nonetheless, they are still both significantly reduced compared to the 60 kDa protein marker. For the specific example of the 20-40-60 construct, it is unknown whether the bacterial cells create individual transcripts for each construct or maintain one long transcript that

encodes all three together. In order to test this, a Northern blot can be performed to determine RNA length(s).

If the results of the Northern blot suggest that multiple transcripts are synthesized, two possible explanations arise. The first is that translational feedback mechanisms are causing unequal expression of the different constructs. Although there are many types of translational feedback, this discussion will focus on the most prominent types, translational feedback of ribosomal proteins and feedback regulation of rRNA synthesis. During ribosome biogenesis, the rate of ribosomal protein synthesis is dictated by the activity of ongoing rRNA expression. In *E. coli,* there are 19 r-proteins operons. These r-protein operons are responsible for controlling the autogenous feedback regulation. This mechanism works through one of the r-proteins in the operon that binds directly to the target rRNA and to its own mRNA at a site called the operator. The operator can be located either upstream of the first gene or in between genes in the operon.[26] Regulation occurs when the repressor r-protein is in excess and binds to the operator to inhibit translation of the proteins encoded in the mRNA. Furthermore, the repressor has the ability to either compete with the ribosomes for the mRNA or upon binding to the message, induce formation of structures to prevent the ribosome from accessing the translation start site.[26] This mechanism could be occurring when expressing our polycistronic expression plasmid. A repressor r-protein might bind to an operator that is located after the 40 kDa transcript, effectively repressing the expression of the 20 and 40 kDa construct. An RNA-seq experiment can be performed to determine the sequence of the RNA transcripts and determine where the operator(s) reside. This result will be useful in figuring out if any ribosomes are prevented from accessing the translation start site of the 20 and 40 kDa constructs.

The other feedback mechanism, known as feedback regulation of rRNA synthesis, attempts to explain how the number of ribosomes is regulated to meet the cell's demand for protein synthesis. During rapid growth, which the cells are subjected when performing a small-scale expression, several rounds of replication of the chromosome are initiated simultaneously. The targets during feedback are known as *rnn* promoters which control the production of ribosomes. It has been shown that deletion of four *rnn* operons induces expression from the remaining three by several times and this this increase is induced at the level of transcription initiation.[26] A feedback phenomenon causing a reduction in the number of ribosomes present, can influence a phenomenon known as a translational maximum.

A translational maximum simply refers to the maximum translation possible in a cell, as there is a finite number of amino acids. Whether it be a single transcript or multiple, the main goal of the small-scale expression is to express as much soluble protein as we can. Based on the results, the addition of the 60 kDa construct reduced the expression of the 20 and 40 kDa construct. Using the possible explanation of a translational maximum, the 60 kDa construct (in the last position of the polycistronic) seemingly conjured more resources due to its position for reasons still not understood. However, it makes sense that for whatever reason the 60 kDa was able to acquire a higher expression, the other constructs were consequently reduced based on the notion that there is a finite number of amino acids (translational maximum).

Lastly, a possible explanation to explain our results is instability of the transcript. mRNA is the least stable compared to other RNA variants. It is possible for single stranded mRNA to engage in self hydrolysis if the 2-hydroxy group makes a phosphodiester bond. Translation can also affect mRNA stability depending on codon composition. It has been demonstrated that the regulatory information affecting mRNA stability is encoded in codons and not in nucleotides.[27]

Stabilizing codons tend to be associated with higher tRNA levels and higher charged/total tRNA ratios while mRNAs with destabilizing codons possess shorter poly(A)-tails. An experiment utilizing reverse transcription followed by quantitative polymerase chain reaction (RT-qPCR) and RNA-seq can be performed to measure the RNA and determine if there are destabilizing components in the transcripts synthesized in our polycistronic expression.

4.2 Optimization of the PSU Ladder

The PSU Ladder can be further optimized with future experiments. A doublet is present in the 20 kDa marker in the 20-40-60-80 polycistronic plasmid. The second band is most likely caused by a degradation product. In order to reduce the likelihood of this protein degradation, an experiment can be performed by adding PMSF, a serine protease inhibitor. Furthermore, the metal affinity chromatography purification parameters can be altered to incubate the sample with talon resin for 20 minutes at 4°C (compared to room temperature). This modification will most likely remove the doublet band from the final ladder.

We will also have to optimize the small-scale expression time. We expressed all plasmids at 21°C for roughly 16 hours (overnight). Although we achieved optimal expression, 16 hours might not be needed to express ample amounts of the ladder. Experiments should be conducted to re-express both plasmids and take hourly time points up to 16 hours. All time points should be analyzed via SDS-PAGE to determine appropriate expression times.

Both expression plasmids will also need to be tested on a Western blot. Due to time constraints, we were not able to perform a final Western blot of our completed ladder. However, we are confident that no further problems will arise because we tested each marker individually.

Optimizing the purity of the individual proteins can also create a more aesthetically pleasing ladder without any stray protein fragments. This is especially important for optimizing the larger protein markers such as the 150 and 250 kDa constructs that contained numerous degradation products. Szu-Yu Kuan, one of our laboratory technicians, has already demonstrated that the 250, 150, 100, 80, 60, 50, 40, and 30 kDa constructs can be effectively purified via ion exchange and hydrophobic contact chromatography. The results yielded pure protein sample evidenced by an HPLC chromatogram. The purified samples can be subcloned into the modular design of the polycistronic vector system and re-expressed.

Lastly, we plan to experiment with textile dyes to create a pre-stained ladder. Proteins have the ability to bind dyes allowing them to be visualized while they are being electrophoresed. This can be beneficial when monitoring a protein sample and/or performing a Western blot. Pre-stained markers offer another advantage in that they do not require the Coomassie staining process.

4.3 Implications

Our markers are superior to the current commercially available ones for several reasons. First, they contain components such as the HST tag and dual binding domain for easy metal affinity purification and Western blot application, respectively. Second, they are well-designed with strict protein criteria for high expression and accurate migration. Third, they are better suited for "home-made" 18% acrylamide gels rather than commercially available gradient gels. Finally, they are economical. The cost of our 10-100 kDa protein ladder will be approximately \$0.01/lane or 1/100th the cost of commercial ladders.

We plan to distribute our protein molecular weight ladder alongside our DNA version on Addgene. Addgene is nonprofit online vector database that contains a digital collection of vector backbones from publications and commercially available sources. This allows researchers across the globe to gain access to our plasmids for a small fee. We do not plan to license these plasmids for academic studies because we believe a simple and inexpensive means of producing protein molecular weight markers has the potential to impact the wider scientific community.

Chapter 5 **Appendix A**

Penn State Protein Ladder Poster

PennState Eberly College of Science

The Penn State Protein Ladders:

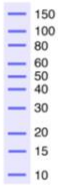
an inexpensive source of protein molecular weight markers appropriate for both Coomassie staining and Western blotting

Ryan Santilli*, Jack Williamson*, Rosalie Sowers*, Yoshitaka Shibata* and Song Tan

* Undergraduates

Abstract

We have designed and prepared recombinant proteins of defined molecular weights from 10 to 150 kDa that migrate appropriately on SDS-PAGE gel. Each protein was selected for high level expression in E. coli and contains a HIS-tag for efficient metal-affinity purification. Each protein contains an IgG binding domain and is therefore visible on Western blots without specific second antibodies. The Penn State protein ladder system should provide an inexpensive method to produce protein molecular weight markers for the research laboratory.



Experimental Design

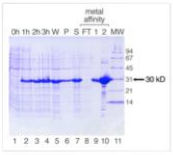
Criteria for molecular weight ladder proteins:

1. Express solubly at high levels in *E. coli*
2. Purify efficiently by metal affinity chromatography
3. Migrate appropriately on SDS-PAGE gel
4. Recognized by second antibodies in Western blotting

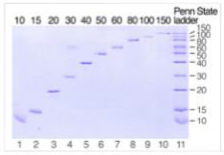
Methods:

1. Construct expression plasmids containing 10xHIS tag and IgG binding domain(s)
2. Express recombinant proteins in *E. coli*
3. Purify HIS-tagged proteins by metal affinity chromatography
4. Analyze protein mobility on 18% and 4-20% gradient gels
5. Analyze proteins by Western blotting

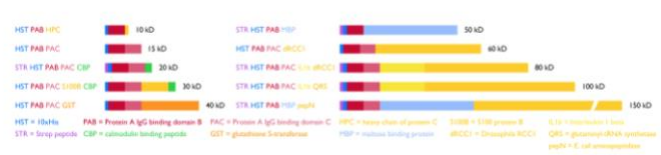
Results



Sample expression and purification of ladder proteins from 40 ml of culture

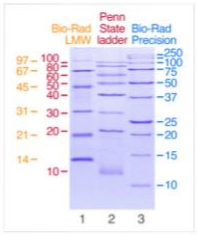


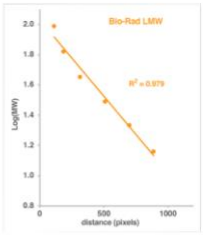
Individual purified ladder proteins

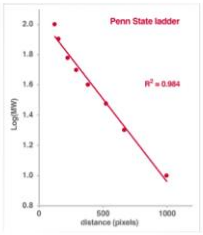


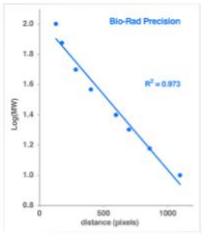
Schematic of recombinant ladder proteins

18% acrylamide gel

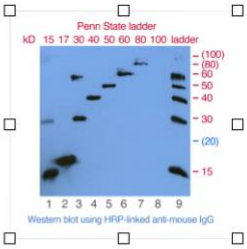




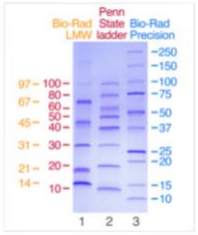


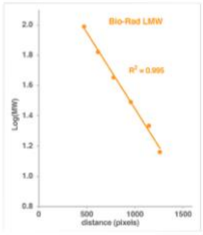


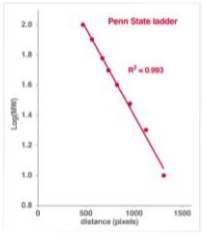
150 and 250 kD bands not plotted

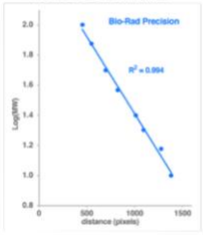


4-20% gradient gel









Please ask us about our DNA ladder plasmids

Status: We have expressed and purified 10, 15, 20, 30, 40, 50, 60, 80, 100 & 150 kD ladder proteins. 100 ml of E. coli culture can provide individual ladder proteins for 1000 lanes

Current work:

1. Optimize 20 kD protein for Western blotting
2. Create polycistronic expression plasmids to coexpress multiple proteins (2 plasmids each expressing 4 proteins)
3. Optimize coexpression conditions
4. Explore use of textile dyes to prepared prestained proteins

We plan to make our protein ladder expression plasmids available through Addgene, as we have done with our DNA ladder plasmids

This work was supported by Penn State Eberly College of Science Undergraduate Research grants and NIH grants R01GM088236, R01GM111651 and R35 GM127034 to S.T.

BIBLIOGRAPHY

1. What are proteins and what do they do? - Genetics Home Reference - NIH. (n.d.). Retrieved from https://ghr.nlm.nih.gov/primer/howgeneswork/protein
2. Mott, V. (n.d.). Introduction to Chemistry. Retrieved from https://courses.lumenlearning.com/introchem/chapter/protein-structure/
3. Dalton (unit). (2020, March 8). Retrieved from https://en.wikipedia.org/wiki/Dalton_(unit)
4. Determination of Protein Molecular Weight.(2005). *Edvotek* [PDF file]. Retrieved from https://www.edvotek.com/153.111202.pdf
5. Lin, W.-W., Chen, I.-J., Cheng, T.-C., Tung, Y.-C., Chu, P.-Y., Chuang, C.-H., … Cheng, T.-L. (2016). A Secondary Antibody-Detecting Molecular Weight Marker with Mouse and Rabbit IgG Fc Linear Epitopes for Western Blot Analysis. *Plos One*, *11*(8). doi: 10.1371/journal.pone.0160418
6. Kao, C.-H., Cheng, C.-M., Chuang, K.-H., Chuang, C.-H., Tzou, S.-C., Cheng, T.-C., … Cheng, T.-L. (2012). A regularly spaced and self-revealing protein ladder for anti-tag Western blot analysis. *Analytical Biochemistry*, *431*(1), 1–3. doi: 10.1016/j.ab.2012.08.001
7. Doucet, A., & Beauregard, M. (2001). A Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis Protein Ladder Made of Disulfide-Bridged Proteins. *Analytical Biochemistry*, *292*(2), 296–297. doi: 10.1006/abio.2001.5089
8. CHF, L. S. at. (2015, November 12). The Invention of Recombinant DNA Technology. Retrieved from https://medium.com/lsf-magazine/the-invention-of-recombinant-dna-technology-e040a8a1fa22
9. Protein A. (2020, March 4). Retrieved from https://en.wikipedia.org/wiki/Protein_A
10. Nilsson et al (1987). "A synthetic IgG-binding domain based on staphylococcal Protein A", Protein Engineering 1(2): 107-113.
11. Gallagher, S., & Chakavarti, D. (2008). Immunoblot Analysis. *Journal of Visualized Experiments*, (16). doi: 10.3791/759
12. Fleischman, Andrew (2010) Western blot markers with universal immunoglobulin binding domain, Penn State Schreyer Honors thesis
13. Tan, S., R.C. Kern and W. Selleck (2005) The pST44 polycistronic expression system for producing protein complexes in Escherichia coli, Prot. Expr. Purif., 40:385-395.
14. Tan, S. (2001) A Modular Polycistronic Expression System for Overexpressing Protein Complexes in Escherichia Coli, Prot. Expr. Purif., 21: 224-234
15. Bolanos-Garcia, V. M., & Davies, O. R. (2006). Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica Et Biophysica Acta (BBA) - General Subjects*, *1760*(9), 1304–1313. doi: 10.1016/j.bbagen.2006.03.027
16. Kimple, M. E., Brill, A. L., & Pasker, R. L. (2013). Overview of Affinity Tags for Protein Purification. *Current Protocols in Protein Science*, *73*(1). doi: 10.1002/0471140864.ps0909s73

17. Brown, N. L., Bottomley, S. P., Scawen, M. D., & Gore, M. G. (1998). A study of the interactions between an IgG-binding domain based on the B domain of staphylococcal protein a and rabbit IgG. Molecular Biotechnology, 10(1), 9–16. doi: 10.1007/bf02745859
18. Harper, S., & Speicher, D. W. (2010). Purification of Proteins Fused to Glutathione S-Transferase. *Methods in Molecular Biology Protein Chromatography*, 259–280. doi: 10.1007/978-1-60761-913-0_14
19. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification*, *41*(1), 98–105. doi: 10.1016/j.pep.2005.01.019
20. Varshavsky, A. (1996). The N-end rule: functions, mysteries, uses. *Proceedings of the National Academy of Sciences*, *93*(22), 12142–12149. doi: 10.1073/pnas.93.22.12142
21. (n.d.). Retrieved from http://www.cryst.bbk.ac.uk/PPS2/projects/pauly/proline/struc.html
22. Kapust, R. B., & Waugh, D. S. (1999). Escherichia colimaltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science*, *8*(8), 1668–1674. doi: 10.1110/ps.8.8.1668
23. Shi, W.-Y., & Skeath, J. B. (2004). The Drosophila RCC1 homolog, Bj1 , regulates nucleocytoplasmic transport and neural differentiation during Drosophila development. *Developmental Biology*, *270*(1), 106–121. doi: 10.1016/j.ydbio.2004.02.011
24. Kim, T., Park, S. G., Kim, J. E., Seol, W., Ko, Y.-G., & Kim, S. (2000). Catalytic Peptide of Human Glutaminyl-tRNA Synthetase Is Essential for Its Assembly to the Aminoacyl-tRNA Synthetase Complex. *Journal of Biological Chemistry*, *275*(28), 21768–21772. doi: 10.1074/jbc.m002404200
25. Chandu, D., & Nandi, D. (2003). PepN is the major aminopeptidase in Escherichia coli: insights on substrate specificity and role during sodium-salicylate-induced stress. *Microbiology*, *149*(12), 3437–3447. doi: 10.1099/mic.0.26518-0
26. Kaczanowska, M., & Rydén-Aulin Monica. (2007). Ribosome Biogenesis and the Translation Process in Escherichia coli. *Microbiology and Molecular Biology Reviews*, *71*(3), 477–494. doi: 10.1128/mmbr.00013-07
27. Wu, Q., Medina, S. G., Kushawah, G., Devore, M. L., Castellano, L. A., Hand, J. M., … Bazzini, A. A. (2019). Author response: Translation affects mRNA stability in a codon dependent manner in human cells. doi: 10.7554/elife.45396.03

ACADEMIC VITA

Ryan Santilli

Education

The Pennsylvania State University| The Schreyer Honors College University Park, PA
The Eberly College of Science Class of May 2020
Major: Biology- Vertebrate Physiology Option
Minor: Psychological Science

Honors & Awards

Schreyer Honors College Scholar	**Jun 2018**
University Fellowships & Phi Kappa Phi Peter T. Luckie Award for Outstanding Juniors (Science and Engineering)	**Apr 2019**
Anita M. Collins Undergraduate Student Research Fund in Biology	**Jun 2019**
Virginia L. Corson Headings Scholarship in the Eberly College of Science	**Jul 2019**
Global Experiences Scholarship from Eberly College of Science	**Feb 2020**
Schreyer Honors College Grant for Experiences Abroad	**Feb 2020**

Research

Structural Biology of Gene Regulation Laboratory
The Pennsylvania State University, University Park, PA **Jan 2018 – May 2020**

- Research the activity and binding domains of HLSD1 enzyme
- Produce protein molecular weight markers appropriate for both SDS-PAGE and for Western Blotting
- Create Polycistronic vectors
- ~25 hours per week

Ecology of Infectious Diseases Laboratory
The Pennsylvania State University, University Park, PA **Sept 2019- Dec 2019**

- Extrapolate data from published research on helminths
- Bioinformatics
- Statistical Analysis and modeling
- ~6 hours per week

Extracurricular

Global Medical Brigades
The Pennsylvania State University, University Park, PA **Aug 2016 – May 2020**

- Helped fundraise for brigades
- Volunteered at 15 BJC events

Alpha Epsilon Delta
The Pennsylvania State University, University Park, PA **Jan 2017 – May 2020**
- Weekly meetings with guest speakers
- Workshops

Volunteer at Hospital of University of Pennsylvania
Hospital of University of Pennsylvania, Philadelphia, PA **May 2017 – Aug 2017**
- Assisted nurses in protocols
- Answered call bells
- Transported patients

Saint Mary's Hospital PreMed Program
St. Mary's Medical Center, Langhorne, PA **Jun 2018 – Aug 2018**
- 100 hours of patient interaction
- 50 hours of shadowing

Penn State Biology Teaching Assistant
The Pennsylvania State University, University Park, PA **Aug 2019- May 2020**
- Teaching students Anatomy and Physiology
- 8 hours per week

The Penn State Red Cross Club
The Pennsylvania State University, University Park, PA **Aug 2019- May 2020**
- Volunteer at the Blood Drive

Life Link PSU Mentors
The Pennsylvania State University, University Park, PA **Aug 2019- May 2020**
- Volunteer to help mental disabled students

Employment History

Aug 2018- May 2019	Penn State Learning Assistant	The Pennsylvania State University, University Park, PA
Jan 2019- May 2019	Penn State Organic Chemistry Grader	The Pennsylvania State University, University Park, PA
May 2019- Aug 2019	Research Assistant	The Pennsylvania State University, University Park, PA

References
References are available upon request.