DESIGN OF ESCHERICHIA COLI HOST STRAINS FOR IMPROVED RECOMBINANT PROTEIN PURIFICATION: AN APPROACH THAT BRIDGES THE UPSTREAM AND DOWNSTREAM REALMS OF BIOPROCESSING

by

Patrick Robert Bartlow

B.S. Biosystems Engineering, Clemson University, 2006

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Patrick R. Bartlow

It was defended on

September 26, 2011

and approved by

Ipsita Banerjee, Ph.D., Assistant Professor, Chemical and Petroleum Engineering Department

Richard R. Koepsel, Ph.D., Associate Professor, Chemical and Petroleum Engineering Department

Badie I. Morsi, Ph.D., Professor, Chemical and Petroleum Engineering Department

William R. Wagner, Ph.D, Professor, Bioengineering Department

Dissertation Director: Mohammad M. Ataai, Ph.D., Professor, Chemical and Petroleum Engineering Department

DESIGN OF ESCHERICHIA COLI HOST STRAINS FOR IMPROVED RECOMBINANT PROTEIN PURIFICATION: AN APPROACH THAT BRIDGES THE UPSTREAM AND DOWNSTREAM REALMS OF BIOPROCESSING

Patrick R. Bartlow, PhD

University of Pittsburgh, 2011

Escherichia coli is a favored host for rapid, scalable expression of recombinant proteins for academic, commercial or therapeutic use. To maximize its economic advantages, however, it must be coupled with robust downstream processes. These are typically composed of three or fewer chromatography steps that remove the majority of the host proteins to achieve a reasonable degree of purification. Removal of remaining impurities is often very difficult and costly due similarity among physicochemical properties with the target. This work implements a novel approach that overcomes such limitations by bridging upstream and downstream realms, that is, by subjecting lysed cell material to a variety of purification procedures, identifying individual impurities and mitigating their removal by genetic modification.

Successful knockout of three prominent contaminants of immobilized metal affinity chromatography (IMAC) without detriment to cell growth or recombinant protein expression was demonstrated. Elution of the recombinant target was strategically manipulated with peptide tags that allowed purification to virtual homogeneity. Additional IMAC studies focused on the most problematic host proteins, those that retained binding affinity under stringent conditions. Two-dimensional difference gel electrophoresis discerned variation in the soluble extract pools loaded in IMAC and the subsequent impurities, with respect to varied levels of recombinant protein expression. Peptidyl-prolyl isomerase SlyD and catabolite activator protein were shown to be the most persistent contaminants and had greater prevalence at low target protein

expression. Since genetic removal of the transcription activator would negatively impact cellular function, I substituted specific residues to eliminate its IMAC affinity with minimal impact on its activity.

I applied this integrative strategy seeking to improve performance of cheaper, non-affinity based processes. Phosphoenolpyruvate carboxykinase and peptidase D were significant contaminants during serial purification of a target by hydrophobic interaction and anion exchange chromatography. Ribosomal protein L25 dominated non-target binding of a polyarginine tagged recombinant on cation exchange resin.

With the development of comprehensive genomic manipulation in higher order species, such integrative approaches will be conventional in the development of coupled expression systems for the production of complex biologics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

To the reader, my solitary authorship of this dissertation is not reflective of the nature of the work, which indeed would have been an impossibility without enriching, extraordinary collaboration. For this, I first and foremost thank Dr. Mohammad Ataai. An accomplished scholar, versatile researcher and gifted thinker, he is unquestionably all of these things. But to me his extraordinary mentorship has had the greatest influence. Before our working relationship even began, his welcoming, modest demeanor during the conversations that followed classroom instruction struck me. He was not at all dismissive of a lowly, unproven recruit struggling to gain traction in the program (at the time, an entirely fair description), but rather he engaged me as a respectable student in whom he saw great potential. Early on and throughout the progress of our work, during our fruitful discussions I perceived he was not doubtful I could propose experiments designed to yield reputable results and complete them. I believe this attribute of his mentorship has helped me in ways I can only begin to understand. I am far from a seasoned researcher, but I have him to thank for my development and preparedness for the road ahead. Indeed I am not a special case in this regard, as I have come to know others he mentored before me that share the same conviction. This is further testament to his character, for which I am most grateful.

Bob Beitle is also worthy of special recognition and unending praise. Throughout our collaboration, he has provided me with numerous opportunities to advance my fledgling scientific career for which I am indebted. He is solely responsible for arranging each of my

conference podium presentations to date and has entrusted me to contribute to funding proposals and manuscripts. His contributions and critical evaluations undoubtedly heightened the value of my work. I would be honored to be considered his peer and cannot speak highly enough of his generosity or hold him in higher regard.

I have been fortunate in having many other mentors and educators who positively shaped my development as a young scientist. In doing so, they have taught me that educating the inexperienced goes hand in hand with reaching new heights, scientific or otherwise. I salute Michael Domach, Richard Burgess, Matthew Ohland, Mani Balasubramani, John Cardamone, Guy Uechi, Tamanna Sultana, Zhu Liu, Ram Trivedi, Parvez Akhtar, Kyle Grant, Drew Cunningham and Lynette Spataro. Special thanks are due to committee members Ipsita Banerjee, Richard Koepsel, Badie Morsi and William Wagner. I thank Eric Beckman and Laura Schaefer for their affording me an opportunity to pursue this degree and reminding that scientific and engineering endeavors, no matter the application, must exist in the context of sustainable innovation. I am also in the debt of Kathryn Weatherhead, Caye Drapcho and David Brune for sparking my general interest in the biological sciences and specifically in humankind's burgeoning ability to hijack life's processes to realize novel and righteous goals, which is the essential basis of this work.

To my grandmother Mary, parents Bob and Stacey, brothers Jordan and Grant, close friends Shawn and Andy and the always lovely Erin, I express my deepest sincere gratitude. Any modicum of success was and will be built on the foundation of your unending love and support.

1.0 INTRODUCTION

In recent decades, advances in upstream technologies have dramatically improved cell culture scale, yield, and performance. Recombinant DNA techniques, hybridoma technologies, mammalian cell culturing, metabolic engineering, and fermentation improvements have permitted large-scale production of biological products, whether it be proteins for diagnostic or research interest, vaccine products, monoclonal antibodies, or biologic therapeutics.[1-3] These advances were in part a response to the ever-increasing market for biologics, which was valued at $149 billion in 2010 and projected to reach $239 billion by 2015.[4] As a result, a perceived shift has occurred where the manufacturing steps limiting productivity are now in the downstream phase, in which chromatographic separation remains the cornerstone of bioprocessing in both industry and academia. While affinity chromatography methods in particular are renowned for their exceptional resolving power, their efficiency suffers at production scales required to process the high volumes and product titers attained upstream.[5] Research efforts have focused on cutting material costs, improving productivity at large scale, and developing robust, generic separation steps in an effort to quicken time-to-clinic and market.[6, 7] Since downstream processes account for 50-80% of total manufacturing costs, efforts to ameliorate purification of high-value, high-quality products are critical to success in the biopharmaceutical industry.[8-11]

1.1 GENERAL BACKGROUND AND PRINCIPLES OF PROTEIN PURIFICATION

For comprehensive understanding of protein purification and its principles, downstream processing itself must be viewed in the full context of manufacturing. Generation of complex cellular material occurs in the upstream phase, in which cell lines are engineered to produce a protein of interest and methodically cultivated in successively larger volumes of growth media. The goal is to maintain the conditions that generate the highest possible number of viable cells in the shortest time possible to achieve an optimum rate of protein productivity. This is accomplished through a variety of means. Controlled nutrient feeding strategies can mitigate formation of unwanted byproducts like acetic acid to reach high cell densities and product formation.[12] The cell line or bacterial strain itself is customarily selected for superior gene expression and designed to be deficient in proteases that can inactivate the product. The product may be expressed as part of a protein fusion with signal peptides or chaperonins/foldases to promote its export from the cell membrane or to ensure folding into its proper soluble form, respectively. With empirical testing, these upstream strategies can be implemented to select for the ideal host system. Conversely, downstream processing relies on the exploitation of differences among physicochemical properties of the cell components and cannot be selected for in a similar manner.[13] Instead, general principles of purification guide the investigator through careful deliberation in choosing, testing, and refining a separation scheme.

The initial considerations address the general goals of the purification. What is the intended use of the protein? What is the nature of the source material? What is the desired purity? Are there safety concerns related to the source material that might require complete separation of a component (e.g. bacterial endotoxins)? Answers to these questions dictate the desired level of purity and shed light on the number and nature of the purification steps. The

specific number of steps is critical for one reason: some amount of product will be lost in each step of the process. If each step has a yield of 80% for example, in three steps the total amount of product recovered will be about half of what was made upstream. For high-value products, step yields integrally affect cost efficiency. To maximize resolving power of each step, they should exploit the disparities among *different* physical or chemical properties, such as size, surface charge distribution, hydrophobicity, or affinities to specific ligands or functional groups. Such steps are said to be orthogonal to one another. With optimized orthogonal separations, proteins can typically be resolved to the desired purity with high yield in four or less steps.

After concluding the upstream phase, when the cultures have run their course, the working material for separation must be prepared. At this stage, sample extraction, clarification and stabilization are the primary goals. As is often the case, the protein of interest may reside within the membrane of the cell, so the sample must be extracted. For proteins that are exported into the culture medium, cell disruption is obviously unwanted. When cells are lysed by enzyme treatment, pressure or sonication for example, particulate matter in the solution must be thoroughly clarified from the sample since it is not compatible with column separation techniques later in the scheme. Clarification by centrifugation then filtration is common at small scales of production. The product may also be clarified by precipitation, whether by ammonium sulfate, polyethyleneimine or other means. Knowledge of the product with respect to temperature, pH and ionic strength stability is important in preventing permanent inactivation.

The protein of interest is next captured from the clarified, complex cell lysate and simultaneously separates from the most plentiful contaminants. It is commonplace to pass this material directly over a compatible chromatography resin that emphasizes high product yield and capacity, often to the sacrifice of resolution. For affinity chromatography methods with superior

resolving power like immobilized metal affinity chromatography (IMAC), however, this step can purify the target protein to an intermediate degree as well. In bind-and-elute mode, the total elution volume that contains the protein of interest should be less than the loading material, which effectively concentrates it.

Schemes that capture the target with less resolving power, as with ion exchange chromatography (IEX) for example, must use a complementary step for enhanced purification. Hydrophobic interaction (HIC), affinity, and IEX of the opposite charge if used again (cation or anion exchange) are the principal intermediate stages in downstream processing. This is chiefly due to their ability to resolve differences in physicochemical properties, often even among differing protein species from the same gene. If the desired purity level is not met by this stage, a final polishing step is needed to remove any remaining trace contaminants, often on the basis of size (gel filtration) or surface charge.

IMAC was introduced by Porath and colleagues in 1975.[14] A complex, clarified protein mixture is passed over a packed column of porous matrix, typically highly cross-linked agarose, with covalently functionalized metal chelating groups such as nitrilotriacetic acid (NTA). Immobilized nickel resins are widely commercial, although purifications are also carried out on immobilized cobalt, zinc, and copper. While immobilized, the metal ion is still free to covalently coordinate with an electron-donating base, such as the aromatic nitrogen in the imidazole group of a histidine residue, the strongest retaining amino acid.[15] For this reason, loading material is typically prepared in the presence of a low concentration of imidazole and a sufficient level of sodium chloride to curtail weak and nonspecific interactions, respectively. Peptide tags containing histidine are frequently used to endow a target protein with metal-binding affinity for rapid purification, the nuances of which have been extensively reviewed.[16-30] Bound proteins are

typically eluted by increasing imidazole concentrations in the mobile phase or by lowering the pH. Since imidazole competes with histidines for binding sites, higher imidazole concentrations shift the equilibrium of binding such that proteins will elute. Decreasing the pH changes protein residue ionization, lowering its potential to donate an electron and thus interact with the immobilized ion. High target protein capacity, specificity, and column operation under mild conditions are just few of IMAC's many advantages. Further details for discussion on IMAC characterization and applications may be found here.[14, 15, 31-39]

Ion exchange chromatography is perhaps the most widely utilized chromatography operation for protein separation. Electrostatic interactions between proteins and the charged matrix give separation based on differences in their net surface charges and their distribution. The mechanism is therefore not only determined by protein sequence and structure, but also highly influenced by matrix type and mobile phase pH and ionic strength.[40-42] Anion exchange chromatography (AEX), which captures negatively charged species, and cation exchange (CEX) that captures positive species, are versatile in their use and purpose as they often can be useful at any stage in purification.[2, 9, 43-48] A CEX application of particular interest is the use of polyarginine peptide tags to endow a protein with strong, localized positive charges, allowing protein retention at a pH that would otherwise result in its flow-through.[49-59] Bound proteins are typically eluted by increasing the salt concentration. Although this technique has been successful, it is more protein-dependent than histidine-tagged IMAC separations since retention is fundamentally reliant on electrostatic interactions.

Hydrophobic interaction is a chromatography method that permits interaction between non-polar residues of a protein with a porous matrix functionalized with hydrophobic groups, making it technique orthogonal to both affinity and ion exchange. The amount of a kosmotropic

agent in solution, that is a salt such as ammonium sulfate that promotes stable intermolecular interactions, determines the binding and elution conditions. Fundamentally, under low salt conditions, water molecules that surround proteins and non-polar ligands have a "shielding effect" on their interaction. As the buffer salt concentration increases, however, the salts strongly interact with the water molecules, leaving less water available to shield protein-ligand interactions. It is then thermodynamically favorable for hydrophobic residues to interact with the ligands since there is a gain in entropy from the change in water molecule structure. Desorption is thus achieved by reducing the salt concentration. The discovery of these HIC principles are reviewed in a comprehensive guide put forth by Amersham Pharmacia.[60] Mobile phase pH is less influential on HIC than in IEX or IMAC, having only subtle effect on protein retention within the pH range 5-8.5.[60, 61] Therefore at room temperature with a given column and salt type, salt concentration is the variable for tailoring protein purification. The technique is commonly used to separate protein aggregate species from monomeric forms,[62] serum proteins,[63] membrane-bound proteins,[64] and recombinant proteins[65] in industrial or academic laboratories.[66, 67]

Although purification strategies have long favored packed-bed chromatography methods, they are limited by high costs, low throughput and cumbersome scale-up.[68] To this end, some researchers are fervently seeking alternatives to chromatographic separations,[5] such as charged ultrafiltration membranes,[69, 70] crystallization,[71, 72] aqueous two-phase partitioning.[73, 74] But it is the applications that require the utmost resolving power to produce highly pure biologics such as those for therapeutic use when chromatography is preferred, though formidable technical challenges must be overcome. As mentioned, the field of biotechnology was spurred by further understanding of host expression organisms and recombinant DNA technologies. Accordingly, some proteins may be expressed to very high levels in *E. coli*.[12, 75-89] Exceptionally high levels

will result in the formation of insoluble inclusion bodies that require renaturation procedures that diminish recovery, putting emphasis on expressing soluble, stable target proteins. But often high-value proteins are present only at low concentrations in cell lysate mixtures composed of a myriad of impurities.[3] Target proteins present at low feed concentrations in complex mixtures are challenging to resolve. The critical operating parameters of each downstream step must be systematically optimized. Maintaining conditions within a target's 'window of stability' with regard to pH, temperature, ionic strength and exposure to chemical substrates presents more constraints. And integrating a combination of downstream operations is itself an undertaking. But perhaps the most challenging aspect of bioprocess development is the presence of host cell proteins with physicochemical properties that are nearly indistinguishable from the target protein, making their separation exceptionally difficult. This nearly always occurs in a given purification even when the target protein is endowed with unusual or non-native properties. The general lack of reliable and relevant information on such contaminants is thus a costly shortcoming. This dearth of information is largely due to the fact that the contaminants themselves are inherently specific to the host organism itself, the downstream processes selected, and to a lesser extent the target expressed.

It is precisely this latter challenge that the work presented in this dissertation addresses. An ideal plan of action would be one that not only uncovers the comprehensive physicochemical information about the contaminating host proteins relevant to specific downstream process, but one that also entirely *eliminates* their downstream burden without additional steps. For many applications, this is an ambitious goal indeed, and one that will require unique approaches.

1.2 A RATIONAL PURIFICATION STRATEGY THAT BRIDGES THE UPSTREAM AND DOWNSTREAM REALMS

In the last decade, techniques common in the field of proteomics, the large-scale study of an organism's proteins, have become more readily utilized to address certain upstream biotechnology considerations such as how the host cell metabolism fluctuates in response to recombinant protein expression or changing fermentation conditions. But only in the last few years have techniques that couple high throughput protein resolution and mass spectrometry been applied to enhance downstream process development. It is important to understand these technologies and how they can provide powerful insight to both the upstream and downstream realms. Only then can they be appropriately utilized to spur novel solutions to difficult problems that arise in protein purification.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a powerful resolving technique that couples isoelectric focusing (IEF) that separates proteins on the basis of net charge in the first dimension with conventional electrophoresis for separation by mass in the second dimension. From a sample taken from the culture containing a living host undergoing fermentation for example, 2D-PAGE can give a comprehensive view of the physiological state of the cell. In 2001, Champion and colleagues performed this analysis of *E. coli* at the completion of four different high-cell-density fermentation processes in which differing biopharmaceutical products were expressed.[90] Proteins were visualized and identified to assess how the cell responds to the fermentation conditions at the end of the growth cycle when much of the recombinant products are expressed. Reproducibility of the protein profile, for example, would lend credence to the robustness of the fermentation. While this might be interpreted as an upstream concern, this analysis also addresses if the majority of immunogenic host proteins are

consistently present in all fermentation processes. If so, it would be feasible to develop a single, generic immunoassay to quantitatively determine sample content of host cell protein (HCP). Clearance of which is a critical measure of downstream separation performance. Indeed, the study determined that the protein profiles were 85-90% similar across all fermentations. This demonstrates that understanding protein profile differences upon changes in upstream conditions is of vital interest to downstream process development.

Dürrschmid and colleagues coupled RNA microarray analysis with difference gel electrophoresis to monitor the transcriptome and proteome to gain insight into *E. coli* stress response mechanisms, the impact of inducer IPTG and again the profile consistency after expression of two different recombinant products.[91] Difference gel electrophoresis (DIGE) is a technique similar to 2D-PAGE except protein samples from different sources are independently labeled with fluorescent dyes, pooled together and then co-migrated on the same gel to eliminate variations. The dyes themselves only minimally influence migration behavior and upon imaging using excitation wavelengths pertaining to each dye coupled with statistical analysis, changes in individual protein expression across conditions can be determined. Here it was important that the investigators used tightly controlled fermentation conditions to give the best representative, homogeneous cell material for accurate transcript and proteomic data. The observed decrease in energy-generating and respiratory chain genes lead to the conclusion that electron transfer via the membrane was impaired due to high expression of recombinant products; and further that breakdown of the proton motive force induced expression of cytoplasmic heat shock proteins. In all, 62 genes and only seven proteins were consistently expressed due to recombinant production. This not only gives rise to needed stress markers for upstream monitoring, but also gives insight on the proteins that one may encounter downstream, regardless of the biologic product.

In a similar study carried out by Aldor and colleagues, proteomic analysis of high-cell-density *E. coli* fermentation helped improve the yield of a recombinant humanized antibody fragment.[92] In *E. coli*, antibody fragments are expressed in the cytosol, but secreted into the periplasm for assembly into a soluble form. Distinguishing from changes in protein expression due to culture conditions with those from biologic production, the investigators showed that synthesis of the stress protein PspA was strongly correlated with recombinant synthesis. The cell poorly assembled the antibody, with much of the heavy and light chain products accumulating independently. By inducing another plasmid expressing PspA before induction of the antibody in a preventative maintenance approach, a moderate increase in soluble antibody fragment yield was observed, which they speculated was due to PspA increasing transport efficiency through an alternative pathway. The downstream implications here are also important. Cells that produce biologics less prone to aggregation and are more efficiently assembled will not burden separation steps with as much malformed protein with properties similar to the target.

It must be stressed again that these techniques, which effectively and rapidly expose cellular processes and their components, have major influences on the success of recombinant biologic production. They present targets for bioprocess improvement that could not be readily apparent to an investigator a priori. This power has only recently been harnessed to validate development of purifications schemes directly.

Jin and collaborators applied DIGE analysis in Chinese Hamster Ovary (CHO) cells to detect profile changes in secreted HCPs with respect to: culture media; bioreactor control strategies like temperature shifting; feeding strategy; and cell viability.[93] In mammalian hosts, biologics like monoclonal antibodies are secreted into the culture medium along with other proteins and the cytosolic contents of lysed cells. Incidentally, HCP differences between

monoclonal antibody producing and null cell cultures were minimal; confirming that use of cell not producing an antibody to generate immunoassay reagents (e.g. for the detection of HCP) is acceptable. It should be noted that while immunoassays provide sensitive measurements regarding unwanted protein content, 2D-PAGE or DIGE coupled with identification tools reveal physicochemical properties of individual protein species. Not surprisingly, differences in cell viability generated the most significant changes to the HCP profile. This was supported by a study by Grzeskowiak and colleagues, which found that cell viability had more influence on HCP profile than the CHO expression clone itself.[94] In that work, DIGE was also used to confirm that the chosen IgG_1 antibody purification scheme was robust enough to compensate for variations upstream, demonstrating that DIGE was a complementary tool to semi-quantitatively track impurities and monitor product quality downstream.

The use of DIGE as a quantitative quality control measure provides a nice segue to a comment on "design space." The term is defined by the FDA as "the multi-dimensional combination and interaction of input variables that have been demonstrated to provide an assurance of quality."[95] Understanding of the design space is one of the primary components of the "Quality by Design" (QbD) paradigm, which is an imperative adopted by the FDA that expects quality to be built into a process rather than testing product quality after a process. Other QbD components are understanding the relationship between the process and the critical quality attributes (CQAs) of the product and the influence CQAs have on the product's clinical properties.[96] The proteome-defining technologies discussed in this chapter can directly contribute to QbD by defining the design space of upstream or downstream processes.

Modifications to 2D-PAGE can provide useful predictors of downstream behavior as well. Despite how well established IEX is as a purification tool, for example, specific knowledge

of individual protein behavior on the chromatographic media is still limited. IEX interactions are complicated by hydrophobic and hydrogen bonding interactions, or protein-protein associations that may lead to unexpected behavior. Tools such as titration curve analysis have done little to predict elution behavior. Now, in general practice, IEF uses reducing and denaturing conditions so proteins can be migrated in a linear form. But IEF can be modified to allow resolution under native, non-reducing conditions akin to though used during IEX.[97] Migration of proteins based on their "functional p*I*" could roughly anticipate its chromatographic behavior.

The development of predictor tools for the downstream, however, is not new. Not long after the advent of therapeutic recombinant human insulin and modern biotechnology itself, some efforts were focused on developing predictive, or expert systems to optimize downstream processing trains not just based on traditional heuristics, but the physicochemical properties of prevalent contaminants. Leser and Asenjo gave an excellent review of this early work in 1992.[98] They stressed the need to define the "biochemical and thermodynamic properties of the major contaminants" in the starting material. Basic properties of thirteen prevalent contaminants in the *E. coli* lysate were catalogued and used as model inputs to estimate non-affinity column retention times; thus predicting which steps to use and in what order. Although these expert systems displayed robustness to slight variations in physicochemical data, later experimental validation of target purity significantly deviated from predicted values.[99, 100]

While there may not be a current need to improve the practicality of expert systems in today's industry, the demand to establish the *process-relevant* physicochemical properties of host cell proteins remains. The proteomic technologies discussed are filling that void. Prior works by our group and collaborators have used 2D-PAGE and mass spectrometry to define HCPs that contaminate different column operations. Cai and colleagues emphasized the use of proteomic

routines to impact bioprocess development by characterizing HCPs in cobalt IMAC [101]. Their group has also identified *E. coli* HCPs that bind to both IMAC and HIC columns under relevant conditions.[102] Similar work has elucidated proteins that commonly diminish IMAC performance.[103, 104] Careful structural analysis of individual protein contaminants can only result from such profiling work. The IMAC binder triosephosphate isomerase (TPI) is an enzyme that plays an important role in glycolysis since without it, half of the available carbon from glucose would divert away from central metabolism. Unpublished work has established that three point mutations in TPI can abolish its affinity for IMAC metals while maintaining glycolysis (Beitle, personal communication). Work by Humphreys and collaborators on the expression and purification of Fab' antibody fragments from *E. coli* periplasm had a similar outcome.[55] Under the IEX conditions, co-purification of the periplasmic phosphate binding protein PhoS/PstS presented a significant challenge. While retaining cellular function, the IEX affinity was sufficiently altered by the addition of six C-terminal aspartic acid residues and select lysine residues substituted with glutamic acid, ultimately removing the need for additional purification steps.

In this thesis I address specific challenges in protein purification using an integrated approach (**Figure 1.1**), of applying the high resolving power, semi-quantitative analysis, and in-depth protein information gained by 2D-PAGE, DIGE and mass spectrometry to the study of host cell protein clearance during recombinant protein purification in *E. coli*.

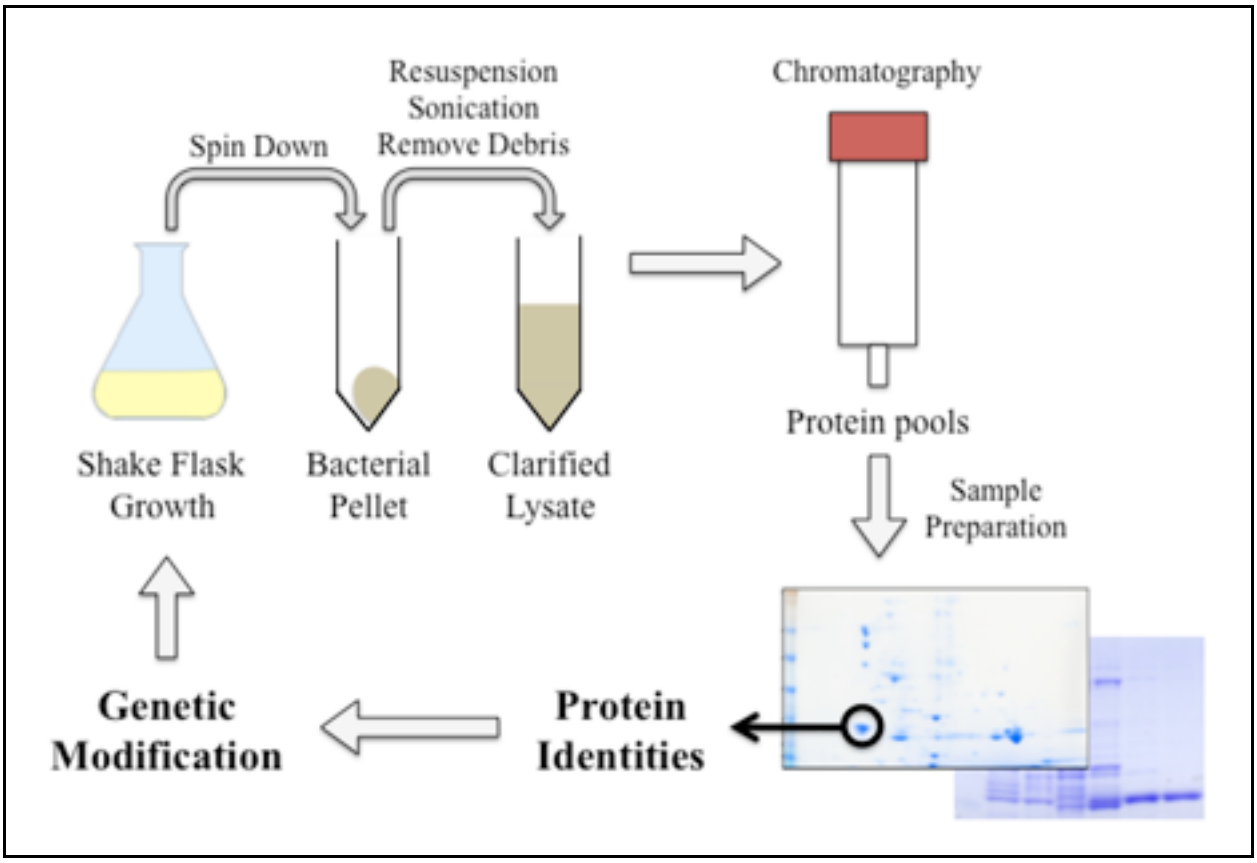

Figure 1.1. Schematic of the general workflow used to modify host strains for improved downstream efficiency.
Bacterial pellets are grown in shake flask cultures, centrifuged, then resuspended in a suitable buffer for cell lysis and chromatography. Whole cells are lysed by sonication then debris is centrifuged. The clarified lysate (supernatant) is used for the chromatography operation of choice and fractionated protein pools are prepared for 1D or 2D gel analysis. Protein identities are determined from the gel spots by MALDI-TOF-MS and information on the contaminants is used to make rational modifications to the host genome.

2.0 MATERIALS AND METHODS

2.1 IMAC STUDIES

2.1.1 DNA techniques, strains and plasmids

E. coli BL21 was used for chromosomal knockout of *cyoA*, *yfbG* and *adhP*. Mutants were generated using the Quick and Easy *E. coli* Gene Deletion Kit (Gene Bridges, Heidelberg, Germany) according to the manual. Strains, plasmids and primers are reported in Appendix A. To obtain the triple mutant (BL21 Δ*cyoA*Δ*yfbG*Δ*adhP*), FRT-PGKgb2-neo-FRT template DNA was used as the PCR template. To knock out cyoA, the primer set cyoA F1 and cyoA R1 was used to amplify a 1737 base pair (bp) DNA fragment containing a kanamycin cassette. The PCR product was purified and stored at −20 °C. Initially, to prepare cells for homologous recombination, electrocompetent BL21 cells were mixed with the plasmid pRedET and placed in an Eppendorf Electroporator 2510 (Fisher Scientific, Pittsburgh, PA, USA). The transformants were selected at 30 °C on Luria-Bertani (LB) agar plates with 3 μg/ml tetracycline. A clone was inoculated in 1 mL LB, and incubated at 30 °C until the A_{660} was about 0.3. Then, it was induced with L-arabinose to a final concentration of 0.4% at 37 °C for 1 h. Subsequently the induced competent cells were transformed with 2 μL (500 ng) of the prepared linear PCR product. The culture was centrifuged and streaked out on LB agar containing 15 μg/mL kanamycin.

For analyzing positive colonies by PCR, candidates were picked and resuspended in 30 μl of distilled water. Samples were boiled at 98 °C for 5 min and employed as templates for the PCR reaction. The primers Kan F/*cyoA* R2 were added to amplify a 451 bp fragment. A positive colony carrying the flipase recognition target site (FRT)-flanked kanamycin cassette in place of *cyoA* was transformed with 706-FLP at 30 °C, then incubated at 37 °C overnight. Then cultures were incubated on fresh LB agar at 37 °C for 10 h. Colonies were replica-plated on two LB agar plates, one containing 15 μg/ml kanamycin, the other one without antibiotic. The clones that grew only on LB agar without kanamycin were chosen for PCR identification with the primers *cyoA* F2 and *cyoA* R2. A 203 bp DNA fragment was visible as the *cyoA* gene was knocked out. This strain was designated BL21 Δ*cyoA*. The same methodology was used in series to knock out *yfbG*, and *adhP* respectively, by repeating homologous recombination, identification, and removal of kanamycin resistance. Confirmation of mutants was by Southern blotting using standard techniques and probes listed in Appendix A.

For the construction of WHHHPH-GFPuv, plasmid GFPuv was obtained (Clontech, Palo Alto, CA). The primers F (5-GCC AAG CTT GTG GCA TCA TCA TCC GCA TAT GAG TAA AGG AGA AGA ACT TTT C-3) and R (5-TTG GAA TTC ATT ATT TGT AGA GCT-3) contained *Hind*III and *Eco*RI sites (underlined correspondingly), were used to polymerase chain reaction (PCR) amplify GFPuv. GFPuv was digested and ligated with the same enzyme sites in pGFPuv plasmid. BL21 (DE3) was transformed with plasmid WHHHPH-GFPuv. Transformation was according to the Promega PureYield plasmid miniprep system (Promega, Madison, WI, USA).

E. coli DH5α expressing his_6-tagged GFPuv was used for IMAC breakthrough experiments. The strain was constructed as described in a prior work.[105] For subsequent analysis of host cell proteins by DIGE and peptide mass fingerprinting, *E. coli* BL21 (DE3) (Stratagene, La Jolla, CA, USA) expressing the same plasmid was utilized.

2.1.2 Cell growth and medium

For studies pertaining to knockout of host proteins and the "low-background" elution region of IMAC, overnight cultures of BL21 in 25 g/L LB were grown from single colonies on LB agar plates containing 50 μg/mL ampicillin and used for inoculating 500 mL cultures in M9 supplemented with 10 g/L glucose. For strains expressing his_6- or WHHHPH-GFP and for all succeeding IMAC studies, overnight cultures in 25 g/L LB, supplemented with 50 μg/mL ampicillin, were diluted 1:25 in 500 mL fresh LB in a 2-L flask. Cells were induced with 1 mM IPTG at 1-2 hrs when A_{660} was approximately 0.4. Fermentations were carried out in a shaker at 37°C and 200 rpm for 6-8 h until the A_{660} indicated strains were in stationary phase. Cell pellets were collected by centrifugation at 5000 x *g* and stored at -80°C prior to lysis. *E. coli* BL21 (DE3) without plasmid was also grown as described without antibiotic or induction.

2.1.3 Sample preparation

Cell pellets were suspended in 20 mL of the appropriate chromatography starting buffer, then supplemented with Triton X-100 to 0.5%, 0.4 mM $MgCl_2$, 1 mM phenylmethylsulphonyl fluoride (PMSF) and 0.5 mg/mL lysozyme. Contents were stirred with a glass rod and incubated on ice for 10 min prior and during sonication at 7 W (RMS) for 15 min using a Vibra cell

ultrasonifier (Thermo Fisher Scientific, Rockford, IL, USA). Cell debris was pelleted at 5000 rpm for 20 min at 4°C and supernatants were passed through a 0.45 μm filter prior to column loading. When more sample volume was required, filtered supernatants from multiple pellets were combined.

2.1.4 Chromatography

Unless otherwise noted, chromatography steps were carried out with an ÄKTAprime Plus and 1 mL HisTrap FF columns (GE Healthcare, Piscataway, NJ, USA). Columns were equilibrated with 10 column volumes (CV) of dH_2O, followed by starting buffer at 1 mL/min. For IMAC breakthrough trials, lysate solutions containing 20 or 50 mM imidazole were applied to the column until outlet GFP concentration reached a plateau. Subsequent IMAC purifications for the purpose of analyzing bound proteins were standardized so sample loading volume and mass were equivalent to 10% GFP breakthrough. After sample loading, columns were equilibrated with starting buffer until the A_{280} reached baseline and stripped using an EDTA buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 0.1 M EDTA, pH 8.0). Fractions of 1 CV were collected as needed. Columns were equilibrated with dH_2O, cleaned by passing through 5 CV of 1 M NaOH, then regenerated and stored according to the manufacturer's specifications.

2.1.5 Protein concentration, assessment, and SDS-PAGE

Select fractions were concentrated using Amicon ultrafiltration cartridges with a MWCO of 3 kDa (Millipore, Bedford, MA, USA). Excessively dilute protein fractions were concentrated by trichloroacetic acid/deoxycholate precipitation, dried, and stored at -80°C. Total protein

concentrations were determined by the Bio-Rad protein assay using BSA standards (Bio-Rad Laboratories, Hercules, CA, USA). Soluble samples were mixed 1:1 with 3X sample buffer, incubated in a water bath for 3 min at 75°C. Precipitated samples were resuspended in 3X sample buffer and incubated in the water bath at least 6 min for adequate solubilization. Samples were loaded into NuPAGE 4-12% Bis-Tris gels and run at 120 V for approximately 1.5 h in MOPS running buffer (Invitrogen, Carlsbad, CA, USA). Protein bands were visualized with Coomassie G-250 GelCode Blue stain reagent (Thermo Fisher). Gels were scanned at 2400 dpi with an Epson Perfection 4490 photo scanner (Long Beach, CA, USA). Densitometry to estimate relative protein concentrations and GFPuv purity was performed with ImageJ software (http://rsb.info.nih.gov/ij). Image background was subtracted using the "rolling ball" algorithm set at 1200 pixels, or half the image resolution, and the sampling box to acquire the density profile was approximately one-third the lane width, according to Gassmann *et al.*[106]

2.1.6 Two-dimensional difference gel electrophoresis

Lysate samples of wild-type strain BL21 (DE3) and those expressing his_6-GFP uninduced and induced with IPTG were compared. A separate comparison was made with these samples among their corresponding proteins that bound during IMAC with 50 mM imidazole loading. Protein fractions were desalted using Zeba spin columns (7 kDa MWCO, Thermo Fisher) and resuspended in labeling buffer (7 M urea, 2 M thiourea, 4% CHAPS, 25 mM DTT, 30 mM tris, pH 8.5). CyDye DIGE fluor minimal dyes (GE Healthcare) were used according to instructions to label protein pools. ReadyStrip IPG strips (17 cm, pH range 3-10, nonlinear, Bio-Rad) were used for overnight rehydration and isoelectric focusing. After sequential 15 min equilibrations in IPG buffers containing DTT and iodoacetamide, the second dimension was run on 17 cm, 10%

acrylamide gels in Bio-Rad TGS running buffer (2X in top chamber, 1X in bottom) at 32 mA for 45 min, 50 mA for 90 min, then 75 mA until the dye front was 2.5 cm from the gel bottom. After fixing, gel images were captured using a scientific-grade, cooled CCD imager, custom built by Jonathan Minden's group. Images were imported into Delta2D software (DECODON, Greifswald, Germany) for dual-channel image warping and spot fusion, quantitation and labeling. Individual samples were reciprocally labeled in six gels, serving as an internal control for improved statistical power in spot quantitation. Resolved protein spots have an associated average volume, calculated from two independent gels. Average spot volumes with variance greater than one standard deviation from the mean were excluded from the quantitative analysis. Further detail about the software is available in a previous report.[107]

2.1.7 Peptide mass fingerprinting analysis

Gel spots of interest were manually picked and digested using the methods of Shevchenko *et al.* with a few modifications.[108] Briefly, plugs were cut from gel bands and destained. Reduction and alkylation utilized 2.5 mM *tris*(2-carboxyethyl)phosphine (TCEP) and 3.75 mM iodoacetamide, respectively, in 100 mM ammonium bicarbonate. Plugs were reduced during incubation for 15 min at 65°C in an air thermostat and alkylated according to Shevchenko.[108] Trypsin Gold (Promega) was prepared in 25 mM ammonium bicarbonate. Tryptic peptides (1 μL) were deposited on a MALDI plate, dried, and then crystallized with 1 μL α-cyano-4-hydroxycinnamic acid (CHCA). Spots were washed by pipetting down then up with 1 μL 0.1% formic acid then allowed to dry. MALDI-TOF-MS spectra were acquired with an Applied Biosystems Voyager DE Pro (Framingham, MA, USA) with delayed ion extraction and reflection mode. Spectra were analyzed using Data Explorer software version 4.6 (Applied Biosystems) by setting the signal-to-

noise ratio to 5 or higher and filtering the mass list for monoisotopic peaks and charge states of one. All identifications were confirmed by searching the SwissProt database (release 2011_02; 525207 sequences; 185522689 residues) using Mascot search engine (Matrix Science Inc., Boston, MA, USA).[109] Trypsin was set as the proteolytic enzyme allowing for one missed cleavage, oxidation of methionines and carbamidomethyl cysteines selected as variable modifications, and peptide mass tolerance set to +/- 0.5 Da or less.

When needed, simultaneous identification of many spots was done using the same general procedure but with equipment capable of high throughput. Gel spots were picked with a ProPic II (Digilab, Holliston, MA, USA) following overnight gel staining with SYPRO Ruby stain reagent (Molecular Probes, Eugene, OR, USA). Digestion and quadruplicate spotting on a 384-spot MALDI plate was automated with a Digest Pro MS (Intavis Bioanalytical Instruments AG, Cologne, Germany). Mass spectra were acquired with an ABI 4800 MALDI-TOF/TOF Analyzer (operation of digestion robot and ABI 4800 were as detailed in a prior report [110]).

2.2 CATABOLITE ACTIVATOR PROTEIN STUDIES

2.2.1 DNA techniques, strains and plasmids

The *crp* gene was amplified by PCR from *E. coli* BL21 (DE3). Forward primer (5'-GG CAT ATG GTG CTT GGC AAA CCG C-3') was designed to include *Nde*I restriction site. Reverse primer (5'-CC GGATCC TTA ACG AGT GCC GTA AAC-3') was designed to include *Bam*HI restriction site. The PCR product was visualized on a 1% agarose gel, extracted and then cloned into electrocompetant *E. coli* TOP10 from the TOPO TA cloning kit (Invitrogen, Carlsbad, CA,

USA) with a Bio-Rad Gene Pulser apparatus. Colonies were grown from which plasmid preparations made using the PureYield plasmid miniprep system (Promega). Double digestion products were generated with from plasmids with *Nde*I and *Bam*HI, visualized on agarose gel and confirmed by sequencing (Genewiz, South Plainfield, NJ, USA). Preparations of TOPO vectors encoding *crp* and pET21a plasmid (Novagen, Madison, WI, USA) were independently digested with *Nde*I and *Bam*HI. Digestion products were extracted from agarose gels, ligated to form pWT*crp*, and then transformed into chemically competent DH5α using the calcium chloride method. Purified plasmid from DH5α was likewise transformed into BL21 (DE3).

Site-directed mutagenesis of the *crp* gene was performed using the QuikChange kit according to the manual (Stratagene). Plasmid pWT*crp* was used as the template for single mutations. Sets of complementary oligonucleotide primers that contained each desired mutation were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA). Primers used to substitute alanine for His17 were 5'-CG ACT CTC GAA TGG TTC TTG TCT **GCC** TGC CAC ATT CAT AAG TAC CCA TC-3' and the reverse-complement, generating pH17A*crp* (mutated bases are indicated in boldface type). Primers used to substitute tyrosine for His19 were 5'-GG TTC TTG TCT CAT TGC **TAT** ATT CAT AAG TAC CCA TCC-3' and the reverse-complement, generating pH19Y*crp*. To make the double histidine mutation (pH17A-H19Y*crp*), pH17A*crp* was used as a template with the following primer and its reverse-complement: 5'-GG TTC TTG TCT GCC TGC **TAT** ATT CAT AAG TAC CCA TCC-3'. The double mutation plasmid was used as a template to create a triple mutant by substituting asparagine for Lys52 using the following primer and its reverse-complement: 5'-C TCT GTG GCA GTG CTG ATC **GAT** GAC GAA GAG GGT AAA GAA ATG ATC C-3'. All plasmid mutations were confirmed by sequencing and independently transformed into *E. coli* BL21 (DE3).

2.2.2 Cell growth and medium

Overnight cultures of BL21 in 25 g/L LB supplemented with 100 μg/ml ampicillin were diluted 1:25 in 500 ml fresh LB in a 2-L flask. Cells were induced with 1 mM IPTG at 1-2 hrs when A_{660} was approximately 0.4. Fermentations were carried out in a shaker at 37°C and 200 rpm for 6-8 h until the A_{660} indicated strains were in stationary phase. Cell pellets were collected by centrifugation at 5000 x *g* and stored at -80°C prior to lysis. *E. coli* BL21(DE3) without plasmid was also grown as described without antibiotic or induction.

2.2.3 Sample preparation

Carried out as described in section 2.1.3.

2.2.4 Chromatography

Clarified lysates were prepared in either 20 or 50 mM imidazole buffer and purified using 1 mL nickel-immobilized HisTrap FF columns. Twenty CV of lysate was passed over column prior to equilibration and elution, either with a single EDTA stripping, or with a linear imidazole gradient to 250 mM followed by EDTA. Relevant fractions were collected for analysis.

2.2.5 Protein concentration, assessment and SDS-PAGE

Carried out as described in section 2.1.5.

2.3 NON-AFFINITY-CHROMATOGRAPHY STUDIES

2.3.1 Strains and plasmids

Escherichia coli BL21 (DE3) harboring pGFPuv (Clontech, Palo Alto, CA, USA) was used for both HIC and AEX purifications and ammonium sulfate precipitations. BL21 (DE3) expressing GFPuv with an N-terminus arg_6 tag was used for CEX purification. For cloning arg_6-GFP, in pGEM3Z as a first step, PCR was carried out using forward and reverse primers and pGFPuv as template. Forward primer (5'-ATA AAG CTT CAT GAG TCG TCG TCG TCG TCG TCG TAA AGG AGA A-3') was designed to include *Hin*dIII restriction site and six arginine residues. Reverse primer (5'-TTG AAT TCA TTA TTT GTA GAG CTC ATC CAT GCC ATG TGT AAT CC-3') was designed to include *Eco*R1 restriction site. The resultant PCR fragment was ligated into *Sma*I restricted pGEM3Z, to produce blunt end ligated arg_6-GFP-pGEM3Z. The sequence of arg_6-GFP-pGEM3Z was confirmed by the DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock, AR). The arg_6-GFP fragment was then subcloned by digesting arg_6-GFP-pGEM3Z and pGFPuv with *Hind*III and *Eco*RI, followed by ligation to produce arg_6-pGFPuv. The sequence of arg_6-pGFPuv was again confirmed and protein expression was verified by Western blotting and fluorescence determination. Strains, plasmids and primers used in this study are reported in Appendix A.

2.3.2 Cell growth and medium

Fermentations were carried out as stated in section 2.2.2.

2.3.3 Sample preparation

Samples were prepared as described in section 2.1.3 with one exception. Lysate was prepared without lysozyme in later experiments that identified host proteins that co-purified with arg_6-GFPuv.

2.3.4 Purification of non-tagged GFPuv

All column purifications were performed with an ÄKTAprimePlus purification system with real-time monitoring of total protein absorbance at 280 nm and pre-packed columns from GE Healthcare (Piscataway, NJ, USA) operating at 1 mL/min and room temperature. Operating conditions of AEX and HIC purifications were independently optimized using soluble lysate from BL21 (DE3) pGFPuv. For AEX optimization, 1 mL HiTrap Q FF columns (GE Healthcare) were readied with five CV of start buffer, five CV of high salt elution buffer, then equilibrated with start buffer again before loading five CV of soluble extract in varying conditions. Initially, optimal pH was determined from a series of experiments where extracts were prepared as described above using 20 mM Bis Tris starting buffer containing no salt, then pH adjusted to either 5.4, 5.7, 6.0, 6.2, 6.7 or 7.2. After loading, columns were equilibrated with start buffer until A_{280} reached baseline before applying a 20 CV elution gradient from 0 mM to 1 M NaCl, followed by five CV of 1 M NaCl before re-equilibration with start buffer. Fractions of one CV were collected. Offline, each fraction was assayed for GFPuv fluorescence in triplicate using a Tecan Infinite M200 96-well plate reader with excitation/emission spectra set to 395/509 nm with subtraction of background fluorescence of non-GFPuv proteins. Extracts were adjusted to

pH 6.2 for subsequent runs to optimize the starting ionic strength. AEX runs using start buffers containing 50 or 100 mM NaCl were compared in the same manner.

For HIC optimization, 1 mL HiTrap Phenyl FF low-sub and high-sub columns were compared, as were starting ammonium sulfate concentrations. Extracts were initially prepared in HIC elution buffer, 20 mM NaH_2PO_4, pH 7.2. Then 5 mL samples of which were adjusted with elution buffers lacking or containing 4 M $(NH_4)_2SO_4$ to 8 mL, creating a series of extracts in starting buffers ranging from 0.7 to 1.5 M $(NH_4)_2SO_4$ with equal protein concentrations. After spin down of any precipitated material and filtration, 5 mL samples were initially applied to the low-sub matrix after washing with five CV elution buffer (low salt) and equilibration with start buffer (high salt). Sample loading was followed by start buffer equilibration to A_{280} baseline then elution with a 15 CV gradient and five CV equilibration with low salt buffer. Columns were regenerated with five CV of distilled water followed by five CV of start buffer. This scheme with 1.5 M $(NH_4)_2SO_4$ starting buffer was used to evaluate the high-sub matrix. Fractions of one CV were collected and assayed as noted above.

Final purification for proteomic analysis consisted of initial precipitation, in which clarified lysate was adjusted to 1.5 M ammonium sulfate and allowed to incubate on ice for 10 min prior to centrifugation of precipitated proteins and cell debris (10,000 *g*, 30 min, 4 °C). Supernatant was loaded on low-sub a HIC column and bound proteins were pooled (13 mL), sealed in dialysis tubing (MWCO 3500 Da, Thermo Fisher Scientific, Waltham, MA, USA), and floated in a graduated cylinder containing 1 L AEX start buffer (20 mM Bis Tris, 50 mM NaCl, pH 6.2) with gentle stirring. Buffer was refreshed once after several hours. Five mL of this sample was applied to the AEX column using the scheme described above. For comparison, the use of buffers, dialysis and order of columns was reversed using the same method.

2.3.5 Ammonium sulfate precipitation

To observe the precipitation behavior of host proteins with respect to varying concentrations of ammonium sulfate prior to column application, a soluble extract containing GFPuv was prepared as above and eight 1-mL aliquots were placed in 1.5 mL eppendorf tubes. A stock solution of saturated ammonium sulfate (541.8 g/L, 25 °C) was prepared and added to each aliquot to create a series of concentrations from 0.8-1.5 M in 0.1 M increments. Aliquots were incubated on ice for 30 min then spun 30 min at 10,000 *g*, 4 °C. Supernatants were carefully transferred to new tubes without disturbing protein pellets. Stock solution was again added to all aliquots except for the 1.5 M tube according to the formula provided by Encore Biotechnology (Gainesville, FL, USA, http://www.encorbio.com/protocols/AM-SO4.htm) to achieve final concentrations of 0.9-1.5 M ammonium sulfate. Solutions were incubated, spun, and separated as before. Pellets were washed with 50 μL 20 mM bistros buffer (50 mM NaCl, pH 8.5) before final spin down for 15 min at 15,000 *g*, 4 °C. Resulting pellets were resuspended in 3X sample buffer for gel analysis.

2.3.6 Cation exchange chromatography of arg_6-GFPuv

Operating conditions of CEX purifications were determined using soluble extract from BL21 (DE3) expressing arg_6-GFPuv. Extracts were prepared, in this instance, without the use of chicken egg lysozyme (p*I* 11.35, Sigma, St. Louis, MO, USA) as it interfered with target binding in initial trials. Column operation and optimization was similar to that of AEX, using 50 mM MES start buffer and extract pHs adjusted to range from 5.5 to 6.0. Five mL of extract was applied to HiTrap SP FF columns before start buffer equilibration and NaCl gradient or step elations. Ultimate isolation of proteins was carried out at pH 5.7 with a 15 CV gradient from 0 to

1 M NaCl. This was immediately followed by a step to 2 M NaCl for ten CV to ensure total protein desorption and column regeneration. For comparison, the identical procedure was carried out with BL21 (DE3) pGFPuv.

2.3.7 Protein concentration, assessment and SDS-PAGE

Carried out as described in section 2.1.5.

2.3.8 Two-dimensional PAGE

Concentrated pooled proteins were buffer exchanged with rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 25 mM DTT, 30 mM Tris, 0.2% Bio-Rad 3/10 carrier ampholyte, pH 8.5) using Zebra spin columns. Protein solution was applied to a 7 cm Bio-Rad IPG strip (pH 3-10, nonlinear) in a focusing tray and covered with a layer of mineral oil. Active rehydration was programmed for 12 h in the Bio-Rad Protean IEF Cell. Paper wicks were applied to cover the tray electrodes before focusing under preset default conditions (with linear voltage ramping and a sustained holding voltage at completion to prevent diffusion of focused proteins). Focused strips were carefully washed with distilled water then submerged in equilibration buffer (6 M urea, 0.375 M Tris-HCl, 2% SDS, 20% glycerol, pH 8.8) with gentle shaking under reducing (buffer with 2% DTT) then alkylating (2.5% iodoacetamide) conditions separately for 15 min each. NuPAGE Novel 4-12% ZOOM gels (Invitrogen) were used to resolve proteins in the second dimension at 120 V for about 1.5 h in MOPS running buffer. Protein spots were visualized with Coomassie G-250 GelCode Blue stain reagent.

2.3.9 Peptide mass fingerprinting analysis

Carried out as described in section 2.1.7 without the need for automated, high-throughput equipment.

3.0 EVALUATION OF *ESCHERICHIA COLI* PROTEINS WITH AFFINITY FOR NICKEL IMMOBILIZED AFFINITY CHROMATOGRAPHY

3.1 INTRODUCTION

Immobilized metal affinity chromatography (IMAC) is a primary resolving tool in protein biochemistry due in large part to its superb selectivity.[14, 104] Target proteins endowed with metal binding affinity via polyhistidine peptide tags separate well from proteins in complex, crude lysate.[16, 20] The relatively few host proteins with functional metal binding sites or clusters of surface-exposed histidine residues, however, do co-purify in IMAC and are problematic. Heat-shock proteins, expressed in reaction to nutrient starvation, oxidative stress, and/or elevated recombinant protein production, are among the common contributors to IMAC contamination.[103] After IMAC, removal of these and other co-eluting host proteins from the target protein is often problematic due to similarities among the physicochemical properties. Understanding the identity of these proteins and the nature of their IMAC binding with respect to target protein expression level are critical to formulating a successful strategy that mitigates their removal without added purification steps.

Previous studies within the research group were distinctive for selection and fusion of optimum affinity tags to recombinant proteins for improved binding and elution behavior. Beitle and Ataai fused an octapeptide derived from angiotensin I with native metal binding affinity to

TEM-β-lactamase and successfully purified the fusion in one step with immobilized zinc matrix from resolubilized inclusion body material.[111] Zinc, instead of the more popular nickel, was chosen for the column since its interaction with peptides is more stringent than with almost all transition metals used in IMAC; therefore higher purification can be achieved since fewer cellular proteins can bind. Aside from conferring zinc binding affinity to the target, this fusion peptide had two other advantages over traditional fusion proteins that serve as purification handles: (1) it contains only eight amino acids (two of which are histidine), so its minimal size imposes almost no metabolic burden on the cell, bypassing the problems experienced when recombinant proteins are fused to other full-form proteins and over-expressed; and (2) it is derived from the natural peptide angiotensin I, so expression of the target would not be confounded by stability, solubility or yield issues sometimes experienced with unnatural handles like hexahistidine (his_6).[112-114] Later, Pasquinelli et al. enhanced this approach with zinc by reporting on several affinity tags (notably HPHHGG) that were discovered to direct the elution of a recombinant protein to "low-background" regions of the elution profile.[105] The recognition and exploitation of the "low-background" region, or the specific window of the elution profile containing the fewest host protein, is critical to this and similar studies.

Patwardhan and Ataai detailed the effect of loading pH on binding capacity and target purity by mathematically modeling the parameters that govern purification from E. coli extract using immobilized copper.[33] Predictions of the model were evaluated experimentally to confirm that loading extract at an intermediate pH of 6.5 to 6, as opposed to a high pH of 7.25, resulted in decreased binding of impurities and higher binding capacities of the target protein (i.e. the "low-background" region). Next, optimum affinity tags were selected from phage-displayed peptide libraries that expressed a diverse array of hexamers as a viral coat protein (this method also

generated the peptides identified by Pasquinelli mentioned above). Copper-immobilized beads were put in suspension with the phage library at a desired pH and centrifuged, allowing selection for hexamers that only remained bound to the beads at a pH relevant to the "low-background" region.[24] After repeating the experiment with nickel resin,[115] 24 hexamers were identified as having optimal affinity (one of which being WHHHPH for nickel, which is a tag that will be addressed in the next section). This technique was later applied by Jiang *et al.* to identify a 19-mer peptide with optimal affinity for cobalt matrices that allowed purification of active herpes simplex virus type 1 based gene therapy vectors.[116]

Described in the next section is the first exploration into host cell design, and how it proved successful in demonstrating that comprehensive data of co-purifying proteins could be used to *enhance* the "low-background" elution region by removing unessential genes that express contaminating proteins. This recalls the theme mentioned in the opening paragraph of this chapter.

The results presented and discussed in the subsequent sections showed progression of my preliminary approach with identification of all IMAC-binding proteins above a specified affinity threshold, and thus are relevant to purification of any generically expressed, histidine-tagged recombinant protein. I also explored how contaminating protein pools change as recombinant protein expression is altered. It was hypothesized that strong-binding contaminants become more prevalent relative to the target protein as its expression diminishes. Two-dimensional difference gel electrophoresis (DIGE) – invented by Jonathan Minden[117] and advantageous for circumventing the gel-to-gel variation of two-dimension gel electrophoresis by independent labeling of two or three protein samples with fluorescent dyes and resolving them on a single gel[118, 119] – was employed to monitor such profile changes in the proteome and IMAC-binding

pool. This tool supplemented my proteomic-based approach to streamline bioseparation. The results led me to put forth a recently conceived strategy that mitigates IMAC contamination by essential gene products with protein surface modifications, which is the subject of chapter four.

3.2 IDENTIFICATION AND KNOCKOUT OF HOST PROTEINS THAT CONTAMINATE A STRATEGIC ELUTION REGION IN NICKEL IMAC

To examine nickel-NTA (nitrilotriacetic acid), the most frequently used IMAC resin, cultures of BL21 grown overnight in LB were used to inoculate 500 mL of M9 minimal media supplemented with 10 g/L glucose. In a manner described in the methods, pellets were produced and resuspended with IMAC start buffer (50 mM NaH_2PO_4, 500 mM NaCl, pH 8.0) without imidazole. Filtered lysate was applied to 4 mL ProBond nickel-chelating resin in an open column and after loading and equilibration with start buffer, step elutions were carried out with the same buffer at the following imidazole concentrations: 60 mM, 80 mM, 100 mM, 120 mM, and 200 mM. This was followed by column stripping with 500 mM EDTA. The elution volumes for each step were 24 ml, or 6 CV at a flow rate of 0.5 mL/min.

Figure 3.1 illustrates the protein concentrations in each fraction normalized to the total protein used for column loading (note that the scale of normalized protein concentration is logarithmic). Containing the lowest amount of host proteins, the fraction at 120 mM imidazole was deemed the "low-background" region of the elution profile, so proteins pooled from that peak were of keen interest for further study. SDS-PAGE and LC-MS/MS were used to identify the cellular proteins present in the "low-background" region (in this instance, mass spectrometry was performed independently by the University of Pittsburgh Genomics and Proteomics Core

Laboratory). Briefly, three spots were excised from each band and each digested with trypsin. Peptides were separated by liquid chromatography (LC) and then identified by tandem mass spectrometry (MS/MS) fragmented by collision-induced dissociation. The LC/MS data were searched against the MSDB database (20040106, 1319480 sequences) with MASCOT v. 2.1. For positive identification, spectral data from each of the three spots matched.

A total of 18 proteins were identified, presented in **Table 3.1**. They were classified based on their functionality and essentiality to the cell. Genes are essential when strains deficient in the gene cannot survive. Strains survive the deletion of non-essential genes, but are not necessarily suitable for a biotechnology application. From the identifications, nine genes are non-essential, eight are essential, and one is unknown. Functional classification are as follows: eleven are associated with DNA binding, RNA binding and protein biosynthesis; five contained metal ion binding sites; one involved in carbohydrate metabolism; and one was of unknown function. Four of the proteins, DnaK, YfbG, Fur and SlyD, were previously reported as impurities in IMAC purification schemes.[103, 120] Further inspection of data in **Table 3.1** reveals that the four heaviest proteins, CyoA, YfbG, AdhP, and DnaK, are all nonessential. To demonstrate the cellular engineering approach to improve protein purification, focus was given to the four proteins with the larger molecular weights. Mutants lacking *dnaK* do survive, but have been reported to grow poorly.[121] Thus, a triple knockout mutant that does not express *cyoA*, *yfbG*, and *adhP* was constructed.

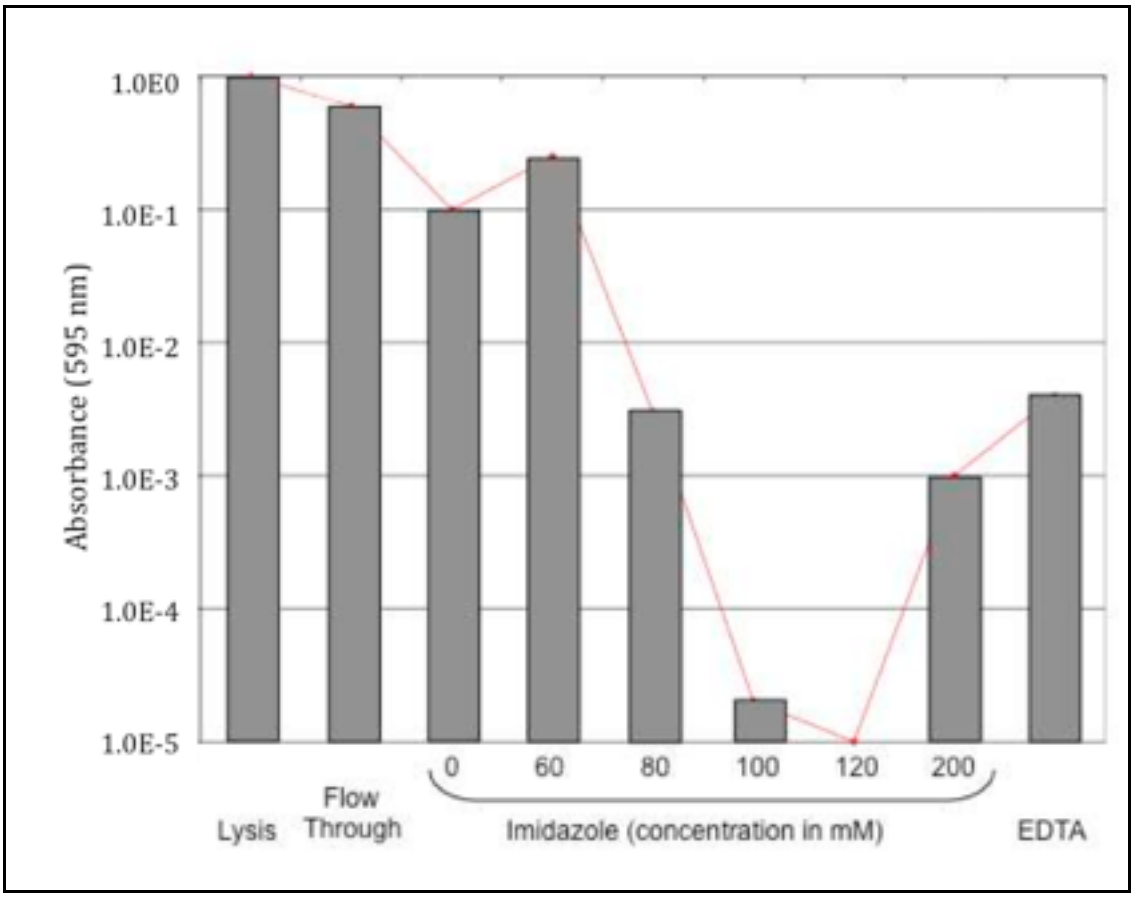

Figure 3.1. Normalized protein concentration of each fraction during nickel IMAC purification. Clarified E. coli lysate was passed through an immobilized nickel column and eluted by step-gradient increases of imidazole. A "low-background region" was observed in the 120 mM elution fraction, in which the smallest amount of host protein eluted from the column relative to the loading material, showing here on a logarithmic scale. This region is ideal for directing the elution of a theoretical target protein.

Table 3.1. Proteins identified to elute in the "low-background" region of nickel IMAC. Proteins were identified by LC-MS/MS and their essentiality to the cell is referenced as 'E' (essential) or 'N' (non-essential).

No.	Gene	E/N	Length (AA's)	p*I* (Calc.)	MW (Calc.)	Annotations
1	*dnaK*	N	638	4.83	68,984	Replicative DNA helicase / nucleotide and protein binding
2	*yfbG*	N	660	6.39	74,158	Nucleoside-diphosphate-sugar epimerases / fused UDP-L-Ara4N formyltransferase and UDP-GlcA C-4'-decarboxylase / surface histidine protein.
3	*adhP*	N	336	5.94	35,249	alcohol dehydrogenase / Zn-dependent alcohol dehydrogenases / metal ion binding
4	*cyoA*	N	315	6.76	34,780	Ubiquinol oxidase subunit 2/ Native metal-binding proteins(copper ion binding)
5	*rplB*	E	273	10.93	29,729	50S ribosomal protein L2 / RNA binding
6	*slyD*	N	196	4.92	20,722	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD / metal ion binding
7	*nagD*	N	250	5.18	27,032	UMP phosphatase /N-acetylglucosamine metabolism / carbohydrate metabolism
8	*ahpC*	N	187	5.03	20,630	alkyl hydroperoxide reductase C22 subunit / ATP-binding
9	*rpsG*	E	156	10.3	17,473	30S ribosomal protein S7 / RNA binding
10	*rplO*	E	144	11.18	14,835	50S ribosomal protein L15 / RNA binding
11	*rpsE*	E	167	10.11	17,472	30S ribosomal protein S5 / RNA binding
12	*rplM*	E	167	9.68	18,645	50S ribosomal subunit protein L13 / protein biosynthesis
13	*Fur*	N	148	5.68	16,664	ferric uptake regulator / COG0735: Fe2+/Zn2+ uptake regulation proteins / metal ion binding
14	--	--	183	4.68	20,269	Putative uncharacterized protein ECs2542
15	*rplJ*	E	165	9.04	17,711	50S ribosomal subunit protein L10 / RNA binding
16	*rpsL*	E	124	11.01	13,765	30S ribosomal protein S12 / RNA binding
17	*Hns*	N	137	5.44	15,409	DNA-binding protein H-NS / global DNA-binding transcriptional dual regulator H-NS / DNA binding and transcription
18	*rplL*	E	121	4.6	12,164	50S ribosomal protein L7/L12 / protein biosynthesis

Knockout strains were prepared using Red/ET homologous recombination and confirmed by southern blotting and SDS-PAGE, shown below in **Figure 3.2**. Probes for the three host proteins were used to detect the presence of their DNA sequences in the host genome, forming bands when hybridization occurs. The absence of bands indicates successful knockout of the sequence, which is shown in the first panel. Lysate from each mutant was subjected to IMAC again and their "low-background" regions were compared in the second panel. Observing the lanes from left to right, three distinct bands are consecutively removed from the protein pool. The gel lane of triple mutant appears significantly clearer than its wild-type parent strain.

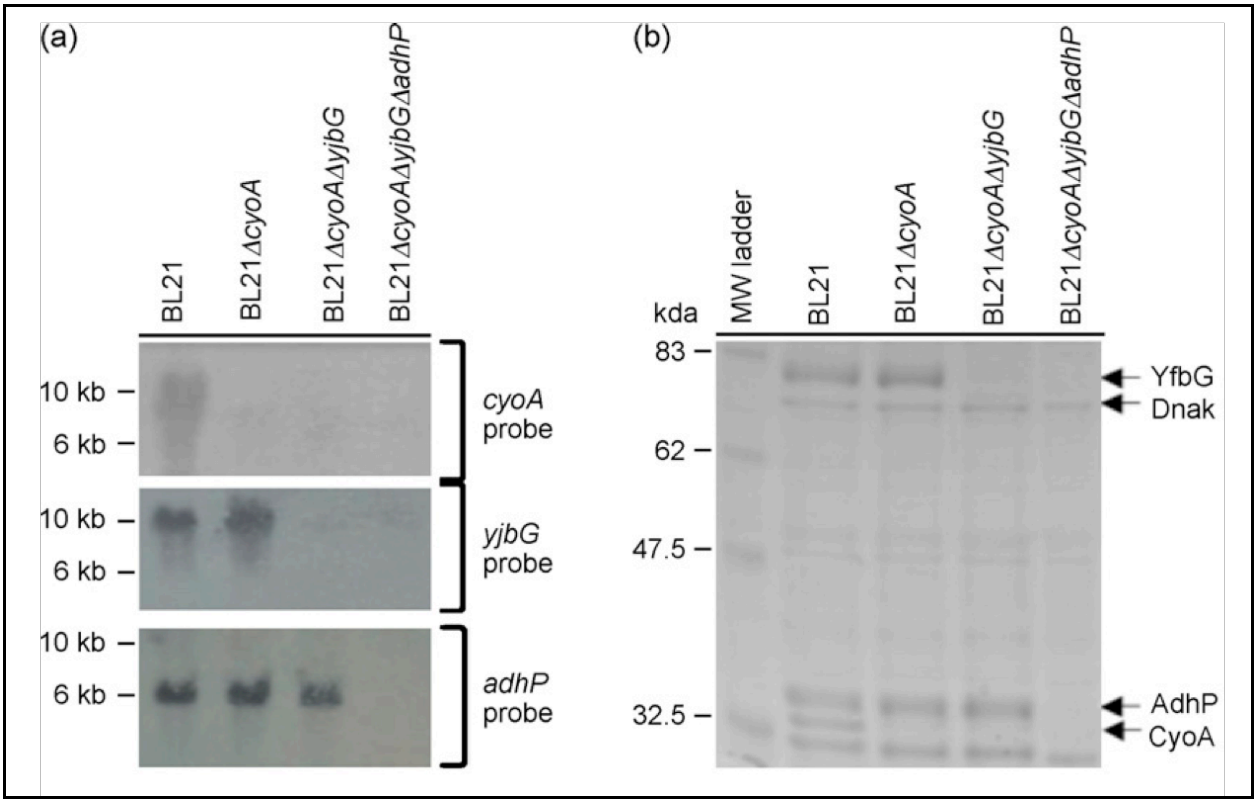

Figure 3.2. Confirmation of knockout strains.
(A) Confirmation by Southern blot is apparent when complimentary oligonucleotide probes do not interact with their corresponding genes. When the gene is present in the harvested DNA, bands appear in the gel. (B) Confirmation by SDS-PAGE.

An important feature of an engineered host for protein production is that it be able to grow at rates comparable to that of the strain from which it was derived. Additionally, the expression of exogenous protein should be comparable. The growth rates and protein expression of plasmid-encoded GFP were compared between the triple mutant and parent strains (**Figure 3.3**). In M9 minimal media supplemented with 10 g/L glucose, both strains grew at the same rate (shown in panel A). Strains were also assayed for GFP expression by specific fluorescence intensity, where every sample was diluted to an A_{660} of 0.1 before measurement. The triple knockout mutant displayed better GFP expression than the wild-type strain (panel B). The data indicates that removal of the three host proteins was without any ill effects.

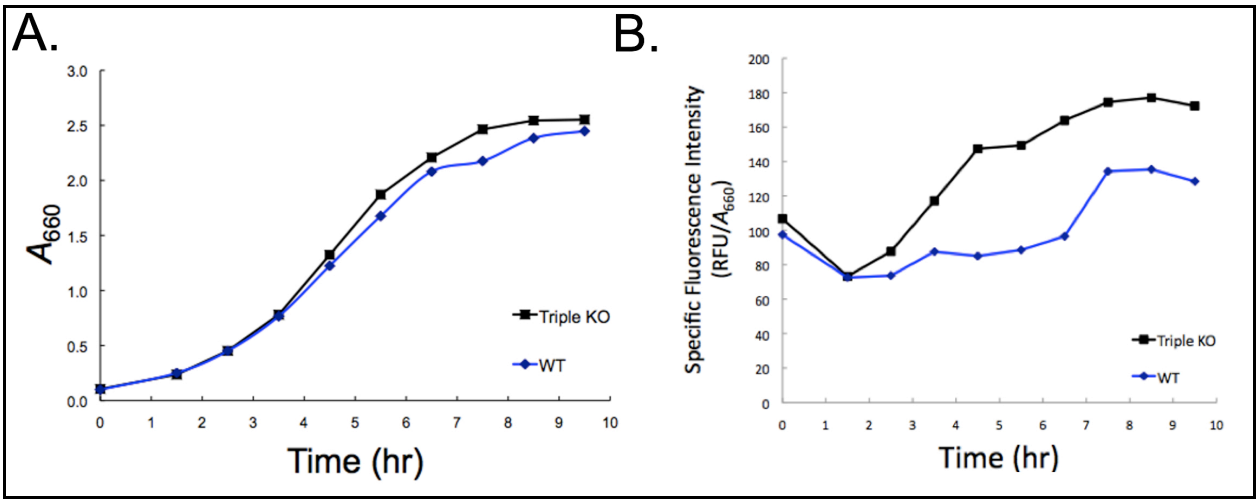

Figure 3.3. Growth rate and recombinant GFP expression of wild-type and triple mutant strains. (A) The time course of growth for the parent (WT) and triple KO strains. (B) The time-dependent expression of plasmid-encoded GFP, normalized to the corresponding A660 value.

To illustrate the feasibility of directing the elution of a target protein to a region of few natural contaminants, native GFPuv was selected as a model protein. The GFPuv affinity for Ni(II)-NTA is low and thus will require fusion to an affinity tag to direct its elution to 120 mM elution window. The elution behavior of native GFPuv and his_6-GFPuv were compared. His_6 is the most often used affinity tag for purification of proteins on immobilized nickel. The affinity tag WHHHPH was also compared, which is a tag slightly weaker than his_6 that was identified in the phage display library experiment previously mentioned. The chromatogram in **Figure 3.4** shows how his_6 directed GFPuv to elute in the now enhanced "low-background" region of the IMAC elution profile. **Figure 3.5** shows WHHHPH-GFPuv breakthrough during loading and elution at 80 mM imidazole.

The results have demonstrated that significant benefits could be gained in downstream purification by applying information derived from proteomic-based tools to implement rational changes to the host genome without detrimental effects. Here, the strategy was to create a clean, discrete elution window in which a recombinant protein could be directed using his_6 or other tags. The multi-step elution strategy, however, was operationally cumbersome. The next steps were to designed to simplify the use of IMAC and to generalize the strategy by identifying all IMAC-relevant proteins under conditions that allow a single bind-and-elute step. Further, I wanted to observe how the relative host protein concentrations changed as target protein expression fluctuated. The remainder of this chapter addresses those issues and proposes an unconventional strategy to mitigate target protein contamination.

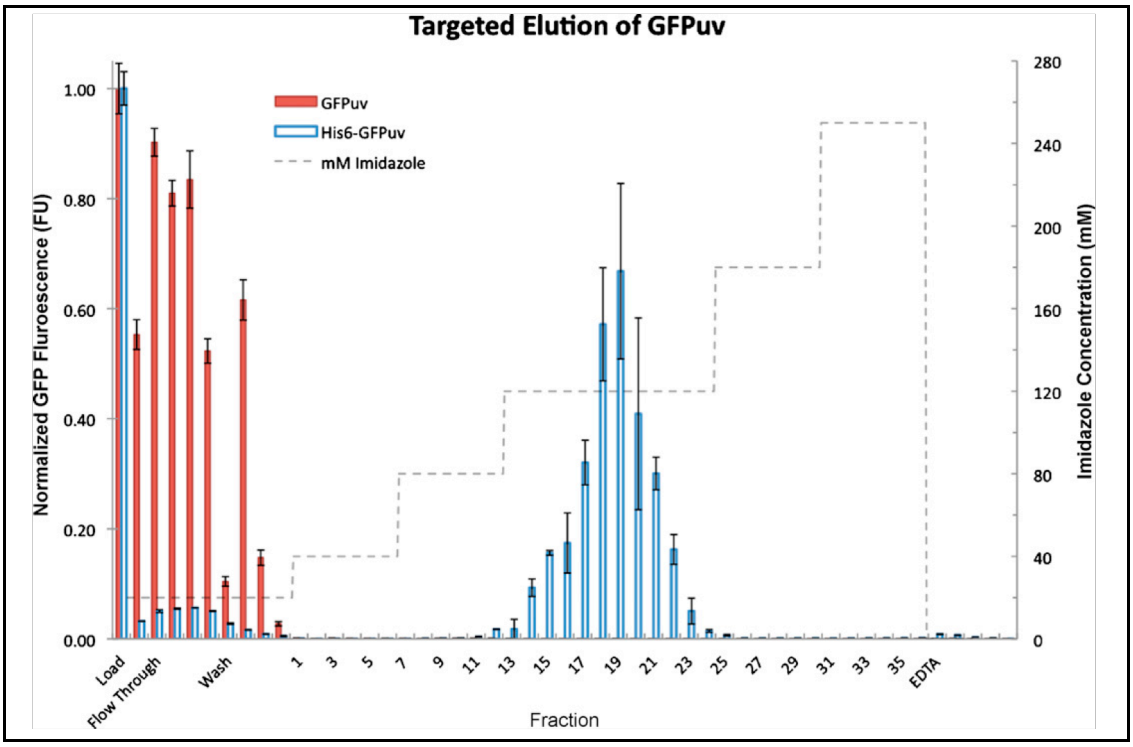

Figure 3.4. Directed elution of GFPuv using a his_6 peptide tag**.**
Complex lysate with either native (red) or his-tagged GPFuv (blue) were applied to nickel IMAC column and eluted by step-wise increases of imidazole.

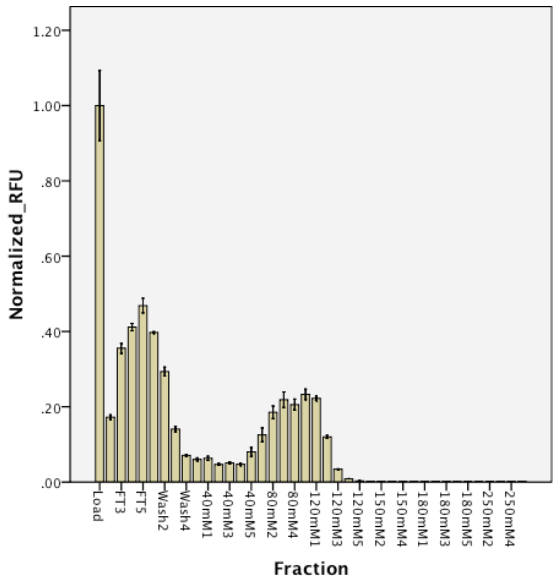

Figure 3.5. Directed elution of GFPuv using a novel affinity tag, WHHHPH.
Fractions labeled along the x-axis indicates that the WHHHPH conferred GFP elution at 80 mM.

3.3 IDENTIFICATION OF NATIVE ESCHERICHIA COLI PROTEINS THAT BIND TO IMAC UNDER HIGH IMIDAZOLE CONDITIONS AND USE OF 2D-DIGE TO EVALUATE CONTAMINATION POOLS WITH RESPECT TO RECOMBINANT PROTEIN EXPRESSION LEVEL

To assess the effect of increasing the imidazole loading concentration, I compared breakthrough curves of *E. coli* soluble lysates containing his_6-GFP in 20 mM and 50 mM imidazole IMAC buffers and observed the nickel-binding proteins under each condition (**Figure 3.6)**. For the breakthroughs, which illustrate the presence of GFP exiting the column as it reaches capacity (**Figure 3.6A**), exit GFP fluorescence was measured in triplicate and normalized to the input fluorescence corresponding to the soluble extract (RFU_{out}/RFU_{in}). Elution volumes for 10% GFP breakthrough were 17.2 and 24.6 ml for 50 mM and 20 mM imidazole, respectively. Column capacity expectedly diminished with higher imidazole loading. At 10% target breakthrough, total column capacity was 38 mg/ml and 85 mg/ml for the 50 mM and 20 mM imidazole runs, respectively. GFP yield was 98-99% in both cases. IMAC purifications of each imidazole concentration were carried out, with data summarized in **Table 3.2** (data sets correspond to 10% GFP breakthrough using the curve areas in **Figure 3.6A** and GFP purity is estimated by densitometry from the SDS-PAGE gel **Figure 3.6B**).

Table 3.2. Purification table comparing use of 20 mM and 50 mM imidazole in nickel IMAC.

Purification step	Total Protein (mg)	GFP (mg)	GFP Yield (%)	GFP Purity (%)	Purification Fold
20 mM imidazole					
Soluble extract	350	53	100	15	1
HisTrap FF	85	52	99	61	4.1
50 mM imidazole					
Soluble extract	246	37	100	15	1
HisTrap FF	38	36	98	95	6.3

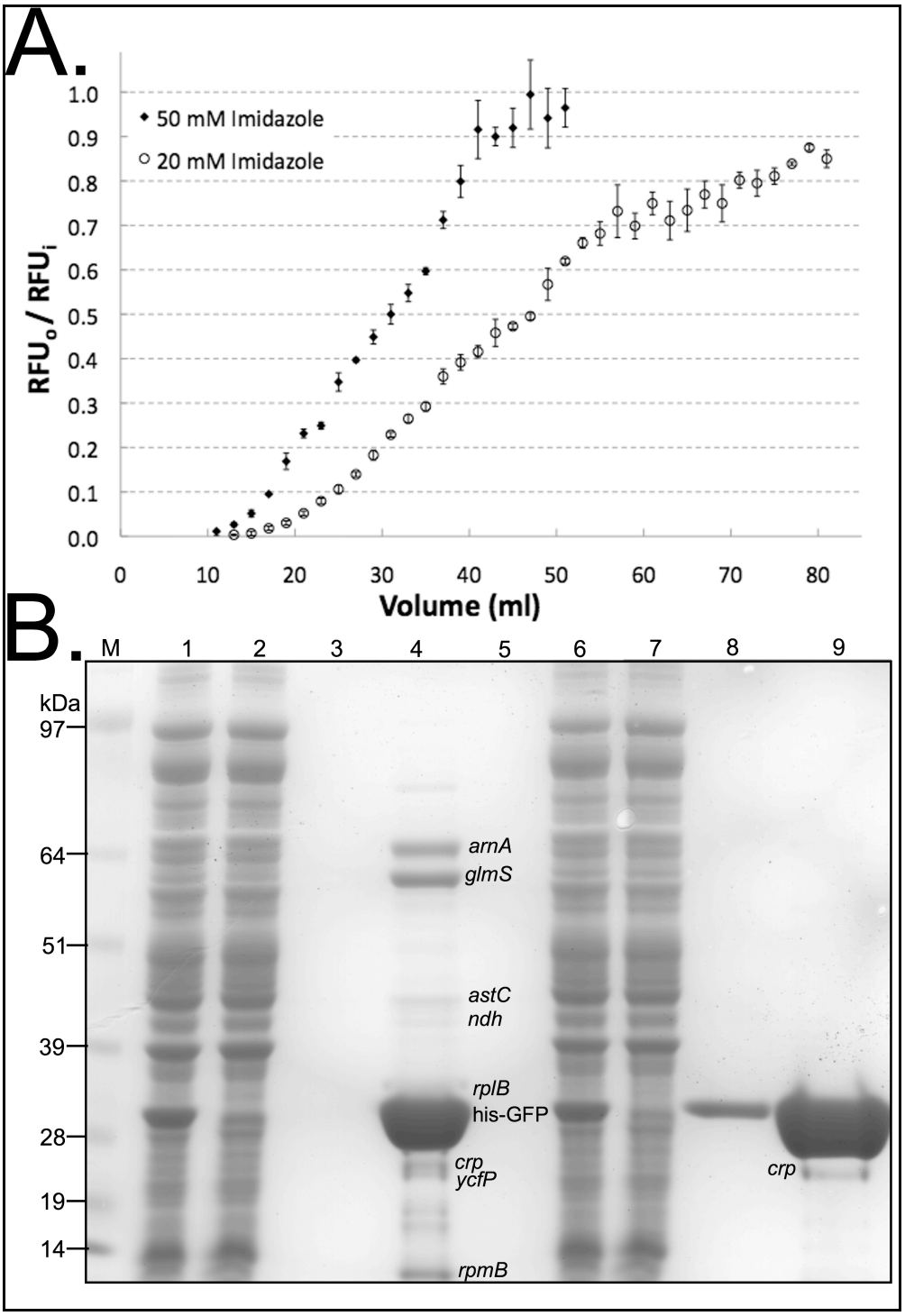

Figure 3.6. His_6-GFP purifications using 20 mM and 50 mM imidazole loading conditions. (A) Breakthrough curves of complex lysate containing his6-GFP in loading buffer with either 20 or 50 mM imidazole. (B) SDS-PAGE gel of comparing the bound protein pools at 10% GFP breakthrough. Lanes are: molecular weight ladder (M); 20 mM imidazole load, flow through, wash, bound proteins (1-4, repsectively); 50 mM imidazole load, flow through, wash, bound protein (6-9, respectively).

In the purification runs, columns were loaded with cell extracts in either 20 mM or 50 mM imidazole to 10% GFP breakthrough, equilibrated and stripped with EDTA. Proteins were concentrated, separated by SDS-PAGE, and identified by MALDI-TOF-MS as indicated by gene name in **Figure 3.6B** (Note: extract and flow-through lanes were loaded with 25 μg protein. Samples from column wash fractions were excessively dilute, containing approximately 2 μg protein in the lane for the 50 mM case. Bound protein lanes were overloaded with 50 μg protein). As evident from the gel, elevated imidazole concentration translated to higher nickel binding selectivity, noting the estimated GFP purity of 95% using 50 mM imidazole versus 61% for 20 mM. A minute amount of his_6-GFP was also found in the wash fraction at 50 mM.

For the 20 mM case, several proteins significantly compete with his-tagged proteins; among them include bifunctional polymyxin resistance protein (*arnA*, previously known as *yfbG*), glucosamine-fructose-6-phosphate aminotransferase (*glmS*), catabolite gene activator protein CAP otherwise known as catabolite repressor protein (*crp*), and 50S ribosomal subunit L28 (*rpmB*). By densitometry, gene products from *arnA* and *glmS* accounted for nearly half of the non-GFP bound protein mass, whereas they do not bind at 50 mM. In the latter case, while small proteins were observed in trace amounts on the gel, only *crp* was present in amounts that allowed identification by MALDI-TOF-MS.

DIGE was used to observe changes in relative protein concentration in soluble extracts and IMAC-relevant proteins as GFP expression level was altered. Here, IMAC with 50 mM imidazole loading was used to generate the proteins pools. Comparisons were made among strains not transformed with the his_6-GFP plasmid, here denoted wild type (WT), and those expressing GFP either with or without IPTG induction (induced/uninduced). **Table 3.3** illustrates the experimental design employed to compare protein groups.

Table 3.3. DIGE experimental design comparing soluble lysates and IMAC bound proteins.

	Cy3	*Cy5*
Lysate Gels		
A	WT	Uninduced
B	Uninduced	Induced
C	Induced	WT
IMAC Gels		
A	WT	Uninduced
B	Uninduced	Induced
C	Induced	WT

A brief note is needed on the differential expression of GFP between induced and uninduced cases. BL21 (DE3) does not control expression as tightly as other expression vectors such as BL21 or BL21 (DE3)pLysS.[86] For more direct comparison, it was decided that one strain be used to produce all samples. Without induction, GFP was expressed to about 10% soluble protein as determined by densitometry of 1D SDS-PAGE gels (not shown).

Gels were loaded with CyDye-labeled protein pools as indicated in **Table 3.3** and spot determination and quantitation were resolved with DECODON Delta2D software. Images captured for each dye are overlayed so protein spots with differing abundances are illustrated by a blue or orange shade accordingly, whereas proteins with similar abundances appear black. Gels are shown with associated graphs that correlate these individual protein abundances in each sample, giving a quick indication of likeness between protein pools.

Qualitative and quantitative analyses of the DIGE comparing lysate pools, based on 250-350 protein spots, revealed that they are considerably similar (**Figure 3.7**). Expectedly, plasmid-encoded his_6-GFP isoforms and β-lactamase (*blaT*) stand out in the gels, especially those that include the WT pool. In the WT vs. uninduced comparison (**Figure 3.7A**, where relative abundance of spots in WT sample appears orange and uninduced appear blue), GFP and β-

lactamase were among only 12 resolved spots with average protein volumes that differed by more than 1.5-fold. Two spots appear to be cleaved fragments of outer membrane protein A (*ompA*) and another was DNA protection during starvation protein (*dps*). Other differing spots identified were: transcriptional regulatory protein OmpR, NADP-dependent L-serine/L-allo-threonine dehydrogenase (*ydfG*), molybdenum cofactor biosynthesis protein B (*moaB*), and L-fucose mutarotase (*fucU*).

A comparable number of proteins differed in the uninduced (orange) vs. induced (blue) DIGE (**Figure 3.7B**). Among them include UPF0337 protein (*yjbJ*), galactose-6-phosphate isomerase subunit A (*lacA*) and glutamine-binding periplasmic protein (*glnH*). Elevated GFP expression in the induced sample and the emergence of suspected cleaved or incompletely transcribed GFP fragments accounted for other differing spots (marked by unlabeled arrows), which are more explicit in the third gel (**Figure 3.7C**) comparing induced (blue) and WT (orange) groups. Phosphoglyceromutase (*gpmA*) was distinctly over-expressed in the induced pool. While the quantitative comparison of spot volume indicates the induced pool is more similar to the WT than uninduced, qualitatively that is not apparent from the gel images. This finding is likely in part due to the exclusion of reciprocally labeled spots with exceedingly variable calculated volumes, resulting in a tighter correlation. In general, results indicating high similarity among groups were expected since they were produced in identical culture conditions and that DIGE of soluble lysate is primarily a comparison of genome-encoded proteins not relevant to recombinant protein production or stress response.

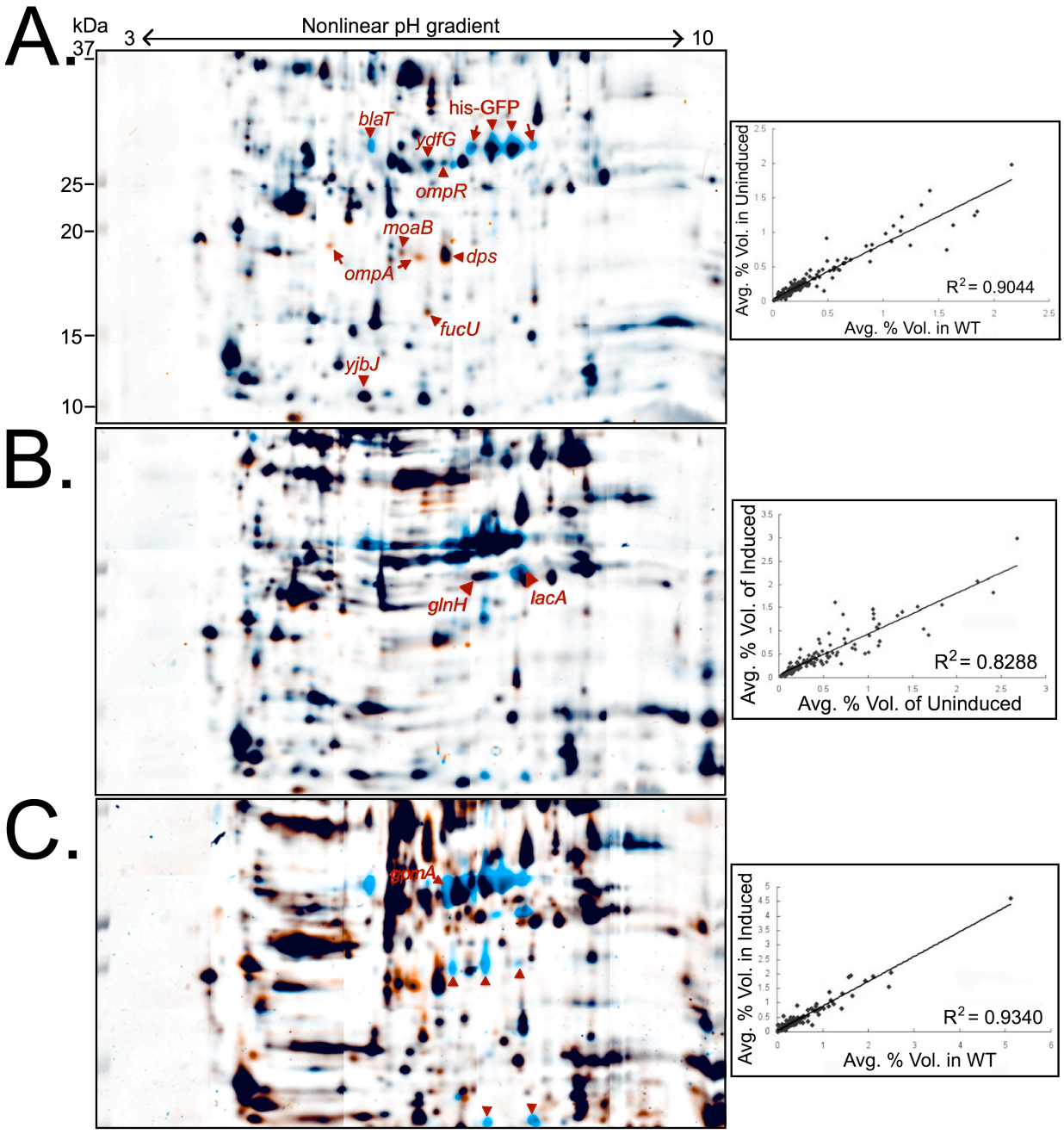

Figure 3.7. Two-dimensional fluorescence difference gel electrophoresis of soluble extracts. Panels at the right correlate individual spot volumes of corresponding samples, offering a fast qualitative comparison. Black gel spots indicate equivalent volumes in the samples. (A) Wild-type (orange) and Uninduced (blue). (B) Uninduced (orange) and Induced (blue). (C) Wild-type (orange) and Induced (blue).

The DIGE analysis of IMAC-relevant proteins at varying GFP expression levels is shown in **Figure 3.8**. The gel of WT (orange) and uninduced (blue) pools effectively shows the contrasting quantity of IMAC proteins, bound under 50 mM imidazole loading, as expression changes (**Figure 3.8A**). In the WT group are common IMAC contaminants such as *crp*, host factor-I protein (*hfq*), and ferric uptake regulator (*fur*). As expected, FKBP-type peptidyl-prolyl cis-trans isomerases (PPIases) were present, which in UniProt were attributed to SlyD. While these appear to be different PPIases, both spots yielded peptide-level evidence of domains homologous with SlyD**.** All other nickel-binding proteins identified also contain histidine clusters: ribosomal small subunit pseudouridine synthase A (*rsuA*), 3-ketoacyl-CoA thiolase (*fadA*), and universal stress protein G (*uspG*). One important note, this gel displays black-shaded areas of spot overlap that suggest presence of his_6-GFP in the WT pool. It is believed these areas are a result of signal bleed-through due to slight overlapping of fluorophore spectra as no GFP is present in standalone gels of WT protein stained with coomassie (not shown). Further, what appear as evident "streaks" above the GFP spots are likely artifactual or free fluorescent label in the gel. This is based on the observation that after overnight SYPRO Ruby staining, no protein spots were discernable with the ProPic II imager; and further, gel plugs cut from those regions did not yield spectra with MALDI-TOF/TOF-MS.

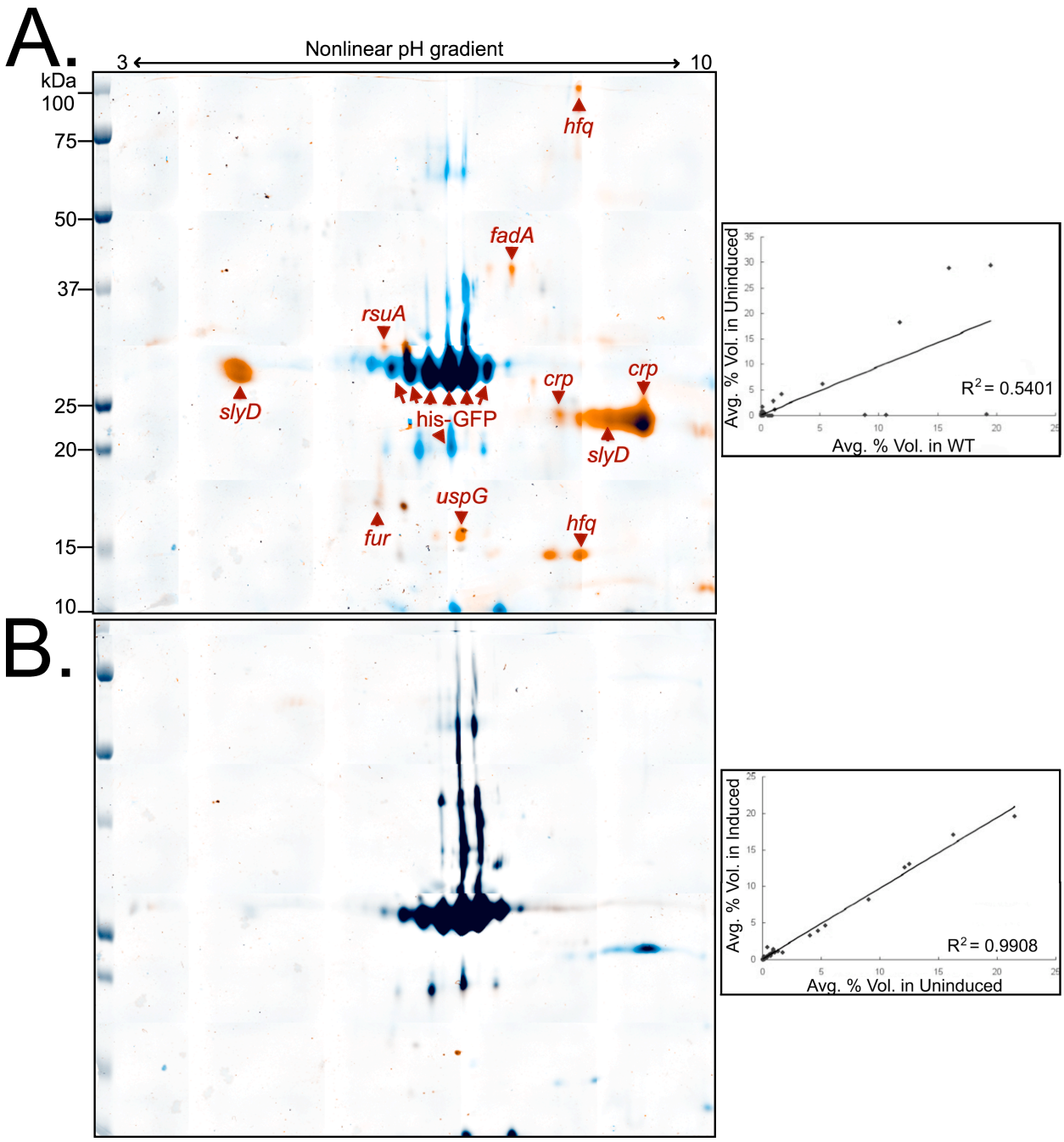

Figure 3.8. Two-dimensional fluorescence difference gel electrophoresis of IMAC proteins. Panels at the right correlate individual spot volumes of corresponding samples, offering a fast qualitative comparison. Black gel spots indicate equivalent volumes in the samples. (A) Wild-type (orange) and Uninduced (blue). (B) Uninduced and Induced (blue). The evident streaks in panel B are due to residual fluorescent dye in the gel and does not represent the presence of protein (see text for details).

The uninduced pool essentially appears as a two-dimensional, fluorescent-labeled representation of Lane 9 in **Figure 3.6B** as it almost entirely consists of his_6-GFP and *crp*. This is more clearly illustrated in **Figure 3.8B**. The quantitative results from its comparison with the induced pool (blue) show virtually no differences in average spot volume. While it is expected that these groups be mostly comparable, this degree of equivalence is perhaps artificially high due to the mentioned overlapping fluorophore spectra. This idea is based on more distinctly varied levels of GFP expression determined by densitometry on SDS-PAGE gels. The DIGE gel comparing induced and WT IMAC samples was omitted since it did not noticeably differ from **Figure 3.8A**. Further, one suspected GFP fragment, also present in the GFP-containing lysate pools, was confirmed by MALDI-TOF/TOF-MS. A list of all identified proteins is presented in **Table 3.4**. High confidence Mascot scores determined protein identities. Where available, fold expression values are given for spots evaluated by DIGE. For clarity, GFP isoforms are omitted from the list, though all corresponding labeled gel spots met the minimum significant confidence Mascot score.

The effect that target expression level has on contamination, observable with DIGE, was illustrated with another model protein. This work was carried out by collaborators in Arkansas, Robert Beitle and colleagues, and is briefly presented here since it neatly corroborates the findings just discussed. A recent construct by Hatefi et al. used for gene therapy applications is expressed at only microgram levels in *E. coli* as a result of its composition and biophysical properties.[122] This construct, a N-terminal extension of fibroblast growth factor 2 (FGF2) contains a sequence of lysine and histidine necessary for therapeutic DNA condensation and endosomal uptake, and as such, the rather high histidine content of $(KKKHHHHKKK)_6$-FGF2 permits recovery via metal affinity. The protein fusion was expressed in *E. coli* BL21 growing

aerobically using an exponential fed batch strategy in an Applikon Bioreactor (Foster City, CA, USA). Base media was minimal, consisting of 10 g/L glucose, 15 g/L K_2HPO_4, 7.5 g/L KH_2PO_4, 2 g/L citric acid, 2.5 g/L $(NH_4)_2SO_4$, 2 g/L $MgSO_4$-$7H_2O$ and 1 mL of trace element solution per liter. Trace elements per liter of 1 M hydrochloric acid were 2.8 g $FeSO_4$-$7H_2O$, 2 g $MnCl_2$-$4H_2O$, 2.8 g $CoCl_2$-$7H_2O$, 1.5 g $CaCl_2$-$2H_2O$, 0.2 g $CuCl_2$-$2H_2O$ and 0.3 g $ZnSO_4$-$7H_2O$. An exponential feeding strategy was designed to supplement a solution containing 700 g/L of glucose, 20 g/L of $MgSO_4$-$7H_2O$, 1 ml of trace element solution, 5 g/L yeast extract, and 35 ml/L of 2x LB. Note also that antifoaming agent and arabinose (inducer) were appropriately added during cultivation. This protein fusion itself was further endowed with a C-terminal his_6 tag and purified by IMAC from complex cell extract, shown in **Figure 3.9**. Sequence data from mass spectroscopy (not shown) confirmed bands corresponding to a FKBP-type PPIase (again denoted SlyD) and $(KKKHHHHKKK)_6$-FGF2-his_6. As such, both proteins were present in the pool fraction represented by Lane 1. Pure SlyD is shown in Lane 2. The results indicated that without care, the majority of the elution pool could in fact be comprised of the *E. coli* protein and not the target.

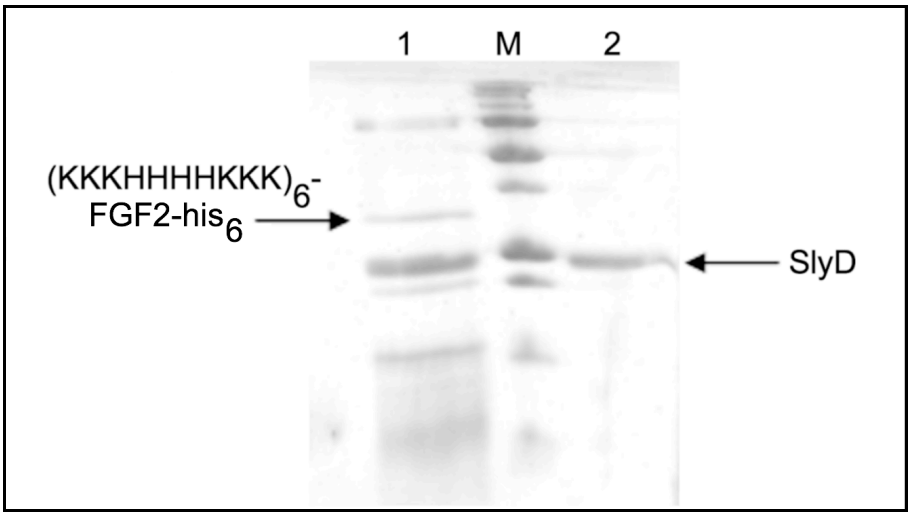

Figure 3.9. SDS-PAGE gel comparing a pooled aliquot of IMAC-purified $(KKKHHHHHKKK)_6$-FGF2-his_6 and pure SlyD.
Lanes are: 1, purified protein sample; M, molecular weight ladder (sizes are arbitrary for this comparison and are not show); 2, pure SlyD.

Table 3.4. List of proteins identified in this study relevant to IMAC.

	Gene	*Protein Name*	*UniProtKB Accession No.*	*Sequence Coverage*	*Mascot Score*	*Fold Expression*
Fig. 3.6B	*arnA*	Bifunctional polymyxin resistance protein	**P77398**	23	122	
	glmS	Glucosamine-fructose-6-phosphate aminotransferase	**P17169**	45	220	
	astC	Succinylornithine transaminase	**P77581**	31	98	
	ndh	NADH dehydrogenase	**P00393**	16	61	
	rplB	50S ribosomal protein L2	**P60422**	36	127	
	crp	Catabolite gene activator	**P0ACJ8**	32	93	
	ycfP	UPF0227 protein	**P0A8E1**	35	69	
	rpmB	50S ribosomal protein L28	**P0A7M2**	44	62	
Fig. 3.7A	*ompR*	Transcriptional regulatory protein	**P0AA16**	33	99	2.0*
	ompA	Outer membrane protein A	**P0A910**	25[a] 44[b]	91[a] 202[b]	-3.7[a] -3.1[b], -3.3*
	dps	DNA protection during starvation protein	**P0ABT2**	72	552	-2.1
	moaB	Molybdenum cofactor biosynthesis protein B	**P0AEZ9**	40	72	-1.8, -1.8*
	blaT	Beta-lactamase TEM	**P62593**	41	220	---
	ydfG	NADP-dependent 3-hydroxy acid dehydrogenase	**P39831**	29	156	---
	fucU	L-fucose mutarotase	**P0AEN8**	26	59	---
B	*lacA*	Galactoside O-acetyltransferase	**P07464**	17	76	1.9
	yjbJ	UPF0337 protein	**P68206**	68	77	-1.9, 4.1*
	glnH	Glutamine-binding periplasmic protein	**P0AEQ3**	26	131	---

Table 3.4. (continued).

C	*gpmA*	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	**P62707**	35	302	---
Fig. 3.8A	*slyD*	FKBP-type peptidyl-prolyl cis-trans isomerase	**P0A9K9**	47[a] 28[b]	436[a] 97[b]	-78[a] -110[b]
	hfq	Host factor-I protein	**P0A6X3**	38[h] 24[s]	222[h] 119[s]	--- -96[s]
	crp	Catabolite gene activator	**P0ACJ8**	53[a] 50[b]	299[a] 126[b]	--- -79[b]
	fadA	3-ketoacyl-CoA thiolase	**P21151**	39	391	---
	rsuA	16S pseudouridine 516 synthase	**P0AA43**	34	172	---
	fur	Ferric uptake regulator	**P0A9A9**	32	193	---
	uspG	Universal stress protein G	**P39177**	52	77	---

***** Denotes fold expression of redundant spot found in **Figure 3.2C**.
[a] Denotes an *acidic* isoform or fragment.
[b] Denotes a *basic* isoform or fragment.
[h] Indicates the relatively *heavy* Hfq spot.
[s] Indicates the relatively small Hfq spot

3.4 DISCUSSION

The IMAC strategy was simplified to a single bind-and-elute step to capture and identify all contaminants with binding retained above a certain imidazole threshold. This eliminates the need to consider the precise binding strengths of various his-tagged proteins, e.g. the pH or imidazole concentration at which they elute in traditional bind and gradient-elute schemes. Breakthrough curves (**Figure 3.6A**) show that use of 50 mM imidazole lowers the total protein binding capacity relative to 20 mM, in this case from about 80 to 38 mg/ml (**Table 3.2**). However upon loading to 10% target breakthrough, his_6-GFP binding was only reduced by 30% (52 mg at 20 mM imidazole versus 36 mg at 50 mM) and consequentially the purity was improved.

Several of the proteins identified are familiar IMAC contaminants (**Fig. 3.6B**). SlyD, Fur, and ArnA were mentioned in the discussion on the prior IMAC work, in which *arnA* was knocked out without ill effects to cell growth or recombinant protein expression.[123] In addition to these proteins, host factor-I protein (HFQ), CAP, and GlmS were previously discussed by Bolanos-Garcia & Davies.[103] Of these contaminants, the most substantial were ArnA and GlmS. GlmS catalyzes the synthesis of D-glucosamine-6-phosphate, a precursor of amino sugars *N*-acetyl-D-glucosamine and D-glucosamine, which are essential building blocks of the cell envelope. Strains deficient in *glmS* require sufficient exogenous sources of these amino sugars to maintain cell viability.[124, 125] GlmS monomers contain 24 histidine residues, making it an unfavorable candidate for site-directed mutagenesis as well; therefore, use of higher imidazole concentration is the best current approach to eliminate its presence in target protein pools.

DIGE analyses of the lysate pools primarily attest to their similarity, irrespective of recombinant protein concentration. This was in large part due to analysis of only the soluble extract of the cell lysate, an approach favored for two reasons. Primarily, my focus was to identify IMAC-relevant proteins encountered during purification of a target with ideal, soluble expression. I also expected to minimize the capture of some stress-response proteins and chaperones such as DnaK/J, ClpB, IbpA/B, GrpE, GroEL/S, and SlyD. Indeed they were largely unseen in the DIGE, whereas with an approach that includes analysis of insoluble protein bodies, I would have expected higher presence of these proteins and differing expression with respect to IPTG induction.[91] Plasmid-encoded proteins aside, the proteins that differed most across the samples have widely varying functions and appear unrelated.

One point to note is the prominence of SlyD at low recombinant protein expression (**Figure 3.8A**). Certainly SlyD is a troublesome contaminant when steps are taken to resolubilize inclusion bodies to purify aggregated recombinant protein. Additionally, its presence in IMAC purified fractions increases as his-tagged protein expression is lessened due to competition for column binding sites. This was clearly observed upon examining the recovered pools of purified $(KKKHHHHKKK)_6$-FGF2-his_6, where SlyD and the target are difficult to resolve (**Figure 3.9**). The key issue raised in this contribution by Beitle and colleagues is that under such circumstances, owing to the close molecular weight of this example, SlyD could be mistaken for the true target product because of reactivity with anti-his tag antibodies used in Western blot analysis. Similar issues could be encountered with other desired recombinant proteins of similar molecular weight tagged with enhanced histidine content, *viz*. when SlyD displays similar elution and migration behavior, the possibility of providing false specific activity values exist because the prominent band is thought to be the true material both due to biological activity (assay) and

cross reactivity (Western). Therefore, the absence of spectroscopy data combined with both direct observation of the gel (the figure is an enlarged photograph of a small format gel) and nonspecific anti-his Western data would tempt one to identify the faster migrating band as the target. For the case of $(KKKHHHHKKK)_6$-FGF2-his_6 this incorrect migration would be rationalized based on the significant charge localization of sixty combined histidine and lysine residues, with purity and activity determination severely compromised.

In the GFP-containing pools where target protein expression levels were at least 10% soluble protein, however, many common IMAC contaminants are not relevant with 50 mM imidazole loading. CAP (the gene product of *crp*) is the only co-purifying protein in this IMAC scheme that could be confirmed by MALDI-TOF/TOF-MS. Once again, CAP is known to interact with immobilized nickel matrices, but this case has shown it to be the most tenaciously bound host protein.

This finding lends to the hypothesis that with all other things being equal (an ideal rarely if ever experienced in cell biology work), a host lacking CAP would allow IMAC purification of a sufficiently expressed recombinant target to homogeneity and would thus be strikingly advantageous for such an application. Understanding CAP's cellular role, a transcription activator that upon binding with cyclic AMP interacts with RNA polymerase and the promoter regions of over 100 genes, it comes as no surprise that mutants lacking CAP exhibit slow growth and greater sensitivity to stress conditions.[126, 127] It is hypothesized, however, that its nickel-binding affinity is due to only three histidine residues located in close proximity near the N-terminus at positions 17, 19 and 21. This convenience provided the fortuitous opportunity to test both stated hypotheses. And that is why CAP and the nature of its binding to familiar partners and IMAC columns is the subject of the following chapter.

4.0 RATIONAL MUTAGENESIS OF CATABOLITE GENE ACTIVATOR PROTEIN MITIGATES IMAC CONTAMINATION WHILE PRESERVING ITS CELLULAR FUNCTION

4.1 INTRODUCTION

As a facultative anaerobe, *E. coli* is capable of surviving on a wide variety of substrates with or without the presence of oxygen. It preferentially utilizes one carbon source until depletion before utilizing the next, doing so by repressing expression at other catabolic operons in a mechanism known as catabolite repression. In absence of the preferred carbon source glucose, high cytoplasmic levels of cyclic AMP (cAMP) promote allosteric binding to CAP, which regulates transcription at over 100 CAP-dependent promoters. The protein is a 47-kDa homodimer that binds to specific DNA regions, enhancing RNA polymerase (RNAP) binding and transcription. It was the first transcription activator to be isolated,[128, 129] to have its three-dimensional structure determined,[130] and has since been the focus of thorough biochemical, biophysical, genetic research as described in previous reviews.[131-133] It is known from my previous studies that CAP tenaciously binds to IMAC resins, and given its integral cellular function, it is an unsuitable candidate for genomic knockout. Indeed, Δ*crp* exhibits significantly slower growth[126] and greater sensitivity to stress conditions such as recombinant protein production.[127] To uncover an

alternative method to mitigate its IMAC contamination, we must first gain a deeper understanding of CAP at each distinct level of protein structure.

Upon allosteric binding of camp, CAP undergoes a conformational change and forms a complex with DNA at a specific region near CAP-dependent promoters. The crystallographic structure of the CAP-DNA complex determined by Schultz *et al.* is shown in **Figure 4.1** (protein databank entry 1cgp modified with space-filled binding determinants as presented by Busby and Ebright). [131, 134] The C-terminal domain of each CAP subunit contains a helix-turn-helix DNA binding motif that interacts with the major groove, bending the DNA (represented in red as a wire-surface) by about 90°. Histidine residues at positions 17, 19 and 21 are shown as light blue space-filled residues in **Figure 4.1A**. Their close proximity is likely related to their role in IMAC contamination.

At class I CAP-dependent promoters such as *lac*, transcription activation requires direct protein-protein interaction with the C-terminal domain (CTD) of the RNAP α subunit via residues 156-164 of the downstream CAP subunit known as "activating region 1" (AR1; multi-colored space-filled region in **Figure 4.1A** and indicated in the cartoon of Class I promoter interaction in **Figure 4.1B**).[135, 136] This interaction tethers RNAP to the DNA, allowing subsequent recognition of promoter elements by RNAP σ^{70} subunit and catalytic activity by β and β' subunits. Substitution of any residue within AR1, particularly Thr158, will reduce or eliminate activation without affecting CAP-DNA binding.[137, 138]

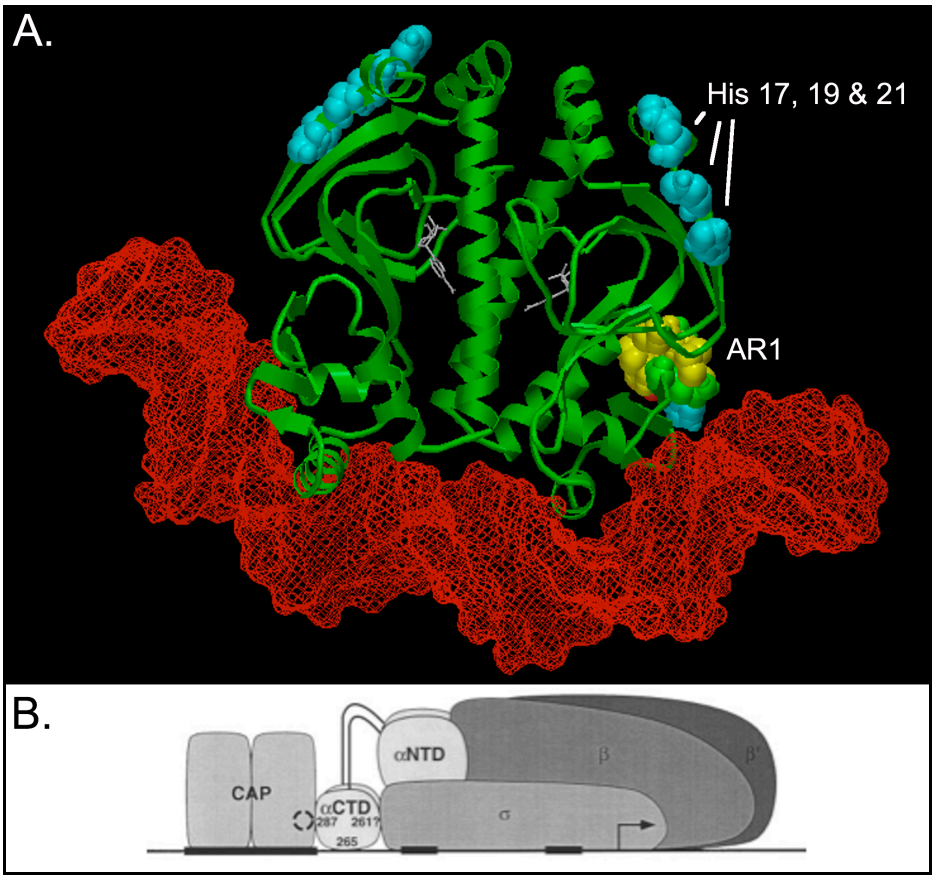

Figure 4.1. The CAP-DNA complex at Class I promoters.
(A) Crystallographic structure of the CAP-DNA complex, modified from Schultz et al. to highlight residues that interact with RNAP at Class I promoters. (B) Illustration of CAP interaction with RNAP at Class I promoters.

At class II CAP-dependent promoters like *galP1*, however, the DNA site for CAP and RNAP overlap (**Figure 4.2**, featuring protein databank entry 1cgp).[134, 139] This requires that AR1 of the *upstream* stream subunit of CAP interact with the CTD of αRNAP (**Figure 4.2B**).[140] In addition, CAP residues His19, His21, Glu96 and Lys101 of the downstream subunit interact with the NTD of αRNAP. These residues make up "activating region 2" (AR2) and are required for class II promoter activation. AR2 is not required for RNAP-DNA binding, but rather to increase the rate of RNAP-promoter isomerization of closed complex to open.[131] Mutational studies indicate that the most important structural feature of AR2 is a net-positive charge, which is sensible since the interacting residues of α-NTD (162-165) are negative.[141] Those studies also showed that alanine substitution at His17 (H17A) did not inhibit class II promoter activation and tyrosine substitution at His19 (H19Y) retained 40% activation, indicating that they could be opportune targets for substitution.[141]

Substitution of Lys52 by a neutral or negatively charged residue presents a third, non-native CAP-RNAP interaction and was found to substantially increase transcription activation at class II promoters but not at class I.[135, 142] Mutation of Lys52 to asparagine (K52N) strengthens CAP interaction with RNAP by 5-fold with respect to the wild-type.[141] Alanine substitution of nearby residues revealed that Glu58 plays a role in this region, dubbed "activating region 3" (AR3).[142] These two residues are visible in **Figure 4.2A** in the downstream CAP subunit. Their close proximity suggests that they form a salt bridge that neutralizes their opposing charges. Upon substitution of Lys52 to a neutrally charged residue, however, the resulting net negative charge in AR3 forms an energetically favorable interaction with the net positive residues 590-600 of RNAP σ^{70}.[143, 144] Their proximity is evident in **Figure 4.2B**. It is therefore believed that

an asparagine substitution at Lys52 would recover lost transcription activation at class II promoters that results from other substitutions, e.g. a tyrosine substitution at His19.[131]

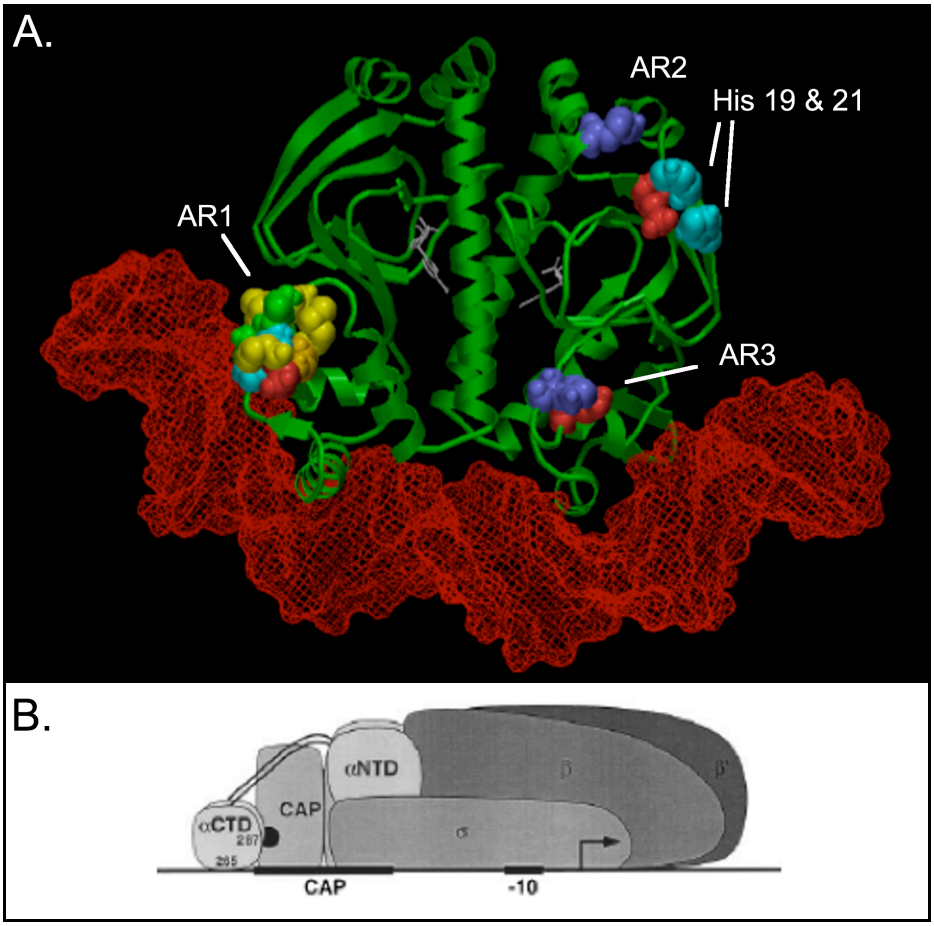

Figure 4.2. The CAP-DNA complex at Class II promoters.
(A) Crystallographic structure of the CAP-DNA complex, modified from Schultz et al. to highlight residues that interact with RNAP at Class II promoters. (B) Illustration of CAP interaction with RNAP at Class II promoters.

Based on the structure and mutation analysis studies and the current understanding of the interactions among CAP and its partners, I postulate that a modified CAP protein (H17A, H19Y, K52N) would have negligible nickel binding affinity and maintain adequate interaction with RNAP. The work presented in this chapter demonstrates that these proposed mutations to CAP indeed eliminate its IMAC binding under the conditions presented in the previous chapter. Implications from the results are discussed in the context of creating an improved host strain for generic recombinant protein expression and purification.

4.2 RESULTS AND DISCUSSION

To study amino acid substitutions in CAP and their influence on nickel-binding, the chromosomal *crp* gene was encoded on a pET21 plasmid, downstream of a T7 promoter and without use of the optional C-terminal his_6 tag, and expressed in BL21 (DE3). Plasmid expression of CAP was necessary not only to achieve high inducible cytosolic concentrations that allows easier characterization on columns, but also to serve as a template for generating mutants of CAP. While gene manipulation could be done with relative ease, intrinsic expression of the chromosomal CAP brought some complication since two variants of CAP were expressed in all mutant strains. This limitation could have been overcome with chromosomal knockout of *crp*, but given our familiarity with its IMAC binding it was thought satisfactory to compare the relative concentrations of bound CAP under identical conditions.

The first consideration was the choice of binding conditions. In the previous chapter, elevated imidazole concentration during loading was effective in preventing host protein binding except for CAP. But such stringent conditions are here not ideal to assess the binding of CAP or its altered forms. With a lower imidazole threshold, less material is needed to yield a representative pool of bound proteins, wild-type CAP included.

The gel in **Figure 4.3** displays the IMAC purification steps of wtCAP loaded with 20 mM imidazole (note that lanes correspond with loading material, insoluble pellet, flow-through, wash, and elution fractions). What is apparent is the lack of target protein found in the flow-through fraction and its heavy presence in the subsequent wash and elution fractions. It was found to be 78% pure from simply loading at 20 mM imidazole and stripping without the need for an affinity tail. This finding is a validation of wild-type CAP's proclivity to bind with nickel.

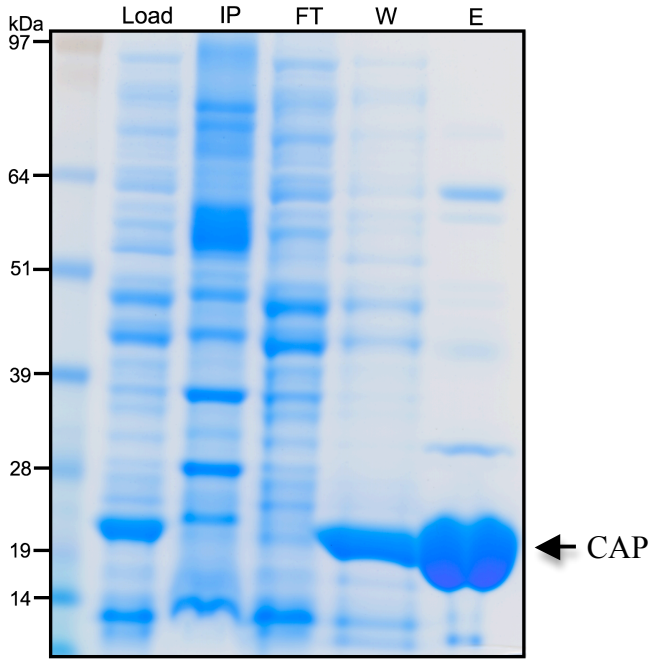

Figure 4.3. IMAC purification of wild-type CAP.
Lanes correspond to the complex lysate (Load), insoluble pellet (IP), flow through (FT), wash (W) and elution (E) fractions.

Figure 4.4A shows a gel produced in an identical manner, but from strain expression pH17A*crp*. Three details support the notion that alanine substitution of His17 results in diminished IMAC binding. First, a noticeable band corresponding to the mutant CAP now appears in the flow-through lane where it had not with the wild-type case, indicative of CAP breakthrough during loading. Given that the amount of CAP in the loading was identical in each case, CAP breakthrough is symptomatic of reduced affinity for nickel. This is akin to the breakthrough curves of his_6-GFP in buffers of different loading imidazole concentrations shown in the previous chapter (**Figure 3.6A**). Second, the mutant CAP more readily washes from the column, as the corresponding band was shown by densitometry to be twice as intense relative to the wild-type gel. Third, host proteins in the elution fraction have a higher relative concentration, as CAP was purified to just 47%. These findings agree with the near identical gel in **Figure 4.4B** that was produced from strain expression pH19Y*crp*. While this evidence shows that a single histidine substitution in CAP reduces its affinity to nickel, the amount of mutant CAP that ultimately remains bound through equilibration and elutes with other host proteins remains unclear.

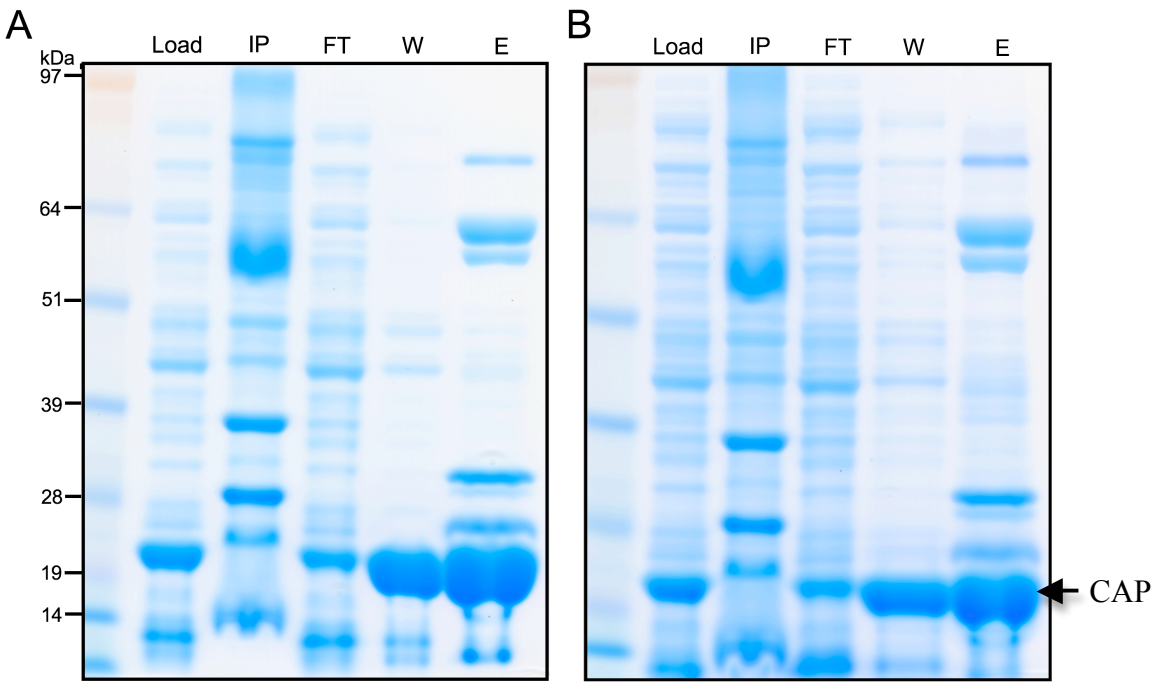

Figure 4.4. IMAC purifications of (A) H17A and (B) H19Y mutant CAP.
Lanes correspond to the complex lysate (Load), insoluble pellet (IP), flow through (FT), wash (W) and elution (E) fractions.

Substitution of two histidine residues resulted in greater nickel affinity loss (**Figure 4.5**). The amount of CAP in the flow-through is identical relative to the load material, unlike what is found with prior cases. Its presence is also limited in the wash fraction, which is either the result of residual load material passing over the column or desorption of bound CAP. The best indicator, however, is in the elution fraction, of which CAP only comprises 6%. The relative concentrations of other proteins are equivalently higher. The significantly higher amount of precipitated CAP in the "IP" lane should be addressed. Substitution of two solvent-exposed histidines with polar residues has evidently promoted hydrophobic interactions among CAP dimers, resulting in unwanted aggregation and precipitation. This would be cause for concern if my aim were to express plasmid-encoded CAP or mutants to high concentrations, where such protein interactions are more likely to occur. However, this expression system only serves the

purpose of evaluating column binding. A host cell with basal chromosomal CAP expression would be much less prone to this precipitation; therefore, this issue is not concerning.

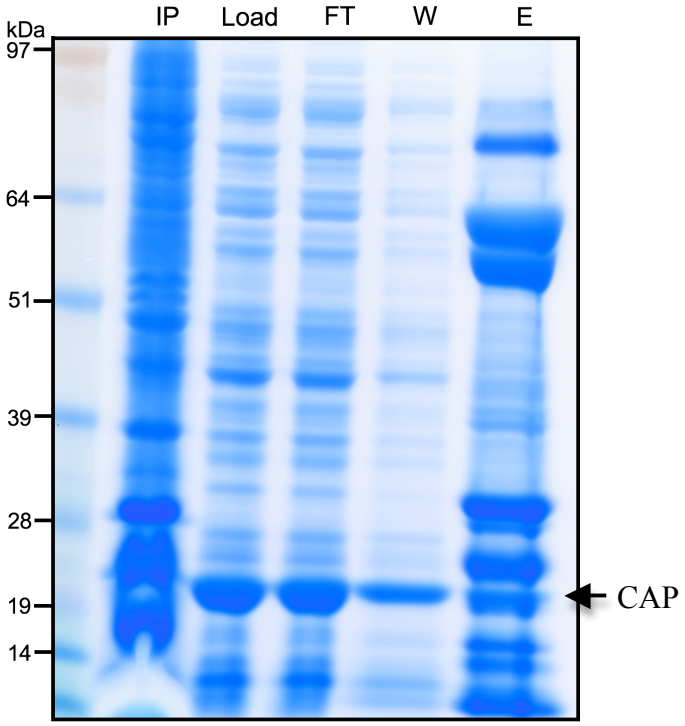

Figure 4.5. IMAC purification of double mutant CAP (H17A, H19Y).
Lanes correspond to the complex lysate (Load), insoluble pellet (IP), flow through (FT), wash (W) and elution (E) fractions.

Figure 4.6A is a gel produced in an identical manner from strain expression the triple mutant CAP (H17A, H19Y, K52N), which has both histidine substitutions and arginine in place of Lys52. Again there is a strong presence of mutant CAP in the flow-through and wash lanes and very little in the elution fraction, of which the mutant again comprised just 6%. **Figure 4.6B** was produced from BL21 (DE3) lysate without expression of plasmid. This is a suitable comparison since only chromosomal CAP is expressed. In this case CAP comprised 8% of the elution pool, which is essentially an identical amount given the error inherent to the gel procedure and densitometry analysis. This evidence strongly suggests that the double histidine substitution eliminates CAP binding even at 20 mM imidazole. Another approach was used to help confirm this finding that only chromosomally expressed CAP is found in the elution fraction of the triple mutant case.

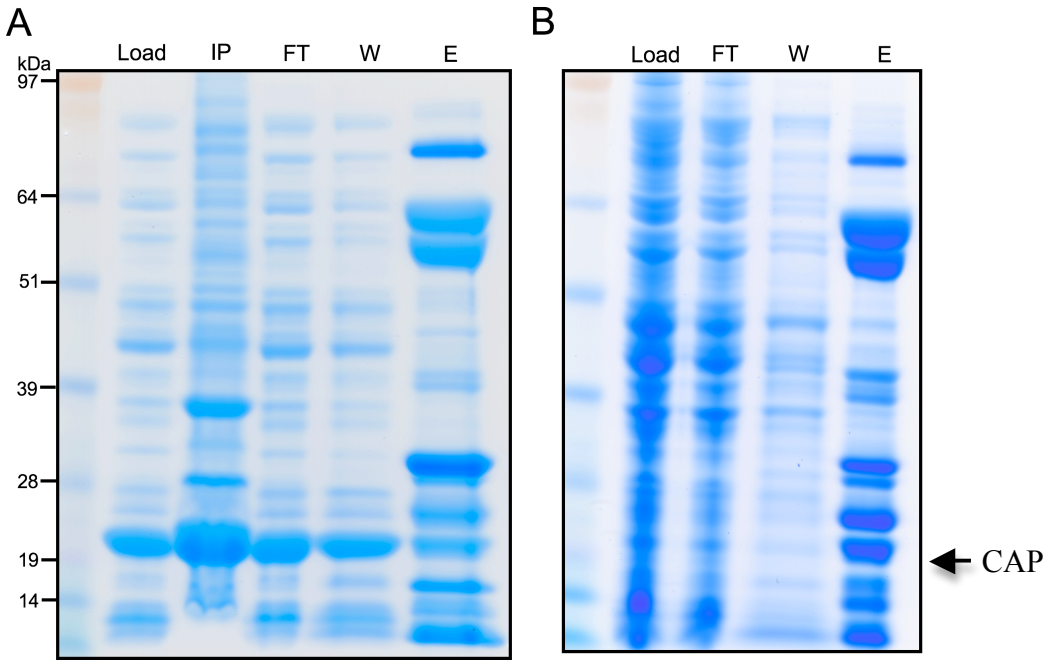

Figure 4.6. IMAC purification of (A) triple mutant CAP and (B) BL21 lysate without plasmid. Lanes correspond to the complex lysate (Load), insoluble pellet (IP), flow through (FT), wash (W) and elution (E) fractions.

It is a safe argument that the wild-type and triple mutant CAP proteins would have *different* affinities for nickel. So if indeed both forms bound to the IMAC resin, they would elute in separate peaks during a gradient elution of imidazole. If only chromosomal CAP is present, it would elute in a single peak. This is exactly what was found in **Figures 4.7** and **4.8**. The total protein curve of the chromatogram reveals that protein was eluted in a single peak (note that the baseline increases in a linear fashion due to the imidazole gradient). The gel shows the wash fraction and elution fractions as the imidazole concentration is linearly increased from 20 to 250 mM. Fractions E1-E6 each represent 1 mL, E7 represent the next 3 mL, E8 the next 3 mL, E9 a 7-mL equilibration at 250 mM and E10 the EDTA pool. No protein was detected in the latter two. The intensity of the CAP band increases from left to right, peaking in the E4 lane, then diminishes until it is no longer detected in E8. It should be further noted that elution of CAP in a single peak supports the explanation of CAP prominence in the wash fraction, that it is from residual lysate passing through the column during column equilibration and not from desorption of weakly bound protein.

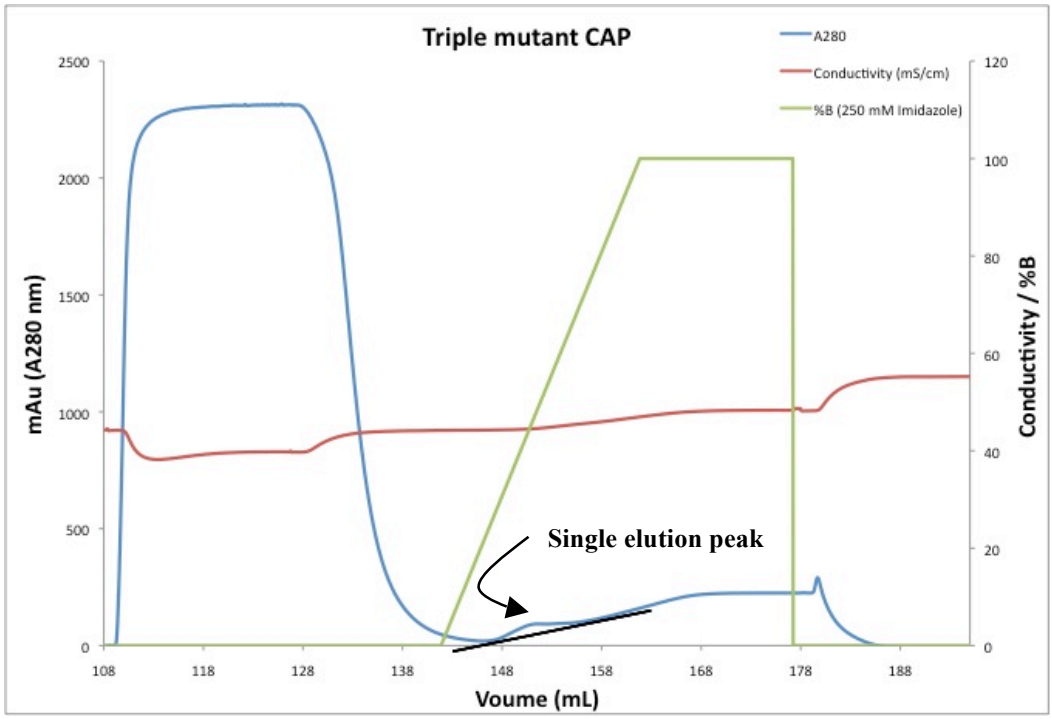

Figure 4.7. Imidazole gradient elution of triple mutant CAP.
Total protein (blue line) indicates a single elution peak followed the gradient imidazole elution (green line). Conductivity is shown in red.

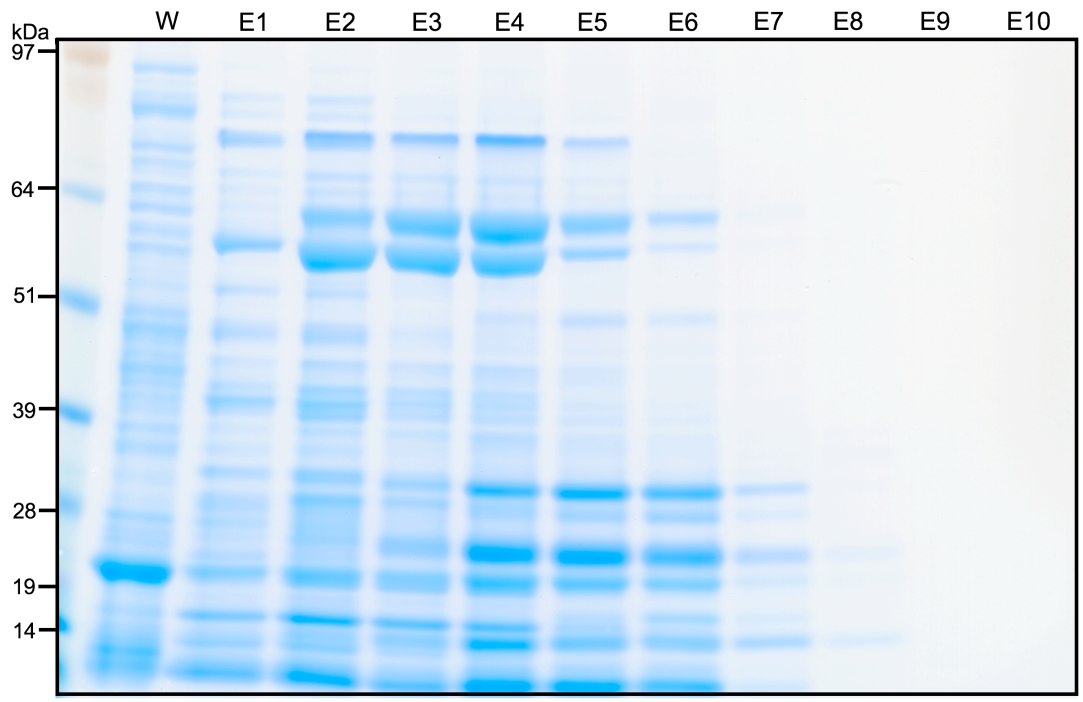

Figure 4.8. Wash and elution fractions of triple mutant CAP purified by imidazole gradient. Each lane indicates a fraction of the gradient elution shown in Figure. 4.7. Lanes correspond to wash (W), the first 6, 1-mL elution fractions (E1-E6), pools of the following 3 mL of the elution fractions (E7 and E8), a pool of a 7-mL equilibration using 250 mL imidazole (E9), and the EDTA fraction (E10).

For a final observation, bound proteins of the triple mutant CAP and wild-type strains were captured after purification using 50 mM imidazole and resolved by 2D-PAGE, similar to the DIGE analysis of the previous chapter except now with Coomassie staining (**Figure 4.9**). Immediately evident is the comparable amount of CAP in each pool (protein spot labeled with an arrow), indicating that only chromosomal CAP was captured from the triple mutant. SlyD, a familiar IMAC contaminant featured in the previous chapter, is found in both pools. Inducible expression of the plasmid-encoded protein resulted in noticeably higher expression of chaperonins DnaK, GroEL and ribosomal subunit protein RS1 (the protein cluster in panel A at p*I* 4.9, 57-69 kDa). Elongation factor Tu (43 kDa, p*I* 5.3), a protein that guides aminoacyl-tRNA into the free site of the ribosome, was also captured by IMAC and is seemingly expressed higher in the plasmid-containing strain.

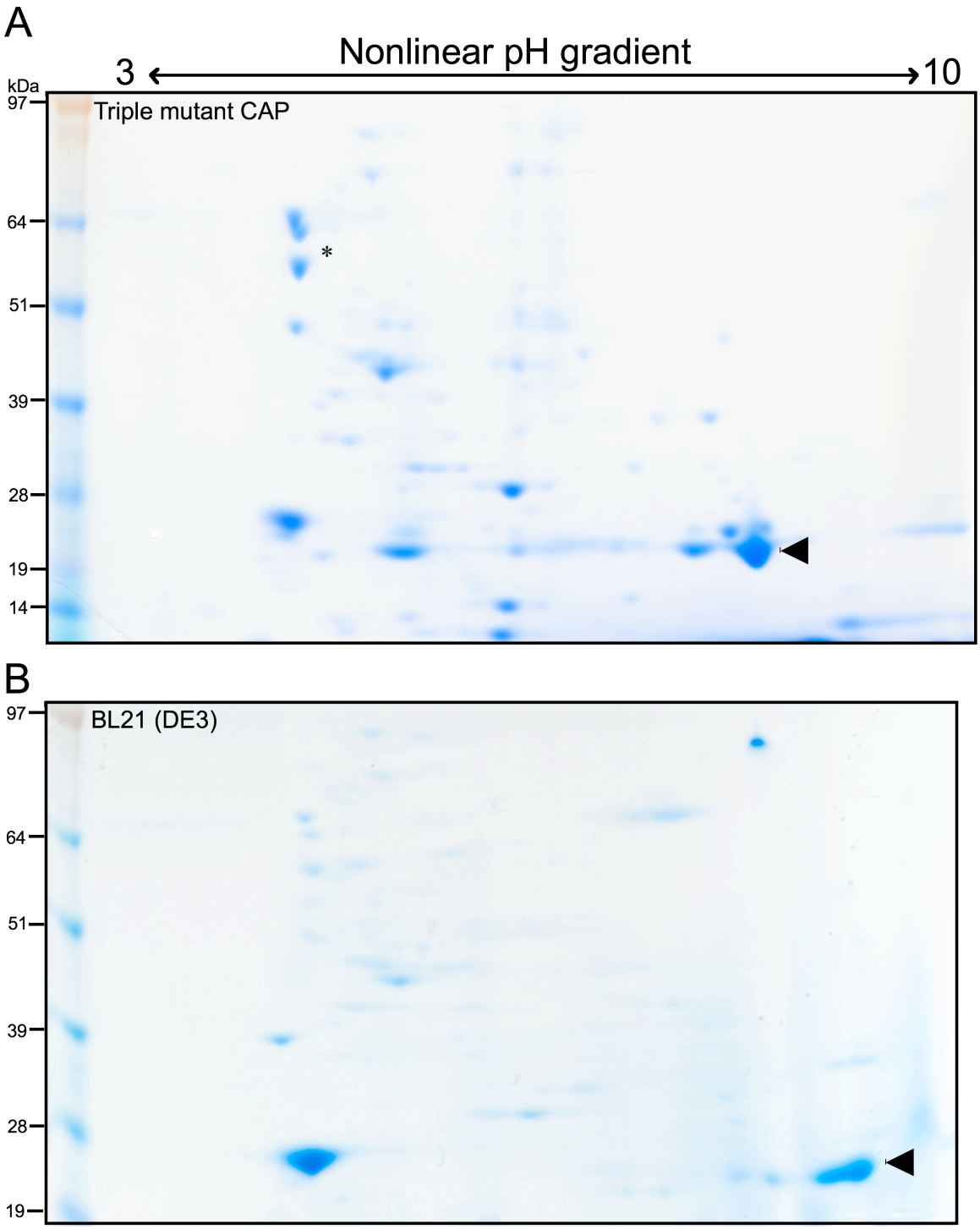

Figure 4.9. 2D-PAGE of IMAC proteins from (A) triple CAP mutant strain and (B) BL21 (DE3) loaded with 50 mM imidazole.
IEF strips were loaded with equal amounts of protein (50 mg) and the 2nd dimension was run in identical conditions. The protein spot corresponding to CAP is marked with an arrow whereas the protein cluster referenced in the text is marked with an asterisk.

One conclusion from the IMAC studies in the previous chapter validated what others have affirmed: that CAP, a protein with low histidine content relative to the other contaminants and no metal binding sites, is among the most tenaciously binding host proteins in *E. coli*, capable of contaminating purifications of highly-expressed, his_6-tagged proteins even in stringent loading conditions.[103] Closer examination of CAP, however, provides a likely explanation for its affinity for immobilized nickel. Three surface-exposed histidines with two residues spaced between can coordinate well with metals. When nickel is immobilized with nitrilotriacetic acid as they are in HisTrap columns, only two coordination sites are available for protein binding. Thus, substitution of two histidine residues (leaving only a single histidine to interact with nickel) was predicted to sufficiently disrupt its affinity to nickel. The results in this chapter have given that confirmation and further study should only be done in the context of a host cell chromosomally-expressing this modified CAP during plasmid-encoded recombinant expression. Thus, this work closes a chapter of research dedicated to optimizing perhaps the most widely applied affinity chromatography technique in bioscience.

There is, however, a trend in the biotechnology industry that seeks cost-efficient alternatives to affinity methods. In the introductory chapter, two other widely used chromatography techniques were discussed, hydrophobic interaction and ion exchange, based on hydrophobic and electrostatic interactions. Application of our upstream-downstream integration approach for these resins would be quite ambitious indeed, given that *all* proteins have at least a *degree* of hydrophobicity or net charge (dependent on pH of course). Nevertheless, this is precisely the topic of the following chapter.

5.0 EVALUATION OF ESCHERICHIA COLI PROTEINS THAT BURDEN NON-AFFINITY-BASED CHROMATOGRAPHY

5.1 INTRODUCTION

As we have seen, affinity chromatography media is favored for its superb selectivity. Media with conjugated proteins for monoclonal antibody capture, such as protein A for example, are among the most selective. The remarkable performance comes at a cost, however, as protein A resin is over 30 times more expensive than typical ion exchange resins.[43] Although resin lifetime is limited by fouling by product or contaminant build-up, loss is primarily due to ligand degradation after repeated cycles of stripping and regeneration.[145] To mitigate costs, the number of investigations that seek to find alternatives to packed-bed chromatography is trending upward. For example, advanced techniques for bulk separation via aqueous two-phase partitioning, three-phase partitioning, specific and non-specific precipitation, crystallization as well as use of uncharged, charged and affinity-based membranes for filtration and chromatography are of high interest to industry (see reviews[5, 68, 146]). Follman and Fahrner characterized seven multi-step chromatography schemes using non-affinity based media, particularly hydrophobic interaction (HIC), anion (AEX) and cation exchange chromatography (CEX), to assess their performance against their customary protein A affinity-CEX-AEX scheme in the purification of Chinese Hamster Ovary-derived monoclonal antibodies.[43] Three such processes were found to be

comparably effective in removing CHO proteins as the more traditional protein A-CEX-AEX process. Effective purification of proteins endowed with positively charged polyarginine tails have also been demonstrated. Sassenfeld and Brewer pioneered the concept by fusing five consecutive arginines to the C-terminus of recombinant human urogastrone.[49] Variants of a positively charged 58-amino acid sequence excised from the B domain of staphylococcal protein A has been successfully used as a fusion tag for protein isolation via CEX.[58, 147] Separation to high purity of cyclodextrin glycosyltransferase fused to polylysine or polyarginine has also been reported.[54]

Use of non-affinity based matrices is favored for their low cost and robustness towards harsh cleaning procedures, but less selectivity means multiple steps are typically needed to remove host contaminants. The solution to this problem can be found in a return to the central theme of this body of work: improved efficiency in the downstream phase can be gained by integrating rational host design upstream. We know from prior discussion that this requires focused examination of host proteins that remain bound to column matrix after equilibration, which reduces capacity for the target protein, as well as proteins that co-elute with the target itself. These proteins are particularly bothersome since they reduce final purity and resolving such proteins (with similar physicochemical properties to the target) complicate the process with additional, orthogonal purification steps that drive up operational costs and detriment overall yield.

Evaluation of such troublesome proteins is the basis of the work presented in this chapter. I now apply our familiar techniques to improve knowledge of host proteins with a likelihood of contaminating two schemes base on hydrophobic and electrostatic interaction, HIC-AEX and polyarginine-aided CEX. Based on the results, preliminary direction for development of tailored host strains with intrinsically improved separation performance is discussed.

5.2 PURIFICATION OF NON-TAGGED GFP

Before describing how the host proteins were captured, a general comment is needed on my general purification strategy and the choice to use GFPuv as a model protein. GFPuv was a well-suited choice for a generic model protein not only for convenient assay, but its medium size (27 kDa) and acidic isoelectric point (p*I* 5.7) are reasonably average in the context of the host proteome. For instance, the distribution of *E. coli* proteins by p*I* is bimodal and the mean p*I* of the acidic protein group is about 5.5-5.7 (the mean p*I* of the basic protein group is about 9.5); while the average size of a typical *E. coli* is 35 kDa.[148, 149] Ideally, a group of model proteins with a spectrum of sizes and net charges would be used. The resulting analysis would perhaps generate a more complete list of host proteins that contaminate the IEX step across a broader range of binding parameters. This could allow design of a host strain with intrinsically better downstream efficiency that is more robust to IEX conditions. In light of this idea, process development that established only the most fundamental conditions for GFPuv purification via HIC and AEX was carried out. I chose this approach so these processes, with minimal alteration, could be applicable for purification of a wider range of target proteins. Further, by not fully optimizing the binding and elution conditions for GFPuv, the intent was to put less restriction on

the likelihood that any identified contaminants could appear in similar processes in which a protein of interest is purified. Ultimately, however, the overall approach can be applied to cases where a target requires binding or elution conditions that significantly deviate from those reported here. Contaminants relevant to that specific process could then be identified.

5.2.1 Optimization trials on HIC and AEX columns

The variables pertaining to HIC were first explored. The effect of pH on target retention in HIC is complicated. While it can be said generally that increasing pH weakens hydrophobic interactions and decreasing pH increases them, pH is specifically a modulator of amino acid ionization. Its effect on retention is thus amplified if the ionization states of the residues at the site of interaction with the support matrix can be altered. This suggests that both hydrophilicity and hydrophobicity moderate retention. Further, variation in target retention is considerably dampened in the range pH 5-8.5.[150] Given these considerations and noting again that purification of GFPuv is not the specific intention, pH was not varied in the loading or elution buffers.

Columns of 1 mL HiTrap Phenyl FF that contain phenyl groups at low density (low-substitution) and high density (high-substitution) were compared, as were starting ammonium sulfate concentrations. Extracts were initially prepared in HIC elution buffer, 20 mM NaH_2PO_4, pH 7.2. Five ml samples were adjusted with elution buffers either with or without 4 M $(NH_4)_2SO_4$ to 8 ml total volume, creating a series of extracts in starting buffers ranging from 0.7 to 1.5 M $(NH_4)_2SO_4$ with equal protein concentrations. Incubation on ice preceded centrifugation of any precipitated material. A final filtration resulted in five ml samples which were applied to the low-sub matrix. Sample loading was followed by start buffer equilibration until protein desorption completed. Elution used a 20 CV gradient and a final five CV equilibration with low

salt buffer to remove any residual protein. Fractions of one CV were collected and assayed in routine fashion. Columns were regenerated with five CV of distilled water followed by five CV of start buffer.

With samples containing less than or equal to 1 M ammonium sulfate, GFP was not retained. For the 1 M trial, the target was visibly present in the flow through and equilibration fractions. Trials of 1.2 M and 1.5 M ammonium sulfate gave GFP breakthrough at completion of loading of 31% and 6%, respectively, when normalized to the lysate fluorescence. Thus, this series of column runs determined that 1.5 M was the lowest ammonium sulfate concentration that allowed the desired target retention. Since 6% breakthrough is lower than commonly referenced "threshold" of allowable protein loss of around 10%, and higher ammonium sulfate concentrations would only increase the likelihood of excessive precipitation, further tuning of the salt concentration was deemed unnecessary. Chromatograms of these trials are shown in **Figures 5.1** and **5.2**. (Note that lines representing total protein concentration and conductivity correspond to vertical axes. The data series in green represents relative GFP fluorescence with markers corresponding to GFP assays of each fraction, whereas the connecting line is purely for display purposes. Further, the GFP curves do not correspond to a given vertical axes, but are merely scaled and superimposed on the charts).

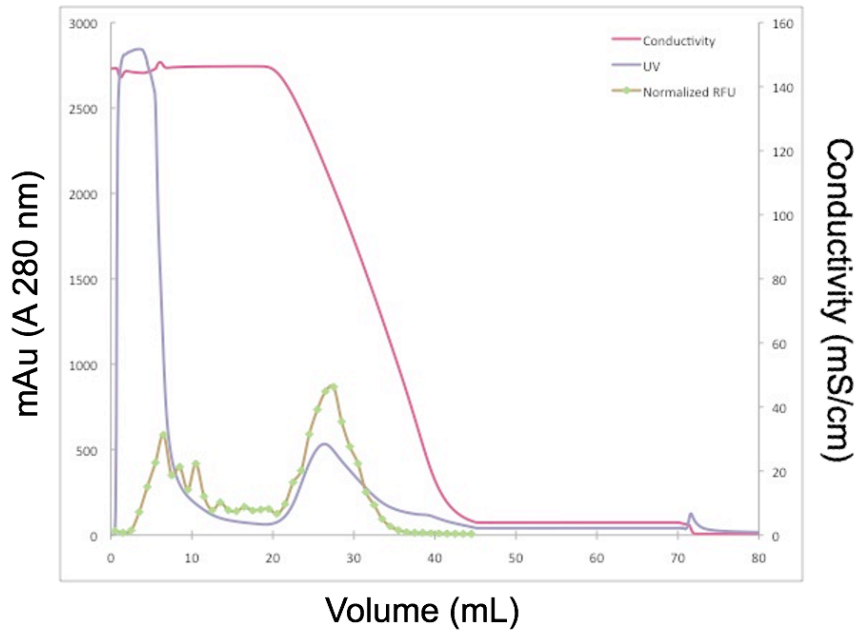

Figure 5.1. Chromatogram of HIC loaded with 1.2 M ammonium sulfate.
GFP assays (green points with connecting line) indicate that GFP had 31% breakthrough at the end of loading.

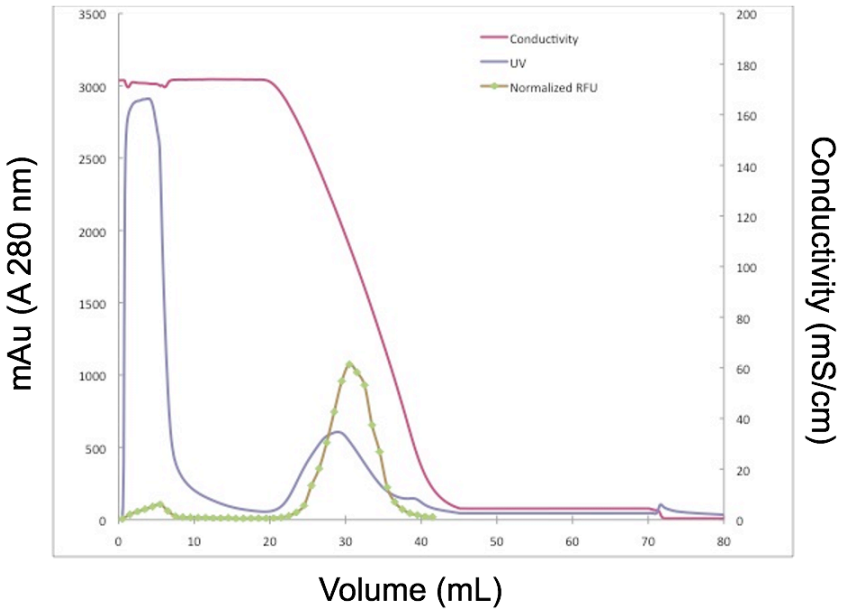

Figure 5.2. Chromatogram of HIC loaded with 1.5 M ammonium sulfate.
GFP assays (green points with connecting line) indicate that GFP had 6% breakthrough at the end of loading.

Starting buffer adjusted to 1.5 M ammonium sulfate was also used to evaluate the high-sub matrix. The chromatogram in **Figure 5.3** shows a significantly wider elution peak relative to the low-sub column (the sharp second peak contains trace hydrophobic proteins that only elute during wash with pure distilled water post elution). GFPuv was visually detectable in the fractions that corresponded with the total protein peak. The apparent decrease in resolution, and thus the more cumbersome buffer exchange step for ensuing column steps, lead to the decision to focus exclusively on low-substitution matrix.

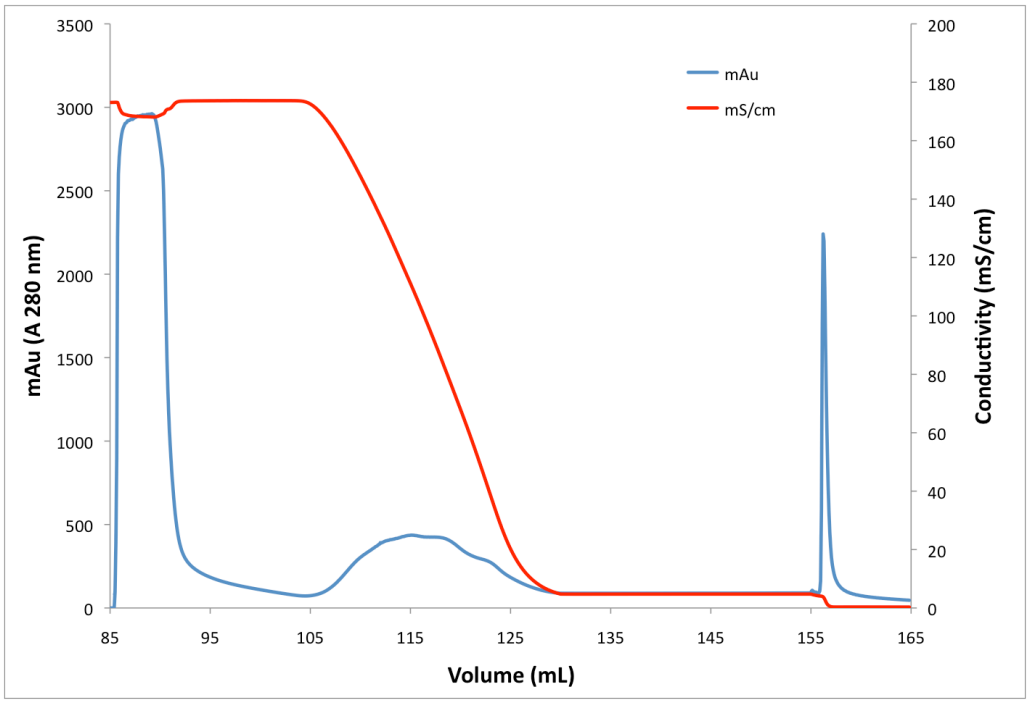

Figure 5.3. Chromatogram of HIC on phenyl high-substitution matrix.
Total protein (blue line) indicates a wide GFP elution peak relative to the low-substitution matrix and residual bound protein after all salt (red line) was washed from the column.

A series of AEX trials was performed to determine the optimum operating pH and loading salt concentration. Varying the pH is the primary way to influence selectivity in IEX applications, that is to say, the order of elution and resolution of peaks. As is commonly done, a series of starting buffers with varying pH and no salt are initially used for bind-and-elute purifications. After model protein retention is determined in each condition, the optimum pH is selected. A new series of buffers, each at the chosen pH, contain varying amounts of NaCl before the optimization process is repeated. Since this generalized methodology is more reflective of an IEX used as a capture step, as opposed to a polishing step requiring fine-tuning of these and other conditions for highest selectivity, it was the favored strategy. One alternative to this would have been to skip optimization of the pH entirely, opting to load at say pH 8.0 to promote relatively more host protein binding and relatively less of GFP, for example. While this may cast a wider net in the exploration of co-purifying host proteins, even a modest attempt to purify any target from complex material by IEX would require tuning of the pH in a practical setting. Further, I anticipated that proteins of relative abundance in any stage of the purification process, not just in the GFP elution fraction, would be scrutinized and perhaps identified from SDS-PAGE gels. This indeed was the case and will be discussed in the next section of results.

To briefly summarize the methods, *E. coli* BL21 (DE3) pGFPuv pellets were resuspended in AEX start buffers, consisting of 20 mM Bis Tris at pH 5.4, 5.7, 6.0, 6.2, 6.7 and 7.2. Columns were loaded with five CV of lysate and eluted with a NaCl gradient (0-1 M) over 20 CV. Noting that the p*I* of GFP is 5.7, and from prior reports concluding that an optimal pH should be at least 0.3 units more basic,[151] it was hypothesized that pH 6.0 or 6.2 would be optimal. Indeed, curves generated under such conditions show unfavorable binding or elution characteristics. Trials at pH 5.4 and 7.2 showed excessive GFP breakthrough, indicative of a net positive charge in the former

and diminished binding capacity due to host protein competition in the latter. Trials between pH 5.7 and 6.7 yielded similar results, though GFP retention was slightly improved at pH 6.0 and 6.2. Operation at pH 5.7 would likely prevent addition of little or any salt in the loading, as the target would surely wash out given its p*I* of 5.7. Binding at pH 6.7 was satisfactory but was inferior to the pH 6.0 and 6.2 conditions. The convention of a capture step is that the minimal pH that removes the most host protein is optimal.

Trials at pH 6.0 and 6.2 were more carefully examined with SDS-PAGE shown in **Figure 5.4** (Lanes are labeled within the figure and were loaded with 25 μg protein. Note that "Peak" refers to the single fraction containing the highest relative amount of GFP, whereas "Pooled" refers to a sample mixed from every elution fraction that contained GFP).

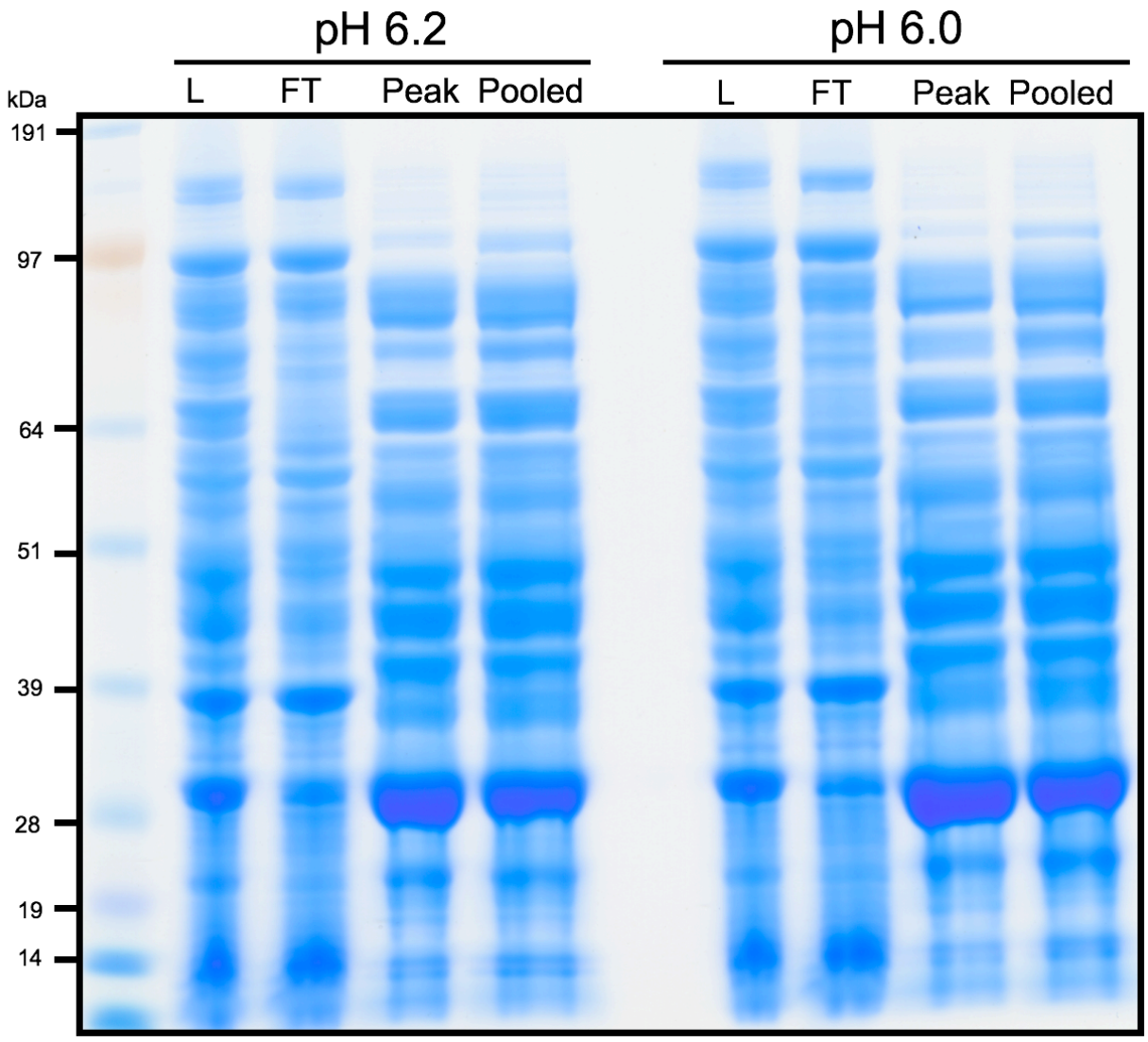

Figure 5.4. SDS-PAGE comparing AEX runs at pH 6.0 and 6.2.
The first lane is the molecular weight ladder. Other lanes correspond to the load material, flow through, single 1-ml fraction containing highest amount of GFP (peak), and the bound proteins pooled for each pH conditions.

Gel fractions in each case look virtually identical, so either pH could be used without significant difference. Nevertheless, densitometry that determined relative protein concentration and GFP purity of each trial gave evidence that pH 6.2 was slightly more advantageous. Relative GFP purity in the pooled fractions was 21.5% and 18.7% for pH 6.2 and 6.0, respectively. A chromatogram from the pH 6.2 run is shown in **Figure 5.5** (Remember that the GFP curve presented here was prepared as described in **Figures 5.1** and **5.2**. Labels of peak relative fluorescence after loading and during elution are given).

Two additional trials were carried out to determine the initial salt concentration. Buffers at pH 6.2 contained either 50 or 100 mM NaCl were compared. GFP did not retain binding at 100 mM, whereas all conditions were satisfactory at 50 mM (shown in **Figure 5.6**).

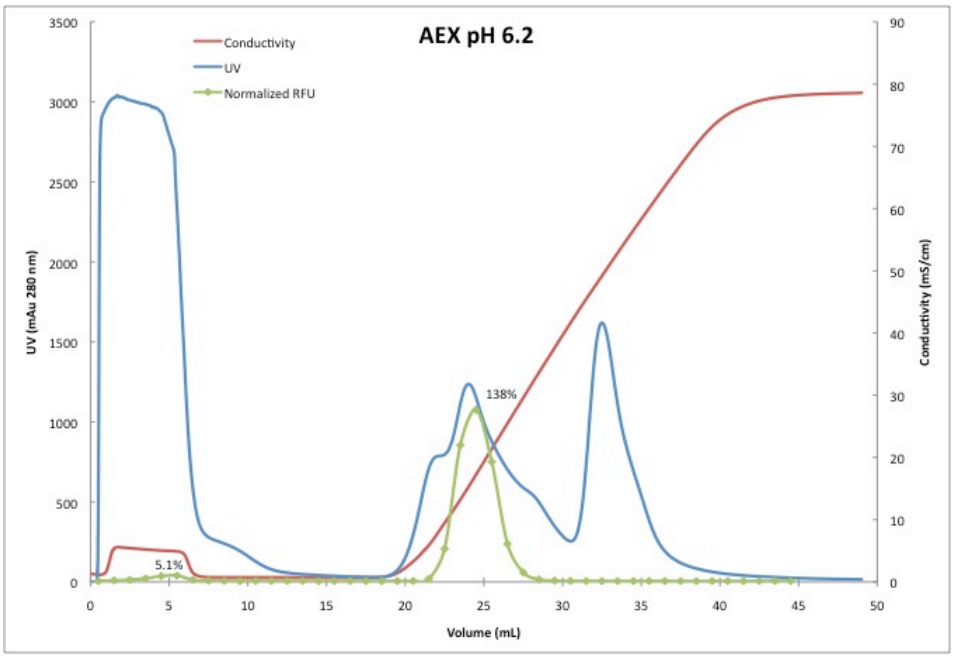

Figure 5.5. Chromatogram of the AEX trial with pH 6.2, 0 M NaCl loading buffer.
GFP assays (green points with connecting line) indicate that GFP had 5.1% breakthrough at the end of loading and eluted in a single peak. Total protein (blue line) indicated two elution peaks during salt gradient elution (red line). The second peak is indicative of residual DNA in the sample as no proteins were observed from these fractions via SDS-PAGE (not shown).

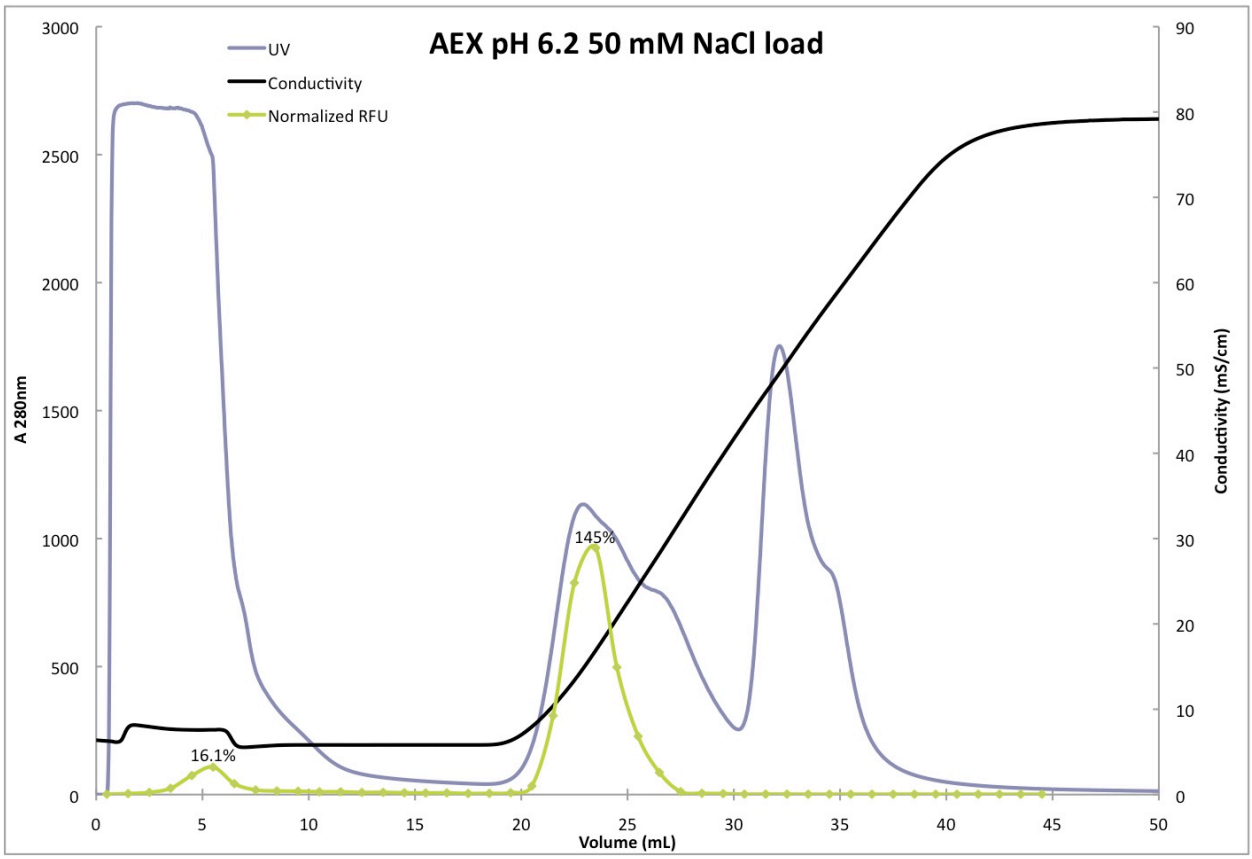

Figure 5.6. Chromatogram of the AEX trial with pH 6.2, 50 mM NaCl loading buffer. GFP assays (green points with connecting line) indicate that GFP had 16% breakthrough at the end of loading and eluted in a single peak. Total protein (blue line) indicated two elution peaks during salt gradient elution (black line). The second peak is indicative of residual DNA in the sample as no proteins were observed from these fractions via SDS-PAGE (not shown).

While GFPuv flow through was a bit higher than ideal for the 50 mM NaCl case (16%), the marginal increase in flow through would not detract from subsequent analysis of the bound protein pool. GFPuv desorbed from the column at the forefront of the first elution peak, which was anticipated from both prior results and column manufacturer literature stating that when proteins are loaded in buffers with a pH of 0.5 units more basic, they typically begin to elute at 0.1 M NaCl. An SDS-PAGE gel of this purification (not shown) did not differ noticeably from the "pH 6.2" bands in **Figure 5.4**. Based on that outcome and the recommendation by GE to find optimal ionic strengths at increments of 0.05 M NaCl in a series of buffers, it was felt that pursuit of trials in the range of 0−50 mM would be of little consequence.

5.2.2 Purification analysis and characterization of co-purifying proteins

A schematic of the purification strategy is given in **Figure 5.7**. *E. coli* BL21 pellets were resuspended with HIC elution buffer (20 mM NaH_2PO_4, pH 7.2) and then adjusted to 1.5 M ammonium sulfate. The solution was incubated on ice prior to spin down of precipitated material, filtration, and column loading. Initial experiments on HiTrap Phenyl FF (low-substitution) columns revealed that this procedure yielded adequate GFP retention. GFPuv breakthrough at the end of the loading phase, measured by the fluorescence exiting the column normalized to the cleared lysate, was 6% with 1.5 M ammonium sulfate. The elevated concentration resulted in significant precipitation of other strongly expressed proteins, easing binding competition with the target. This is more clearly illustrated in **Figure 5.8**, which shows sequential "cuts" of an ammonium sulfate precipitation (The lane labeled "IP" is the insoluble pellet centrifuged after sonication. Lanes 2 through 9 are labeled with molar concentration of ammonium sulfate. The lane labeled "Sup." is the resulting supernatant after precipitation with

1.5 M ammonium sulfate, and is representative of the loading material for the HIC step). Outer membran protein OmpF is labeled with an arrow in the figure. OmpF forms a homotrimeric porin that allows passive diffusion of ions and small molecules into the cell and was identified among those that were preferentially reduced prior to HIC column loading (confirmed by PMF; Mascot score 116, e- value 1.3e-6, 9 peptides with 33% sequence coverage).

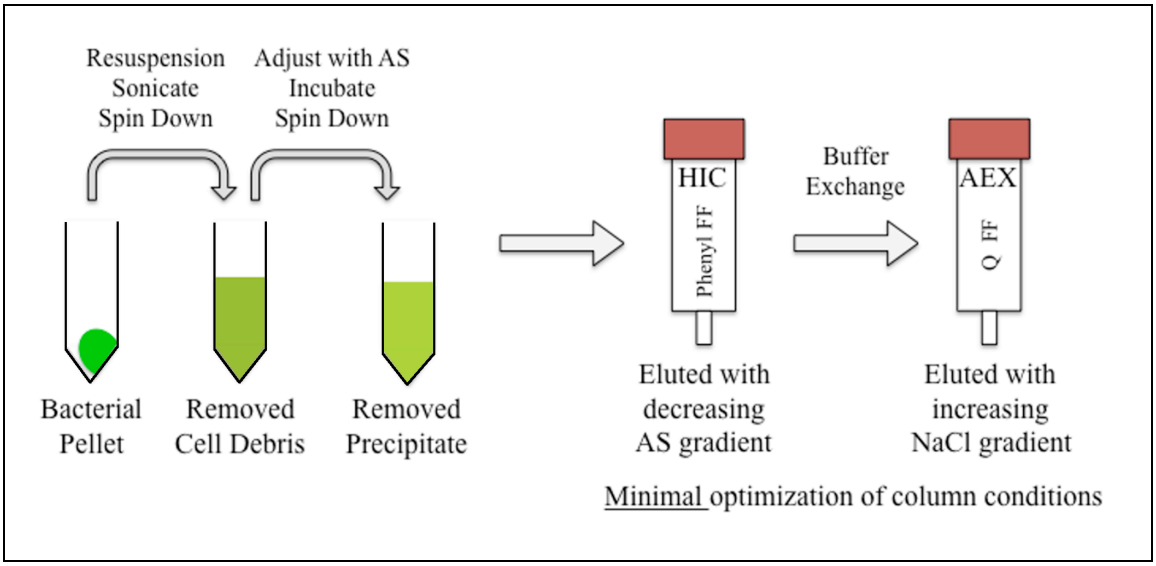

Figure 5.7. Schematic of the workflow used to purify un-tagged GFP and host proteins. The bacterial pellet was resuspended in HIC elution buffer (20 mM NaH2PO4, pH 7.2), sonicated, and then centrifuged to remove debris. The supernatant was adjusted with ammonium sulfate to 1.5 M, incubated on ice for 10 min, and then centrifuged again to remove precipitated protein. The remaining supernatant was loaded on the HIC column. After equilibration, bound proteins were eluted by decreasing ammonium sulfate gradient and buffer exchanged (20 mM Bis Tris, 50 mM NaCl, pH 6.2) prior to loading on the AEX column. Bound proteins were eluted with increasing NaCl gradient. Protein fractions from each step were saved for gel analysis.

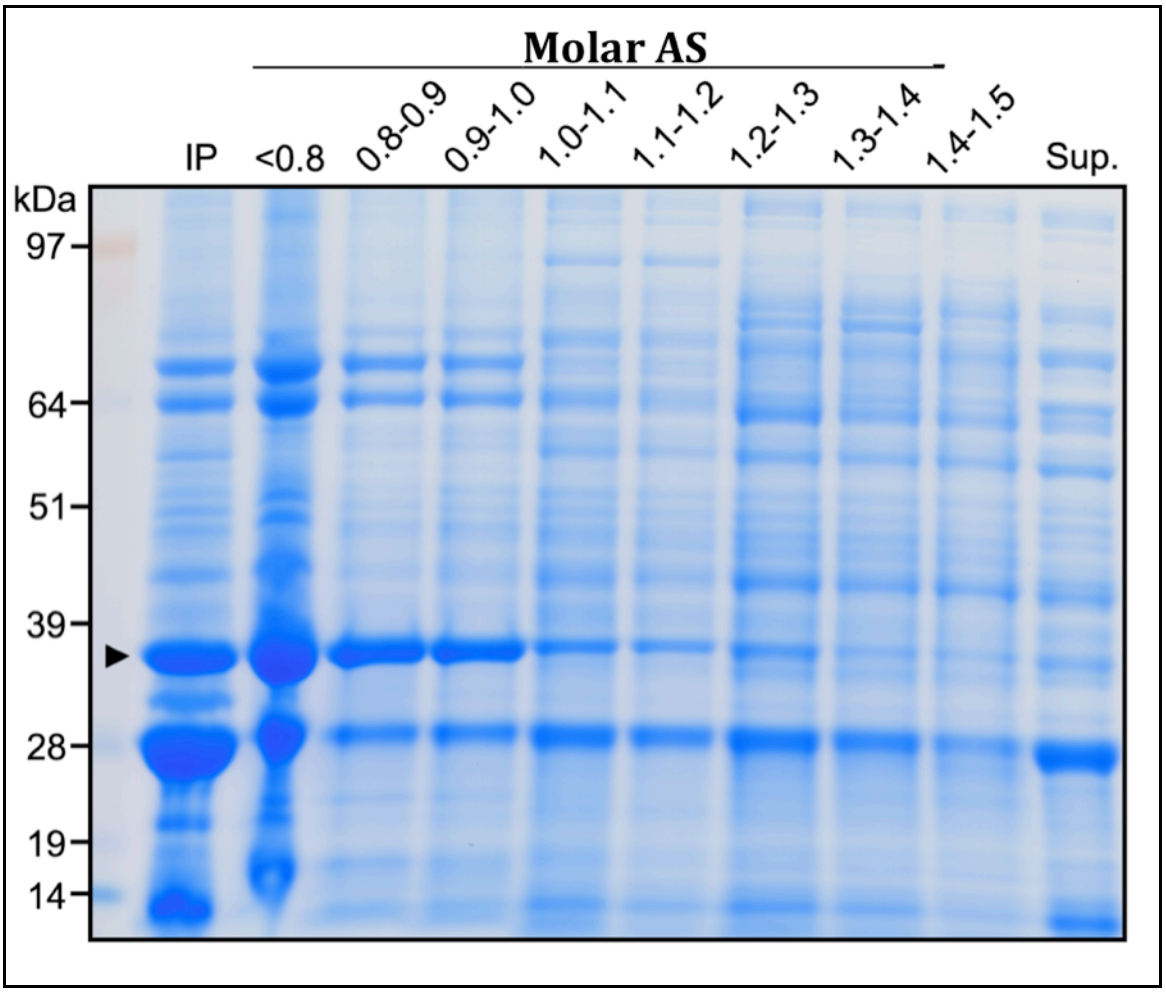

Figure 5.8. Ammonium sulfate precipitation of proteins expressed in BL21 (DE3) pGFPuv. The "IP" lane shows proteins contained in the insoluble pellet (cell debris and inclusion body proteins). The following lanes show protein precipitated at their corresponding ammonium sulfate concentrations, known as "cuts" of the precipitation. The soluble protein (supernatant, labeled "Sup." lane) after incubation with 1.5 M ammonium sulfate was the loading material for the HIC column. The arrow corresponds to outer membrane protein OmpF, which is a highly-expressed host protein that was preferentially removed during precipitation. GFP is the prominent band aligned with the 28 kDa standard.

Proteins eluted from the HIC column were pooled, buffer exchanged, and applied to the AEX column. Initial trials on Q columns determined that an optimal loading buffer was 20 mM Bis-Tris, 50 mM NaCl at pH 6.2. GFPuv breakthrough was not detected during column loading as it was during optimization using complex lysate due to the fact that AEX is here used as the second column step. The preliminary trial runs were conducted without anticipating which order the columns would be used. Indeed, a purification using the reverse of the scheme presented here was conducted with inferior results and will be discussed in Section 5.5.

Proteins recovered in all steps of the purification are illustrated in **Figure 5.9**. Co-purifying proteins were confirmed by PMF and the most prominent bands are labeled by gene name. GroEL (57 kDa) and elongation factor Tu (43 kDa) are marked by arrows next to lane 4 and are discussed later. To briefly review the methods used to generate the gel, soluble extract was loaded on the HIC column using a buffer containing 1.5 M ammonium sulfate, equilibrated, then eluted with a gradient down to 0 M. Fractions containing GFPuv were dialyzed against AEX buffer and 5 ml was loaded on the Q column at pH 6.2. Bound proteins were eluted with a 20 CV linear salt gradient to 1 M. Lanes were loaded with 25 μg protein, except for lanes 7-9, which contained about 10 μg. Samples for the HIC flow through, and AEX flow through and first elution peak were prepared from TCA/DOC precipitated proteins. The accompanying process data is given in **Table 5.1**.

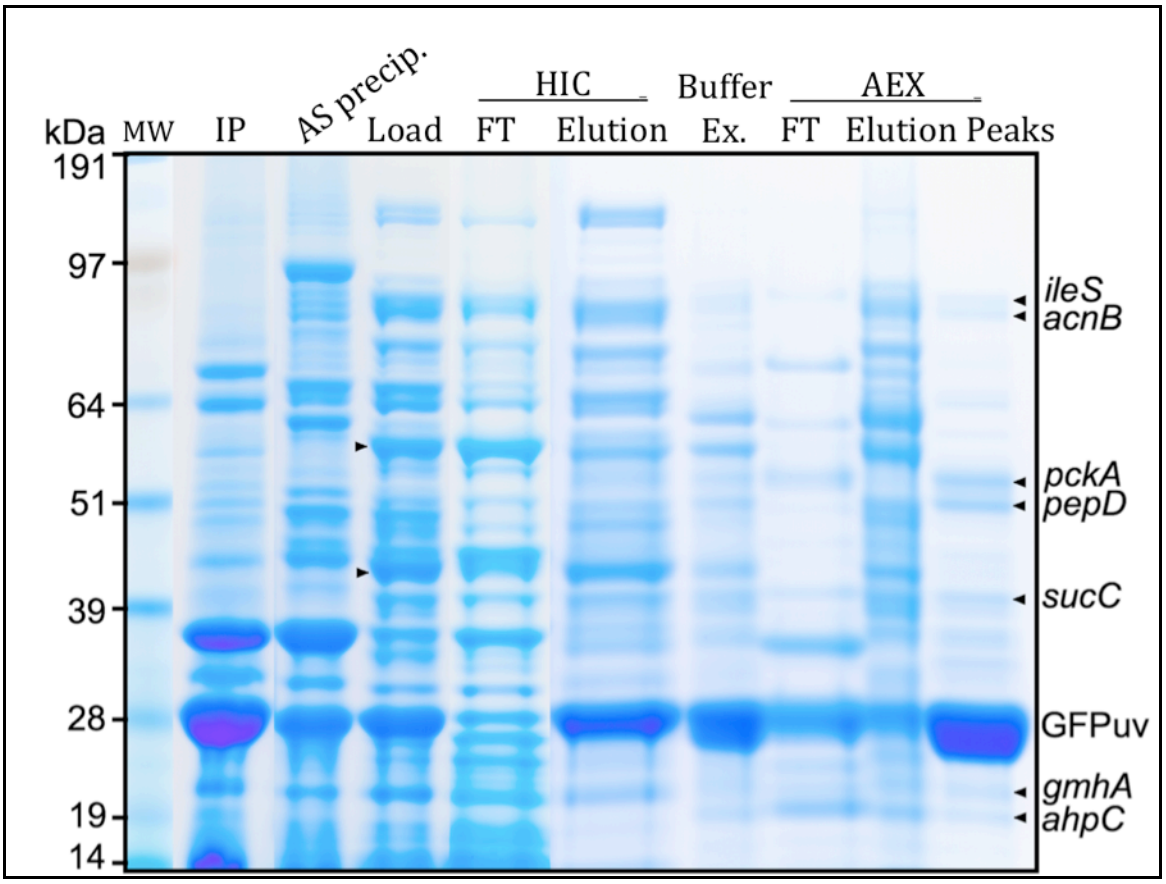

Figure 5.9. SDS-PAGE of GFPuv purified by HIC-AEX.
Protein fractions were taken from each step of the purification process, as illustrated in Figure 5.7. Lanes are: MW, molecular weight ladder; IP, insoluble pellet; AS precip., protein precipitated by 1.5 M ammonium sulfate; Load, HIC load material; HIC flow through; HIC elution pool; Buffer Ex., protein pool after buffer exchange; AEX flow through; first AEX elution peak; second AEX elution peak. Arrows in HIC load lane correspond to GroEL (57 kDa) and elongation factor Tu (43 kDa).

Table 5.1. Relevant purification data for each downstream process.

Purification Step	Total Protein (mg)	GFP (mg)	Yield (%)	Purity (%)	Purification Fold
HIC-AEX					
Soluble Extract	42	8.6	100	21	1
HiTrap Phenyl FF	29	8.0	92	27	1.3
Post-dialysis	4.5	1.9	---	42	2.0
HiTrap Q FF	2.4	1.6	84	66	3.2
CEX					
Soluble Extract	47	8.8	100	19	1
HiTrap SP FF	16	7.9	90	50	2.6

This scheme purified over-expressed GFPuv by 3.2-fold using non-affinity based matrices, noting that this is out of a maximum purification fold of 4.8. Overall GFPuv yield attained from the column steps was 77% and was purified to approximately 66% purity, determined by gel densitometry. Calculation of the process yield did not consider the HIC bound proteins that were dialyzed but not ultimately loaded on the Q column, as only five mL were used for the final column step; however when accounting for actual GFP loss during dialysis, overall yield was 55%. While my aim was not centered on yield optimization, other dialysis methods could have mitigated loss between column steps. Minor flow through in the HIC step and the GFPuv fraction pooling strategy contribute to diminishing the yield during chromatography, but overall the yield that I obtained is comparable to that reported for multi-step non-affinity schemes.[43]

Purification of a target protein to homogeneity using this or similar scheme, without further fine-tuning of operational or column conditions, would likely necessitate a third polishing column step. This presents an opportunity to improve efficiency using an approach that uses proteomic tools to suggest rational genetic modifications specifically advantageous for downstream processing. Thirteen proteins that co-eluted with GFPuv from the AEX column are presented in **Table 5.2**.

Table 5.2. Proteins that co-purified with GFPuv during purification via ammonium sulfate precipitation, HIC, and AEX.

Gene	Protein Name	Protein Mass (Da)	Protein p*I*	UniProtKB Accession No.	Matched Peptides	Sequence Coverage %	Mascot Score
ileS	Isoleucyl-tRNA synthetase	104231	5.66	**P00956**	11	10	92
acnB	Aconitate hydratase 2	93439	5.24	**P36683**	18	21	74
pckA	Phosphoenolpyruvate carboxykinase	59606	5.46	**P22259**	16	38	91
pepD	Aminoacyl-histidine dipeptidase	52882	5.20	**P15288**	9	19	74
gltA	Citrate synthase	47984	6.21	**P0ABH7**	6	12	57
sucC	Succinyl-CoA ligase [ADP-forming] subunit beta	41367	5.37	**P0A836**	13	34	96
gcvT	Aminomethyltransferase	40121	5.36	**P27248**	6	17	76
talB	Transaldolase B	35197	5.11	**P0A870**	8	33	84
cysK	Cysteine synthase A	34468	5.83	**P0ABK5**	9	41	70
tsf	Elongation factor Ts	30404	5.22	**P0A6P1**	7	25	83
dapD	Tetrahydrodipicolinate N-succinyltransferase	29873	5.56	**P0A9D8**	7	25	79
gmhA	Phosphoheptose isomerase	20802	5.97	**P63224**	5	22	67
ahpC	Alkyl hydroperoxide reductase subunit C	20748	5.03	**P0AE08**	4	32	77

Several of the identified proteins in the final pool were present in only small amounts. Included in this group are TCA cycle enzymes: aconitase B (*acnB*) catalyzes the reversible isomerization of citrate and isocitrate;[152] the beta subunit of ADP-forming succinyl-coA ligase (*sucC*); and citrate synthase (*gltA*), which produces citrate and CoA from acetyl-CoA and oxaloacetate. Mutants lacking *sucC* and *gltA* are undesirable for biotechnology applications since they have been shown to produce elevated acetate levels.[153, 154] Enzymes involved in amino acid biosynthesis are tetrahydrodipicolinate N-succinyltransferase (*dapD*), an enzyme involved in lysine synthesis via the diaminopimelate pathway; and cysteine synthase A (*cysK*), which converts *O*-acetylserine to cysteine and acetate. Others include the essential class I isoleucyl-tRNA synthetase (*ileS*);[155, 156] the T subunit of the glycine cleavage complex aminomethyltransferase (*gcvT*); transaldolase B (*talB*), important for balancing metabolite concentrations in the pentose-phosphate pathway and has been previously purified by ammonium sulfate precipitation and AEX;[157] alkyl hydroperoxide reductase subunit C (*ahpC*); phosphoheptose isomerase (*gmhA*) that is involved in LPS biosynthesis; and elongation factor Ts (*tsf*), which associates with the EF-Tu-GDP complex at the ribosome and induces exchange of GDP to GTP.[158] Nonessential proteins in this group total less than 8% of the final pool by densitometry; and thus knockout of such genes, if not detrimental to the cell, would not be expected to significantly enhance purification efficiency.

Other gene products that provide opportunities for deletion are those that burden the purification scheme, but do not ultimately co-purify with the protein of interest. For column steps, for example, such proteins hinder target binding capacity without necessarily reducing final purity. Indeed, proteins bands of interest were identified regardless of where they appeared in the process. Here, clarification with ammonium sulfate prevented some such proteins from

burdening column steps. Outer membrane porin OmpF (~37 kDa) contributed significantly to the insoluble and soluble protein fractions, but was preferentially removed with ammonium sulfate, shown in **Figure 5.8** and lane 3 of **Figure 5.9**. Based on this case, knockout of *ompF* may be desirable to ease the metabolic burden on a host expressing recombinant proteins at high levels. Although it was reported that mutants lacking OmpF have reduced susceptibility to β-lactam and tetracycline antibiotics, this was found at low concentrations (1-2 μg/mL ampicillin) relevant to natural environments and not at those used to provide selection pressure (50-100 μg/mL).[159] Based on this case, knockout of *ompF* may be desirable to ease the metabolic burden on a host expressing recombinant proteins at high levels.

While knockout of *ompF* could be advantageous, its dramatic precipitation relative to the target prior to any column step nevertheless lends credence to use of this scheme that dissipates the burden of host protein removal over multiple non-affinity-based steps. This is also neatly illustrated by the separation of elongation factor Tu (EF-Tu, *tufA*) and chaperonin GroEL (*groL*) over both column operations. Shown in lanes 4-7 and 9 of **Figure 5.9** are prominent bands of 43 kDa and 57 kDa, associated with EF-Tu and GroEL respectively, which depicts their removal in both the flow through during HIC and elution in a peak distinct from the target during AEX (EF-Tu/GroEL were confirmed by PMF; corresponding scores 92/83, e-values 3.4e-4/2.6e-3, 10/9 peptides, 26%/18% sequence coverage). Their ultimate binding to the AEX column is unfortunate given the importance of GroEL and EF-Tu to protein folding and translation, respectively. Mutants deficient in *groL* are not viable and those deficient in *tufA* have a slow-growth phenotype in complex media.[160, 161]

A combined proteomic and host cell engineering approach offers the most significant gains in downstream processing efficiency by genetically removing the most burdensome co-

eluting proteins. Phosphoenolpyruvate carboxykinase (PEPCK; encoded by *pckA*) and aminoacyl-histidine dipeptidase (PepD) were the prominent contaminants, accounting for about 15% of co-eluting proteins by mass. PepD is a metallo-dipeptidase that acts on a broad spectrum of substrates with unblocked N termini, including unusual dipeptides like L-carnosine.[162, 163] It has been hypothesized that a significant portion of PepD activity lies outside the cytoplasmic membrane, perhaps aiding the degradation of lengthy peptides for transport into the cell.[164] *E. coli* posses many peptidases, however, and *pepD*-deficient mutants can still utilize a diverse array of peptides.[162] PEPCK is a gluconeogenic enzyme that requires divalent cation(s) to catalyze the reversible decarboxylation and mononucleotide-dependent phosphorylation of oxaloacetate, yielding phosphoenolpyruvate (PEP), CO_2, and the corresponding nucleoside diphosphate.[165] PEPCK and PEP carboxylase, which catalyzes the reverse reaction, act at a critical node of carbon metabolism by maintaining the PEP:oxaloacetate ratio and concentrations of TCA intermediates.[166] PEP carboxylase exclusively catalyzes the anaplerotic reaction from C3 metabolites in *E. coli*. Inactivation of *pckA* has been shown by Yang et al. to cause a reciprocal drop in PEP carboxylation and increased carbon flux through glyoxylate shunt, replenishing TCA cycle intermediates, namely oxaloacetate.[167] It was also noted that this flux pattern was similarly found in the low acetate producer BL21,[168] and further that *pckA* mutants showed significantly increased biomass yields and reduced CO_2 evolution rates compared with wild-type *E. coli* during slow growth in glucose-limited media.[167] These reports lend credence to a plausible use for Δ*pckA* strains in biotechnology applications.

Given the strong presence of PEPCK and PepD in the co-purification pool and the prior findings discussed, there are grounds for further development of a Δ*pckA* Δ*pepD* strain intrinsically advantageous for purification of select recombinant products via HIC-AEX schemes.

5.3 PURIFICATION OF ARG_6-GFP BY CATION EXCHANGE CHROMATOGRAPHY

As before, little process optimization was carried out for the CEX step. It was intended to use the minimal operating pH that ensured sufficient retention of the target while the vast majority of E. coli host proteins, carrying a net negative charge, flowed through. Analysis of co-purifying proteins could thus be applicable to a variety of elution strategies. Initial preparations included loading buffers at pH 5.5 and 6.0, but proved problematic with either residual arg6-GFPuv bound to the column after elution with high salt or less-than-desired retention, respectively. Soluble lysates were ultimately prepared in a loading buffer of 50 mM MES and adjusted to pH 5.7, the p*I* of untagged GFPuv. Lysozyme was not used in the sample preparation, anticipating its dominating presence in the CEX eluate (this was actually observed in preliminary experiments discussed later in section 5.4). Five CV of sample was passed through a 1 ml HiTrap SP FF column. Fluorescence detected arg_6-GFPuv breakthrough of 8%; whereas lysate containing untagged GFPuv in the same conditions expectedly showed no target retention. Proteins were eluted with a 15 CV linear NaCl gradient to 1 M, followed by a step gradient to 2 M. No proteins were recovered in the subsequent step elution to 2 M in either case. Chromatograms that compare behavior of the native and arg_6-GFPuv are presented in **Figures 5.10** and **5.11**.

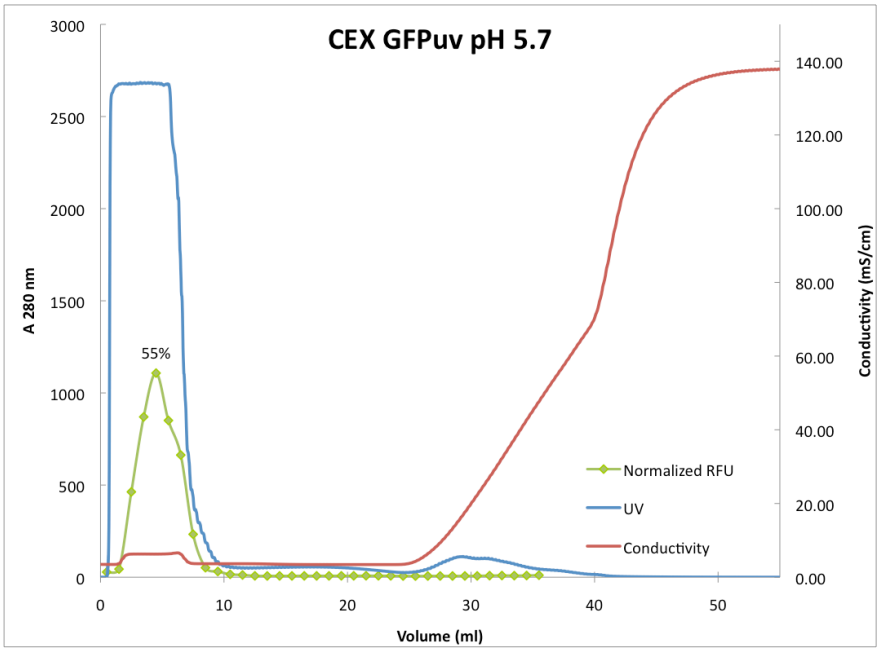

Figure 5.10. Chromatogram of CEX of untagged GFPuv with loading at pH 5.7.
GFP assays (green dots with connecting line) indicate that untagged GFP showed no column retention, flowing through the column during loading.

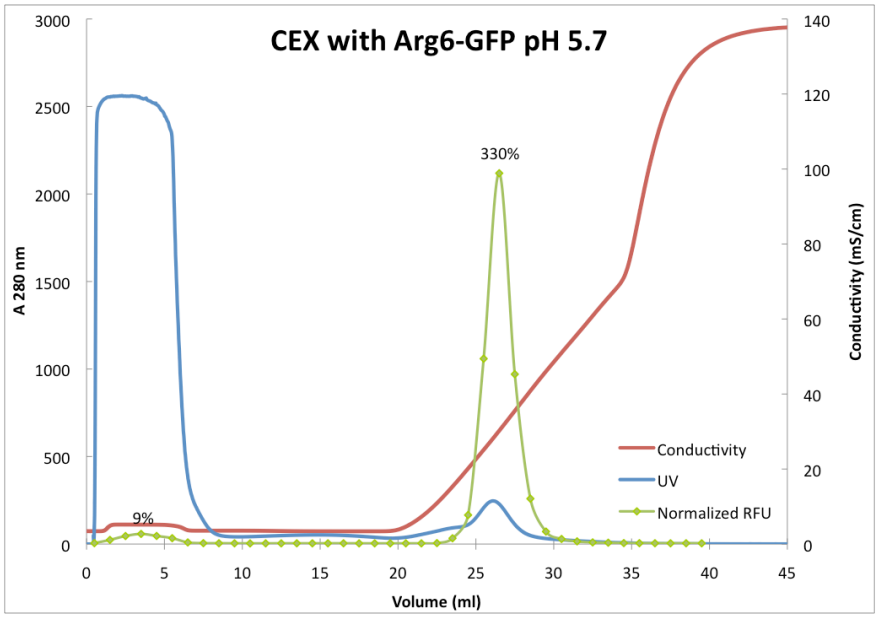

Figure 5.11. Chromatogram of CEX of arg_6-GFPuv with loading at pH 5.7.
GFP assays (green dots with connecting line) indicate that arg6-GFP had 9% breakthrough at the end of loading and eluted in a single peaking during salt gradient elution (red line).

Proteins were sampled at each step in the process and fractionated by SDS-PAGE and presented in **Figure 5.12**, with corresponding purification data shown in **Table 5.1**. The target was purified to about 50% purity, a 2.6-fold refinement, at 90% yield. Note that while this CEX scheme did not outperform the HIC-AEX scheme, the column step was superior to either individual step in the prior strategy, presumably due to the endowed localized charge. The purity achieved may seem underwhelming in comparison to other studies where multiple steps were used,[49] but the process yield compares very favorably to previous successful uses of this method.[54] However, it is noted again this purification was intentionally not optimized for maximum performance. Though it was troubling to find a high amount of insoluble target protein production (lane 2), even more so than in the un-tagged strain cultured identically in spite of evidence that polyarginine tags improve solubility.[169] This effect could possibly have been mitigated with the addition of NaCl during resuspension.[56] The focus, however, was specifically on soluble host proteins that challenged the column step. Note that OmpF, in this case again highly expressed, flowed through the column under the given conditions (lane 4). Also found in the flow through are the characteristic bands of RNA polymerase β/ β′ subunits (~155 kDa).

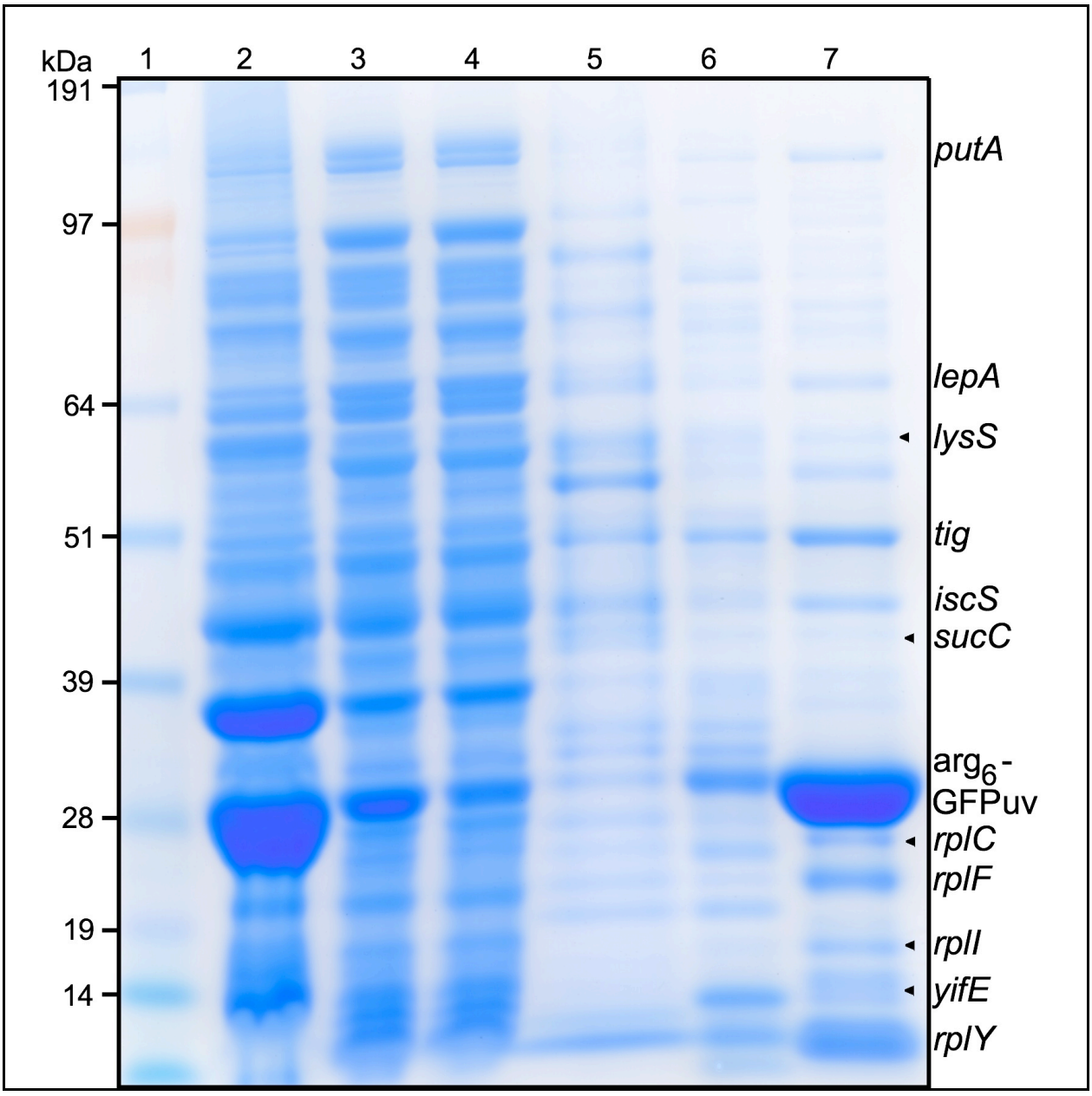

Figure 5.12. SDS-PAGE of arg_6-GFPuv purification by CEX.
Lanes were loaded with 25 μg protein. Lanes are: (1) molecular weight ladder, (2) insoluble pellet, (3) CEX load, (4) flow through, (5) wash fraction, (6) first elution peak, (7) elution peak containing arg6-GFPuv. Identified bands in the final lane correspond to the gene names on the right.

Eleven proteins confirmed to co-elute with arg_6-GFPuv are presented in **Table 5.3**. Some lesser relevant contaminating proteins are lysyl-tRNA synthetase (*lysS*), bifunctional protein PutA (*putA*) is a proline catabolic enzyme and an autogenous transcriptional repressor of proline utilization genes,[170] and an uncharacterized UPF0438 protein (*yifE*). Cysteine desulfurase (*iscS*) is also present. It acts in iron-sulfur clusters after assembly with scaffold homolog IscU and possibly has a role in regulation of genes associated with pyrimidine metabolism.[171] Mutants deficient in *iscS* have been characterized to have a slow growth phenotype.[172]

One prominent contaminating protein is trigger factor (*tig*). Trigger factor is a well-studied ribosome-associated chaperone, binding to nascent peptides at the exit site, exhibits peptidyl-prolyl cis/trans isomerase activity, and play a role in Sec- and SRP-dependent protein export.[173] While Δ*tig* strains have no obvious growth phenotype or increased protein aggregation, they must be compensated by the DnaK/J and GroEL/S systems;[174, 175] therefore *tig* mutants are ill-suited for biotech applications that utilize over-expression of recombinant proteins. For further details, reviews can be consulted.[173, 176]

As opposed to trigger factor, elongation factor 4 (*lepA*) is a good candidate for removal. It is a translational GTPase with an apparent ability to catalyze reverse translocation *in vitro*[177] and was thought to play a role in translation fidelity. Recent evidence, however, not only confirms prior reports that mutants lacking *lepA* exhibit no deleterious effect on growth rate,[178] but also suggests that its gene product does not contribute to translation fidelity.[179]

Table 5.3. Proteins that co-purified with arg6-GFPuv during purification by CEX.

Gene	Protein Name	Protein Mass (Da)	Protein p*I*	UniProtKB Accession No.	Matched Peptides	Sequence Coverage %	Mascot Score
putA	Bifunctional protein putA	143725	5.69	**P09546**	15	10	96
lepA	Elongation factor 4	66528	5.40	**P60785**	14	27	77
lysS	Lysyl-tRNA synthetase	57567	5.11	**P0A8N3**	10	18	88
tig	Trigger factor	48163	4.83	**P0A850**	9	22	70
iscS	Cysteine desulfurase	45061	5.94	**P0A6B7**	10	27	84
sucC	Succinyl-CoA ligase [ADP-forming] subunit beta	41367	5.37	**P0A836**	10	25	76
rplC	50S ribosomal protein L3	22230	9.91	**P60438**	6	33	84
rplF	50S ribosomal protein L6	18892	9.71	**P0AG55**	6	36	87
rplI	50S ribosomal protein L9	15759	6.17	**P0A7R1**	5	28	87
yifE	UPF0438 protein yifE	13125	6.10	**P0ADN2**	7	56	81
rplY	50S ribosomal protein L25	10687	9.60	**P68919**	6	70	65

Four out of the thirteen identified contaminants are 50S ribosomal proteins L3 (*rplC*), L6 (*rplF*), L9 (*rplI*), and the particularly abundant L25 (*rplY*). Excluding GFPuv, L25 accounted for about a third of bound protein by densitometry. The former three are essential for *E. coli* cell viability,[180-182] though L25 deserves further consideration. It was the first discovered representative the CTC (catabolite controlled) family in 1972.[183] Proteins of this class are associated with stress conditions and characteristically bind to the designated loop E region of 5S rRNA in the 50S subunit,[184-186] a domain strictly conserved in bacteria[187] and distinctive for consecutive non-Watson-Crick base pairs (i.e. the nucleotides are not entirely complementary). Indeed, evidence has shown that all CTC proteins posses a domain, homologous to L25, responsible for 5S binding at this site, their only known target for interaction.[188, 189] Despite the wealth of biochemical, interaction, and structural studies, questions remain concerning their cellular function in bacteria and role in translation. L25, like others of the CTC class, is thought of as an evolutionary feature of translation exclusive to bacteria, and is hypothesized to have a possible role in ribosome recycling.[190] In *E. coli*, cells survive knockout of L25 but exhibit inhibited growth and protein translation which could be recovered by L25 expression from plasmid;[190] nevertheless such a strategy is unsuitable for biotechnology applications.

For an alternative approach, it is necessary to discuss the conservation of residues among CTC proteins. Work by Stoldt,[191] Wöhnert and colleagues,[192] and Gongadze's group [188, 193] has utilized sequence alignment of 300 CTC proteins across all bacterial taxons to reveal conserved residues. Given the specific nature in which CTC proteins associate with the ribosome, they have surprisingly low sequence homology and only five strictly conserved residues. In L25, these are R9, R21, Y31, H88 and D90. It has been shown that these residues interact primarily with the phosphoribose backbone via hydrogen bonding, inaccessible to the solvent. Importantly, this

low-homology characteristic differs from typical proteins with sequence-specific DNA-binding, in which protein side chains and the nucleotides with which they interact are highly conserved among bacterial species.[193] Mutation experiments with TL5, an L25 homologue in *Thermus thermophilus*, found that alanine substitution of non-conserved hydrogen bond-forming residues does not significantly attenuate 5S rRNA complex formation.[189, 194] This is attributed to the accessibility the residues have with the solvent. While these substitutions may exact an entropic cost, it is compensated by new hydrogen bonding with the solvent; whereas with any single conserved, inaccessible residue, such mutation does not allow complex formation.[189]

In contrast to affinity techniques, CEX interaction is based on electrostatic charge. Modifications of proteins that result in reduction of retention are thus not likely limited to one or two select residues, especially with large proteins. In contrast, *E. coli* L25 is a good candidate for residue modification due to its small size and few instances of consecutive positively charged residues. In yet unpublished results, our collaborators have used this surface modification strategy to substitute two histidines of triosephosphate isomerase (TPI) to eliminate retention on immobilized metal columns (Beitle, personal communication). Humphreys et al. introduced surface mutations and fused six aspartic acid residues to the C-terminus of PhoS/PstS protein, substantially shifting its 'functional p*I*' to prevent its co-elution with a recombinant Fab' fragment.[55] It is hypothesized that an altered L25 with substitutions to some of its non-conserved lysine and arginine residues, particularly consecutive arginines 18 and 19 and/or lysines 68, 71 and 73, would have vastly reduced affinity for CEX matrices. Further, use of a host strain deficient in *lepA* and with altered L25 could have significant, intrinsic advantages for CEX separation of polyarginine tagged proteins.

5.4 CONCLUSIONS

Reported in this study are the preliminary efforts to integrate rational upstream host design and downstream efficiency using proteomic tools to evaluate co-purifying host proteins that most significantly burden two non-affinity based processes. Phosphoenolpyruvate carboxykinase and peptidase D dominated non-target binding after HIC-AEX, accounting for almost half of the host protein that co-eluted with GFPuv. Previous inactivation of *pckA* and *pepD* has not revealed evidence for growth or expression-associated problems that concern biotechnology applications. L25 may alternatively be modified to lessen or eliminate its column burden to significantly improve target purity. This analysis serves an expressed need for process integration based on improved knowledge on physicochemical properties of relevant host proteins, in an effort to rival efficiency of affinity separation methods with inexpensive resins. As this knowledge expands with purification of other model proteins, host strains more robust to contamination can be developed.

5.5 SUPPLEMENTAL FINDINGS

Some of the results obtained during the development of the research described herein were thought to be secondary to the more cohesive story. They are, nonetheless, interesting and should be discussed here, starting with bioseparation via CEX.

Preliminary work with CEX was carried out using lysate derived from stains containing no plasmids. Without monitoring the binding and elution behavior of a target protein, buffer parameters could not be optimized as described above. Instead, the buffers purposefully

mimicked those reported by Sassenfeld and Brewer in their CEX purification of polyarginine-tagged human recombinant urogastrone, which was 50 mM MES, pH 5.5.[49] Twenty CV of lysate prepared in this buffer were applied to a HiTrap SP FF column, equilibrated with buffer, then eluted stepwise to 200 mM and 1 M NaCl. The chromatogram is presented in **Figure 5.13**. Proteins eluted during the second step were pooled and subjected to 2D-PAGE, shown in **Figure 5.14** with labels denoting protein spots analyzed by PMF. It is immediately noticeable from the gel image that a group of low molecular weight, high p*I* proteins has amassed at the lower right corner. The largest contributor to this cluster was identified as lysozyme, which as noted in **Table 5.4** is the list of co-purifying proteins identified from this and a subsequent two-dimensional gel produced with the same protein pool. Preceding sonication, ten mg of chicken egg white lysozyme had customarily been added to the resolubilization buffer to hydrolyze peptidoglycans in the bacterial cell wall to encourage release of cellular proteins. With a p*I* of 9.37, it follows that it would bind the CEX matrix and thus hinder interaction between the matrix and host proteins I seek to identify. This result formed the basis for my decision to forego use of lysozyme in the preparation of lysate in the prior chapter.

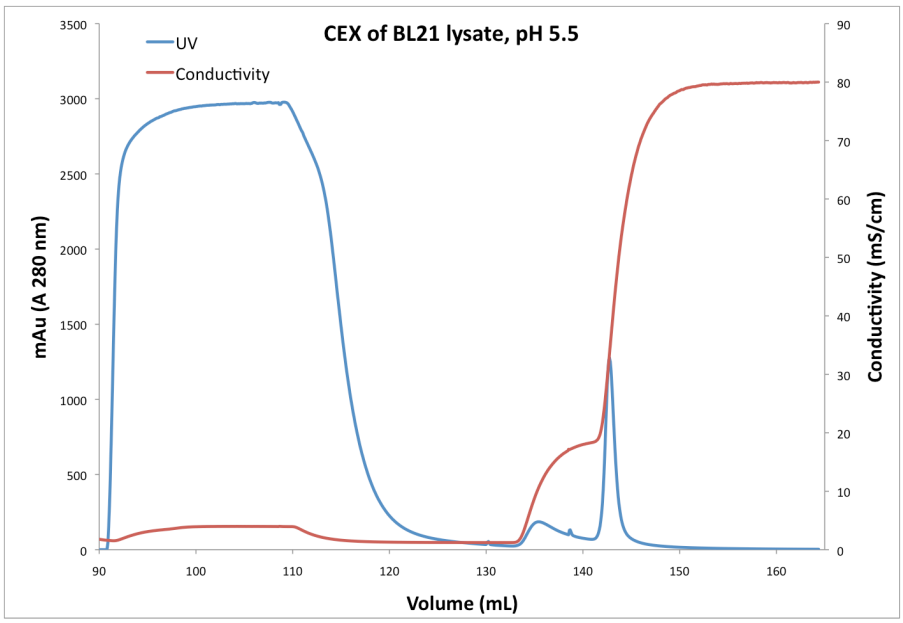

Figure 5.13. Chromatogram of CEX purification of BL21 cellular proteins.
Total protein (blue line) indicates the elution of two peaks that correspond with step increases of NaCl (red line).

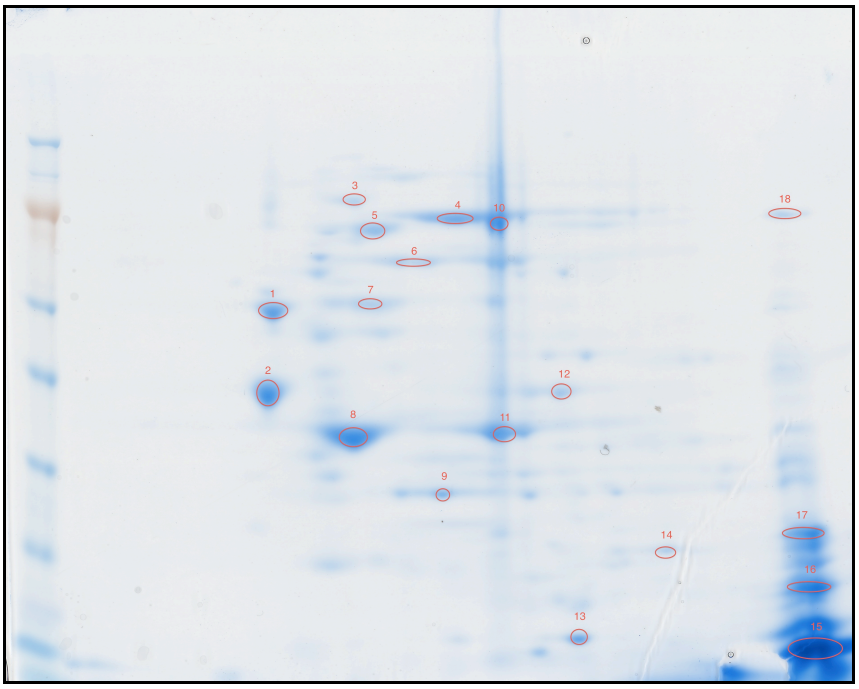

Figure 5.14. 2D-PAGE of BL21 proteins in the second elution peak of the CEX purification.
Protein spots were excised from the gel for identification by MALDI-TOF-MS.

Table 5.4. BL21 cellular proteins captured by CEX that correspond to the gel in Figure 5.13.

Spot	Gene	Protein Name	Protein Mass (Da)	Protein p*I*	UniProtKB Accession No.	Matched Peptides	Sequence Coverage %	Mascot Score
1	*rpsA*	30S ribosomal protein S1	61121	4.89	**P0AG67**	14	32	100
2	*tig*	Trigger factor	48163	4.83	**P0A850**	19	55	129
4/10	*infB*	Translation factor IF-2	97290	5.80	**P0A705**	22	25	114
5	*secA*	Protein translocase subunit	101959	5.43	**P10408**	13	14	66
7	*lepA*	Elongation factor 4	66528	5.40	**P60785**	8	14	58
8	*tufA*	Elongation factor Tu	43256	5.30	**P0CE47**	17	51	167
9	*accA*	ACCase subunit alpha	35219	5.76	**P0ABD5**	10	40	93
13	*rplI*	50S ribosomal protein L9	15759	6.17	**P0A7R1**	11	83	155
15	*LYZ*	Lysozyme C	16228	9.37	**P00698**	7	48	76
16	*rplF*	50S ribosomal protein L6	18892	9.71	**P0AG55**	6	38	92
17	*proQ*	ProP effector	25877	9.66	**P45577**	13	59	117

Several of the contaminants listed in **Table 5.4** would later appear in the CEX purification of arg_6-GFPuv such as elongation factor 4 (*lepA*), trigger factor (*tig*), and 50S ribosomal proteins L6 and L9. Others including the abundant 30S ribosomal protein S1 and elongation factor Tu (*tufA*) were curiously not detected in the later experiments. If they were present in the elution peak not containing arg_6-GFPuv and could not be confirmed by PMF, perhaps they could be detected by more sensitive mass spectrometry techniques. Alternatively, the genes that express those proteins may be severely down regulated in the presence of a plasmid-encoded recombinant protein. A comparative technique such as DIGE could provide that answer.

Regarding the purification of untagged-GFPuv by sequential column steps, it is worthwhile to discuss an alternative procedure that was pursued. Before protein capture and identifications were made, the modest optimization of column parameters was performed independently as though either AEX or HIC would be used as the initial capture step. Indeed the preferred scheme was compared with its reverse (i.e. AEX preceding HIC). During the ion exchange step, the five 1-mL fractions containing GFP were dialyzed overnight in HIC elution buffer (containing no ammonium sulfate) and then adjusted to 1.5 M before removal of precipitate material by filtration an HIC column loading. The chromatogram of this HIC step is presented in **Figure 5.15**. Note the scale of the total protein axis on the left. After dialysis and salt adjustment, a minute amount of total protein was ultimately applied to the column, of which only about a fifth flowed directly through the column. In these runs, selectivity was lacking when compared to the precipitation-HIC-AEX scheme. This is further supported by the SDS-PAGE gel of the final protein pool (**Figure 5.16**), here shown next to the final pool presented in section 5.2 for the preferred protocol.

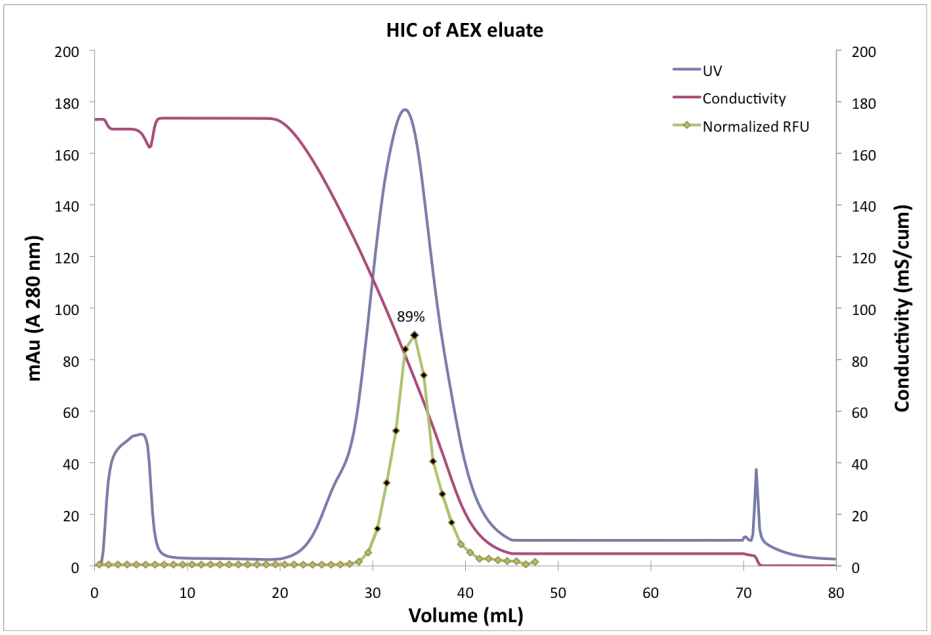

Figure 5.15. Chromatogram of HIC of proteins fractions containing GFP derived by AEX. GFP assays (green points with connecting line) indicates maximum column retention with elution of a single peak during decreasing ammonium sulfate elution (red line).

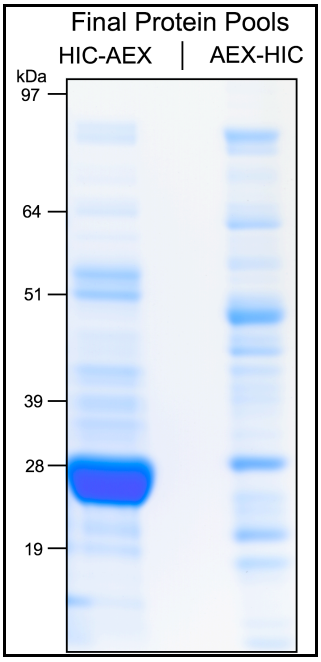

Figure 5.16. Comparison of final protein pools between alternative non-affinity schemes. The first lane shows proteins captured by ammonium sulfate precipitation, hydrophobic interaction and anion exchange chromatography. The proteins in the second lane were captured by anion exchange, precipitation, and then hydrophobic interaction chromatography.

The difference in final GFP purity is immediately evident in the gel figure. Densitometry analysis of the gel lane showed that GFP was 19% pure, well below the 66% purity achieved in the other method. While some target loss was undoubtedly experienced in the dialysis and salt adjustment steps, this only speaks to the ineffective nature of using HIC second. Here, ammonium sulfate precipitation is practically not even utilized since it is being applied to only about five mL of dilute protein solution. In the former method, precipitation selects against much of the full composition of cell lysate, allowing better resolution in the subsequent column steps. This finding is consistent with a basic principle in protein purification: clarifying steps such as precipitation that are characteristically high yield, high capacity and lower resolution should be performed early, if not first, in a purification strategy.

6.0 CONSIDERATIONS FOR FUTURE WORK

As discussed in the introductory chapter, the elution behavior of bacterial proteins during IMAC has been an area of study in our group for more than a decade. The experiments described here have yielded some important findings: (1) I demonstrated that knockout of select, nonessential bacterial genes that encode process-related impurities can improve final purity of a recombinant protein; (2) at low recombinant expression levels, host proteins such as CAP and chaperonins like SlyD become problematic in IMAC even under high imidazole conditions; (3) at high recombinant expression levels, most cellular contaminants are not relevant with high imidazole loading, in fact only CAP measurably contributed to the impurity in my case; and (4) specific CAP histidine residues can be substituted to eliminate its affinity for immobilized nickel even with low imidazole concentration during loading and without detriment to its cellular function. I may then reasonably suggest that further work in this area calls for chromosomal alteration of CAP and demonstration of improved downstream purification of other plasmid-expressed model proteins. Beyond such results, one additional worthwhile frontier in the pursuit of a host strain truly optimized for IMAC purification would be one that overcomes the challenge presented by SlyD. As an important foldase with a large histidine-rich C-terminal domain, a strategy that specifically addresses SlyD binding to nickel IMAC without additional column/polishing steps will be elusive.

Development of multi-knockout strains such as those proposed for the nonaffinity-based purifications is likely not a suitable direction at this stage. Attempting to demonstrate effective use of such strains would first necessitate intensive genetic manipulation and sequence confirmations over multiple months. Perhaps a more efficient use of resources would be to focus on strain robustness by varying culturing conditions/methods and expression of differing model proteins. The resulting identified impurities would allow design of strains with innate downstream efficiency that are more robust to fluctuations in culturing or column parameters.

APPENDIX

STRAINS, PLASMIDS, PRIMERS AND PROBES

Table A.1. E. coli strains and plasmids used in this work.

Strains/Plasmids	Description	Source
E. coli Strains		
BL21 (DE3)	High expression; encodes T7 RNA polymerase under the control of the lacUV5 promoter	Stratagene
DH5α	Cloning; plasmid amplification	Stratagene
TOP10	Cloning; Site-directed mutagenesis	Invitrogen
BL21	General cloning; template for knockout experiments	Stratagene
ΔcyoA	Mutant deficient in cyoA	This study
ΔcyoA_yfbG	Mutant deficient in cyoA and yfbG	This study
ΔcyoA_yfbG_adhP	Mutant deficient in cyoA, yfbG and adhP	This study
Plasmids		
pRedET	Red/ET expression plasmid. Confers ampicillin resistance.	Gene Bridges
706-Flp	FLP expression plasmid. Confers tetracycline resistance.	Gene Bridges
pGFPuv	pUC backbone. Carries a variant of GFP. IPTG inducible vector. Confers ampicillin resistance.	Clontech
pHis$_6$-GFP	Carries GFPuv with six histidines at the N-terminus. Confers ampicillin resistance.	This study
pWHHHPH-GFP	Carries GFPuv with a WHHHPH tag at the N-terminus. Confers ampicillin resistance.	This study
pWT*crp*	Carries wild-type *crp*. Confers ampicillin resistance.	This study
pH17A*crp*	Carries *crp* with the H17A substitution. Confers ampicillin resistance.	This study
pH19Y*crp*	Carries *crp* with the H19Y substitution. Confers ampicillin resistance.	This study
pH17A-H19Y*crp*	Carries *crp* with the H17A and H19Y substitutions. Confers ampicillin resistance.	This study
pH17A-H19Y-K52N*crp*	Carries *crp* with the H17A, H19Y, K52N substitutions. Confers ampicillin resistance.	This study
pArg$_6$-GFP	Carries GFPuv with six arginines at the N-terminus. Confers ampicillin resistance.	Gift from Beitle

Table A.2. Primers used in this work.

Primers [a]	Sequence [b]	Nucleotide location
cyoA F1	5-TGGGATGGTTGTCATTATTTGCAGGCACTGTATTGC TCAGTGGCTGTAATAATTAACCCTCACTAAAGG GCG-3	Complement (450757-450806)
cyoA R1	5-ATGCCTTCCATACCTTCGTGTGCGCTGTGCTCACCT TCTGGCTGGGTCATTAATACGACTCACTATAGGGCTC-3	Complement (499918-499967)
yfbG F1	5-ATATGGGATGCCTCGGTATTGAAGCCCTGCTGGCTGC CGGTTACGAAATTATTAACCCTCACTAAAGGGCG-3	2366089-2366138
yfbG R1	5-CGCAGGAAGAAATCCAGCGTTTCGTCGATGGTTTCCT GCATATCAATTTTTAATACGACTCACTATAGGGCTC-3	2367963-2368012
adhP F1	5-ATCATGTTGACGTTACGTATAAAACACTGCGC TCAC TGAAACATGGCGAAAATTAACCCTCACTAAAGGGCG-3	Complement (1551795-1551834)
adhP R1	5-CGGATTTTGCCTTCTTCCATCTCAGTAAAGATGGTGT TGATGTCCGCTAATAATACGACTCACTATAGGGCTC-3	Complement (1550883-1550932)
cyoA F2	5- GGGATGGTTGTCATTATTTGCA-3	Complement (450784-450805)
cyoA R2	5-GGGCGGATTCCGCGTGGCTC-3	Complement (449863-449882)
yfbG F2	5- GCCTACCACGATATGGGATG-3	2366079-2336098
yfbG R2	5- CCGTAAGATCAACGGTGCGCAG-3	2368008-2368029
adhP F2	5-ATGAAGGCTGCAGTTGTTACG-3	Complement (1551842-1551862)
adhP R2	5-GTGACGGAAATCAATCACCATG-3	Complement (1550855-1550876)
kan F [c]	5-TATCAGGACATAGCGTTGGCTACC-3	
WHHHPH-GFP F	5-GCCAAGCTTGTGGCATCATCATCCGCATATGAGTAA AGGAGAAGAACTTTTC-3	N/A
WHHHPH-GFP R	5-TTGGAATTCATTATTTGTAGAGCT-3	N/A
WTcrp F	5-GGCATATGGTGCTTGGCAAACCGC-3	N/A
WTcrp R	5-CCGGATCCTTAACGAGTGCCGTAAAC-3	N/A
H17Acrp F	5-CGACTCTCGAATGGTTCTTGTCTGCCTGCC ACATTCATAAGTACCCATC-3	N/A
H19Ycrp F	5- GGTTCTTGTCTCATTGCTATATTCATAAGTACCCAT CC-3	N/A
H17A-H19Y crp F	5-GGTTCTTGTCTGCCTGCTATATTCATAAGTACCCAT CC-3	N/A
H17A-H19Y-K52N crp F	5- CTCTGTGGCAGTGCTGATCGATGACGAAGAGGGTAA AGAAATGATCC-3	N/A

a. F indicates the forward primers, and R indicates reverse primers.
b. Underlined sequences were PCR primers for amplification of FRT-PGK-gb2-neo-FRT cassette.
c. The sequence of Kan F was located in the kanamycin cassette.

Table A.3. Southern blot probes used in this work.

Southern blot Probes		Primers	Nucleotide location	Size
cyoA	F3	5-TGGCTGTTGGTTTCGCCTGG-3	Complement (450634-450653)	599 bp
	R3	5-GTATTCGCTAGGCGCGGCC-3	Complement (450055-450073)	
yfbG	F3	5-CCGCTGTGGGTGGAACGC-3	2366253-2366270	
	R3	5-CGCGCAGCAG GCGTTCTG-3	2367055-2367072	820 bp
adhP	F3	5-CCGGCGTAATTCTGGGCC-3	Complement (1551691-1551708)	741 bp
	R3	5-CGGCGGCAAACTGGAAGG-3	Complement (1550968-1550985)	

BIBLIOGRAPHY

1. Langer E, Ranck J. Capacity bottleneck squeezed by downstream processes. *Bioprocess Int.* 2006;4(3):14-18.

2. Tugcu N, Roush DJ, Göklen KE. Maximizing productivity of chromatography steps for purification of monoclonal antibodies. *Biotechnol Bioeng.* 2008;99(3):599-613.

3. Nfor BK, Ahamed T, van Dedem GWK, et al. Design strategies for integrated protein purification processes: challenges, progress and outlook. *J Chem Technol Biotechnol.* 2008;83:124-132.

4. Highsmith J. *Biologic Therapeutic Drugs: Technologies and Global Markets (BIO079A).* Wellesley, MA: BCC Research LLC;2011.

5. Thömmes J, Etzel M. Alternatives to Chromatographic Separations. *Biotechnol Prog.* 2007;23:42-45.

6. Shukla AA, Hinckley P. Host cell protein clearance during Protein A chromatography: development of an improved column wash step. *Biotechnol Prog.* 2008;24:1115-1121.

7. Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. Downstream processing of monoclonal antibodies- development of platform approaches. *J Chromatogr B.* 2007;848:28-39.

8. Aldridge S. Bioprocessing: downstream processing needs a boost. *Genet Eng News.* 2006;26(1):51.

9. Fahrner RL, Knudsen HL, Basey CD, et al. Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes. *Biotechnol Genet Eng Rev.* 2001;18:301-327.

10. Roque ACA, Lowe CR, Taipa MA. Antibodies and genetically engineered related molecules: production and purification. *Biotechnol Prog.* 2004;20(3):639-654.

11. Lightfoot EN, Moscariello JS. Bioseparations. *Biotechnol Bioeng.* 2004;87(3):259-272.

12. Lee SY. High cell-density culture of Escherichia coli. *Trends Biotechnol.* 1996;14:98-105.

13. Gottschalk U. Bioseparation in antibody manufacturing: The good, the bad and the ugly. *Biotechnol Prog.* 2008;24(3):496-503.

14. Porath J, Carlsson J, Olsson J, Belfrage G. Metal chelate affinity chromatogrpahy, a new approach to protein fractionation. *Nature.* 1975;258:598-599.

15. Ueda EKM, Gout PW, Morganti L. Current and prospective applications of metal ion-protein binding. *J Chromatogr A.* 2003;988:1-23.

16. Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif.* 2006;48(2006):1-13.

17. Becker K, Alstine JV, Bülow L. Multipurpose peptide tags for protein isolation. *J Chromatogr A.* 2008;1202:40-46.

18. Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. His-tag impact on structure. *Acta Crystallogr D.* 2007;63:295-301.

19. Hedhammar M, Gräslund T, Hober S. Protein engineering strategies for selective protein purification. *Chem Eng Technol.* 2005;28(11):1315-1325.

20. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein Expr Purif.* 2005;41:98-105.

21. McCluskey AJ, Poon GMK, Gariépy J. A rapid and universal tandem-purification strategy for recombinant proteins. *Protein Sci.* 2007;16:2726-2732.

22. Mohanty AK, Wiener MC. Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expr Purif.* 2004;33:311-325.

23. Nilsson J, Ståhl S, Lundeberg J, Uhlén M, Nygren P-Å. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr Purif.* 1997;11:1-16.

24. Patwardhan AV, Goud GN, Koepsel RR, Ataai MM. Selection of optimum affinity tags from a phage-displayed peptide library. Application to immobilized copper(II) affinity chromatography. *J Chromatogr A.* 1997;787:91-100.

25. Pedersen J, Lauritzen C, Madsen MT, Dahl SW. Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases. *Protein Expr Purif.* 1999;15:389-400.

26. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol.* 1999;17:1030-1032.

27. Steffens MA, Fraga ES, Bogle IDL. Synthesis of purification tags for optimal downstream processing. *Comput Chem Eng.* 2000;24:717-720.

28. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2003;60:523-533.

29. Uhlén M, Forsberg G, Moks T, Hartmanis M, Nilsson B. Fusion proteins in biotechnology. *Curr Opin Biotech.* 1992;3:363-369.

30. Waugh DS. Making the most of affinity tags. *Trends Biotechnol.* 2005;23(6):316-320.

31. Johnson RD, Todd RJ, Arnold FH. Multipoint binding in metal-affinity chromatography II. Effect of pH and imidazole on chromatographic retention of engineered histidine-containing cytochromes c. *J Chromatogr A.* 1996;725:225-235.

32. Patwardhan AV, Ataai MM. IMA chromatography of a cell extract: effect of loading pH on protein purification. *AIChE Journal.* 1997;43(12):3232-3240.

33. Patwardhan AV, Ataai MM. Site accessibility and the pH dependence of the saturation capacity of a highly cross-linked matrix. Immobilized metal affinity chromatography of bovine serum albumin on chelating superose. *J Chromatogr A.* 1997;767:11-23.

34. Todd RJ, Johnson RD, Arnold FH. Multiple-site binding interactions in metal-affinity chromatography I. Equlibrium binding of engineered histidine-containing cytochromes c. *J Chromatogr A.* 1994;662:13-26.

35. Vunnum S, Natarajan V, Cramer S. Immobilized metal affinity chromatography self-sharpening of protein-modular interfaces in frontal chromatography. *J Chromatogr A.* 1998;818:31-41.

36. Chaga GS. Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. *J Biochem Bioph Meth.* 2001;49(1-3):313-334.

37. Gaberc-Porekar V, Menart V. Perspectives of immoblized-metal affinity chromatography. *J Biochem Bioph Meth.* 2001;49(1-3):335-360.

38. Pasquinelli RS, Koepsel RR, Wu N, Ataai MM. Vector engineering anomalies: impact on fusion protein purification performance. *Protein Expr Purif.* 1999;17:449-455.

39. Porath J. Immobilized metal ion affinity chromatography. *Protein Expr Purif.* 1992;3:263-281.

40. *Ion Exchange Chromatography & Chromatofocusing*. AA ed: GE Healthcare; 2004.

41. Dismer F, Petzold M, Hubbuch J. Effects of ionic strength and mobile phase pH on the binding orientation of lysozyme on different ion-exchange adsorbents. *J Chromatogr A.* 2008;1194(1):11-21.

42. Finette GMS, Mao Q-M, Hearn MTW. Comparative studies on the isothermal characteristics of proteins adsorbed under batch equilibrium conditions to ion-exchange, immobilised metal ion affinity and dye affinity matrices with different ionic strength and temperature conditions. *J Chromatogr A.* 1997;763:71-90.

43. Follman DK, Fahrner RL. Factorial screening of antibody purification processes using three chromatography steps without protein A. *J Chromatogr A.* 2004;1024:79-85.

44. Ng MYT, Tan WS, Abdullah N, Ling TC, Tey BT. Direct purification of recombinant hepatitis B core antigen from two different pre-conditioned unclarified Escherichia coli feedstocks via expanded bed adsorption chromatography. *J Chromatogr A.* 2007;1172:47-56.

45. Tsonev LI, Hirsh AG. Theory and applications of a novel ion exchange chromatography technology using controlled pH gradients for separating proteins on anionic and cationic stationary phases. *J Chromatogr A.* 2008;1200:166-182.

46. Yi S, Brickenden A, Choy W-Y. A new protocol for high-yield purification of recombinant human prothymosin α expressed in Escherichia coli for NMR studies. *Protein Expr Purif.* 2008;57:1-8.

47. Londo T, Lynch P, Kehoe T, Meys M, Gordon N. Accelerated recombinant protein purification process development- automated, robotics-based integration of chromatographic purification and analysis. *J Chromatogr A.* 1998;798:73-82.

48. Degerman M, Jakobsson N, Nilsson B. Designing robust preparative purification processes with high performance. *Chem Eng Technol.* 2008;31(6):875-882.

49. Sassenfeld HM, Brewer SJ. A Polypeptide Fusion Designed for the Purification of Recombinant Proteins. *Bio/Technology.* 1984;2:76-81.

50. Brewer SJ, Sassenfeld HM. The purification of recombinant proteins using C-terminal polyarginine fusions. *Trends Biotechnol.* 1985;3(5):119-122.

51. Sassenfeld HM. Engineering proteins for purification. *Trends Biotechnol.* 1990;8:88-93.

52. Stempfer G, Höll-Neugebauer B, Kopetzki E, Rudolph R. A fusion protein designed for noncovalent immobilization stability, enzymatic activity, and use in an enzyme reactor. *Nat Biotechnol.* 1996;14:481-484.

53. Gräslund T, Ehn M, Lundin G, et al. Strategy for highly selective ion-exchange capture using a charge-polarizing fusion partner. *J Chromatogr A.* 2002;942:157-166.

54. Kweon D-H, Lee D-H, Han N-S, Rha CS, Seo J-H. Characterization of polycationic amino acids fusion systems for ion-exchange purification of cyclodextrin glycosyltransferase from recombinant Escherichia coli. *Biotechnol Prog.* 2002;18(2):303-308.

55. Humphreys DP, Heywood SP, King LM, Bowering LC, Turner JP, Lane SE. Engineering of Escherichia coli to improve the purification of periplasmic Fab' fragments: chaning the pI of the chromosomally encoded PhoS/PstS protein. *Protein Expr Purif.* 2004;37:109-118.

56. Fuchs SM, Raines RT. Polyarginine as a multifunctional fusion tag. *Protein Sci.* 2005;14:1538-1544.

57. Kim S-G, Kim J-A, Yu H-A, Lee D-H, Kweon D-H, Seo J-H. Application of poly-arginine fused minichaperone to renaturation of cyclodextrin glycosyltransferase expressed in recombinant Escherichia coli. *Enzyme Microb Tech.* 2006;39:459-465.

58. Hedhammar M, Hober S. Zbasic - A novel purification tag for efficient protein recovery. *J Chromatogr A.* 2007;1161:22-28.

59. Sereikaite J, Statkute A, Morkunas M, et al. Production of Recombinant mink growth hormone in E. coli. *Appl Microbiol Biotechnol.* 2007;74:316-323.

60. *Hydrophobic Interaction Chromatography: Princples and Methods*. AB ed: Amersham Pharmacia Biotech; 2000.

61. Hjertén S, Yao K, Eriksson K-O, Johansson B. Some general aspects of hydrophobic interaction chromatography. *J Chromatogr.* 1973;87:325-331.

62. McCue JT, Engel P, Ng A, Macniven R, Thömmes J. Modeling of protein monomer/aggregate purification and separation using hydrophobic interaction chromatography. *Bioprocess Biosyst Eng.* 2008;31:261-275.

63. Hrkal Z, Rejnkova J. Hydrophobic interaction chromatography of serum proteins on Phenyl-Sepharose CL-4B. *J Chromatogr.* 1982;242:385-388.

64. McNair RD, Kenny AJ. Proteins of the kidney microvillar membrane. The amphipathic form of dipeptidyl peptidase IV. *J Biochem.* 1979;179:379-395.

65. Lefort S, Ferrara P. Hydrophobic adsorbants for the isolation and purification of biosynthetics human growth hormone from crude fermentation mixtures. *J Chromatogr.* 1986;361:209-216.

66. Jennissen HP. Hydrophobic interaction chromatography. *Encyclopedia of Life Sciences.* 2002;9:353-361.

67. Lienqueo ME, Mahn A, Vásquez L, Asenjo JA. Methodology for predicting the separation of proteins by hydrophobic interaction chromatography and its application to a cell extract. *J Chromatogr A.* 2003;1009:189-196.

68. Przybycien TM, Pujar NS, Steele LM. Alternative bioseparation operations: life beyond packed-bed chromatography. *Curr Opin Biotech.* 2004;15(5):469-478.

69. Zeman L, Zydney AL. *Microfiltration and ultrafiltration: principles and applications*. New York: Marcel Dekker; 1996.

70. van Reis R, Gadam S, Frautschy L, et al. High performace tangential flow filtration. *Biotechnol Bioeng.* 1997;56(1):71-82.

71. Etzel M. Bulk protein crystallization- principles and methods. *Process Scale Bioseparations for the Biopharamaceutical Industry*. Boca Raton, FL: Taylor & Francis; 2006.

72. Peters J, Minuth T, Schröder W. Implementation of a crystallization step into the purification process of a recombinant protein. *Protein Expr Purif.* 2005;39:43-53.

73. Schmidt AS, Ventom AM, Asenjo JA. Partitioning and purification of α-amylase in aqueous two-phase systems. *Enzyme Microb Tech.* 1994;16:131-142.

74. Hart R, Lester PM, Reifsnyder DH, Ogez JR, Builder SE. Large scale, in situ isolation of periplasmic IGF-I from E. coli. *Bio/Technology.* 1994;12:1113-1117.

75. Makrides SC. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol Rev.* 1996;60(3):512-538.

76. Miroux B, Walker JE. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol.* 1996;260:289-298.

77. Murby M, Uhlén M, Ståhl S. Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli. *Protein Expr Purif.* 1996;7:129-136.

78. Baneyx F. Recombinant protein expression in Escherichia coli. *Curr Opin Biotech.* 1999;10:411-421.

79. Andersen DC, Krummen L. Recombinant protein expression for therapeutic applications. *Curr Opin Biotech.* 2002;13:117-123.

80. Shokri A, Sandén AM, Larsson G. Cell and process design for targeting of recombinant protein in the culture medium of Escherichia coli. *Appl Microbiol Biotechnol.* 2003;60:654-665.

81. Choi JH, Lee SY. Secretory and extracellular production of recombinant proteins using Escherichia coli. *Appl Microbiol Biotechnol.* 2004;64:625-635.

82. Schmidt FR. Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biotechnol.* 2004;65:363-372.

83. Georgiou G, Segatori L. Preparative expression of secreted proteins in bacteria status report and future prospects. *Curr Opin Biotech.* 2005;16:538-545.

84. Jana S, Deb JK. Strategies for the efficient production of heterologous proteins in Escherichia coli. *Appl Microbiol Biotechnol.* 2005;67:289-298.

85. Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J Biotechnol.* 2005;115:113-128.

86. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2006;72:211-222.

87. Tolia NH, Jashua-Tor L. Strategies for protein coexpression in Escherichia coli. *Nat Methods.* 2006;3(1):55-64.

88. Peti W, Page R. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. *Protein Expr Purif.* 2007;51:1-10.

89. Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem.* 2008;307:249-264.

90. Champion KM, Nishihara JC, Joly JC, Arnott D. Similarity of the Escherichia coli proteome upon completion of different biopharmaceutical fermentation processes. *Proteomics.* 2001;1:113-1148.

91. Duerrschmid K, Reischer H, Schmidt-Heck W, et al. Monitoring of transcriptome and proteome profiles to investigate the cellular response of E. coli towards recombinant protein expression under defined chemostat conditions. *J Biotechnol.* 2008;135(1):34-44.

92. Aldor IS, Krawitz DC, Forrest W, et al. Proteomic profiling of recombinant Escherichia coli in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. *Appl Environ Microbiol.* 2005;71(4):1717-1728.

93. Jin M, Szapiel N, Zhang J, Hickey J, Ghose S. Profiling of Host Cell Proteins by Two-Dimensional Difference Gel Electrophoresis (2D-DIGE): Implications for Downstream Process Development. *Biotechnol Bioeng.* 2010;105(2):306-316.

94. Grzeskowiak JK, Tscheliessnig A, Toh PC, et al. 2-D DIGE to expedite downstream process development for human monoclonal antibody purification. *Protein Expr Purif.* 2009;66(1):58-65.

95. Administration UFaD. Guidance for industry; Q8 pharmaceutical development. In: HHS, ed2006.

96. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol.* 2009;27(1):26-34.

97. Cabrera R, Zhelyazkova P, Galvis L, Fernandez-Lahore M. Tailoring orthogonal proteomic routines to understand separation during ion exchange chromatography. *J Sep Sci.* 2008;31:2500-2510.

98. Leser EW, Asenjo JA. Rational design for purification processes for recombinant proteins. *J Chromatogr.* 1992;584:43-57.

99. Lienqueo ME, Salgado JC, Asenjo JA. An expert system for selection of protein purification processes: experimental validation. *J Chem Technol Biot.* 1999;74:293-299.

100. Lienqueo ME, Asenjo JA. Use of expert systems for the synthesis of downstream protein processes. *Comput Chem Eng.* 2000;24:2339-2350.

101. Cai Y, Moore M, Goforth R, Henry R, Beitle R. Genomic data for alternate production strategies. I. Identification of major contaminating species for cobalt immobilized metal affinity chromatography. *Biotechnol Bioeng.* 2004;88(1):77-83.

102. Tiwari N, Woods L, Haley R, et al. Identification and characterization of native proteins of Escherichia coli BL-21 that display affinity towards Immobilized Metal Affinity Chromatography and Hydrophobic Interaction Chromatography Matrices. *Protein Expr Purif.* 2010;70(2):191-195.

103. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta.* 2006;1760:1304-1313.

104. Edwards A, Gräslund S, Nordlund P, et al. Protein production and purification. *Nat Methods.* 2008;5(2):135-146.

105. Pasquinelli RS, Shepherd RE, Koepsel RR, Zhao A, Ataai MM. Design of affinity tags for one-step protein purification from immobilized zinc columns. *Biotechnol Prog.* 2000;16:86-91.

106. Gassmann M, Grenacher B, Rohde B, Vogel J. Quantifying Western blots: Pitfalls of densitometry. *Electrophoresis.* 2009;30(11):1845-1855.

107. Luhn S, Berth M, Hecker M, Bernhardt J. Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images. *Proteomics.* 2003;3(7):1117-1127.

108. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc.* 2006;1(6):2856-2860.

109. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 1999;20(18):3551-3567.

110. Balasubramani M, Nakao C, Uechi GT, et al. Characterization and detection of cellular and proteomic alterations in stable stathmin-overexpressing, taxol-resistant BT549 breast cancer cells using offgel IEF/PAGE difference gel electrophoresis. *Mutat Res-Genet Toxicol Environ Mutag.* 2011;722(2):154-164.

111. Beitle R, Ataai MM. One-step purification of a model periplasmic protein from inclusion bodies by its fusion to an effective metal-binding peptide. *Biotechnol Prog.* 1993;9:64-69.

112. Lilius G, Persson M, Bulow L, Mosbach K. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. *Eur J Biochem.* 1991;198(2):499-504.

113. Goel A, Colcher D, Koo JS, Booth BJM, Pavlinkova G, Batra SK. Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *BBA-Gen Subjects.* 2000;1523(1):13-20.

114. Lawrence D, Shahrokh Z, Marsters S, et al. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nat Med.* 2001;7(4):383-385.

115. Patwardhan AV, Goud GN, Pasquinelli RS, Koepsel RR, Ataai MM. Phage-displayed libraries for the selection of optimal affinity peptides for protein purification using Ni-nitrilotriacetic acid chromatography. *Biotechnol Tech.* 1998;12(6):421-424.

116. Jiang CP, Wechuck JB, Goins WF, et al. Immobilized cobalt affinity chromatography provides a novel, efficient method for herpes simplex virus type 1 gene vector purification. *J Virol.* 2004;78(17):8994-9006.

117. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis.* 1997;18(11):2071-2077.

118. Rozanas CR, Loyland SA. Capabilities using 2-D DIGE in proteomics research - The new cold standard for 2-D gel electrophoresis. *Methods Mol Biol.* 2008:1-18.

119. Minden JS, Dowd SR, Meyer HE, Stuhler K. Difference gel electrophoresis. *Electrophoresis.* 2009;30:S156-S161.

120. Wülfing C, Lombardero J, Plückthun A. An Escherichia coli protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif. *J Biol Chem.* 1994;269(4):2895-2901.

121. Blum P, Ory J, Bauernfeind J, Krska J. Physiological consequences of DnaK and DnaJ overproduction in Escherichia coli. *J Bacteriol.* 1992;174(22):7436-7444.

122. Hatefi A, Megeed Z, Ghandehari H. Recombinant polymer-protein fusion: a promising approach towards efficient and targeted gene delivery. *J Gene Med.* 2006;8(4):468-476.

123. Liu Z, Bartlow P, Varakala R, Beitle R, Koepsel RR, Ataai MM. Use of proteomics for design of a tailored host cell for highly efficient protein purification. *J Chromatogr A.* 2009;1216(12):2433-2438.

124. Sarvas M. Mutant of Escherichia coli K-12 defective in d-glucosamine biosynthesis. *J Bacteriol.* 1971;105(2):467-471.

125. Wu HC, Wu TC. Isolation and characterization of a glucosamine-requiring mutant of Escherichia coli K-12 defective in glucosamine-6-phosphate synthetase. *J Bacteriol.* 1971;105(2):455-466.

126. Perrenoud A, Sauer U. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. *J Bacteriol.* 2005;187(9):3171-3179.

127. Gosset G, Zhang ZG, Nayyar SN, Cuevas WA, Saier MH. Transcriptome analysis of Crp-dependent catabolite control of gene expression in Escherichia coli. *J Bacteriol.* 2004;186(11):3516-3524.

128. Zubay G, Schwartz D, Beckwith J. Mechanism of activation of catabolite-sensitive genes: a positive control system. *P Natl Acad Sci USA.* 1970;66(1):104-111.

129. Emmer M, deCrombrugghe B, Pastan I, Perlman R. Cyclic AMP receptor protein of E. coli: its role in the synthesis of inducible enzymes. *P Natl Acad Sci USA.* 1970;66(2):480-487.

130. McKay DB, Steitz TA. Structure of catabolite gene activator protein at 2.9 Angstroms resultion suggests binding to left-handed *B*-DNA. *Nature.* 1981;290(5809):744-749.

131. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol.* Oct 22 1999;293(2):199-213.

132. Harman JG. Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta.* 2001;1547(1):1-17.

133. Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebright RH. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol.* 2004;14(1):10-20.

134. Schultz SC, Shields GC, Steitz TA. Crystal structure of the CAP-DNA complex - the DNA is bent by 90 degrees. *Science.* 1991;253(5023):1001-1007.

135. Bell A, Gaston K, Williams R, et al. Mutations that alter the ability of the Escherichia coli cyclic-amp receptor protein to activate transcription. *Nucleic Acids Res.* 1990;18(24):7243-7250.

136. Eschenlauer AC, Reznikoff WS. Escherichia coli catabolite gene activator protein mutants defective in positive control of Lac operon transcription. *J Bacteriol.* 1991;173(16):5024-5029.

137. Zhou YH, Zhang XP, Ebright RH. Identification of the activating region of catabolite gene activator protein (CAP) - Isolation and characterization of mutants of CAP specifically defective in transcription activation. *P Natl Acad Sci USA.* 1993;90(13):6081-6085.

138. Niu W, Zhou YH, Dong QP, Ebright YW, Ebright RH. Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). 1. Saturation and alanine-scanning mutagenesis. *J Mol Biol.* 1994;243(4):595-602.

139. Gaston K, Bell A, Kolb A, Buc H, Busby S. Stringent spacing requirements for transcription activation by CRP. *Cell.* 1990;62(4):733-743.

140. Williams RM, Rhodius VA, Bell AI, Kolb A, Bushy SJW. Orientation of functional activating regions in the Escherichia coli CRP protein during transcription activation at class II promoters. *Nucleic Acids Res.* 1996;24(6):1112-1118.

141. Niu W, Kim Y, Tau G, Heyduk T, Ebright RH. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell.* 1996;87(6):1123-1134.

142. Williams R, Bell A, Sims G, Busby S. The role of 2 surface exposed loops in transcription activation by the Escherichia coli CRP and FNR proteins. *Nucleic Acids Res.* 1991;19(24):6705-6712.

143. Lonetto MA, Rhodius V, Lamberg K, Kiley P, Busby S, Gross C. Identification of a contact site for different transcription activators in region 4 of the Escherichia coli RNA polymerase sigma(70) subunit. *J Mol Biol.* 1998;284(5):1353-1365.

144. Rhodius VA, Busby SJW. Interactions between activating region 3 of the Escherichia coli cyclic AMP receptor protein and region 4 of the RNA polymerase sigma(70) subunit: Application of suppression genetics. *J Mol Biol.* Jun 2 2000;299(2):311-324.

145. Jiang CP, Liu J, Rubacha M, Shukla AA. A mechanistic study of Protein A chromatography resin lifetime. *J Chromatogr A.* 2009;1216(31):5849-5855.

146. Shukla AA, Thommes J. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends Biotechnol.* 2010;28(5):253-261.

147. Graslund T, Lundin G, Uhlen M, Nygren PA, Hober S. Charge engineering of a protein domain to allow efficient ion-exchange recovery. *Protein Eng.* 2000;13(10):703-709.

148. Kiraga J, Mackiewicz P, Mackiewicz D, et al. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics.* 2007;8:163.

149. Cagney G, Amiri S, Premawaradena T, Lindo M, Emili A. In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Science.* 2003;1.

150. Hjerten S, Yao KQ, Eriksson KO, Johansson B. Gradient and isocratic high-performance hydrophobic interaction chromatography of proteins on agarose columns. *J Chromatogr.* May 30 1986;359:99-109.

151. Ahamed T, Chilamkurthi S, Nfor BK, et al. Selection of pH-related parameters in ion-exchange chromatography using pH-gradient operations. *J Chromatogr A.* 2008;1194(1):22-29.

152. Bradbury AJ, Gruer MJ, Rudd KE, Guest JR. The second aconitase (AcnB) of Escherichia coli. *Microbiol-UK.* 1996;142:389-400.

153. Li M, Ho PY, Yao SJ, Shimizu K. Effect of sucA or sucC gene knockout on the metabolism in Escherichia coli based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by C-13-labeling experiments. *Biochem Eng J.* 2006;30(3):286-296.

154. Lee J, Goel A, Ataai MM, Domach MM. Flux adaptations of citrate synthase deficient Escherichia coli. *Biochemical Engineering Viii.* 1994;745:35-50.

155. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of transfer-RNA synthetases into 2 classes based on mutually exclusive sets of sequence motifs. *Nature.* 1990;347(6289):203-206.

156. Antonio M, McFerran N, Pallen MJ. Mutations affecting the Rossman fold of isoleucyl-tRNA synthetase are correlated with low-level mupirocin resistance in Staphylococcus aureus. *Antimicrob Agents Ch.* 2002;46(2):438-442.

157. Sprenger GA, Schorken U, Sprenger G, Sahm H. Transaldolase B of Escherichia coli K-12 - cloning of its gene, talB, and characterization of the enzyme from recombinant strains. *J Bacteriol.* 1995;177(20):5930-5936.

158. Miller DL, Weissbac.H. Interactions between elongation factors: displacement of GDP from Tu-GDP complex by factor Ts. *Biochem Biophys Res Commun.* 1970;38(6):1016-1022.

159. Duval V, Nicoloff H, Levy SB. Combined Inactivation of lon and ycgE Decreases Multidrug Susceptibility by Reducing the Amount of OmpF Porin in Escherichia coli. *Antimicrob Agents Ch.* 2009;53(11):4944-4948.

160. Fayet O, Ziegelhoffer T, Georgopoulos C. The GroES and GroEL heat-shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. *J Bacteriol.* 1989;171(3):1379-1385.

161. Zuurmond AM, Rundlof AK, Kraal B. Either of the chromosomal tuf genes of E-coli K-12 can be deleted without loss of cell viability. *Mol Gen Genet.* 1999;260(6):603-607.

162. Miller CG, Schwartz G. Peptidase-deficient mutants of Escherichia coli. *J Bacteriol.* 1978;135(2):603-611.

163. Schroeder U, Henrich B, Fink J, Plapp R. Peptidase D of Escherichia coli K-12, a metallopeptidase of low substrate specificity. *FEMS Microbiol Lett.* 1994;123(1-2):153-159.

164. Henrich B, Monnerjahn U, Plapp R. Peptidase D gene (pepD) of Escherichia coli K-12: nucleotide sequence, transcript mapping, and comparison with other peptidase genes. *J Bacteriol.* 1990;172(8):4641-4651.

165. Utter MF, Kolenbrander HM. *The Enzymes.* Vol 6. 3rd ed. New York and London: Academic Press, Inc.; 1972.

166. Sauer U, Eikmanns BJ. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev.* 2005;29(4):765-794.

167. Yang C, Hua Q, Baba T, Mori H, Shimizu K. Analysis of Escherichia coli anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout. *Biotechnol Bioeng.* 2003;84(2):129-144.

168. Noronha SB, Yeh HJC, Spande TF, Shiloach J. Investigation of the TCA cycle and the glyoxylate shunt in Escherichia coli BL21 and JM109 using C-13-NMR/MS. *Biotechnol Bioeng.* 2000;68(3):316-327.

169. Kato A, Maki K, Ebina T, Kuwajima K, Soda K, Kuroda Y. Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers.* 2007;85(1):12-18.

170. Gu D, Zhou YZ, Kallhoff V, Baban B, Tanner JJ, Becker DF. Identification and characterization of the DNA-binding domain of the multifunctional PutA flavoenzyme. *J Biol Chem.* 2004;279(30):31171-31176.

171. Mihara H, Hidese R, Yamane M, Kurihara T, Esaki N. The iscS gene deficiency affects the expression of pyrimidine metabolism genes. *Biochem Biophys Res Commun.* 2008;372(3):407-411.

172. Schwartz CJ, Djaman O, Imlay JA, Kiley PJ. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. *P Natl Acad Sci USA.* 2000;97(16):9009-9014.

173. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol.* 2004;22(11):1399-1408.

174. Teter SA, Houry WA, Ang D, et al. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell.* 1999;97(6):755-765.

175. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature.* 1999;400(6745):693-696.

176. Hoffmann A, Bukau B, Kramer G. Structure and function of the molecular chaperone Trigger Factor. *BBA-Mol Cell Res.* 2010;1803(6):650-661.

177. Qin Y, Polacek N, Vesper O, et al. The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosorne. *Cell.* 2006;127(4):721-733.

178. Dibb NJ, Wolfe PB. Lep operon proximal gene is not required for growth or secretion by Escherichia coli. *J Bacteriol.* 1986;166(1):83-87.

179. Shoji S, Janssen BD, Hayes CS, Fredrick K. Translation factor LepA contributes to tellurite resistance in Escherichia coli but plays no apparent role in the fidelity of protein synthesis. *Biochimie.* 2010;92(2):157-163.

180. Schulze H, Nierhaus KH. Minimal set of ribosomal components for reconstitution of the peptidyl transferase activity. *EMBO J.* 1982;1(5):609-613.

181. Herbst KL, Nichols LM, Gesteland RF, Weiss RB. A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene 60 from bacteriophage T4. *P Natl Acad Sci USA.* 1994;91(26):12525-12529.

182. Brodersen DE, Nissen P. The social life of ribosomal proteins. *FEBS J.* 2005;272(9):2098-2108.

183. Horne JR, Erdmann VA. Isolation and characterization of 5S RNA-protein complexes from Bacillus stearothermophilus and Escherichia coli ribosomes. *Mol Gen Genet.* 1972;119(4):337-344.

184. Harms J, Schluenzen F, Zarivach R, et al. High resolution structure of the large ribosomal subunit from a mesophilic Eubacterium. *Cell.* 2001;107(5):679-688.

185. Douthwaite S, Garrett RA, Wagner R, Feunteun J. Ribonuclease-resistant region of the 5S RNA and its relation to the RNA-binding sites of proteins L18 and L25. *Nucleic Acids Res.* 1979;6(7):2453-2470.

186. Gongadze GM, Meshcheryakov VA, Serganov AA, et al. N-terminal domain, residues 1-91, of ribosomal protein TL5 from Thermus thermophilus binds specifically and strongly to the region of 5S rRNA containing loop E. *FEBS Lett.* 1999;451(1):51-55.

187. Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J. 5S ribosomal RNA database. *Nucleic Acids Res.* 2002;30(1):176-178.

188. Fedorov R, Meshcheryakov V, Gongadze G, et al. Structure of ribosomal protein TL5 complexed with RNA provides new insights into the CTC family of stress proteins. *Acta Crystallogr D.* 2001;57:968-976.

189. Gongadze GM, Korepanov AP, Stolboushkina EA, et al. The crucial role of conserved intermolecular H-bonds inaccessible to the solvent in formation and stabilization of the TL5 center-5 SrRNA complex. *J Biol Chem.* 2005;280(16):16151-16156.

190. Korepanov AP, Gongadze GM, Garber MB, Court DL, Bubunenko MG. Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in Escherichia coli. *J Mol Biol.* 2007;366(4):1199-1208.

191. Stoldt M, Wohnert J, Ohlenschlager O, Gorlach M, Brown LR. The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition. *EMBO J.* 1999;18(22):6508-6521.

192. Stoldt M, Wohnert J, Gorlach M, Brown LR. The NMR structure of Escherichia coli ribosomal protein L25 shows homology to general stress proteins and glutaminyl-tRNA synthetases. *EMBO J.* 1998;17(21):6377-6384.

193. Lu M, Steitz TA. Structure of Escherichia coli ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Angstrom resolution. *P Natl Acad Sci USA.* 2000;97(5):2023-2028.

194. Korobeinikova AV, Gongadze GM, Korepanov AP, Eliseev BD, Bazhenova MV, Garber MB. 5S rRNA-recognition module of CTC family proteins and its evolution. *Biochemistry-Moscow.* 2008;73(2):156-163.