Theses and Dissertations

12-2011

Proteome Based Development of Novel Affinity Tail for Immobilized Metal Affinity Chromatography and Hydrophobic Interaction Chromatography

Neha Tiwari
University of Arkansas, Fayetteville

PROTEOME BASED DEVELOPMENT OF NOVEL AFFINITY TAIL FOR IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY AND HYDROPHOBIC INTERACTION CHROMATOGRAPHY

PROTEOME BASED DEVELOPMENT OF NOVEL AFFINITY TAIL FOR IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY AND HYDROPHOBIC INTERACTION CHROMATOGRAPHY

**A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology**

By

**Neha Tiwari
Gujarat University
Bachelor of Science in Microbiology, 2003
Gujarat University
Master of Science in Microbiology, 2005**

**December 2011
University of Arkansas**

ABSTRACT

At industrial scale, reducing the steps in purification and recovery is desired; this not only decreases the cost but also increases the yield. Hydrophobic Interaction Chromatography (HIC) and Immobilized Metal Affinity Chromatography (IMAC) are not harsh on biological structure or activity of proteins; also both the techniques are economical and therefore a suitable choice at industrial level. This dissertation comprises of three parts. The purpose for the Part I was to identify and characterize *Escherichia coli* proteins which display affinity towards both IMAC and Hydrophobic Interaction Chromatography (HIC). Co (II) IMAC was chosen as the primary capture step, followed by HIC employing different concentrations of salt to promote adsorption. Nine out of the ten proteins identified have iso-electric values less than six, and half are considered nonessential. These data indicated that the combination of IMAC and HIC could be developed as a potent method for the purification of recombinant proteins by judicious choice of the salt concentration used to promote HIC, the development of *E. coli* strain(s) deficient in certain genomic proteins, and the design of an IMAC - HIC affinity tail for recombinant protein isolation based on the very proteins deleted from the genome.

Part II comprises of study of *E. coli* Alpha Galactosidase (*MelA*) as a candidate for affinity tag for this dual step purification approach. Affinity of *MelA towards* Co (II) IMAC and HIC was confirmed by fusing with a small peptide. Metal binding motif and various salt binding motifs of *MelA* were identified and studied. A *MelA* based affinity tag NT1 was developed by identifying the major salt binding region for HIC.

It is desirable to have a tag which doesn't interfere with solubility or biological activity of the fused protein. Solubility has been the biggest issue with the development of an affinity tag for HIC. In part III, with successful use of novel proteome, we report development of *MelA* as a successful affinity tag for HIC especially as it does not affect the expression, solubility or biological activity of the fused protein, with very few co-eluting proteins as contaminants.

This dissertation is approved for recommendation to the Graduate Council

Dissertation Director:

_____

Dr. Robert R. Beitle

Dissertation Committee:

_____

Dr. Ralph Henry

_____

Dr. Christa Hestekin

_____

Dr. Jim Woo Kim

DISSERTATION DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this dissertation when needed for research and/or scholarship.

Agreed _____
         Neha Tiwari

Refused _____
         Neha Tiwari

ACKNOWLEDGEMENTS

I would first like to thank Dr. Bob Beitle, my mentor, for all the guidance and support he has given me during this work. Without your patience (when I was crazy at times) and excellence as teacher, this work would not have been possible. You are a great advisor, and a wonderful human being. I am fortunate to be one of your graduate students, thank you for everything.

I want to thank all my committee members. You all have been very supportive and encouraging. I am grateful for all the advice and guidance that you have given me during committee meetings and research work. I especially want to thank Dr. Ralph Henry for providing very useful ideas for my project and making his lab available to me for some of my experiments.

I would like to thank Alicia Kight. She was my guide for all the molecular biology work. She has been very kind and patient in answering all my queries and lending reagents for my experiments. I want to thank Dr. Robyn Goforth for providing advice and analysis on lots of topics to me.

Thanks to all the members of Beitle lab, both past and present members. I want to thank Ryan, Tammy, Mckinzie and Ellen for their support and friendship they have given me for all these years. I want to thank all my fellow graduate students in department, with whom I have shared ups and downs of research. I appreciate all the faculties and employees of R. M Department of Chemical Engineering for their support. I am grateful to Dr. Amit Ghosh (India) who gave his continuous support to me to pursue higher studies.

And finally I want to thank my Guru and my family. Guruji has given me strength and stability all the time. I am very grateful to my parents Shashikant and Vijaya Tiwari, my husband Sunny Yadav, and all family members and friends back in India for their valuable support throughout.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURES

LB: Luria Berteni
ECP: Eluting Contaminant Proteins
TCP: Total Contaminant Proteins
HIC: Hydrophobic Interaction Chromatography
IMAC: Immobilized Metal Affinity Chromatograpy
IEX: Ion Exchange
CEX: Cation Exchange
SDS: Sodium Dodecyl Sulphate
PAGE: Poly-Acrylamide Gel Electrophoresis
GST: Glutathione-S-Transferase
MBP: Maltose Binding Protein

CHAPTER 1

Background Studies and Literature Review

1.1 Protein Expression and Purification

A single bacterial cell makes thousands of proteins required for different cellular processes during its life cycle. Many proteins are very important from a therapeutic point of view. Purification and recovery of these products is one of the most important phases of a bioprocess. When dealing with living cells, purification of desired product is all the more challenging (Nfor, Ahamed et al. 2008). In bioprocesses, development of better approaches is always desired for the purification of a desired product. With the development of recombinant DNA technology, it became easy to produce some of human proteins in *E. coli*, thus giving controlled and scalable ways for protein production (Ferrer-Miralles, Domingo-Espin et al. 2009). Amongst the gram-negative bacteria, *E. coli* is the most frequently employed organism for production of recombinant products like pharmacological proteins, enzymes and plasmid DNA. Most of the studies have been done in *E. coli* at genetic and physiology level. At the industrial level, these proteins are produced at a much larger scale and highest recovery and purity of the product is essential. A typical bioprocess can be divided in-to upstream and downstream processes. Upstream processes deal with expression and production of protein while downstream processes are mainly concerned with recovery and purification. Purification of proteins is one of the most challenging steps in bioprocesses at industrial level. About 50-60% of the total cost of production is spent on downstream processes. Along with the purity, it is essential that these bio-products retain their biological role. Thus overall economics of an industrial bioprocess depends a lot on the purification and recovery steps (Kalyanpur 2002). Research in this field has led to

development of better expression and host systems, better mRNA stability, and changes in biochemical pathways at genetic level to reduce byproduct production, post translational modifications, and decreasing protease activity as a part of strain development (Lesley 2001; Mann and Jensen 2003). Various properties of proteins like molecular size, shape, hydrophobicity, solubility, biological activities, and affinity are employed to purify them.

Economic concerns involved in purification prompts to adopt approaches which can reduce the cost involved. Sometimes more purification steps not only increase the cost but also decrease the yield due to loss of product at each step of recovery. Various methods like precipitation, membrane filtration, centrifugation, etc are used for recovery and purification (Charcosset 1998; van Reis and Zydney 2001). A single method is usually not adequate for purification at such a large scale, where low cost and purity is desired. This is why more economical and efficient approaches have been, for all time, a subject of interest for researchers. But chromatographic techniques had come up as a real winner for purification purpose because it exploits different biological or physical properties of proteins for separation and we can exhibit high recovery and purification as such.

1.2 Immobilized Metal Affinity Chromatography

IMAC was introduced by Porath, which is based on the specific property of the proteins showing affinity towards chelated metal ions like Co^{+2}; Ni^{+2} etc (Porath 1988; Porath 1988; Porath 1992). IMAC has emerged as a major technique for protein purification, in past 20 years from research to industrial applications, because of its property of easy scale up for large scale application (Gutierrez, del Valle et al. 2007; Vorackova, Suchanova et al. 2011). Different types of dentate matrices are used for IMAC but most commonly used are tridentates, e.g iminodiacetic acid (IDA) and tetradentates, e.g nitrilotriacetic acid (NTA) due to low cost and easy availability (Figure1-1).
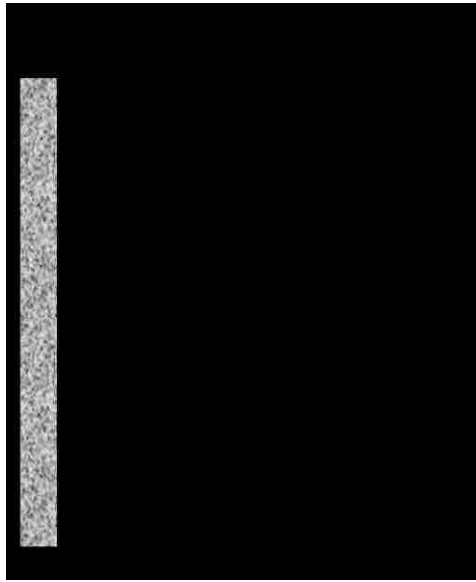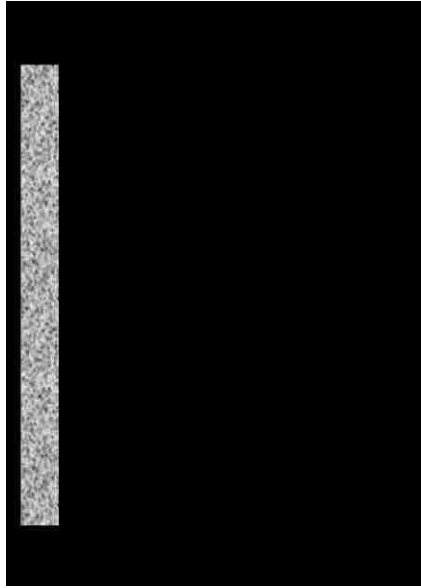
Figure 1-1: Commonly used matrices for IMAC

IMAC is based on the specific coordinate covalent binding of amino acids, particularly histidine to metals ions (Figure 1-2). IMAC is popular technique due to many advantages it possesses. It has high protein loading capacities compared to other affinity chromatographies. It doesn't generally affect the structure of proteins and thus does not hinder biological activity. IMAC is also used to concentrate diluted proteins. A huge advantage of IMAC is that it can be regenerated easily and a charged column can be reused several times.

Sepharose
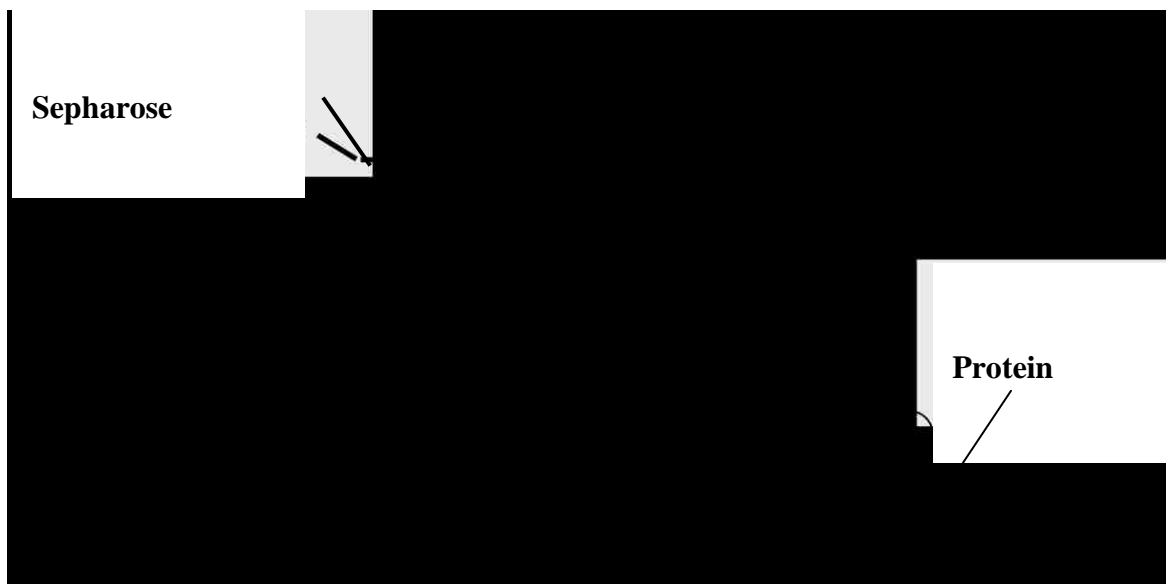
Protein

Figure 1-2: Immobilized Metal Affinity Chromatography

Though tryptophan and cysteine residues also play a role in metal binding, histidine is the key player for IMAC (Hochuli, Dobeli et al. 1987; Miller, Hill et al. 1991). The adsorption of proteins is based on the coordination between metal ion and the electron donor group for the amino acids like histidine on the surface of proteins. Retention behavior of proteins is largely governed by the histidine residues present on the surface. One histidine is enough for weak binding for IDA-Cu (II), but more histidine residues in proximity are needed for strong binding to Zn (II), Co (II) and Ni (II). The bound protein is usually eluted by pH gradient or imidazole challenge.

Hochuli et al showed that six histidine residues can be tagged to the desired protein and the fused protein can be successfully purified using IMAC. Commercial expression vectors are available with His6 and His10 tag, so that desired protein can be easily cloned and purified using IMAC (Hochuli, Dobeli et al. 1987). Although different vectors with His6 and His10 at C-and N-terminal are available, His6 at the N-terminal of protein is most common approach for IMAC purification. It has been shown by phage-displayed library that His2 also can be used for IMAC purification. It has been shown that isolated SPHHGG-phage fusion eluted with less contaminant peptides compared to affinity tails with higher histidine residues (Patwardhan, Goud et al. 1997; Mooney, Fredericks et al. 2011). This opens doors for research to look for sequences with less number of histidine but appropriate orientation, which can be used for purification with more efficient binding and less contaminants. Examples of some of the important proteins which have been effectively purified by IMAC are Thrombin, Immunoglobulins, Lysozyme, human

growth hormone, etc (Aizawa, Winge et al. 2008). His6 tag gives highly efficient purification of the fused protein, but one of the disadvantages is that , there are about 15-20 co-eluting proteins which show natural affinity towards metals due to surface exposed histidine residues in their structure.

1.3 Hydrophobic Interaction Chromatography

Hydrophobic Interaction Chromatography was first introduced in 1973 by *Hjerten et al* (Hjerten 1973; Hjerten, Rosengren et al. 1974), when they tried to separate proteins on carbohydrate gel matrices. Since then with research and development it has been established as a powerful tool for protein separation at both research and industrial scale. Hydrophobic interactions play very important role in protein structure and folding and also play important role in biological activity like enzyme –substrate reaction, antigen-antibody interaction etc. Most of hydrophobic core is usually buried inside the structure but some hydrophobic amino-acids are displayed on the surface of the proteins and they are also responsible for hydrophobic binding. HIC separation and retention is based on the hydrophobicity of proteins. Hydrophobic regions of proteins are generally buried inside the protein structure. In presence of salt buffers, the protein structure opens up a little bit, and the buried hydrophobic regions along with the surface exposed regions are exposed to the surface (Queiroz, Tomaz et al. 2001; Mahn and Asenjo 2005). During HIC, there is a salt promoted hydrophobic interaction between the hydrophobic ligands and the hydrophobic regions of proteins (Figure 1-3). Hydrophobic interactions play an important role in proteins resulting in stable and correct folding of proteins. Van der Waal's forces are main forces behind hydrophobic interactions, and the purification is achieved with minimum structural damage or de-naturation. Hydropathicity of proteins are dependent on the total of individual

7

hydrophobicity of amino-acids. Hydrophobicity of all amino acids was determined by measuring their individual solubility in water and in organic solvents like phenol (Kyte and Doolittle 1982; Kyte 2003).
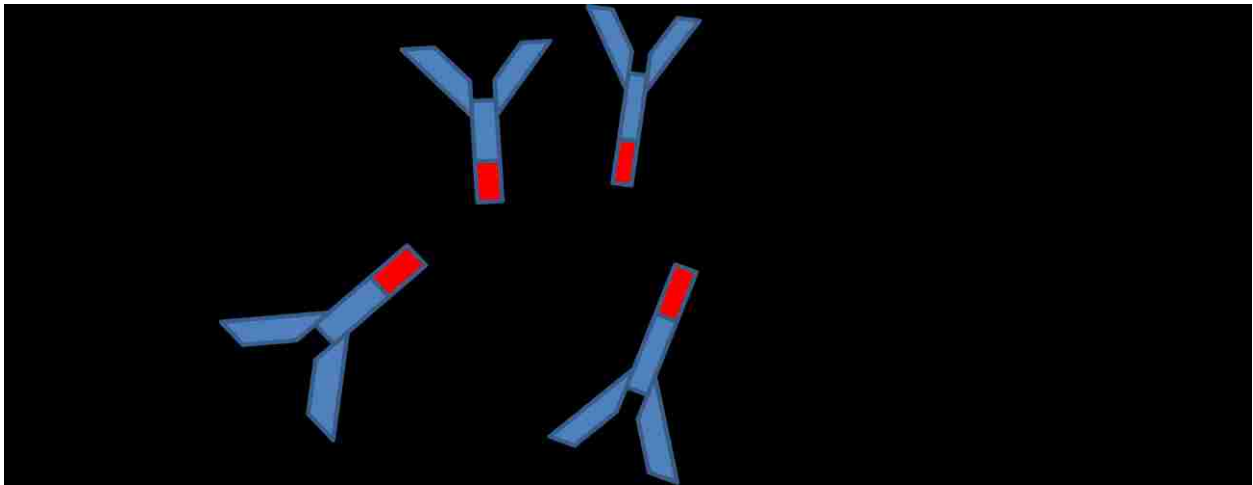
Figure 1-3: Hydrophobic Interaction Chromatography

Different salts like sodium chloride and sodium phosphate are used in the buffer solution to promote hydrophobic interactions. There are various theories behind the binding mechanism of

protein to HIC column. Use of the salt in the binding buffer during HIC allows opening up the protein fold up to some safe level and exposing the hydrophobic region to the column. The bound proteins are eluted by either decreasing salt concentration or in absence of salt. The salts present in binding buffer increase the free energy associated with protein, when the proteins come in contact with hydrophobic stationary phase, and binding to stationary phase decreases the surface contact area between the protein and the salt and thus allows little increase in free energy.

The commonly used HIC ligands are linear chain alkanes with or without a terminal amino group (Figure 1-4). One of the disadvantages is that sometimes these ligands promote very strong interactions with proteins and adsorption is so strong that denaturing conditions or chaotropic agents are used for elution. This can denature the biological activity of proteins; in such cases ligands with intermediate or mild hydrophobicity are more useful. For this project, we used sepharose hydrophobic columns provided by GE Healthcare. HIC is used for purification of biomolecules like serum proteins, nuclear proteins etc. Many times HIC is coupled with other chromatographic techniques like ion exchange chromatography.
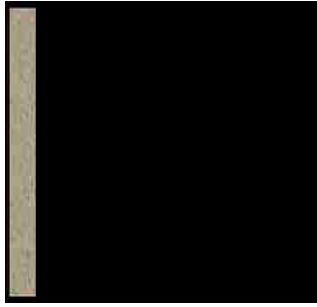
Butyl
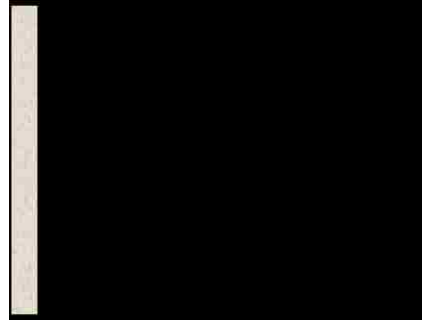
Alkyl

Octyl

PEG

Phenyl
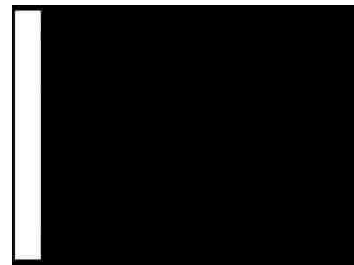
PPG

Figure 1-4: Commonly used ligands for HIC

Development of different stationary phases has increased the use of HIC for separating sensitive bio-molecules like proteins and also DNA. HIC being a mild technique does not hinder the protein structure and thus maintains its biological activity. HIC with all its advantages is a great choice for industrial application for purification. But unlike IMAC there is no affinity tag available for easy purification of desired protein using HIC. Like repeated histidines residues are to develop metal affinity for IMAC to purify the fused protein, efforts have been done to use repeated sequence of hydrophobic amino-acids like tryptophan to work as tag for HIC. But it makes the fused protein insoluble and expression of soluble protein is very low. In this project we have tried to look for a sequence from the native proteins of *E. coli* which can be used to develop an affinity tag for purifying desired protein using HIC with no solubility or expression issues.

1.4 Affinity Tags

Peptide sequences and proteins used to purify other fused proteins are called affinity tags.

These tags aid in purification of the proteins lacking affinity towards the particular column (Figure 1-5). Some tags aid in purification while some aid in solubility of the desired protein. Use of fusion tags has been proved as most effective in improving the expression, solubility. Affinity tags have proved to be an efficient tool for purification of proteins independent of their biochemical characteristics. With high affinity towards ligand and suitable immobilized supports available, small affinity tag sequences allow

proficient purification of fused protein (Waugh 2005; Arnau, Lauritzen et al. 2006; Becker, Van Alstine et al. 2008; Li 2010).

Use of affinity tag is one of the most revolutionary discoveries for protein purification. His6 tag is one of most widely used tag for one step purification of desired fused protein using IMAC. Various fusion tags are available nowadays which employ different modes of interaction with the resins and buffers, like Glutathione S Transferase (GST) (Smith and Johnson 1988; Frangioni and Neel 1993) and Maltose binding protein (MBP) (Diguan, Li et al. 1988; Maina, Riggs et al. 1988) which increase the yield and solubility of recombinant proteins, such tags are called solubility tags (Table1-1). After purification these tags can be easily cleaved by various proteases available e.g thrombin, Factor Xa etc.

Gene

Fuse DNA encoding 'tag' (3' or 5')

Gene fusion

Express   fusion protein

Fused protein with affinity tag

Purify fusion protein using affinity chromatography

Cleave tag from protein if required

Desired protein
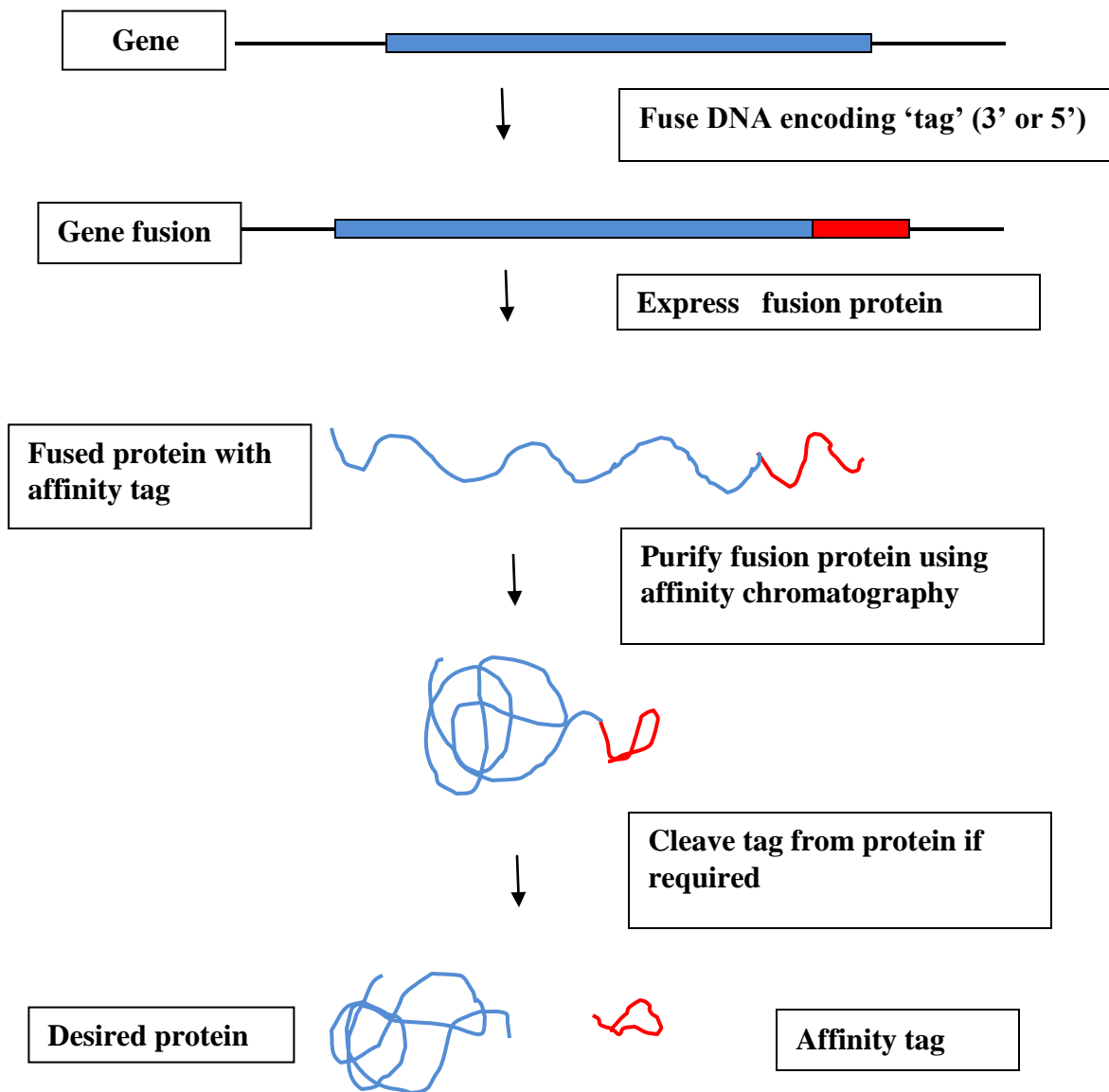
Affinity tag

Figure 1-5: Protein purification using affinity tag

Tags	Size	Uses	Capacity
His tag	6 or 8 amino acid	Purification	5-14mg/ml
FLAG	8 amino acid	Purification	0.6mg/ml
GST	211 amino acid	Purification and solubility	10mg/ml
MBP	396 amino acid	Purification and solubility	3mg/ml
Calmodulin binding domain	26 amino acid	Purification	2mg/ml
Strep II	8 amino acid	Purification	50-100nmol/ml

Table 1 -1: Commonly used tags for purification and solubility

For IMAC poly histidine tags are most commonly used and it can bind with dissociation constant of 10^{-5} to 10^{-7} but there is no successful tag available for HIC till today. Also, there is no affinity tag developed for combined use for Co (II) IMAC and HIC together as method of purification. In this project, we intend to develop a set of affinity tags by using proteome information, for application for HIC individually and combination of HIC and IMAC.

PART I

CHAPTER 2

Identification and Characterization of native proteins of *E.coli* BL-21 that display affinity towards Immobilized Metal Affinity Chromatography and Hydrophobic Interaction Chromatography Matrices

2.1 Introduction

In bioprocess development, design of purification strategies has always been an area of intense study as it requires new and better approaches. Understanding the proteome of the cell can aid in the design for both laboratory and industrial applications. Knowledge of target and undesired protein gives logical tactic to develop methods for better protein purification. The term proteomics refers to a protein level analysis of gene and cellular function in order to better understand the biochemistry of organisms (Lyddiatt and O'Sullivan 1998; Cai, Moore et al. 2004). It may be extended to the analysis of bioprocessing steps in order to gain insight that may be exploited for process improvement. In this study, we report on the major proteins associated with the eluting contaminant pool (ECP) when Immobilized Metal Affinity Chromatography and Hydrophobic Interaction Chromatography are operated in a series configuration. IMAC is one of the most popular techniques for purification of proteins, exploiting the metal binding property of certain amino acids, in particular histidine, at conditions of high sodium chloride concentration (Gaberc-Porekar and Menart 2001; Zatloukalova and Kucerova 2004). Using IMAC as an initial capture step is commonly employed in bio-separations when the protein has native metal binding ability or has been endowed with this property via addition of an affinity tail including N- or C-terminal extensions with six histidine residues (His_6). There has been a desire to pair IMAC with techniques that exploit the hydrophobic nature of protein. HIC exploits non-polar group

interaction under conditions of high salt concentration. Both IMAC and HIC are tolerant of a variety of compatible salts, hence the complementary nature of the two.

It was hypothesized that combining IMAC with HIC would display a small set of adsorbed contaminants that are responsible for both a reduction in column capacity and product specific gradient design. An ECP is defined as proteins which bind and elute under typical operating conditions. Knowledge concerning this protein set can lend insight into new schemes of bioprocessing due to a priori information regarding the contaminant pool and how it may impact choices related to purification. In addition to traditional methods of proper salt choice and combination to minimize ECP adsorption, one can exploit novel methods that include knockout or mutation of certain genes responsible for their expression in order to encounter a smaller ECP, and finally, design of an appropriate affinity tail based on ECP adsorption to shift recovery towards areas of the gradient with few contaminants. The latter, *viz*. design of a host strain with accompanying affinity tag permits a new avenue that is a direct result of the combination of molecular biology and proteomics. Figure 2-1 describes a basic strategy to combine both strain development and affinity tag design by exploiting knowledge of the ECP. Cell lysate, in this case soluble proteins from *Escherichia coli*, is passed through an IMAC column which is developed by reduction in pH, addition of imidazole, or other appropriate method to generate an ECP associated solely with IMAC. After mobile phase adjustment, this pool is then fed to a HIC column, and the final protein pool that elutes from the HIC column is subjected to proteome analysis. After proteins in the final ECP are identified, an assessment based either on available literature or growth data defines a subset which may be deleted from the genome for strain improvement, or altered so they are not captured during the first recovery step. The concept of

gene knockout / modification to provide a minimal contaminant pool has been reported on previously (Liu, Bartlow et al. 2009). A further extension of this concept would be to use proteome-based information in the development of affinity tag(s) based on the ECP. In principle, an entire ECP protein or a portion thereof could be used as an affinity tail. When fused to a recombinant target protein, this protein, originally in the ECP but now absent from the genome by prior knockout, directs purification. This results in exploiting binding properties for *a priori* design of the affinity tail. A complete optimization of an IMAC - HIC system would thus incorporate modification to the genome accompanied by a plasmid expression system incorporating an ECP protein affinity tag. We report on the initial stages of developing an IMAC - HIC affinity system, focusing on proteins found in the ECP that display useful properties (especially Alpha galactosidase).
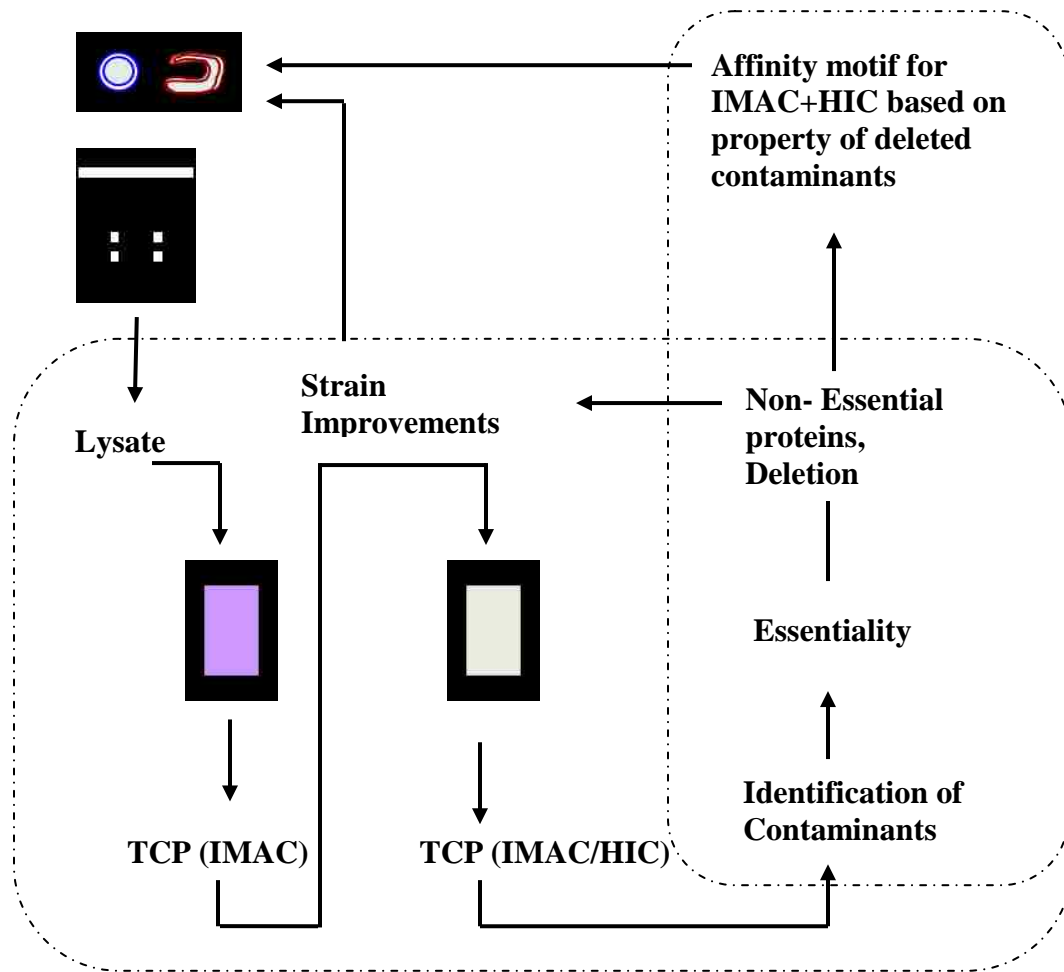
Figure 2-1: Overall strategy to improve bioprocesses using proteome analysis

2.2 Materials and Methods

2.2.1 Strain, culture growth and cell lysate

A single colony of *E .coli* BL21 grown on Luria Bertani (Sigma, St. Louis MO) agar plates was used to inoculate 5 ml of primary culture and grown at 37 ℃ temperature overnight. Next day, a 100 ml secondary culture was inoculated with the primary culture and grown at 37 ℃ until it reached an O.D. of 1-2 measured at 600 nm. The secondary culture was used to inoculate a fermenter containing 1.5 liters of Luria Bertani broth, and the culture was grown at 37 ℃ with sufficient agitation to remain saturated with oxygen. After 24 hours of growth, the cell pellet was recovered by centrifugation at 18,600 g for 15 minutes. The cell pellet was suspended 10 ml of phosphate buffer (0.05 M phosphate, 0.25 M NaCl, pH 7.2), with addition of 100 µl protease inhibitor (Sigma, St Louis, MO). The cell lysate was prepared via four repeated cycles of sonication followed by freezing at -20 ℃. To prepare the final extract for chromatography, the cell suspension was again centrifuged for 24,200 g for 10 minutes to remove cell debris. The supernatant was collected and the sample was stored in -20 ℃ for further analysis.

2.2.2 Protein assay

The quantification of a protein sample was done by spectrophotometry using a Detergent Compatible (DC) assay (BioRad, Hercules CA) which is a colorimetric based, Lowry assay. The absorbance at 750 nm of a protein sample mixed with dye was determined using DU-640 spectrophotometer (Beckman, Fullerton CA) and compared to the adsorption of a series of bovine serum albumin samples.

2.2.3 Immobilized Metal Affinity Chromatography

A 5-ml column (HisTrap HP IMAC column, Amersham Biosciences) was charged with 5 column volume (CV) of $CoCl_2.6H_2O$ (25 mg/ml) using an AKTA FPLC system (GE Healthcare, Piscataway NJ). The IMAC column was washed with 5 CV of distilled water after metal ion charging. During IMAC, the binding buffer (A) contained 1 M NaCl and 0.05 M Na_2HPO_4 (pH 7.4), while the elution buffer (B) had lower pH (5.5). Prior to sample loading, the column was washed and equilibrated with 5 CV of binding buffer. Flow rate was 0.5 ml per minute. Elution was monitored at 280 nm with the integrated UV monitor. After the proteins were loaded, non-binding proteins were flushed from the column with 2 CV of wash buffer. Finally, washing the IMAC column with 4 CV of the low pH, elution buffer generated an intermediate pool of genomic proteins that was loaded to the HIC column.

2.2.4 Hydrophobic Interaction Chromatography

The protein sample eluted during IMAC was applied to a 1-ml HiTrap Octyl FF HIC Column (GE Healthcare, Piscataway, NJ). All buffers contained 0.02 M Na_2HPO_4 to maintain the pH at 7.4. Three different mobile phase (A) Na_2SO_4 concentrations were tested (0.25 M, 0.50 M, and 1M) to investigate the effect on protein adsorption, requiring an appropriate amount of Na_2SO_4 added to the injected sample to match the composition of the HIC mobile phase. Elution buffer (B) contained zero concentration of Na_2SO_4; absence of this salt weakened hydrophobic interaction(s) and caused proteins to elute.

After one of the three Na_2SO_4 concentrations was chosen, an experiment consisted of applying protein sample, binding and washing the column with buffer containing the chosen Na_2SO_4

concentration, and eluting the protein sample (final ECP). The HIC column was washed and equilibrated with 5 CV of binding buffer (A). Samples were applied to a HIC column at 0.5 ml per minute. After the non-binding proteins passed during the initial washing with (A), the column was developed with a step change in elution buffer (B). This final pool of protein displaying affinity towards IMAC and HIC will be referred to as the final ECP when the results are discussed.

2.2.5 2-D electrophoresis and mass spectrometry

The final ECP protein pool was subjected to 2- dimensional gel electrophoresis (2DGE) followed by mass spectroscopy to identify proteins. Isoelectric focusing was performed with an IPGphor instrument (GE Healthcare), using a 7 cm IPG strip and accompanying reagents. The protein pool was desalted using 2D clean up kit (GE Healthcare) and was then applied to an IPG strip (pH 3-10) in total volume of 200 μl (62.5 μg protein in 75 μl + 125 μl of rehydration buffer from GE Healthcare). The sample was allowed to absorb on strip for 15 minutes, and then covered with mineral oil. After a seventeen hour rehydration period, focusing occurred at these conditions: 150 V, 30 mins. 500 V, 30 mins. 1200 V, 40 mins, and 5000 V, 60mins. A strip was then immersed in 10 ml of SDS equilibration buffer (2% (v/v) SDS, 50 mM Tris pH 8, 6M urea, 30% (v/v) glycerol, and 0.002% bromophenol blue) containing 100 mg of DTT for 15 minutes. Finally, the strips were soaked in 10 ml of SDS equilibration buffer with 250 mg of Iodoacetamide for 15 minutes. The second dimension was resolved via electrophoresis using 12.5% SDS PAGE (Protean II, Bio-Rad). Proteins were visualized using Coomassie blue. Spots corresponding to the proteins of high concentration were excised and identified by mass spectrometry by the Mass Spectrometry and Proteomics Facility at The Ohio State University.

2.3 Results

Table 2-1 describes a portion of the protein pool that had the dual property of Co(II) IMAC and HIC retention, *i.e*. the final ECP. Under the conditions chosen for the IMAC capture step, about 13 % of the lysate bound and subsequently eluted from the column due to the step reduction in pH. When this intermediate pool was delivered and eluted from the HIC column, very little protein was collected in the ECP. Indeed, the amount of protein in the final ECP ranged from about 6% down to 2 % (based on original mass of proteins in the *E. coli* lysate), and was dependent on the concentration of Na_2SO_4 used to promote adsorption.

Sample	Concentration in sample (mg/ml)	Total protein (mg/ml)	Percentage of lysate bound
Lysate	5.2	52	-----
IMAC elution peak	0.47	7.5 +/- 2.19	13
ECP final (high salt)	0.30	3+/- 0.73	5.7
ECP final (medium salt)	0.16	1.6 +/- 0.57	3.0
ECP final (low salt)	0.12	1.2 +/- 0.22	2.3

Table 2-1: Summary of Chromatography

SDS polyacrylamide gels were run to examine the molecular weight distribution and number of proteins present in high concentration, recognizing that adsorption via HIC as a function of Na_2SO_4 addition (Figure 2-2). Qualitatively, decreasing the amount of salt used to promote adsorption caused less protein to bind, with lower molecular weight products less favored when less Na_2SO_4 was present. For example, several of the bands representing protein with molecular weights less than 50 kDa substantially decreased in intensity or disappeared. Other proteins in minor quantity moved out of the final ECP pool as the salt concentration was lowered. Referring to Figure 2-2, one rather large molecular weight species with molecular weight of 72 kDa became absent from the final ECP as Na_2SO_4 decreased from 0.5 M to 0.25 M , and four proteins in large concentration appeared to bind regardless of salt concentration. Forming additional qualitative comparisons could have been made but were not necessarily instructive. Since the point of the experiments were to identify some of the proteins in the final ECP present in high concentration with changing adsorption characteristics, a partial analysis would provide the needed data to continue the work.
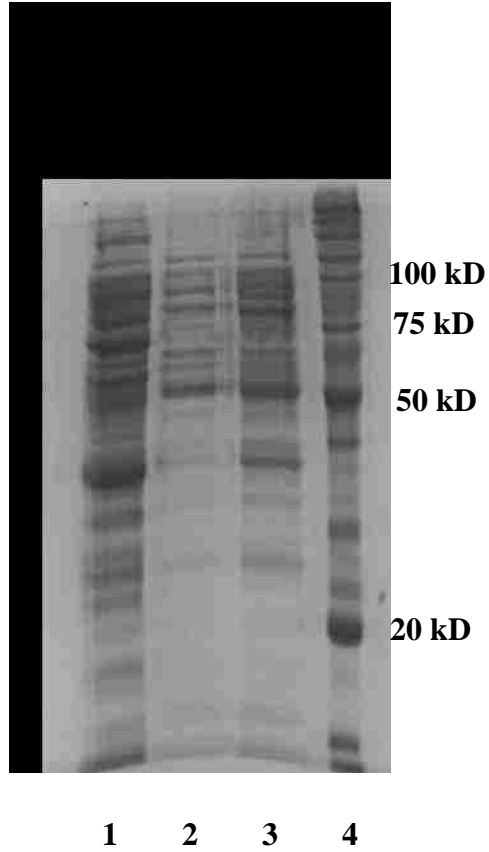
Figure 2-2: 1D SDS PAGE of final ECP for different Na$_2$SO$_4$ concentrations during column loading

Lanes: 1, IMAC/HIC elutant at 1 M Na$_2$SO$_4$salt concentration;
2, IMAC/HIC elutant at 0.25 M Na$_2$SO$_4$; 3, IMAC/HIC elutant at 0.5 M Na$_2$SO$_4$;
4, Invitrogen Bench Marker protein ladder

To further characterize the final ECP, 2DGE was carried out to separate the proteins based on molecular weight and isoelectric point. Figure 2-2 shows the results obtained with 1.0 M Na_2SO_4, which represented the inclusive set of proteins regardless of salt concentration. Figure 2-3 and Figure and 2-4 shows the results for lower salt concentration during HIC.
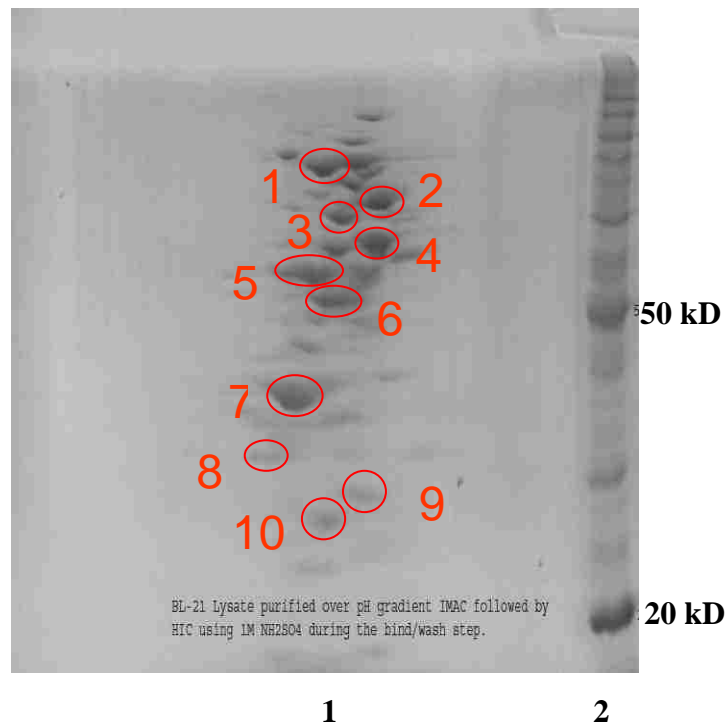
Figure 2-2: 2D SDS PAGE of final ECP from 1.0 M Na_2SO_4

Lanes: 1, IMAC/HIC elutant at 1 M Na_2SO_4 salt concentration; 2, Invitrogen Bench Marker protein ladder
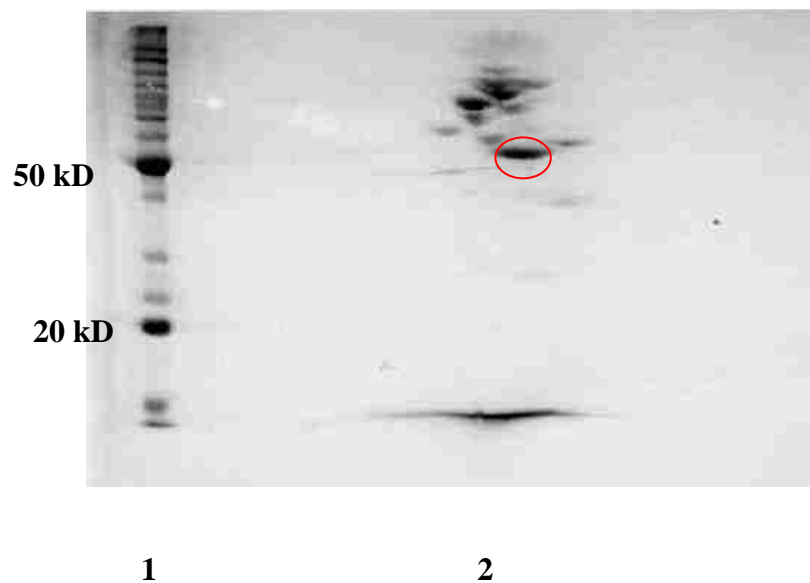
Figure 2-3: 2D SDS PAGE of final ECP from 0.5 M Na$_2$SO$_4$

Lanes: 1, Invitrogen Bench Marker protein ladder; 2, IMAC/HIC elutant at 0.5 M Na$_2$SO$_4$ salt concentration
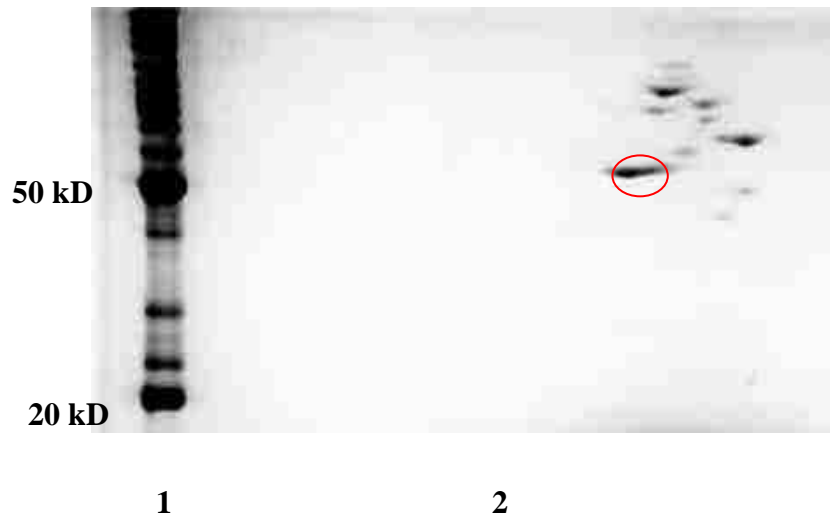
Figure 2-4: 2D SDS PAGE of final ECP from 0.25 M Na$_2$SO$_4$

Lanes: 1, Invitrogen Bench Marker protein ladder; 2, IMAC/HIC elutant at 0.25 M Na$_2$SO$_4$salt concentration

Approximately 18 strong bands were evident, with a subset (Table 2-2) representative of the three salt conditions identified by pI and MW on separate 2D gels. Ten of the eighteen proteins were identified (Table 2-2). All proteins are acidic, and fall within the pI range 5.1 - 6.3, with the requisite property for IMAC retention, i.e. histidine present on the surface of the protein either by inference or crystal structure. For example, the enzyme transketolase I (*Tkt1*) has the largest number of histidine residues (Litlechild, Turner et al. 1995) , while the enzyme triosephosphate isomerase (*Tpi*) has three which have been confirmed to be surface exposed (Noble, Zeelen et al. 1993). Proteins (gene products of *Tkt1*, *Pyk1*, *MelA*, *GatD*, and *Tpi*) in the mixture have been identified as nonessential based on categories defined by Gerdes et al (Gerdes, Scholle et al. 2003). The nonessential pool includes, in addition to the previously mentioned enzymes, enzymes involved in processes principally coupled to energetics. Pyruvate kinase (*Pyk1*), a key enzyme in glycolysis as it converts phosphoenol pyruvate to pyruvate, was accompanied by alpha galactosidase (*MelA*), a catabolic enzyme used in the conversion of mellibiose into glucose and galactose; and galactitol-1-phosphate-5 dehydrogenase (*GatD*), an enzyme responsible for galactitol degradation (Lengeler 1975; Liljestrom and Liljestrom 1987; Mattevi, Valentini et al. 1995; Sauer and Eikmanns 2005).

Presence in final ECP

protein (gene)	Swiss-Prot ID	pI	hydropathicity	(E)ssential or (N)ot essential	(Na$_2$SO$_4$ in loading step) 1.0 M	500 mM	250 mM	MW (kD)
Alanyl-tRNA synthetase (*SyA*)	P00957	5.53	-0.303	E	√	√	√	96.0
Catalase (*CatA*)	P13029	5.14	-0.372	E	√	√	√	80.0
2-Dehydro-3'deoxyphophate aldolase (*KdsA*)	P0A715	6.32	-0.028	E	√			30.83
Lysyl-tRNA synthetase (*Syk2*)	P0A8N5	5.10	-0.399	E	√	√	√	57.82
Aerobic respiration control protein (*ArcA*)	P0A9Q1	5.21	-0.487	E	√			27.29
Triosephosphate isomerase (*Tpi*)	P0A858	5.64	0.010	N	√	√		26.97
Pyruvate kinase (*Pyk1*)	P0AD62	5.77	-0.082	N	√	√		50.72
Alpha galactosidase (*MelA*)	P06720	5.52	-0.182	N	√	√	√	50.65
Galactitol-1-phosphate-5 dehydrogenase (*GatD*)	P0A9S3	5.94	0.139	N	√			37.39
Transketolase I (*Tkt1*)	P27302	5.43	-0.228	N	√	√		72.21

Table 2-2: Identification of representative proteins in the final ECP

Proteins in the final ECP that were categorized as essential include alanyl-tRNA synthetase (*SyA*), lysyl- tRNA synthetase (*Syk2*), catalase (*CatA*), 2-dehydro-3'deoxyphosphate aldolase (*KdsA*), and aerobic respiration control protein (*ArcA*). The first two belong to the class II aminoacyl-tRNA synthetase family since they possess a conserved ATP binding motif. Catalase (*CatA*) and aerobic respiration control protein (*ArcA*) are involved in stress response, with the former involved with oxygen toxicity and the latter involved with repression of aerobic enzymes when *E. coli* is starved of oxygen. The last enzyme (2-dehydro-3'deoxyphosphate aldolase) is involved in lipid biosynthesis in gram negative bacteria (Triggsraine, Doble et al. 1988; Miller, Hill et al. 1991; Onesti, Miller et al. 1995; Carpena, Melik-Adamyan et al. 2004).

2.4 Discussion

Previous studies have demonstrated that separately, IMAC and HIC are effective methods to resolve protein mixtures. One of the major appeals of IMAC is that a specific protein can be isolated based on metal affinity, whereas HIC is more generic since overall hydropathicity dictates adsorption. There are, obviously, limitations to both techniques in that impurities still remain bound to a column, requiring multiple steps in the purification process. In general, by first using a combination of chromatography methods to isolate contaminants followed by proteome analysis, a better understanding of the genomic proteins encountered may be obtained that ultimately may lead to a simpler, generic purification regiment. Such a strategy was employed to understand the combination of IMAC and HIC. Data on the protein set is in agreement with prior studies with regard to the IMAC capture step, since it has been shown that as little as two histidine residues in an appropriate conformation will promote binding to Co (II) IMAC columns. Similarly, data on HIC retention is typically correlated to a negative

hydropathicity value. Hydropathicity of a peptide or a protein is calculated by dividing the summation of hydropathy values for all amino acids by the number of the residues in the sequence. Online tools for calculating hydropathicity are available (http://www.expasy.ch/tools/protparam.html). With the exception of the *GatD* gene product, all have negative values or very close to neutral (zero). As the concentration used to promote HIC adsorption was decreased from 1 M to 250 mM, fewer proteins were retained. The very general strategy of pH reduction (IMAC) and step change by salt depletion (HIC) was capable of providing a small final ECP without the addition of imidazole or gradient design. This supplied a small pool that, while in principle may be further decreased with optimization, provided a reasonable starting set of proteins that display the dual property of IMAC and HIC adsorption.

While the data on the final ECP appears to be descriptive in nature, it points toward a future direction to develop an IMAC - HIC purification system that exploits both an optimized host strain and bioseparation regimen. Deletion of nonessential gene products will further reduce the total number of proteins that carry through the system. Of course, deletions must be made judiciously, because although the enzyme may be deemed nonessential, it may be required from a practical standpoint. A good example is triosephosphate isomerase; databases such as PEC, biocyc, and Gerdes et al. defines this enzyme as nonessential because technically *E. coli* is able to survive with an incomplete EMP pathway but growth is severely compromised. Within the confines of such, characteristics of a finalized strain would include (i) sufficient growth properties such that it is a viable option for bioprocessing, (ii) the ability to effectively express a recombinant product with yields similar to that of established strains, and (iii) limited antibiotic scarring (detritus antibiotic resistance genes that are no longer required for selection). For

example, while inactivation of *Tpi* and *Pyk1* through homologous recombination would presumably aid in downstream processing, it would compromise growth since their gene products are involved heavily in the conversion of glucose to energy and precursors for biosynthesis. Modification of these proteins so they are not captured during IMAC via the deletion of surface exposed histidine and reintroduction of the mutant into genome via recombination would be a more effective strategy for important proteins, as their metabolic function would be retained and they would not carry through the bioseparation steps. In contrast, deletion of genes responsible for unnecessary enzymes like *MelA* and *GatD*, for example, do not affect glucose catabolism and likely would not be deleterious to growth and / or recombinant protein expression. Indeed, experiments with a *MelA*⁻ mutant displayed growth characteristics similar to *E. coli* MG1655 when glucose was used as the sole carbon source (Fry, Zhu et al. 2000) .

Germane to the IMAC - HIC system, data on the final ECP will be used to pick an appropriate fusion partner (affinity tail) used to endow a target protein with the dual ability of IMAC and HIC retention. Glutathione-S-transferase, maltose binding domain, cellulose-binding domain, thioredoxin, and NusA are past examples of fusion partners (proteins) or functional domains that have been used in the past to dictate purification using affinity resins. Results of this study demonstrated that alpha galactosidase or galactitol-1-phosphate-5 dehydrogenase could serve in a similar fashion as fusion partners for an IMAC - HIC process. While inactivation of *MelA* or *GatD* at the chromosomal level removes the protein from an ECP, at the same time constructing an expression vector that employs one of these two genes in an appropriate frame can extend either the N- or C- terminus of a target recombinant protein in order to force retention. The

MelA gene product is encountered with all three Na_2SO_4 concentrations, indicating that the extension of a recombinant protein with alpha galactosidase could guide elution toward salt concentrations that have fewer genomic proteins (e.g. 250 mM). The smaller protein, galactitol-1-phosphate-5 dehydrogenase, was found to bind at 1.0 M salt but disappeared as the salt concentration decreased. Two extremes are thus represented by the data, for the *GatD* gene product being a 37 kDa protein can serve as a smaller, good candidate for development of a fusion tag with lower salt elution requirements, while the *MelA* gene product is present in more stringent washes but is obviously larger. Truncated versions of each protein could, in effect, dictate binding and lead to a shorter affinity motif.

As fusion partners for a combined IMAC - HIC process, neither alpha galactosidase nor galactitol-1-phosphate-5 dehydrogenase currently requires modification because of the native binding properties. This hypothesis can be examined by creating a fusion of alpha galactosidase or galactitol-1-phosphate-5 dehydrogenase with a model target like Green Fluorescent Protein. Alternately, using predictive tools it may be possible to identify a domain in either protein that dominates HIC retention, the more elusive adsorptive properties of the two. Such identification could ultimately result in a smaller fusion tail, and should a HIC domain be identified and tested that accidentally abolishes IMAC retention, His_6 or a similar sequence can be added back to reestablish retention by the primary capture step.

PART II

CHAPTER 3

Study of Expression and Hydrophobic properties of Alpha galactosidase from *E. coli*, as related to the development of novel affinity sequences for Bioseparation

3.1 Introduction

The rather large field of protein purification involves many different challenges and as such, new approaches to achieve yield and purity are always desired. These approaches include the groundbreaking use of affinity tags to dictate adsorption in chromatographic systems, and the application of other novel methods to improve upon this strategy of highly efficient [affinity based] bioseparation. Traditionally, the development of affinity tags has been based on fundamental biophysical principles. For example, the addition of short, histidine rich peptides serves to promote pi bonding during salt promoted adsorption (Immobilized Metal Affinity Chromatography, IMAC); the use of arginine, or other highly charged sequences forces charge – charge interaction (ion exchange chromatography); and the inclusion of functional domains or portions thereof cause binding via known biospecific interactions (protein A). Additionally, the discovery of other, less intuitive affinity tags has occurred through the use of combinatorial libraries and / or phage-display (Patwardhan, Goud et al. 1997; Mooney, Fredericks et al. 2011) . The latter includes the identification of candidate peptides for use in antibody purification and enzymes, sometimes with surprising results as portions of the phage coat sometime complicates interpretation.

Adding to this repertoire of techniques is the use of proteomics to aid the design of candidate affinity tags by first determining protein(s) that interact with a particular chromatography matrix.

In this fashion, the design of the tag is not constrained by *a priori* choices like length or amino acid composition. Rather, proteins with certain binding or elution characteristics are identified and used to guide choices as the affinity tag is developed. In part I of this project, we have identified native *Escherichia coli* proteins that bind both Co (II) IMAC and HIC during salt promoted adsorption. IMAC and HIC individually are two commonly used techniques; the combination of the two recently was proposed for improved purification of recombinant proteins. IMAC and HIC are based on totally different fundamentals when it comes to the binding properties of proteins. Metal binding capacity has been found associated mainly with surface exposed histidine and cysteines in some cases. Alternately, favorable hydrophobic interaction involves adsorption via hydrophobic patches. Thus, proteins showing affinity towards both IMAC and HIC have dual but different properties, presumably decreasing the number of co eluting contaminants. Alpha galactosidase (*MelA*) was identified as one of the major binding proteins during the IMAC / HIC study previously mentioned, and chosen as a protein of interest as it was capable of HIC adsorption even at very low Na_2SO_4 concentrations. *MelA* hydrolyses mellibiose into glucose and galactosidase in *E. coli* and is described as one of the nonessential proteins for cell survival. An analysis of the protein revealed that both hydrophobic and hydrophilic regions are somewhat balanced with mild dominance of hydrophobic regions. This work represents an effort to study *MelA* within the context of developing a dual IMAC / HIC affinity tail. We report on the use of different portions of this protein as a fusion partner with FLAG to identify sequences useful for HIC in particular, and comment on the suitability of these sequences as affinity tags.

3.1.1 Alpha Galactosidase (*MelA)*

The alpha galactosidase (E.C.3.2.1.22) are the hydrolases which breakdown the sugar molecules in smaller subunits. *MelA* gene encodes for the melibiase (α- galactosidase) which hydrolyses mellibiose into glucose and galactosidase (Figure 3-1). The sequence consists of 451 amino acids in *E. coli* with molecular weight of 50.6 kDa as determined on the SDS gel (Liljestrom and Liljestrom 1987). The hydropathic analysis of the gene reveals that both hydrophobic and hydrophilic regions are present with average hydropathcity of -0.18, showing presence of both hydrophobic and hydrophilic regions. The isoelectric point of *MelA* is 5.5. Crystal Structure of *E.coli MelA* has not been reported till now.

```
                    ┌─────────────────────────┐
                    │   MelB GPH transpoter    │
                    └─────────────────────────┘
                                 │
                                 ▼
                        ┌──────────────┐
                        │  Melibiose   │
                        └──────────────┘
                                 ▲
                                 │          ┌──────────────────────────────┐
                                 │          │  Alpha galactosidase: MelA   │
                                 │          │          3.2 .1.22           │
                                 │          └──────────────────────────────┘
                          ┌──────┴──────┐
                          ▼             ▼
              ┌───────────────────┐   ┌────────────────────────┐
              │  β -D- glucose    │   │  β –D- galactosidase   │
              └───────────────────┘   └────────────────────────┘
                       │                         │
                       ▼                         ▼
        ┌───────────────────────────┐   ┌──────────────────────────────────────┐
        │  Glucose and glucose-1-   │   │  UDP-galactose biosynthesis (salvage │
        │  phosphate degradation    │   │  pathway from galactose using UDP-   │
        │                           │   │  glucose)                            │
        └───────────────────────────┘   └──────────────────────────────────────┘
```
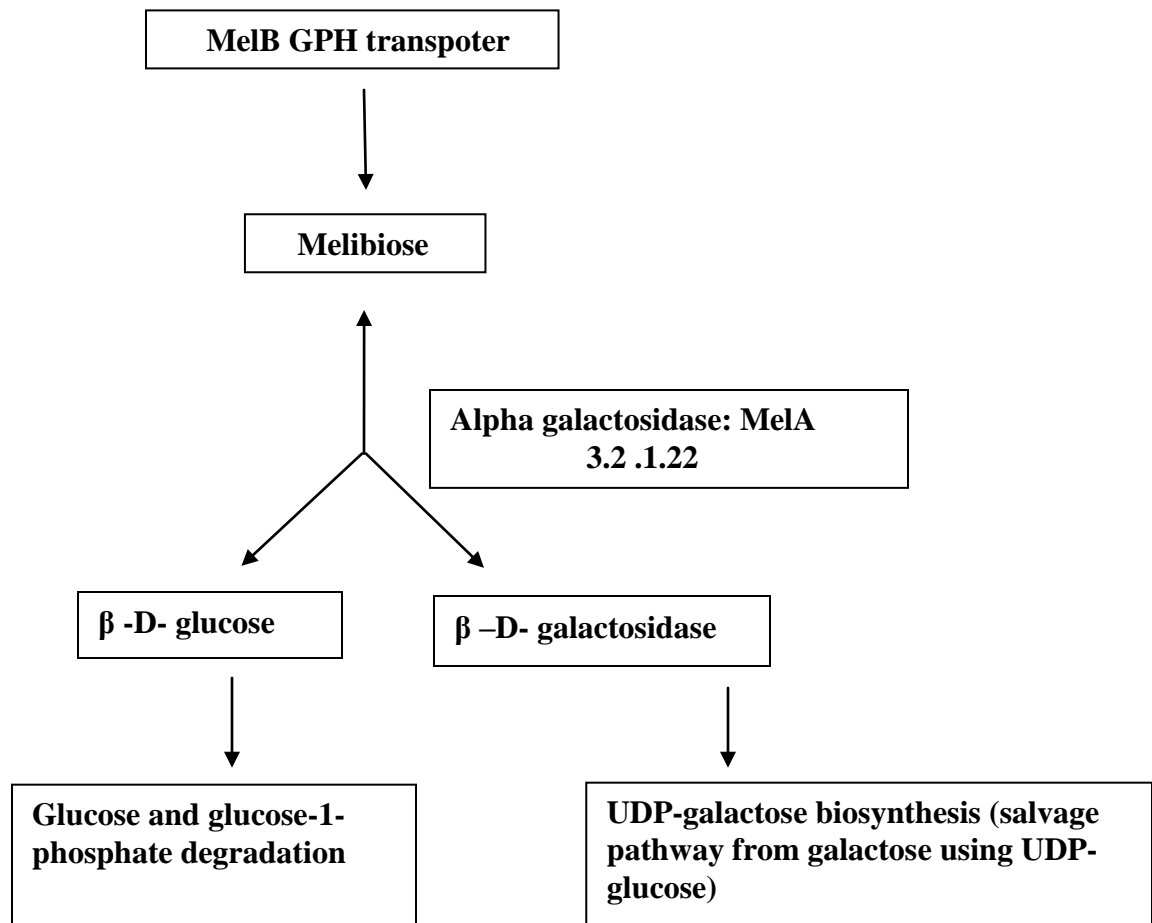
Figure 3-1: Flow chart of Alpha galactosidase (*MelA)* enzyme action

There are various scales developed to study the polarity of proteins based upon the individual amino acid hydropathicity. The Kyte- Dolittle scale is one of the most widely used methods to determine hydropathicity of proteins. This method is based on the amino acid partitioning between two immiscible liquids, water and phenol. As seen in Table 3-1, amino acids with negative values of hydropathicity index are non-polar amino acids; with positive value are polar amino acids. The protein regions with value above 0 are considered hydrophobic regions. Hydrophobic regions of proteins are mostly buried inside of the protein structure.

Amino acid residues	Hydropathicity based on Kyte –Dolittle scale	Side chain Polarity
Isoleucine(Ile)	4.5	Non-polar
Valine(Val)	4.2	Non-polar
Leucine(Leu)	3.8	N Non-polar
Phenylalanine(Phe)	2.8	Non-polar
Cysteine(Cys)	2.5	Polar
Methionine(Met)	1.9	Non-polar
Alanine(Ala)	1.8	Non-polar
Glycine(Gly)	-0.4	Non-polar
Threonine(Thr)	-0.7	Polar
Tryptophan(Trp)	-0.9	Non-polar
Serine(Ser)	-0.8	Polar
Tyrosine(Tyr)	-1.3	Polar
Proline(Pro)	-1.3	Non-polar
Histidine(His)	-1.6	Polar
Glutamic Acid(Glu)	-3.2	Polar
Glutamine(Gln)	-3.5	Polar
Aspartic Acid(Asp)	-3.5	Polar
Asparagine((Asn)	-3.5	Polar
Lysine(Lys)	-3.9	Polar
Arginine(Arg)	-4.5	Polar

Table 3.1: Hydrophobicity of all amino acids based on Kyte-Dolittle scale

As seen in Figure 3-2, when *MelA* sequence was plotted by Kyte – Dolittle method using window size of 19, which makes the hydrophobic region stand out, we found 2 major hydrophobic regions. The regions between amino acid 9-94 and 342-451 were cloned and studied to design a HIC affinity tag.

Figure 3-2: Kyte Dolittle plot of Alpha galactosidase *(MelA)* amino acid sequence

3.1.2 Glutathione –S – Transferase (GST)

Glutathione –S-Transferase (GST) was chosen as a one of the model protein. Glutathione –S-Transferases are set of proteins present in both prokaryotes and eukaryotes, and catalyzes variety of reactions in cytosol, mitochondria and microsomes. GST catalyzes conjugation of reduced glutathione through sulfhydryl group. The protein detoxifies endogenous compounds like xenobiotics, toxins, etc. This enzyme protects cell against the carcinogenic and toxic compounds. In mammals, GST is especially found in liver cells which plays role in detoxification. GST is a very well-studied protein and many other proteins are fused to GST for affinity purification. GST is known to increase solubility of the fused proteins. Also due to its glutathione affinity, it's easy to purify using GST columns. The desired proteins are easily expressed and purified by fusing it with GST and the fusion protein is purified using GST columns. GST was used as control protein in this study. GST with 220 amino acids sequence length and molecular weight of 26kD, has been fused with vectors for expression and purification purposes. Ability to check enzyme activity, high solubility and lack of retention to HIC and IMAC under given condition prompted to select GST as the control protein for this study. Plasmid pGEX-6P-2 provides cloning site to fuse protein on C-terminal of GST. We have fused different versions of developed tags from *MelA* on C-terminal of GST, to study its expression and ability to work as affinity tag.
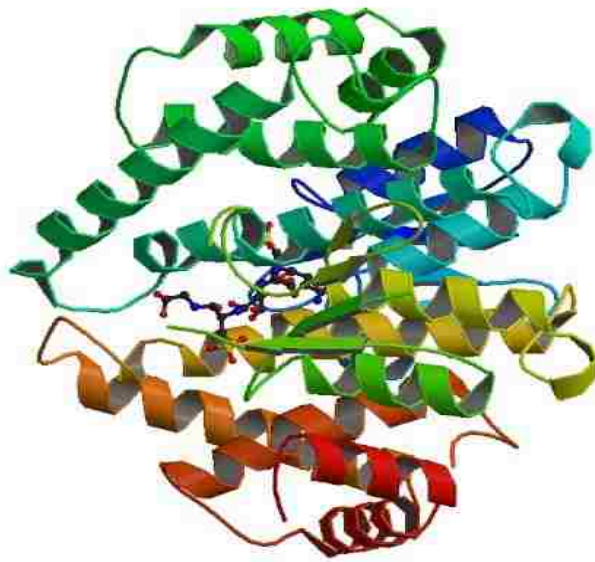
Figure 3-3: Overview of GST structure and substrate binding:

Crystal Structure of a Bacterial Glutathione Transferase from *E. coli* with

Glutathione Sulfonate in the Active Site

3.1.3 Green Fluorescent Protein (GFP)

Green Fluorescent protein is a protein of 238 amino acid residues with molecular weight of approximately 27 kD. It exhibits bright green fluorescence when exposed to ultraviolet light. GFP was first isolated from the jelly fish Aequorea Victoria in 1994. Noble prize in 2008 was awarded based on GFP discovery and development since its invention it has been widely studied for use as a marker due to its easy detection. Today GFP is not only used for bacterial studies but also in higher organisms like mice, fish etc. This is because it is easy to fuse protein of interest with GFP and we can see the fused protein and also track where it goes in the cell. Here is this project, GFP is also used as the control protein, though it binds to HIC column (octyl FF) at high salt concentration (1M and 0.5 M), it does not bind to HIC column (phenyl FF) at lower salt concentrations (0.5 and 0.25M). We have used GFP containing Plasmid pGPuv from Clontech to check N-terminal fusion to GFP of various *MelA* regions.
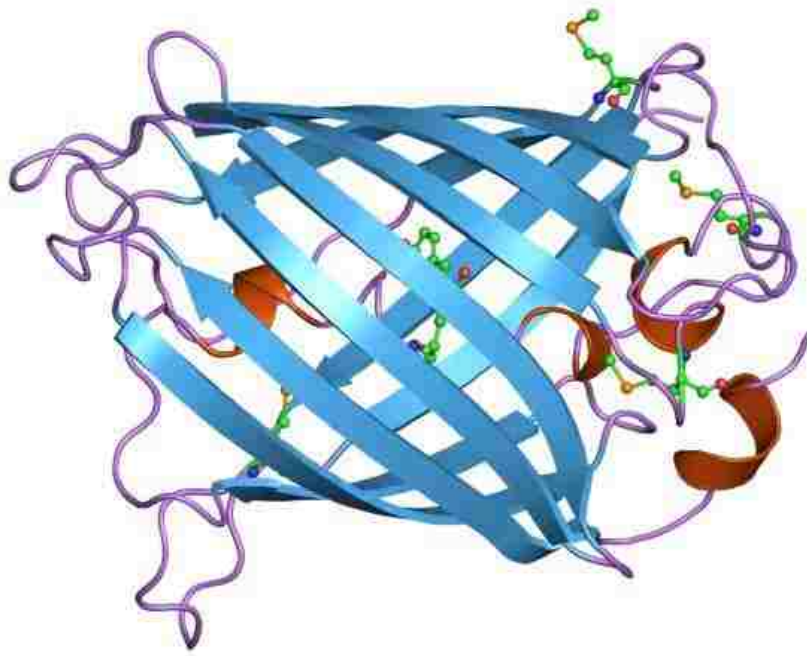
Figure 3-4: Overview of GFP structure

3.2 Materials and Methods

3.2.1 Enzyme digestion, PCR, cleanup and ligation

Procedures, as described below, require the use of several common molecular biology protocols. All enzymatic digestions were carried out using restriction enzymes purchased from New England Biolabs (Ipswich, MA) with an appropriate concentration (100 units) in buffer at 37 °C for 3 hours. PCR reactions were performed using Apex Taq DNA polymerase (Genesee Scientific, San Deigo,CA) at 55ºC. Products from a PCR reaction were examined using 0.8% agarose gel.

Component	Amount per reaction
10 X Enzyme Buffer 2	10µl
pET-22b(M) (100ng/µl)/ PCR product	10µl
EcoRI (20U/µl)	5µl
Hind III (20U/µl)	5µl
Distilled water	20µl
Total reaction volume	**50µl**

Table 3-2: Reaction mixture for Hind III / EcoRI restriction digestion

Component	Amount per reaction
10X Ammonia Buffer	10µl
dNTP's(10mM each)	2µl
Forward primer (10nm/ul)	2.5µl
Reverse primer (10nm/ul)	2.5µl
Template	2.5µl
2.5 mM MgCl$_2$	3µl
DMS0	5µl
Gene choice proof reader Polymerase	2µl
Distilled water	79µl
Total Reaction Volume	**100µl**

Table 3-3: Reaction mixture for PCR amplification

Segment	Number of cycles	Temperature	Duration
1	1	95	5 minutes
2	30	95	1.5 minutes
		55	1.5 minutes
		72	1.5 minutes
3	1	72	5 minutes

Table 3-4: PCR amplification cycles

Component	Amount per reaction
10X T4 DNA ligase buffer	2 μl
Double digested PCR product (100ng/ul)	8μl
Double digested pET22(b)M (50ng/ul)	8μl
T4 DNA ligase	2μl
Total Reaction Volume	**20μl**

Table 3-5: Reaction mixture for ligation

3.2.2 Cloning MelA-FLAG in pET-22b (M)

A cDNA clone for alpha-galactosidase (*MelA*) was obtained by PCR using genomic DNA isolated from *E.coli* strain BL21 (Table I). Forward primer P1 and reverse primer P2, matching the sequence for *MelA* (*E. coli* strain K12 Accession #AAA97019) were designed to include NcoI and XhoI sites, respectively. The reverse primer also included a two amino acid linker coding for serine-alanine (SA) and FLAG (DYKDDDDK). As an intermediate step leading to the final clone, the resulting PCR fragment and pET-28b were cleaved with the aforementioned enzymes, band purified, and ligated. This intermediate plasmid was used to generate a plasmid based on a modified version of pET-22b. Plasmid pET-22b and pET-22b(M) differed by removal of the NdeI − NcoI fragment constituting the DNA sequence of the N-terminal 22 amino acids translated from pET-22b. The restricted plasmid was blunted with Mung Bean Nuclease (New England Biolabs) and religated.

The plasmid used to express MelA-FLAG was constructed by first isolating the MelA-FLAG fusion gene via PCR. Forward and reverse primers were designed to recapture this sequence from the intermediate plasmid, and to extend the gene with flanking sequences for eventual restriction with EcoRI and HindIII. Plasmid pET-22b(M) was cut with EcoRI and HindIII, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection (Figure 3-7). Final sequence for pET-22b(M)MelA-FLAG was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.
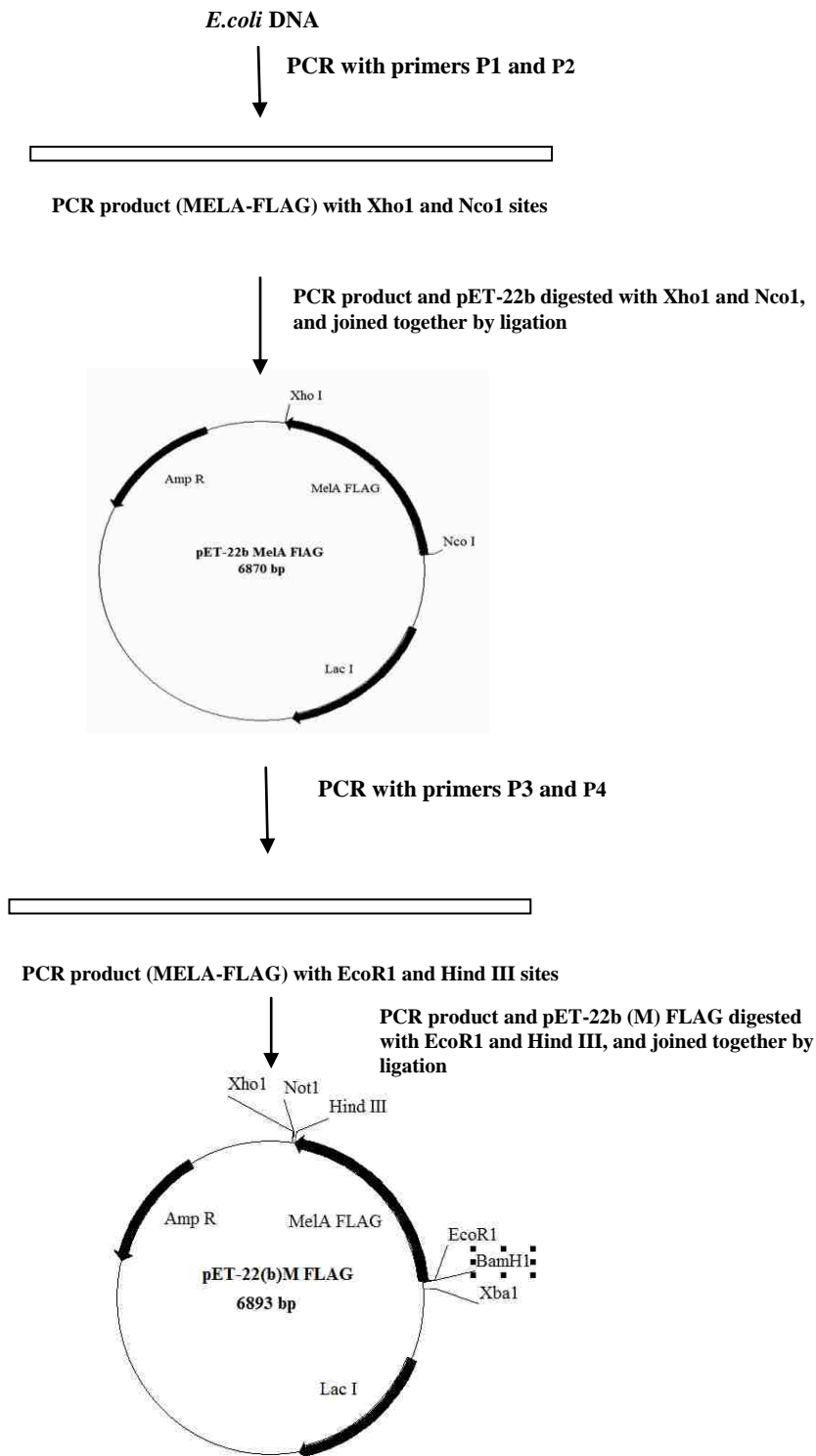
Figure 3-5: Cloning of pET-22b (M) MelA-FLAG
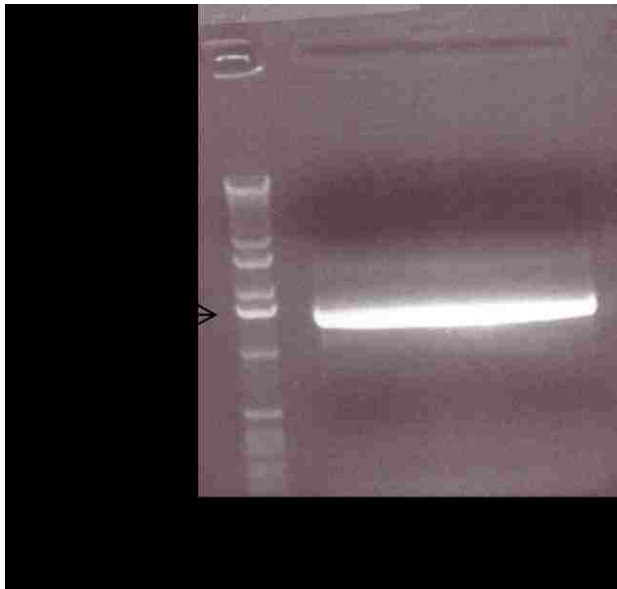
Figure 3-6: 0.8% agarose gel of PCR for amplification of MelA FLAG PCR product.

Lanes: 1, NEB 1Kb DNA ladder; 2, MelA FLAG PCR pdt (1.4 kb)
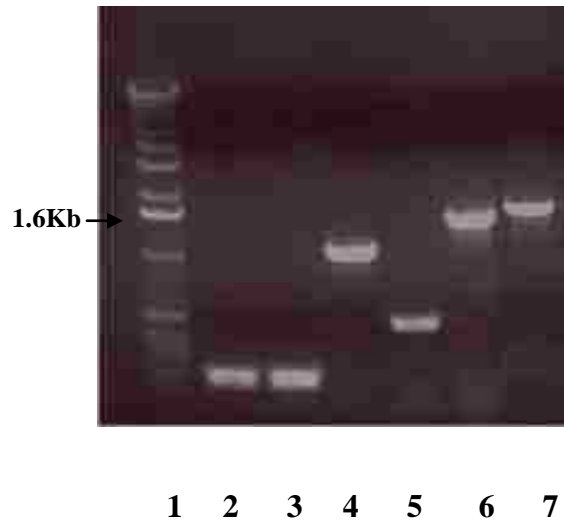
Figure 3-7: 0.8% agarose gel for PCR for selecting positive clones containing MelA FLAG segment

Lanes: 1, NEB 1Kb DNA ladder; 2, 3&5, negative results with absence of appropriate PCR pdt; 4, 6&7, positive results with presence of PCR pdt at 1.4 kb

3.2.3 Cloning NT1-FLAG in pET-22b (M)

The hydrophobic regions of *MelA* were identified using a Kyte-Doolittle plot, with two of the major hydrophobic sequences cloned in pET-22b(M) vector using SOE (splicing by overlap extension). The first PCR was done to amplify the DNA region corresponding to amino acids 9 - 94 by using forward primer P5 and reverse primer P6. Plasmid pET22b (M)MelA-FLAG served as the cloning template. The second PCR used primer P7, reverse primer P8, and pET-22b (M)MelA-FLAG as template to amplify the region corresponding to amino acids 342 - 451. Next, PCR #3 was done using forward primer P5, reverse primer P8, and purified product from the first two rounds of PCR. For the third PCR, a sequence was chosen that corresponded to a thirteen amino acid, hydrophilic sequence, designed to serve as common overlap and linker, respectively. The purified PCR product from the final PCR round was restricted with Hind III and EcoR1 and ligated into a similarly restricted pET-22b (M). After chemical transformation into competent BL-21, a selected clone was confirmed by DNA sequencing.

MAMSAPKITFIGAGSTIFVKNILGDVFHREALKTAHIALMDIDPTRLEESHIVVRKLMDSAGASGKITCHTQQK
EALQDADFVVVAFQIGGYEPCTVTDFEVCKRHGLEQTIADTLGPGGIMRALRTIPHLWQICEDMTEVCPDATM
LNYVNPMAMNTWAMYARYPHIKQVGLCHSVQGTAEELARDLNIDPATLRYRCAGINHMAFYLELERKTADG
SYVNLYPELLAAHDAGQAPKPNIHGNTRCQNIVRYEMFKKLGYFVTESSEHFAEYTPWFIKPGREDLIERYKVP
LDEYPKRCVEQLANWHKELEEYKNASRIDIKPSREYASTIMNAIWTGEPSVIYGNVRNDGLIDNLPQGCCVEVA
CLVDANGIQPTKVGTLPSHLAALMQTNINVQTLLTEAILTENRDRVYHAAMMDPHTAAVLGIDEIYALVDDLIA
AHGDWLPGWLHRSA

Figure 3-8:   Scheme for designing NT1 from *MelA*

pET-22b (M)MelAFLAG

PCR #1 with primers P5 and P6

PCR#2 with primers P7 and P8

PCR product#1 with common hydrophilic linker

PCR product# 2 with common hydrophilic linker

PCR#3 with primers P5 and P8, using SOE product1 & product 2 as template

PCR product#3(NT1FLAG) with HindIII and EcoR1 sites

PCR Pdt #3 and pET-22b (M) digested with Hind III and EcoR1, and joined together by ligation.

Xho1 Not1 Hind III

EcoR1
BamH1
Xba1

NT1 FLAG

Amp R

pET22(b)M-NT1FLAG
6042 bp

Lac I

Figure 3-9: Cloning of pET-22b (M) NT1FLAG using Splicing Overlap

Extension (SOE)   PCR

500bp

1   **2**   **3**   **4**
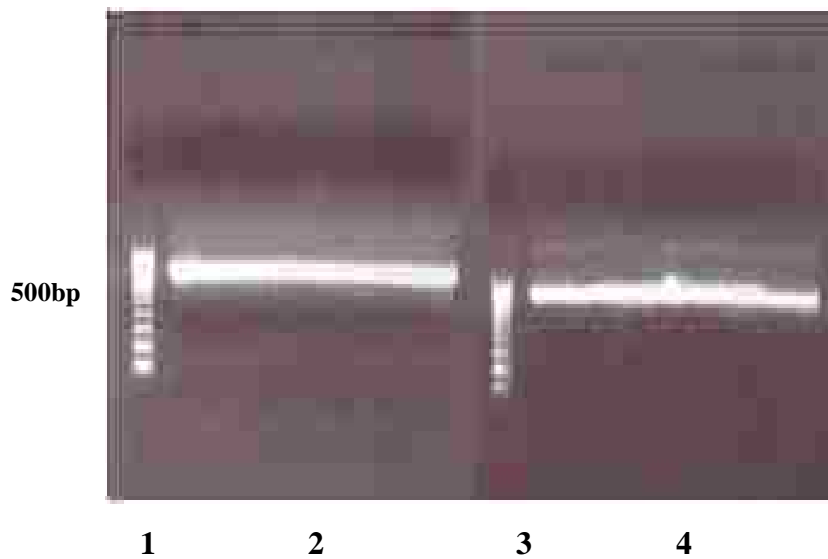
Figure 3-10: Amplification of PCR pdt at 500 b.p detected by 0.8% agarose gel for SOE PCR #1 and PCR #2

Lanes: 1&3, NEB 100 bp DNA ladder; 2, SOE PCR pdt #1; 4, SOE PCR pdt #2

1.0 Kb
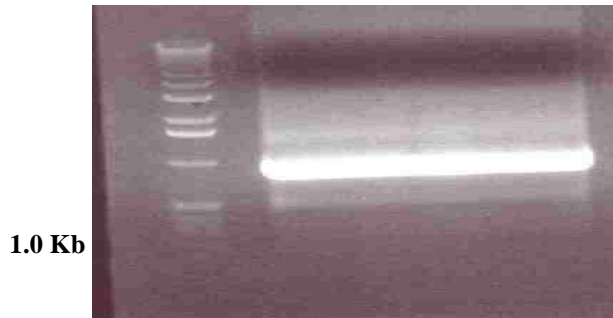
1   **2**

Figure 3-11: Amplification of PCR pdt detected at 800 bp by 0.8% agarose gel for SOE PCR # 3 for amplification of NT1-FLAG

Lanes: 1, NEB 1Kb DNA ladder; 2, SOE PCR pdt #3

1.0 Kb

1   2   3   4   5 6   7   8   9   10
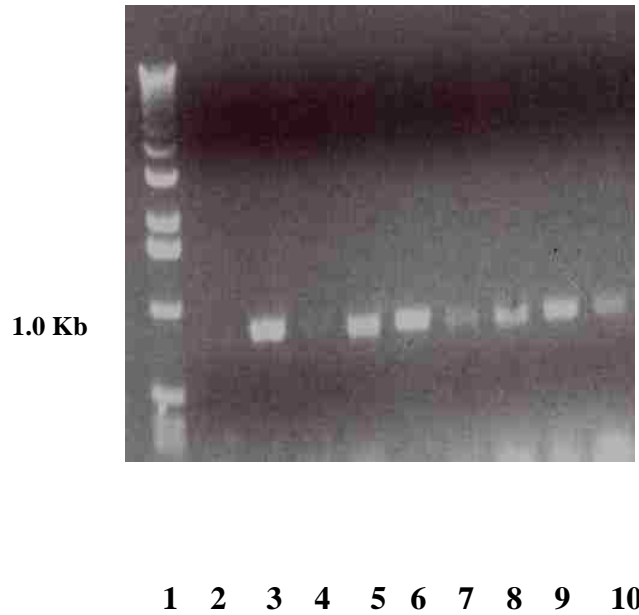
**Figure 3-12: Amplification of PCR pdt at 800b.p 0.8% agarose gel
for PCR for selecting positive clones containing NT1 FLAG
segment**

Lanes: 1, NEB 1Kb DNA ladder; 2,&4, negative results with absence of
appropriate sized PCR pdt;3,5,6,7,8,9,&10, positive results with presence of PCR
pdt

3.2.4 Cloning NT1 –GFP in pGFPuv

The plasmid used to express NT1-GFP was constructed by first isolating the NT1 via PCR. Forward P9 and reverse primers P10 respectively were designed to recapture this sequence from the NT1-FLAG-pET22b(M) and to extend the gene with flanking sequences for eventual restriction with HindIII and SmaI. Plasmid pGFPuv was cut with HindIII and SmaI, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection. Final sequence for pGFPuv-NT1-GFP was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.

3.2.5 Cloning of NTB-GFP in pGFPuv

One of the hydrophobic patches of *MelA* between amino acid regions 342-451 termed NTB was cloned with GFP. The plasmid used to express NTB-GFP was constructed by first isolating the NTB via PCR. Forward P11 and reverse primers P10 were designed to recapture this sequence from the intermediate plasmid, and to extend the gene with flanking sequences for eventual restriction with HindIII and SmaI. Plasmid pGFPuv was cut with HindIII and SmaI, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection. Final sequence for pGFPuv-NTB-GFP was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.

3.2.6 Cloning of GST-NTB in pGEX-6P-2

The plasmid used to express GST-NTB was constructed by first isolating the NTB via PCR. Forward P12 and reverse primer P13 were designed to recapture this sequence from the pGFPuv-NTB-GFP, and to extend the gene with flanking sequences for eventual restriction with BamH1 and EcoRI. Plasmid pGEX-6P-2 was cut with BamH1 and EcoRI, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection. Final sequence for pGEX-6P-2-GST-NTB was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.
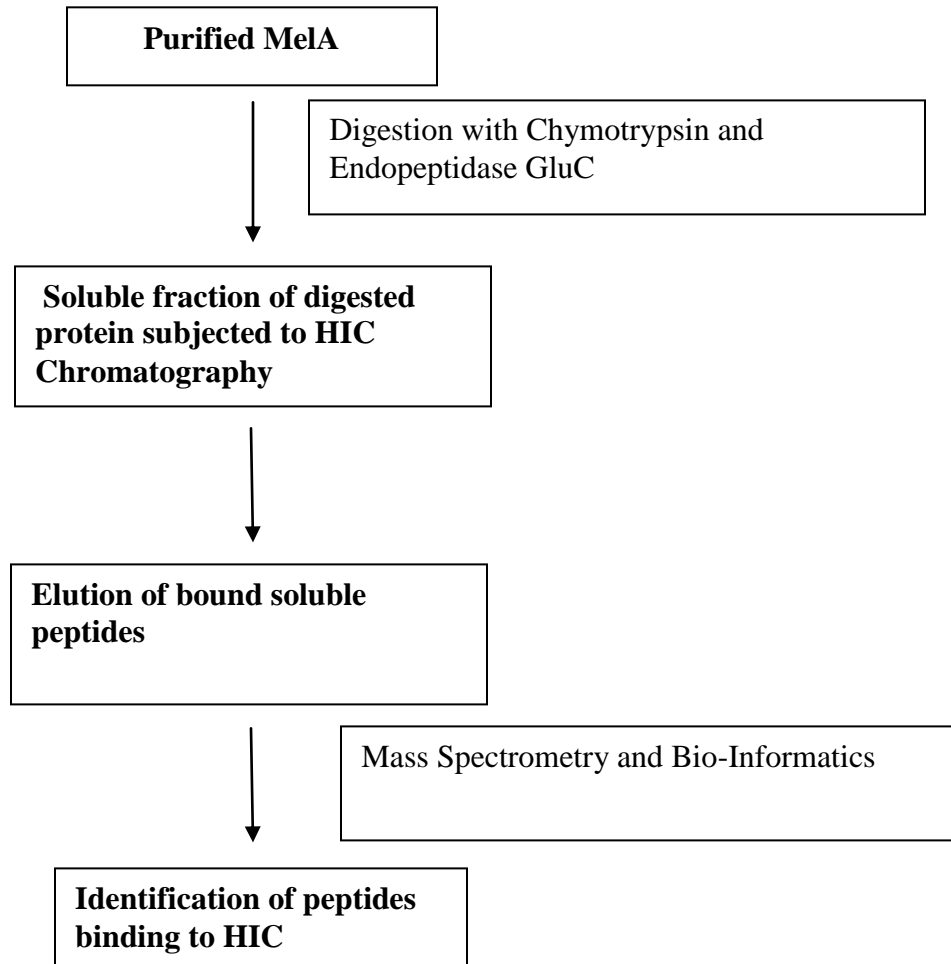
Figure 3-13: Scheme for identifying small soluble peptides from *MelA* with HIC binding property

3.2.7 Cloning of GST-RBNT in pGEX-6P-2

RBNT is 13 amino acid long sequence from *MelA*, which was identified as one of the peptides binding to HIC after protease digestion. The amino acid sequence of RBNT – (EPCTVTDFEVCKRHGL) was changed to (EPHTVTDFEVHKRHGL) to eliminate the cysteines, was fused to GST using pGEX-6P-2 vector. Forward and reverse primers P14 and P15 respectively were designed to recapture the GST sequence from the pGEX-6P-2 plasmid, and to extend the gene with flanking sequences for RBNT sequence and the restriction sites of BstB1 and EcoR1. Plasmid pGex-6P-2 was cut with EcoR1 and BstB1, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection. Final sequence for pGEX-6P-2-GST-RBNT was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.

3.2.8 Cloning of GST-HIC mini in pGEX-6P-2

HIC mini is 9 amino acid long sequence from *MelA*, which was identified as one of the peptides binding to HIC after protease digestion. The amino acid sequence of HIC mini –(S V I Y G N V R N), was fused to GST using pGEX-6P-2 vector. Forward and reverse primers P14 and P16 respectively were designed to recapture the GST sequence from the pGEX-6P-2 plasmid, and to extend the gene with flanking sequences for HIC mini and the restriction sites of BstB1 and EcoR1. Plasmid pGex-6P-2 was cut with EcoR1 and BstB1, and ligated with PCR product recovered via band purification and restricted with the same two enzymes. Positive clones were selected using PCR selection. Final sequence for pGEX-6P-2-GST-HICmini was determined by

DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock). The purified cloned plasmid was purified from overnight culture and transformed into chemical competent BL-21 cells.

Primer	Sequence (5'→3')
P1	GACCATGGCAATGTCTGCACCCAAAATTACATTTATCGGC
P2	GCTCGAGCTTGTCATCGTCGTCCTTGTAGTCAGCGCTACGGTGCAACCAGCCTGG
P3	CGAATTCATGGCAATGTCTGCACCCAAAATTACATTTATC
P4	CGA AGCTTCTACTTGTCATCGTCGTCCTTGTAGTCAGC
P5	TAATACGACTCACTATAGGG
P6	ACAAACTTCAGTCATATCTTCACAAATTTGCCAAAGATGTTCATAACCGCCAATCTGAAATGC
P7	CATCTTTGGCAAATTTGTGAAGATATGACTGAAGTTTGTTGTTGCGTAGAAGTAGCCTGTCTG
P8	GCTAGTTATTGCTCAGCGG
P9	GCCAAGCTTGTCTGCACCCAAAATTTCT
P10	TAATAAATAAATTATAATCCCGGGCGTCAGCGCTACGGTG
P11	GCCAAAGCTTGTGTTGCGTGGAAGTAGCCTGTCTG
P12	AGGATCCTGTTGCGTGGAAGTAGCCTGTCTG
P13	TATACCCGGGATGATGATGGTCAGCGCTACGGTG
P14	AGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACA
P15	AATGAATTCCTTATAAACCATGTCTTTTATGAACTTCAAAATCAGTAACAGTATGAGGTTCTG GGGATCCCAGGGGCCCCTGGAACAG
P16	AATGAATTCCTTAATTACGAACATTACCATAAATAACACTAGGGGATCCCAGGGGCCCCTGG AACAGAACTTCCAG

Table 3-6: Primers used in study

3.2.9 Culture growth and lysate preparation

A single colony of *E .coli* BL21, harboring either cloned plasmid and grown on Luria Berteni (LB, Sigma Chemicals, St. Louis MO) agar plates was used to inoculate 5ml of primary culture. After overnight growth, 500ml LB broth with ampicillin (100 µg/ml) was inoculated with the primary culture and grown at 37 ℃ until it reached an O.D. of 0.4 - 0.6, measured at 600nm. Cultures were induced using 0.5mM IPTG for expression and grown at 37 ℃ overnight with sufficient agitation to remain saturated with air. After 18 hours, cells were obtained as pellet by centrifugation at 10,000 rpm for 30 min (Beckman Model J2-21 M/E). Cell pellet was suspended in 10 mL of phosphate buffer (0.05 M phosphate, 0.25 M NaCl, pH 7.2), with the addition of 100 µL protease inhibitor (Sigma, St Louis MO). Cells were lysed via four repeated cycles of sonication followed by freezing at -20 ℃ with addition of 100µl protease inhibitor (Sigma, St Louis, MO) for 10 ml of cells lysate in phosphate buffer. To remove cell debris, the final extract for chromatography was prepared by centrifugation at 20,000 rpm for 20 min. The supernatant was collected and the sample was stored in -20 ℃ for further analysis. The lyaste was passed through a membrane filter of 0.2µ pore size before applying to column.

3.2.10   Total Protein Quantification

 Quantification of total protein concentration of all samples was done spectrophotometrically (DU-640 spectrophotometer, Beckman Coulter, Fullerton CA) using DC assay technique (Bio Rad, Hercules CA) at 750 nm. This assay is a colorimetric techniques based on Lowry's method. Bovine serum albumin was used as standard.

3.2.11 Immobilized Metal Affinity Chromatography

Multiple rounds of chromatography (n=3) were done using an AKTA FPLC system (GE Healthcare, Piscataway NJ). A 5 ml column (HisTrap HP IMAC column, Amersham Biosciences) was charged with 5 column volume (CV) of 25 mg/ml $CoCl_2$. After charging it with metal ion, the IMAC column was washed with 5 CV of distilled water prior to equilibration with 5 CV of binding buffer. During IMAC, the binding buffer contained 1 M NaCl and 0.05 M Na_2HPO_4 (pH 7.4), while the elution buffer had lower pH (5.5). Flow rate was maintained at 0.5 ml per minute throughout a run. Elution was monitored at 280 nm with the integrated UV monitor. After loading a sample, unbound / lightly retained proteins were washed off the column with 2 CV of buffer. Finally, washing the IMAC column with 5 CV of the low pH elution buffer generated a pool of binding proteins. When appropriate, both elution and flow through were collected for further analysis.

3.2.12 Hydrophobic Interaction Chromatography

Protein samples were applied to a 1 ml HiTrap Octyl FF HIC Column (GE Healthcare, Piscataway NJ). In order to match the composition of the mobile phase, Na_2SO_4 was added to the sample before loading. The binding buffer contained 1M Na_2SO_4, 0.02 M Na_2HPO_4 (pH 7.4). The HIC column was washed and equilibrated with 5 CV of binding buffer. A flow rate of 0.25 ml per minute was maintained throughout the run. After the non-binding proteins passed during the initial washing, the column was developed with a step change in elution buffer (0.025 M Na_2PO_4, pH 7.2). The pool of eluted protein was retained for further analysis, and multiple chromatographic runs (n=3) were performed to demonstrate reproducibility.

3.2.13 SDS PAGE and Western Blotting

Protein samples were separated by 12.5% SDS -polyacrylamide gel (SDS-PAGE). Proteins were transferred from the gel to nitrocellulose membrane at 100v for seventy minutes. Membranes were incubated with anti-FLAG at 1:4000 dilution, anti-GFP at 1:2000 and anti-GST at 1:2000 with 0.2% BSA in TBS-tween overnight. After that, membranes were incubated with secondary antibody at 1:10000 dilution for 3 hours with 0.2% BSA in TBS- tween, then developed using NTB and BCIP in TBS - Tween.

3.3 Results

3.3.1 Expression and Solubility of MelA and NT1 with FLAG tag

Both the entire M*elA* gene product and a novel hydrophobic sequence based on *MelA* were cloned and expressed. Figures 3-8 and 3-9 describe the strategy by which the entire protein, or alternately, the 194 amino acid sequence named NT1 was fused to FLAG. The plasmids, as designed, removed the pelB leader sequence from expressed proteins and placed an antibody-identifiable feature on the C-terminus of the amino acid sequences under investigation. Expression of either MelA-FLAG or NT1 proceeded in batch cultures, as described in the methods section. Typically, overnight growth produced a culture with a final optical density of three – four that served as the source of cell extract. Lysates of either source (MelA-FLAG, or NT1-FLAG) were examined via SDS-PAGE and Western blot to confirm the presence of the target product (Figure 3-14& Figure 3-15). As evident by bands corresponding to the proper molecular weights (51.8 kD and 26.7 kD), both products were expressed in soluble extract of cell lysates.
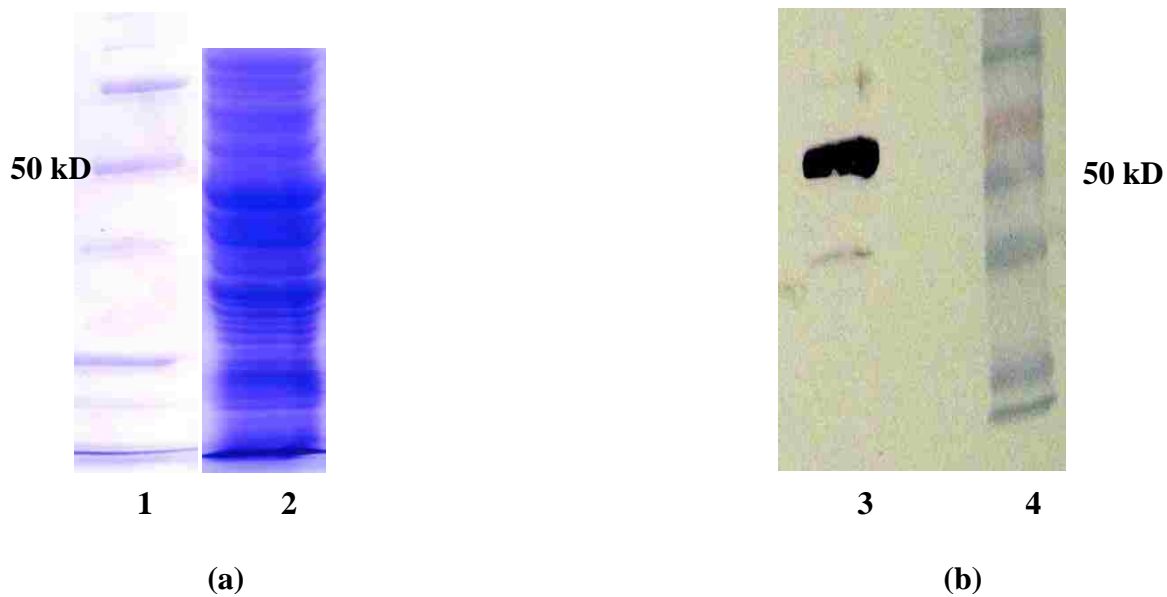
Figure 3-14: Expression of MelA-FLAG in *E.coli* BL-21; (a) 12.5% SDS-Polyacrylamide Gel; (b) Western blotting for Expression of MelA-FLAG with anti-FLAG antibody

Lanes: 1, BioRad Dual Color protein Marker; 2, Cell lysate containing MelA-FLAG; 3, Cell lysate containing MelA-FLAG; 4, Invitrogen prestained protein Marker
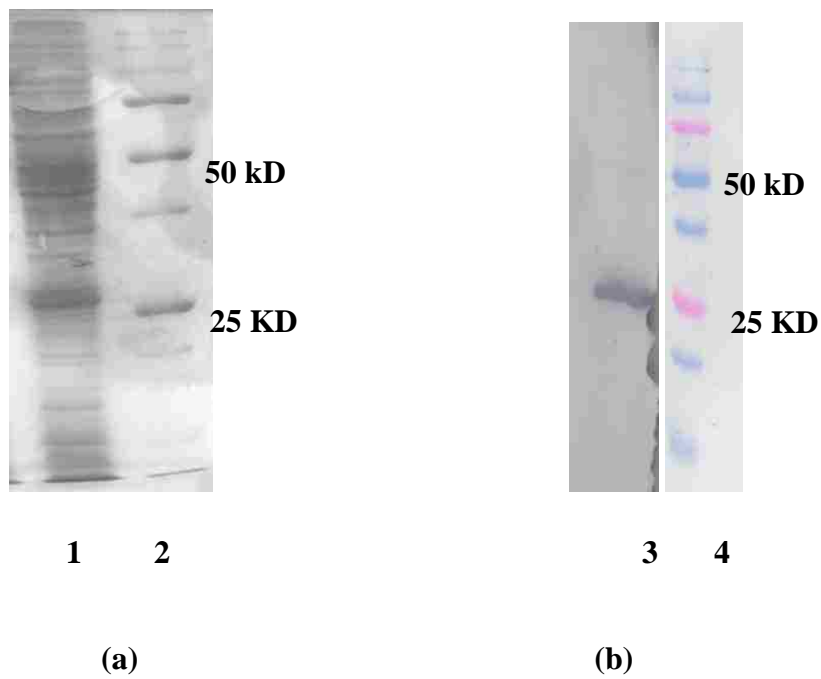
Figure 3-15: Expression of NT1-FLAG in *E.coli* BL-21; (a) 12.5% SDS-Polyacrylamide Gel; (b) Western blotting for Expression of NT1-FLAG with anti-FLAG antibody

Lanes: 1, BioRad Dual Color protein Marker; 2, Cell lysate containing NT1-FLAG; 3, Cell lysate containing NT1-FLAG; 4, Biorad Dual Color protein Marker

3.3.2 Expression and Solubility of GST and GFP with different developed HIC tags

Hydrophobic regions in proteins play major role in expression and solubility of fusion proteins. If a fused protein is not expressed in soluble form, it brings lot of extra cost to the purification and recovery of the target protein. For an economically successful industrial process it is important that expression and solubility of fused protein is good, this not only decreases the steps involved in purification but also increase the yield. The biggest problem with hydrophobic tags till now has been the solubility of the fused protein. So it was very important to study the effect of different developed regions with proteins like GST and GFP and thus study solubility with both N and C terminal fusion of tags to the control protein. Both the entire *MelA* gene product and a novel hydrophobic sequence based on *MelA* were cloned and expressed. Lysates of the designed clones with either GST or GFP were examined via SDS-PAGE and Western blot to confirm the presence of the target product (Figure3-16&Figure3-17). As evident by bands corresponding to the proper molecular weights (NT1-GFP at 51 kDa, NTB-GFP at 37 kDa, GST-NTB at 37 kDa, GST-HIC mini at 27 kDa, GST-RBNT at 27kDa) products were expressed. Expression in soluble fraction of all the fused proteins were compared with expression of GST and GFP, to see the effect on the solubility of protein when fused to developed tag. From the results it was evident that expression and solubility of GST and GFP were decreased when fused with NTB, HIC mini, NT1 and RBNT. But it was found that when entire *MelA* is fused with GST and GFP, solubility of GST is not affected by fusion to *MelA*.
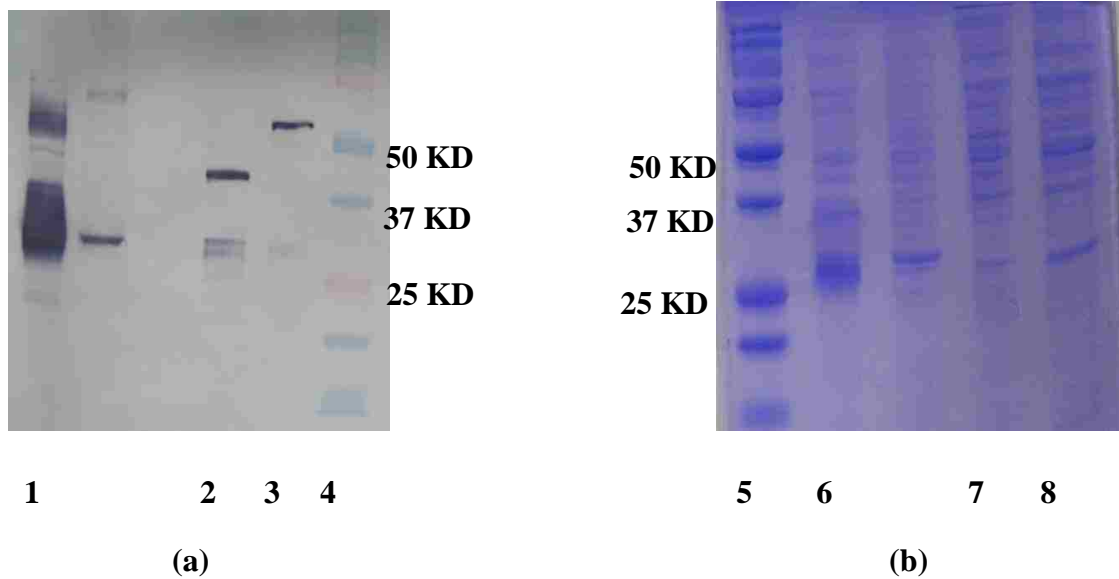
Figure 3-16: Expression of GFP with different developed HIC tags in *E.coli* BL-21; (a) 12.5% SDS-Polyacrylamide Gel; (b) Western blotting for Expression of GFP fused clones with anti-GFP antibody

Lanes: 1&6,Cell lysate containing GFP;2&7,Cell lysate containing NTB-GFP;3&8, Cell lysate containing NT1-GFP; 4&5, Biorad Dual Color protein Marker
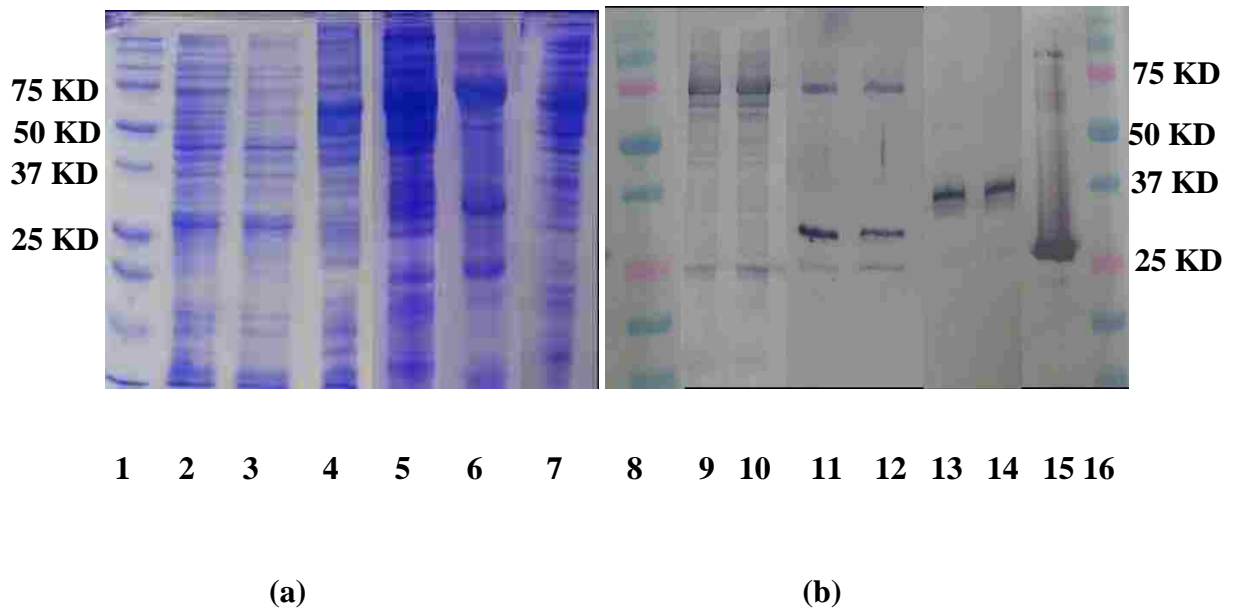
**Figure 3-17: Expression of GST with different developed HIC tags in *E.coli*
BL-21; (a) 12.5% SDS-Polyacrylamide Gel; (b) Western blotting for
Expression of GST fused clones with anti-GST antibody**

Lanes: 1,6&16, Biorad Dual Color protein Marker;2&11, Cell lysate containing
GST-hic mini;4&12,Cell lysate containing GST-RBNT;5&15, Cell lysate
containing GST;6,9,&10, Cell lysate containing GST-MelA;7,13&14, Cell lysate
containing GST-NTB

3.3.3 Hydrophobic Interaction Chromatography and Immobilized Metal affinity chromatography of MELA-FLAG containing lysate

After confirming expression, lysates were subjected to chromatography. With regard to MelA-FLAG, both immobilized cobalt and octyl ligands were capable of capturing the protein. Figure 3-18 shows the elution profile of the IMAC chromatography step. A rather strong signal in the Western blot was associated with the peak that typically eluted from the column during the pH gradient. No signal was associated with any fractions corresponding to unbound / lightly retained material, or any portion of the column wash prior to the start of the gradient. Similarly, MelA-FLAG could be consistently retained and subsequently eluted from HIC columns. These characteristics are evident from the absence and presence of Western blot signals associated with unbound / lightly retained, and adsorbed material (Figures 3-18).
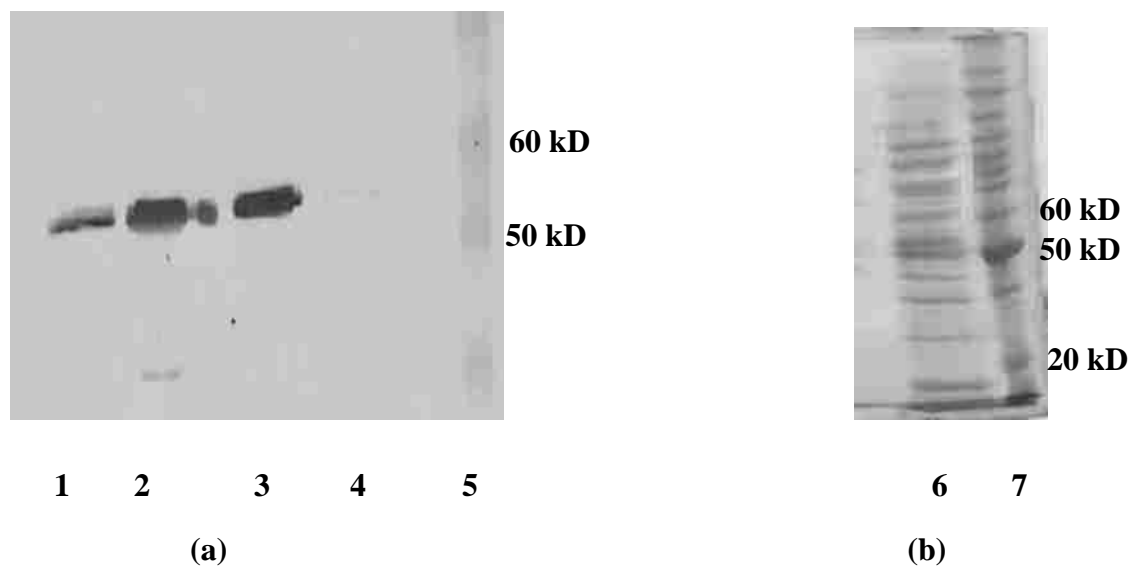
**Figure 3-18: Results of IMAC/HIC of MelA-FLAG containing lysate;
(a) Western blotting results; (b) 12.5% SDS-Polyacrylamide Gel**

Lanes: 1, Cell lysate containing MelA-FLAG; 2,IMAC elutant; 3,HIC elutant; 4, HIC flowthrough;5,Invitrogen prestained protein Marker;6,HIC elutant;7,Invitrogen bench marker.

3.3.4 Hydrophobic Interaction Chromatography and Immobilized Metal affinity Chromatography of NT1-FLAG containing lysate

Results obtained with NT1FLAG indicated that Co (II) IMAC was not able to capture the protein, whereas octyl FF could successfully adsorb and release the recombinant protein derived from *MelA*. Lack of IMAC retention was evident by the presence of a signal associated with the flow through / unbound protein (Figures 3-19). There were no indications of column overload because the Western blot analysis confirmed the lack of NT1FLAG in samples representative of unbound / lightly retained, column washes, or column development. Also, NT1FLAG bound and was subsequently eluted as a single peak from the HIC column. These results were accompanied by no indication of NT1-FLAG in samples other than the elution peak. To study the HIC binding properties of NT1-FLAG, the elution profile was checked at different salt gradient. As seen in Figure 3-20, both western and SDS PAGE showed that NT1-FLAG was not eluted with 50% Buffer B, low signal is seen in 50% buffer B, showing strong binding to HIC, but NT1-FlAG was eluted when column was washed with 100% buffer(absence of Na_2So_4 in buffer B).
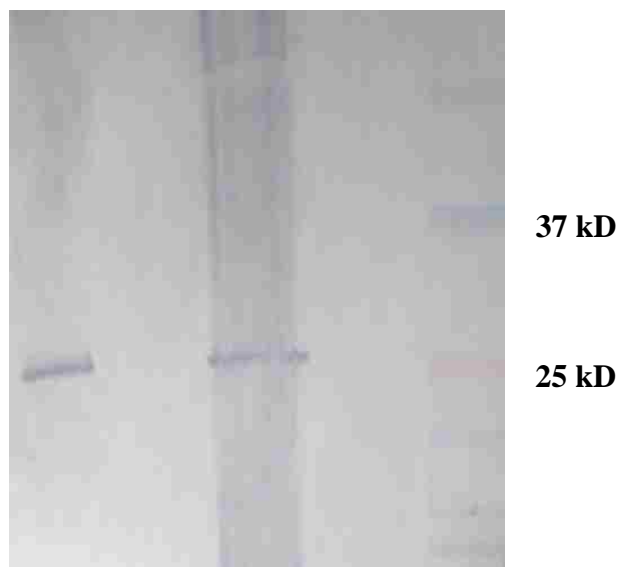
Figure 3-19: Western blot results of IMAC/HIC of   NT1-FLAG containing lysate

Lanes: 1, Cell lysate containing NT1-FLAG; 2, HIC flowthrough; 3, HIC elutant; 4, IMAC elutant; 5, BioRad Dual color Marker
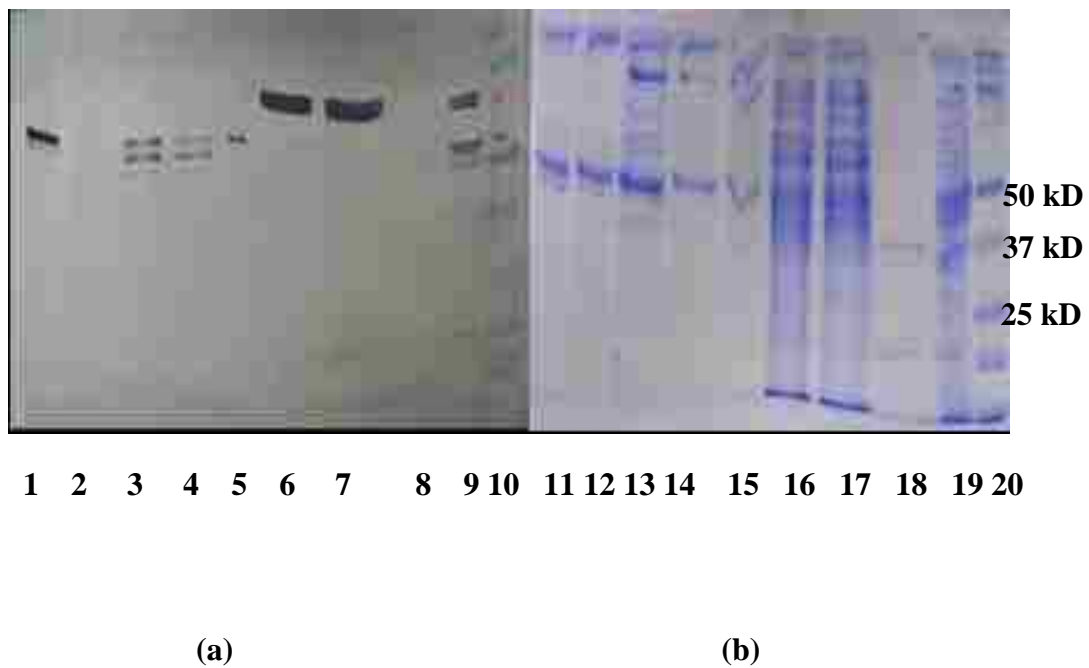
Figure 3-20: Results of HIC of NT1-FLAG containing lysate; (a) Western blotting results; (b) 12.5% SDS-Polyacrylamide Gel

Lanes: 1,2,11&12,Elution at 100 % buffer B;3,4,13&14Elution at 75% buffer B;5&15,wash;6,7,16&17Elution at 50% buffer B;8&18 Flow-through ;9&19 Cell lysate containing NT1-FLAG;10&20, BioRad Dual color Marker

3.3.5 Hydrophobic Interaction Chromatography and Immobilized Metal affinity chromatography with lysate containing GST

Control experiment was first carried out to check if GST can be used as a control protein. It was important that GST didnot bind to HIC and IMAC in given conditions. As seen in result (Figure 3-20). GST was found in low concentration at 1 M salt concentration in binding buffer during HIC. Most of the GST was found in flow through, and a low signal was seen in peak. When tested at salt concentration 0.5 M binding concentration, a strong signal was seen in flow through and there was no signal in elution sample. Similarly, GST containing lysate was subjected to Co (II) IMAC, and it was found that GST did not bind during IMAC and strong signal was present in flow through (Figure 3-20). So GST was selected as control for testing the *MelA* ability to work as tag for IMAC + HIC.
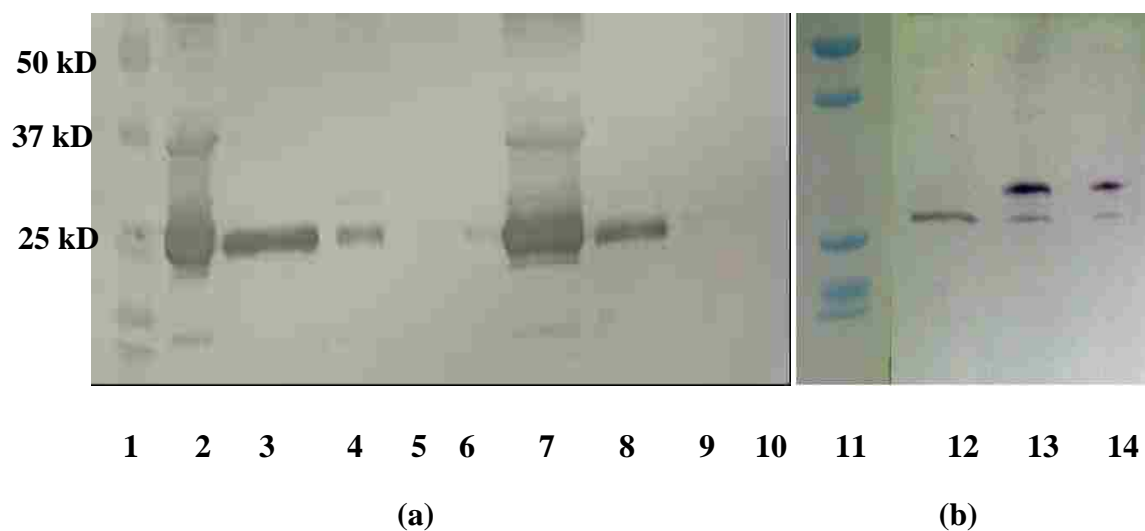
Figure 3-21: Results of HIC and IMAC of GST containing lysate, (a) Western blot of HIC with GST containing lysate (b) Western blot of IMAC with GST containing lysate

Lanes: 1, Bio-Rad Dual color Marker ; 2, Cell lysate containing GST ; 3, flow through at 0.5M salt concentration in binding buffer; 4, Elutant ; 5, Elutant; 6,Elutant; 7, flow through at 0.25 M salt concentration in binding buffer; 8, flow through at 0.25 M salt concentration in binding buffer; 9, Elutant; 10, Elutant

3.3.6 Hydrophobic Interaction Chromatography with lysate containing NT1-GFP

Control experiment was first carried out to check if GFP can be used as a control protein. It was found that GFP was binding at high and medium salt concentration, but not at the lowest salt concentration in binding buffer. It was important that GFT doesn't bind to HIC and IMAC in given conditions. NT1-GFP was subjected to HIC with gradient during elution. As seen in result (Figure 3-21) most of NT1-GFP was found when column was washed with 50% buffer B. there was no NT1-GFP eluted when column was washed with 75% B and 100% B. Most of the NT1-GFP was found when eluted with 50% Buffer Ba and not detected in flow through. These results indicated that NT1-GFP was not very strongly bound HIC column as in case of NT1-FLAG, indicating that folding of NT1-GFP is such that it was not binding to HIC very strongly .

																			75 kD
																			50 kD
																			37 kD
																			25 kD

1   2   3   4   5   6 7   8   9   10   11 12 13 14   15 16 17   18 19 20

(a)   (b)
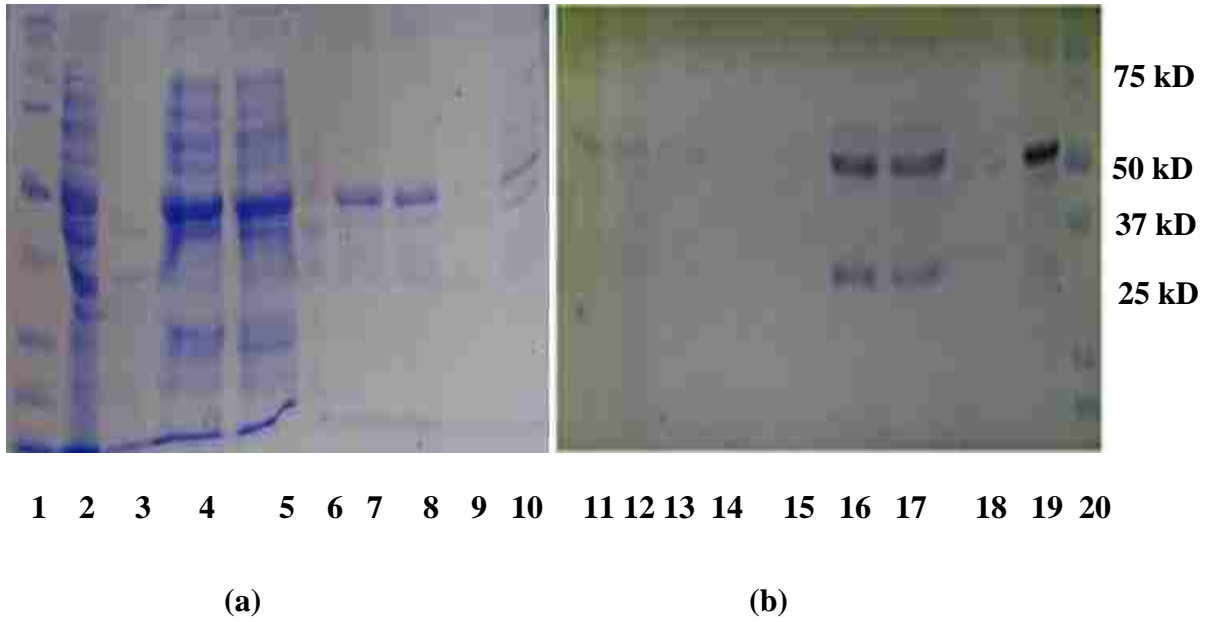
Figure 3-22: Results of HIC of NT1-GFP containing lysate; (a) 12.5% SDS-Polyacrylamide Gel; (b) Western blotting results

Lanes: 1&20, BioRad Dual color Marker; 2&19 Cell lysate containing NT1-GFP; 3&18, Flowthrough;4,5,16&17,Elution at 50 % buffer B;6&15,wash;7,8,13&14Elution at 75% buffer B;9,10,12&11,Elution at 100% buffer B

3.3.7 Hydrophobic Interaction Chromatography with lysate containing GST-NTB and NTB-GFP

NTB is the smaller version of NT1 comprising of *MelA* region between amino acids 342-451. GST-NTB was subjected to HIC at 0.25M salt concentration in binding buffer. The expression of GST-NTB and NTB-GFP decreases compared to GST and GFP in soluble fraction. Both fusion proteins run approximately at 37 kDa. Figure 3-22 and 3-23 show SDS PAGE result of HIC with GST-NTB and NTB-GFP respectively. The elution profile does not show strong presence of GST-NTB or NTB-GFP at these conditions on SDS PAGE.
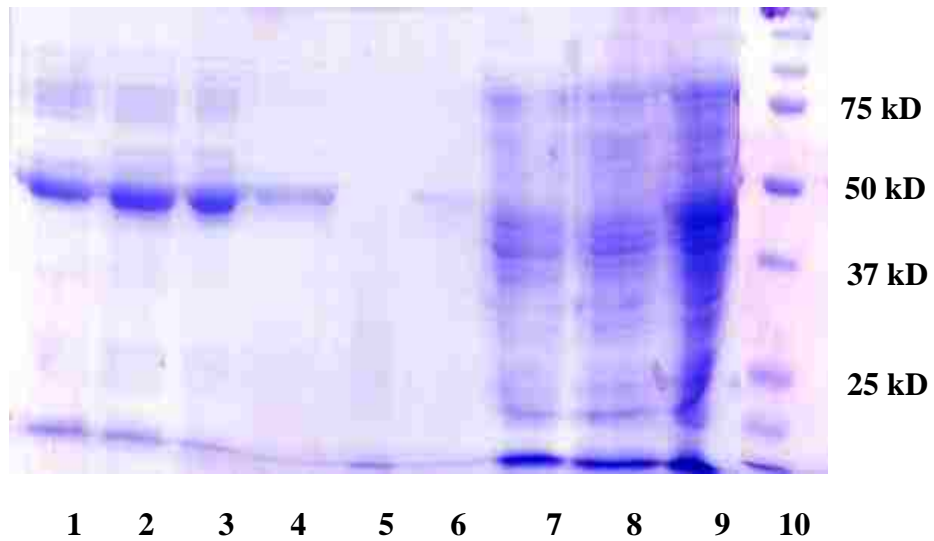
Figure 3-23: 12.5% SDS PAGE results of HIC of GST-NTB containing lysate

Lanes: 1,2,3&4,Elutant;5,6,Wash;7,8,Flowthrough;9,Cell Lysate containing GST-NTB;10, Bio-Rad Dual Color protein Marker

1   2   3   4   5   6   7   8
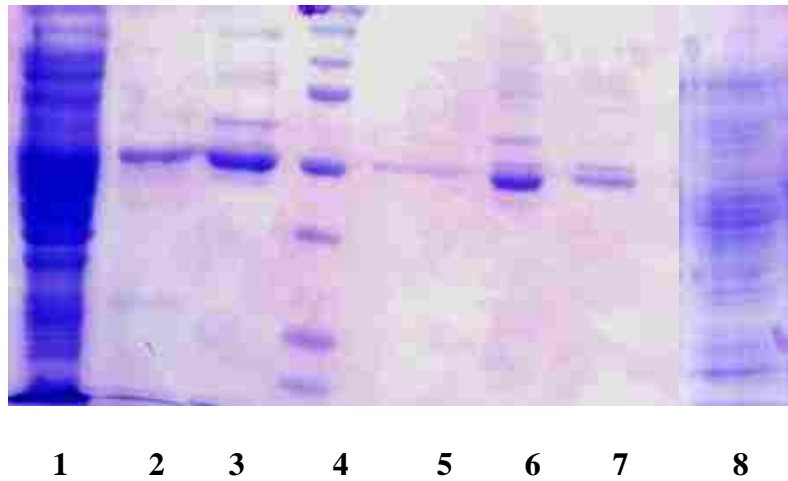
Figure 3-24: 12.5% SDS PAGE results of HIC (0.25M salt in binding buffer) of NTB-GFP containing lysate

Lanes: 1,Cell Lysate containing NTB-GFP;2&3,Elutant;4, Bio-Rad Dual Color protein Marker 5,Wash;6,7,Elutant;8,Flowthrough

3.3.8 Hydrophobic Interaction Chromatography with lysate containing GST-HIC mini and GST-RBNT

HIC mini NTB is the smaller version of NT1 comprising of *MelA* region between amino acids 342-451. GST-NTB was subjected to HIC at 0.25M salt concentration in binding buffer. The expression of GST-NTB and NTB-GFP decreases compared to GST and GFP in soluble fraction. Both fusion proteins run approximately at 37 kD. Figure 3-22 and 3-23 show SDS PAGE result of HIC with GST-NTB and NTB-GFP respectively. The elution profile does not show strong presence of GST-NTB or NTB-GFP at these conditions on SDS PAGE.
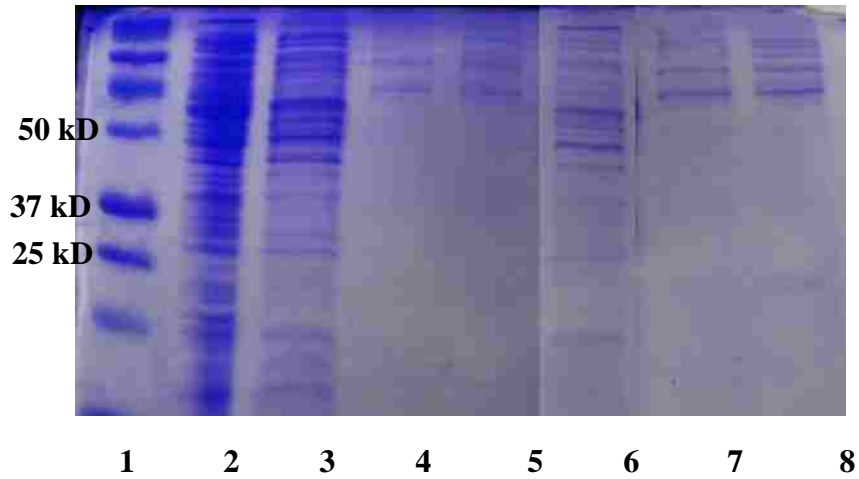
Figure 3-25: 12.5% SDS PAGE results of HIC (0.25M salt in binding buffer) of GST-HIC mini containing lysate

Lanes: 1, BioRad Dual Color Protein Marker; 2, Cell Lysate containing GST-HICmini; 3, Flow-through; 4, 5, 6, 7&8, Elutant
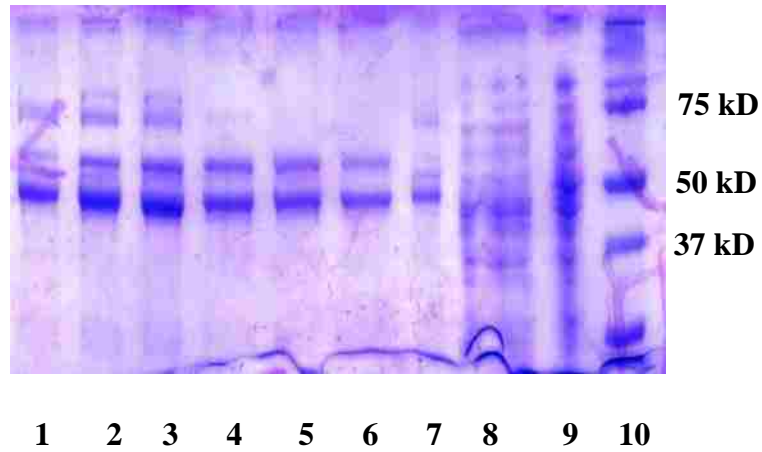
Figure 3-26: 12.5% SDS PAGE results of HIC (0.25M salt in binding buffer) of GST-RBNT containing lysate

Lanes: 1,2,3,4,5,6,7,Elutant;8,Flowthrough;9,Cell Lysate containing GST-RBNT;10, Bio-Rad Dual Color protein Marker

3.4 Discussion

In order to provide a rational design basis for a dual IMAC – HIC affinity tag, the binding properties of alpha galactosidase were examined. HIC is used widely at an industrial scale as it possess several advantages like mild operating conditions, does not affect biological activity of proteins, its easily scalable, and less expensive compared to other affinity chromatographic techniques. But unlike other techniques, there is no affinity tag available for practical use for HIC purification due to solubility issues. NT1, the designed tail by combining hydrophobic regions of *MelA*, fused with FLAG was readily able to bind during HIC, and was found in soluble fraction as discussed in results. Thus NT1 has emerged as potential candidate for the development of first HIC tag that may aid with solubility issues, the subject of our next investigation. NT1 fails to bind to IMAC, but IMAC binding can be easily reestablished by constructing a His$_6$NT1 tag. That way the dual affinity property can be retained for an IMAC+ HIC purification approach.

Both HIC and IMAC binding characteristics could be explained in term of primary amino acid sequences, even in the absence of crystal structure of *E.coli MelA*. The binding properties of NT1-FLAG and analysis of amino acid sequence gave insight into the hydrophobic and metal binding domains of *MelA*. Surface exposed histidines play major roles in IMAC binding, and since metal affinity of *MelA* was lost in NT1, it was interesting to look at the possible IMAC binding sites present in region deleted from the *MelA* (between amino acids 95-341) . It was found that out of total 15 histidine residues present in *MelA*, 8 are present in the region which was deleted from *MelA* to design NT1. Also, a metal binding site for *MelA* which has been identified as Cys173 was deleted in designing NT1. These results leads to the conclusion that the

possible Co(II) IMAC binding motif of *MelA* lies in region between amino acids 95-340 and its deletion leads to loss of Co(II)IMAC binding. In conclusion , the combined hydrophobic regions(NT1) recognized in this study are strong candidates for development of the first affinity tag for hydrophobic interaction chromatography, and when enhanced with a small IMAC tag like His6 will be able to exploit the full binding potential of native *MelA*.

PART III
CHAPTER 4

Proteome-based development of novel affinity tail for Immobilized Metal Affinity Chromatography and Hydrophobic Interaction Chromatography

4.1 Introduction

Based upon the previous results, we know that Alpha galactosidase is able to bind Co (II) IMAC and HIC at all 3 salt concentrations. In this part of the dissertation, we have tested the ability of *MelA* to work as a dual affinity tag to purify the desired protein using Immobilized Metal Chromatography (IMAC) and Hydrophobic Interaction Chromatography (HIC) and also for HIC by itself. Having identified *MelA* as a potential tag by fusing it with FLAG tag, in this part of work, we have tested *MelA* fusion with GST (bigger protein) to purify it using HIC + IMAC and HIC . Here HIC with 0.25M salt concentration in binding buffer was used as the first step, only those proteins with highly hydrophobic regions will be binding at such a low salt concentration and Co (II) IMAC was used as the second step. At 0.25M salt concentration during HIC, there are very few co-eluting proteins and so a successful HIC tag can also be used alone as a purification step, with tailored strain (with gene knock out of the contaminants). This will also check the loss of protein by having more than one step purification.

It is desirable to have a tag which doesn't interfere with solubility or biological activity of the fused protein. Solubility has been the biggest issue with the development of affinity tag for HIC. With *MelA* being a native protein sequence we intend to handle this problem. Here, we have also studied effect of *MelA* fusion to GST solubility and enzyme activity.

4.2 Methods and Materials

4.2.1 Cloning of GST-melA-pGEX-6P-2

Figure 4.2 shows the scheme of GST- *MelA* cloning**.** A cDNA clone for alpha-galactosidase (*MelA*) was obtained by RT-PCR using genomic dna isolated from *E.coli* strain BL21. Forward primer P17 and reverse primer P18 (Table 4.1) matching the sequences for *MelA* (*E. coli* strain K12 Accession #AAA97019) were designed to include BamHI and EcoRI sites, respectively. The resulting PCR fragment (Figure 4.3) was digested using BamHI and EcoRI and ligated into a similarly restricted pGEX-6P-2 (GE Healthcare) to produce the plasmid GST-melA-pGEX-6P-2. The clones containing the *MelA* fragment was screened by doing PCR (Figure 4.4). Final sequence was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock).

E.coli DNA

PCR with primers P1 and P2

PCR product (*MELA*) with BamH1 and EcoR1 sites

PCR product and pGEX-6P2 digested with BamH1
and EcoR1, and joined together by ligation

LacI

GST

BamH1

pGEX-6P2-MelA
5.5 kb

Mel A

Amp R

EcoR1
Not1
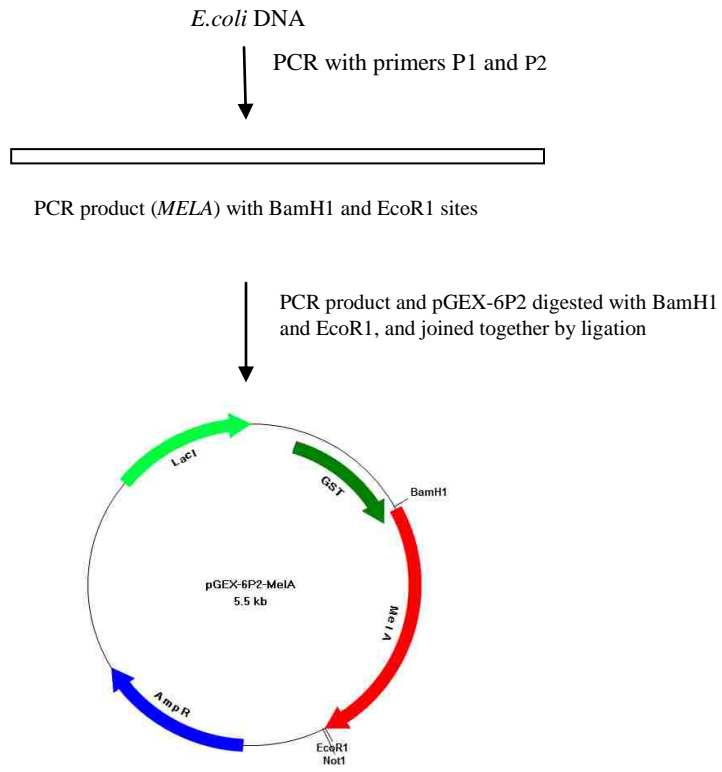
Figure 4.1: Cloning of pGEX-6p-2- MelA

1.5 kb

1   2
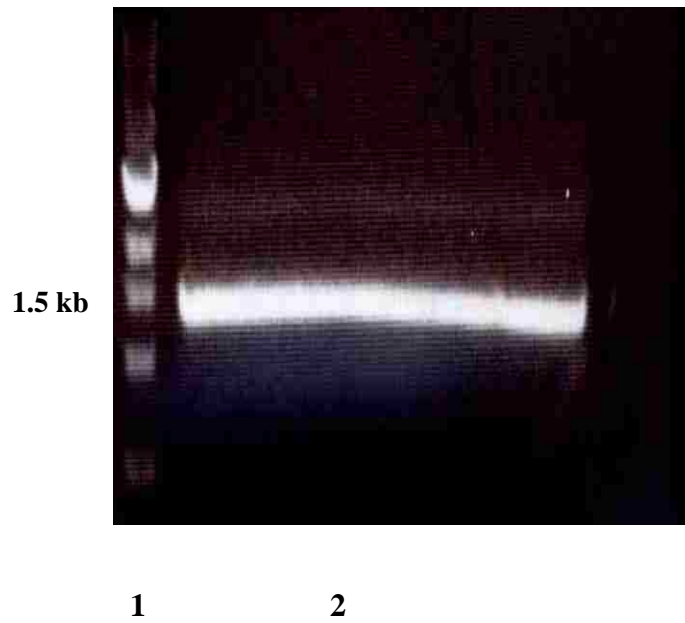
Figure 4-2: 0.8% agarose gel of PCR for amplification of *MelA* PCR product.

Lanes: 1, NEB 1Kb DNA ladder; 2, *MelA* PCR pdt (1.4 kb)

1.5 kb

1 2 3 4 5 6 7 8 9 10
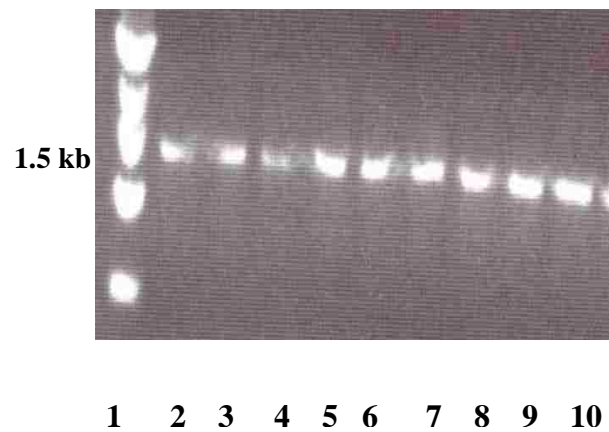
Figure 4-3: 0.8% agarose gel of PCR for amplification of *MelA* PCR product.

Lanes: 1, NEB 1Kb DNA ladder; 2 -10, *MelA* PCR pdt (1.4 kb)

4.2.2 Cloning of pGFPuv-MelA-GFP

A cDNA clone for alpha-galactosidase (*MelA*) was obtained by RT-PCR using genomic dna isolated from *E.coli* strain BL21. Forward primer P19 and reverse primer P20 matching the sequences for *MelA* (*E. coli* strain K12 Accession #AAA97019) were designed to include EcoRI and HindIII sites, respectively. The resulting PCR fragment was digested using EcoRI and HindIII ligated into a similarly restricted plasmid pGEM-3Z to produce the plasmid MelA in pGEM-3Z. The clones containing the pGEM-3Z- MelA fragment was screened by doing PCR. The purified plasmid pGEM-3Z-MelA was double digested with Kpn1 and Hind III and ligated to similarly digested plasmid pGFPuv. The resultant plasmid contained MelA-GFP in pGFPuv. This plasmid was used as template to amplify MelA-GFP using forward primer P21 and Reverse primer P22 with restriction sites Bgl II and HindIII. The amplified pcr product of MelA-GFP was double digested using Bgl II and HindIII and ligated in similarly digested plasmid pBAD-myc-HisA. Final sequence was determined by DNA sequencing (Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock).

Primer	Sequence (5'→3')
P17	CAGGATCCATGATGTCTGCACCCAAAATTACATTTATCGGC
P18	CGAATTCTTAACGGTGCAACCAGCCTGG
P19	GGCCGAAGCTTGATGTCTGCACCCAAAATTACATTTATCGGCG
P20	TGGTACCGCACTTCCGCGGGGTACAAGGTGGTGGTGACGGTGCAACCAGCCTGG
P21	GCAAAGATCTATGTCTGCACCCAAAATTACATTTATCGGC
P22	TGCAAGCTTTTATTTGTAGAGCTCATCCATGCCATGTGTAATC

Table 4-1: Primers used in study

4.2.3 Culture growth and Lysate preparation

Single colony of *E .coli* BL21 with plasmids containing inserts were grown on Luria Berteni (Sigma Chemicals, St. Louis MO) agar plates with 100µg/ml of ampicillin and used to inoculate 5ml of primary culture and grown at 37ºC temperature overnight. Next day, 500 ml LB broth with ampicillin (100µg/ml concentration) was inoculated with the primary culture and grown at 37ºC until it reached an O.D. of 0.4-0.6, measured at 600nm. 0.5mM IPTG was used to induce the expression and the culture was grown at 28ºC with sufficient agitation to remain saturated with oxygen overnight. After 18 hours, O.D of about 2.0 was measured for the culture; cells were obtained as pellet by centrifugation at 10,000 rpm for 30 mins (Beckman Model J2-21 M/E). The cell pellet was suspended in phosphate buffer (0.05 M phosphate, 0.25 M NaCl, pH 7.2), with addition of protease inhibitor (Sigma, St Louis,MO,USA). Cells were lysed via four repeated cycles of sonication followed by freezing at -20ºC with addition of 100µl protease inhibitor (Sigma, St Louis, MO) for 10 ml of cells in phosphate buffer. To remove cell debris, the final extract for chromatography was prepared by high speed centrifugation at 20,000 rpm for 20 mins. The supernatant was collected and the sample was stored in -20ºC for further analysis. The lyaste was passed through a membrane filter of 0.2µ pore size before applying to column.

4.2.4 Total Protein Quantification

 The total protein assay for quantification of all protein samples was done spectrophotometrically (DU-640 spectrophotometer, Beckman Coulter, Fullerton CA) using a DC assay (Bio Rad, Hercules CA) at 750 n.m. The D.C assay is a colorimetric techniques based on Lowry's method. Bovine serum albumin was used as standard.

4.2.5 GST Activity Test

GST activity of fusion protein GST-MelA was tested to check the effect of *MelA* fusion on GST enzyme activity. GST assay kit (Sigma, MO, USA) was used for this purpose. This assay is based on the principle GST mediated catalysis of conjugation of L- glutathione to CDNB (1-Chloro-2, 4-dinitrobenzene) through the thiol group of the glutathione.

$$\text{GSH} + \text{CDNB} \xrightarrow{\text{GST}} \text{GS-DNB Conjugate} + \text{HCl}$$
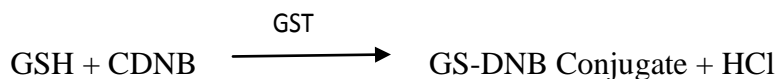
Conjugation of the thiol group of glutathione to the CDNB substrate causes an increase in the absorbance at 340 nm. Rate of increase in the absorbance is directly proportional to the GST activity in sample. Cell lysates containing GST was used as positive control, and H_2O was used to measure blank reading. BL-21 lysate was used as a negative control. All the reactions were tested in 1 ml quartz cuvette. Table 4.2 shows the mixture for testing GST activity.

Reagent	Volume
Dulbecco's Phosphate Buffered Saline	980 µl
200mM L-Glutathione reduced	10µl
100mM CDNB	10µl
sample	2-20µl

Table 4-2: Reaction mixture for GST activity test

4.2.6 Hydrophobic Interaction Chromatography

The protein sample eluted during IMAC or the cell lysate was applied to a 1-ml HiTrap Octyl FF HIC Column (GE Healthcare, Piscataway NJ). Na_2SO_4 was added to the sample before loading to match the composition of the HIC mobile phase. The binding buffer was made of 0.25M Na_2SO_4, 0.02 M Na_2HPO_4 (pH 7.4). The HIC column was washed and equilibrated with 5 CV of binding buffer. Samples were applied to a HIC column at 0.25 ml per minute and maintained throughout the run. After the non-binding proteins passed during the initial washing, the column was developed with a step change in elution buffer (0.025 M Na_2PO_4, pH 7.2). The pool of protein eluted was used for further analysis like Western blotting and protein quantification or loaded to IMAC as the second step.

4.2.5 Immobilized Metal Affinity Chromatography

A 5-ml column (HisTrap HP IMAC column, Amersham Biosciences) was charged using 5 column volume (CV) of 25 mg/ml $CoCl_2$ using an AKTA FPLC system (GE Healthcare, Piscataway NJ). The IMAC column was washed with 5 CV of distilled water after charging it with cobalt. During IMAC, the binding buffer contained 1 M NaCl and 0.05 M Na_2HPO_4 (pH 7.4), while the elution buffer had lower pH (5.5). Before sample loading, the column was washed and equilibrated with 5 CV of binding buffer. Flow rate was maintained at 0.5 ml per minute throughout the run. Elution was monitored at 280 nm with the integrated UV monitor. After loading the proteins, the unbound proteins were washed off the column with 2 CV of wash buffer. Finally, washing the IMAC column with 5 CV of the low pH, elution buffer generated pool of binding proteins that was subjected to western blotting. Both elution and flow through were collected for further analysis.

4.2.6 SDS PAGE and Western Blotting

Protein samples were separated by 12.5% SDS -polyacrylamide gel (SDS-PAGE). Proteins were transferred from the gel to nitrocellulose membrane at 100volts for seventy minutes. Membranes were incubated with anti-GST (GE Healthcare) at 1:4000 dilution and 1:2000 dilution for Anti-GFP (Clontech) with 0.2% BSA in TBS-tween overnight. After that, membranes were incubated with secondary antibody at 1:10000 dilution for 3 hours with 0.2% BSA in TBS- tween, then developed using NTB and BCIP in TBS - Tween.

4.3 Results

4.3.1 Expression and Solubility of GST-MelA
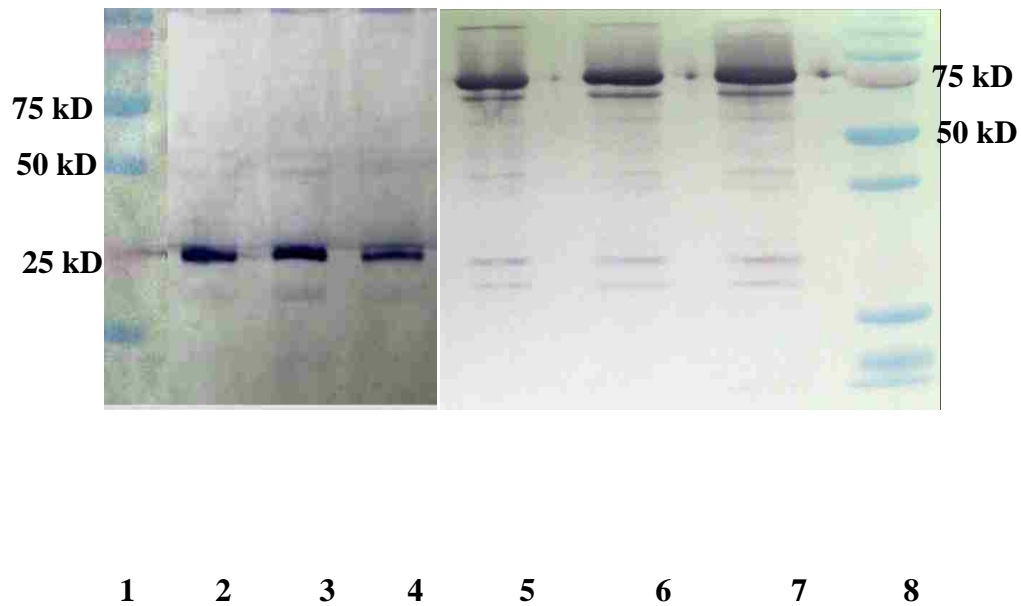
Figure 4-4: Western blot results of expression and solubility of GST and GST-MelA

Lanes: 1, Bio-Rad Dual color marker; 2, total protein fraction of GST; 3, Insoluble fraction of GST; 4, Soluble fraction of GST; 5, total protein fraction of *GST-MelA*; 6, Insoluble fraction of GST-MelA; 7, Soluble fraction of GST-MelA; 8, Bio-Rad Dual color marker
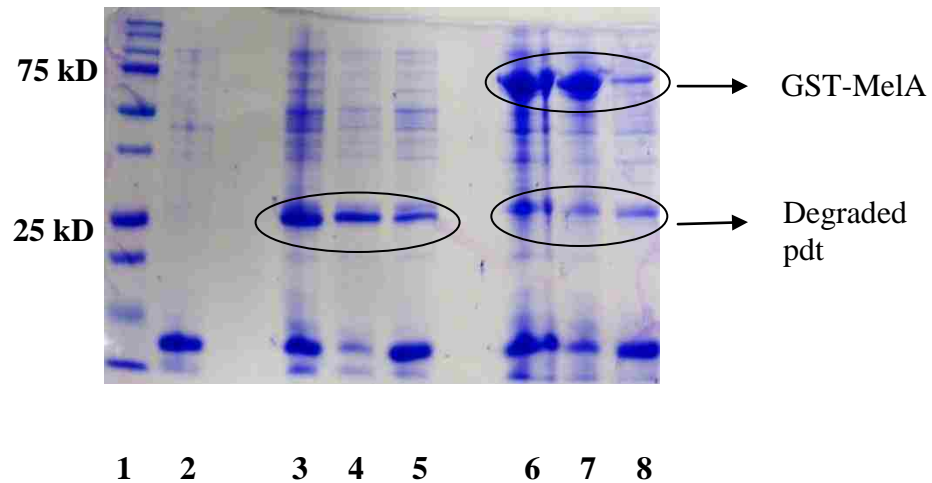
Figure 4-5: 12.5% SDS PAGE of expression and solubility of GST-MelA and GST

Lanes: 1, Bio-Rad Dual color marker; 2, uninduced sample; 3, total protein fraction of GST; 4, Insoluble fraction of GST; 5, Soluble fraction of GST; 6, total protein fraction of GST-MelA; 7, Insoluble fraction of GST-MelA; 8, Soluble fraction of GST-MelA

The expression of GST-MelA was checked for solubility of the fusion protein. As per the western blot and SDS PAGE (Figure 4-5 and Figure 4-6), fusion of *MelA* to GST does not hinder its solubility. Solubility and expression are two big issues with the fusion protein. But we found that *MelA* as a tag is not at all interfering with the GST solubility and the expression of GST-MelA is high in soluble fraction. Both GST and GST-MelA are also present in insoluble fraction, due to high expression not all the protein is folded. But from the results it is clear that amount of GST present in soluble fraction is similar to the GST-MelA present in soluble fraction showing that solubility and expression of GST is not affected by fusion with *MelA*. But in case of GST-MelA, some degraded product is also noticed. We have calculated expression of GST and GST-MelA lysates in total, soluble and insoluble fractions. Densitometry was performed to estimate relative protein concentration using ImageJ Software. As per the table below, approximately 21% of the total protein present in soluble fraction of GST-MelA lysate is GST-MelA (75kD) and the degraded product (26 kD).

Sample	Relative Concentration of GST/Fusion protein in total fraction
GST Lysate (Total protein)	22%
GST Lysate (soluble Fraction)	19.4%
GST Lysate (Insoluble Fraction)	23.87%
GST-MelA Lysate (total protein)	17%
GST-MelA(Soluble Fraction)	21.26
GST-MelA(Insoluble Fraction)	23%

Table 4-3: Table for the relative concentration of protein present in soluble, insoluble and total fraction of lysates

4.3.2 Hydrophobic Interaction Chromatography and Immobilized Metal affinity chromatography with lysate containing GST-MelA and MelA-GFP

Lysate containing GST-MelA was first subjected to HIC (Octly FF) with 0.25M salt concentration in binding buffer as the first step. Unlike GST, GST-MelA bound to HIC even at a low salt concentration. As seen in Figure 4-7, 4-8 and 4-9, GST-MelA binds to HIC at even the lowest salt concentration tested here. Strong signal was observed in HIC elutant and a weak signal is seen in the flow-through. The HIC elutant was then subjected to Co (II) IMAC as the second step to check its binding. As seen in Figure 4-9, the results of Western blotting, no signal was observed in IMAC flow through and strong signals were observed in IMAC elutant.

As seen in Figures, HIC and IMAC samples were loaded to check the total contaminant pool (TCP). Here HIC (0.25M salt concentration in binding buffer) was the first step, and most of the proteins were not binding HIC at such a low salt concentration and present in flow through. About 6-7 proteins are present in HIC elutant along with GST-MelA. When HIC elutant was subjected to IMAC, IMAC elutant shows presence of only two proteins, GST-MelA at 75kD and *MelA* at 50 kD.

Silver staining was performed to identify proteins present even in low concentrations (Figure 4-8). 4-5 more proteins are observed in very low concentrations with silver staining which is more sensitive than Coomassie staining. To check binding of GST-MelA to less hydrophobic column than Octyl FF is Phenyl FF column to check if GST-MelA can bind to it and if contaminant proteins can be eliminated by using less hydrophobic column. It was found that as much as 4-5 proteins were present in HIC elutant.

MelA-GFP containing lysate was subjected to HIC to check C terminal fusion of *MelA* to GFP and check its ability to work as a HIC tag. As seen in Figure 4-12, Western blotting results of samples from HIC of MelA-GFP at 0.25M salt concentration in binding buffer. MelA-GFP low signal was seen in flow-through, but strong signal was observed in MelA-GFP. But when compared the recovery and presence of fusion protein in flow-through, it seems that GST-MelA is binding with higher affinity and more target protein was recovered.
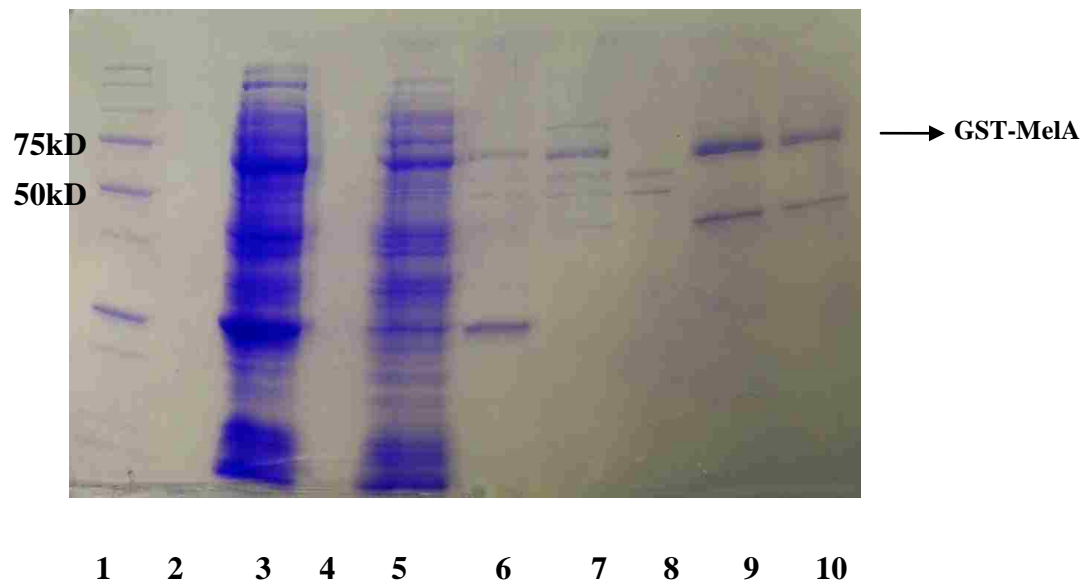
Figure 4-6: Coomassie blue staining of 12.5 %SDS gel of HIC/IMAC with 0.25 M salt in binding buffer of GST-MelA containing lysate

Lanes: 1, Bio-Rad Dual color Marker ; 2, Blank; 3, Cell lysate containing GST-MelA ; 4, Blank; 5, HIC flow-through; 6, HIC elutant ; 7, IMAC flow-through ; 8,IMAC flow-through ; 9, IMAC elutant ; 10, IMAC elutant

75kD

50kD

→ GST-MelA

1   2   3   4   5   6   7   8
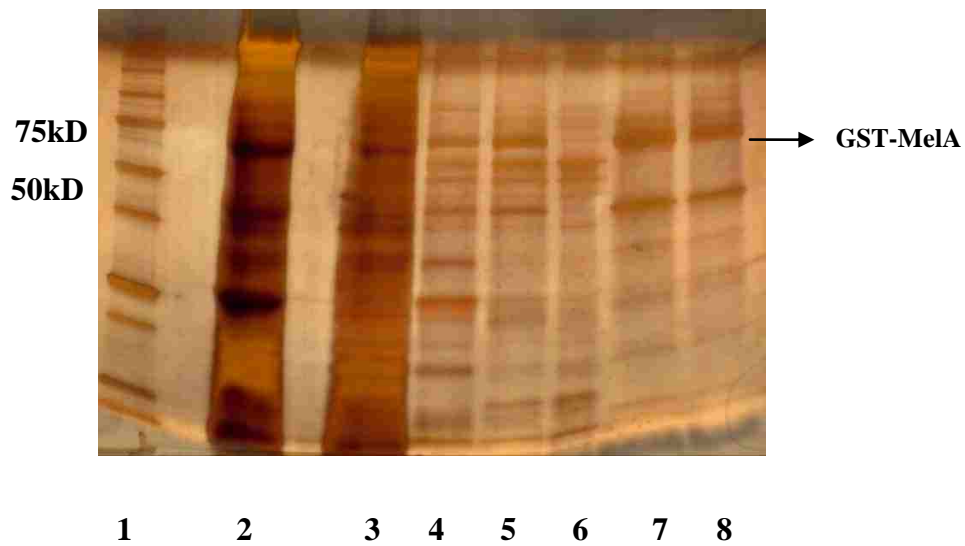
Figure 4-7: Silver staining of 12.5 %SDS gel of HIC/IMAC with 0.25 M salt in binding buffer of   GST-MelA containing lysate

Lanes: 1, Bio-Rad Dual color Marker; 2, Cell lysate containing GST-MelA ; 3, HIC flow-through; 4, HIC elutant ; 5, IMAC flow-through ; 6,IMAC Flow-through; 7, IMAC elutant ; 8, IMAC elutant
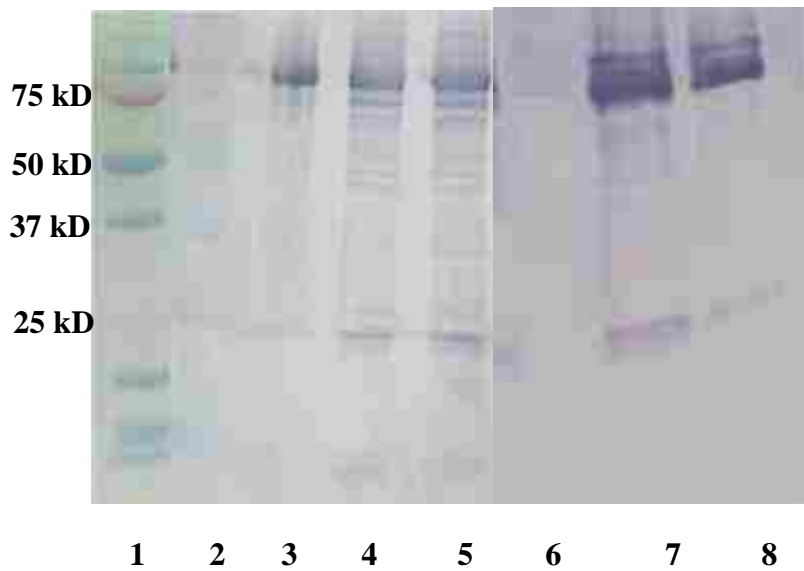
Figure 4-8: Western blot results of HIC with 0.25 M salt in binding buffer of GFP- MelA containing lysate

Lanes: 1, Bio-Rad Dual color Marker; 2, HIC flow through; 3, Cell lysate containing MelA-GFP;4&5 HIC elutant; 6,wash;7&8,HIC elutant
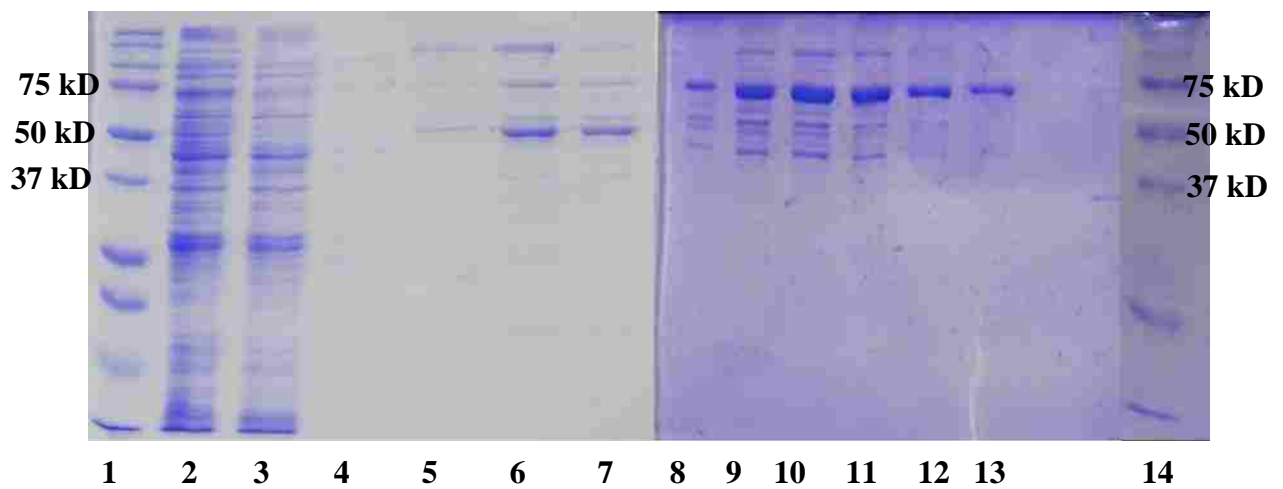
Figure 4-9: 12.5 %SDS gel of HIC using Phenyl FF column with 0.25 M salt in binding buffer of   GST-MelA containing lysate

Lanes: 1&14, Bio-Rad Dual color Marker; 2, Cell lysate containing GST-MelA; 3, flow-through; 4&5, wash; 6-13, HIC Elutant
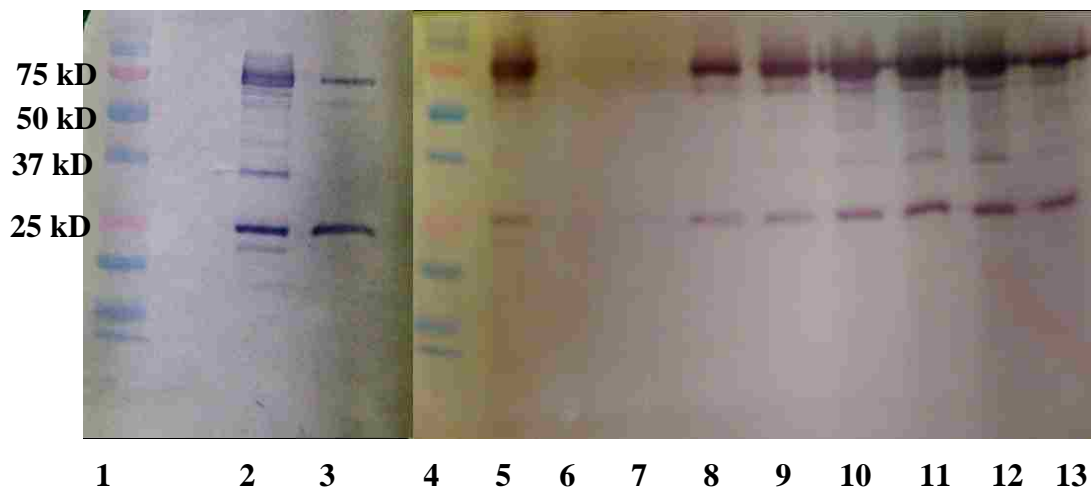
Figure 4-10: Western Blotting of HIC with 0.25 M salt in binding buffer of GST-MelA containing lysate

Lanes: 1&4, Bio-Rad Dual color Marker; 2&5, Cell lysate containing GST-MelA; 3, flow-through; 6&7, HIC flow-through; 8-13, HIC Elutant
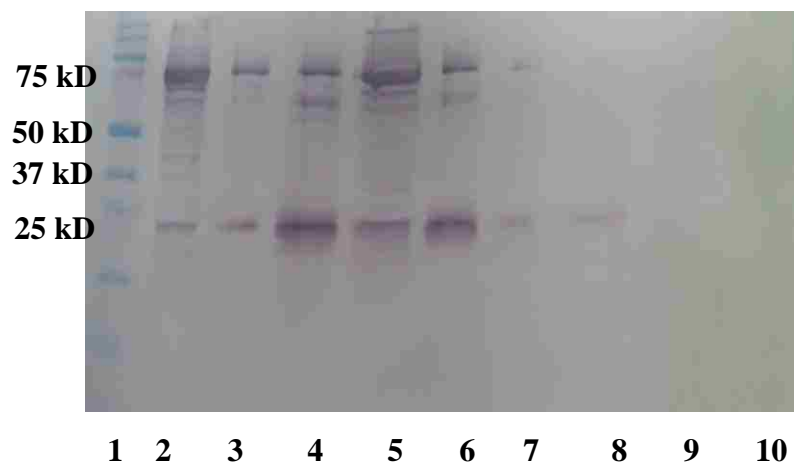
**Figure 4-11: Western Blotting of HIC with 0.25 M salt
in binding buffer of MelA-GFP containing lysate**

Lanes: 1, Bio-Rad Dual color Marker; 2, Cell lysate containing MelA-GFP;
3, flow-through; 4, Wash; 5-10, HIC Elutant

4.3.4 GST Activity Test

GST activity of samples was tested using the kit as mentioned earlier. Blank, BL-21 lysate, GST lysate and GST-MelA lysate were tested. Absorbance at 340 n.m was taken at every 30 seconds. The results of absorbance taken are shown below as graph representation, comparing each samples are shown in Figure 4-10.
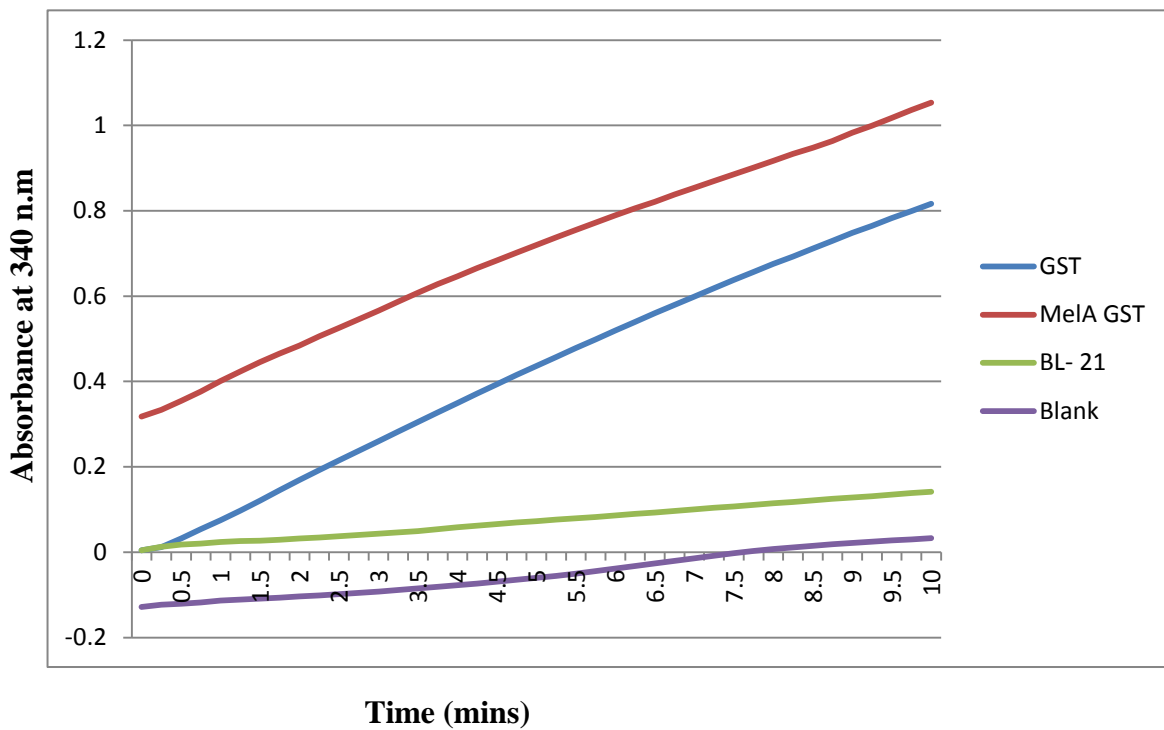
Figure 4-12: Graph representation of absorbance at 340 n.m
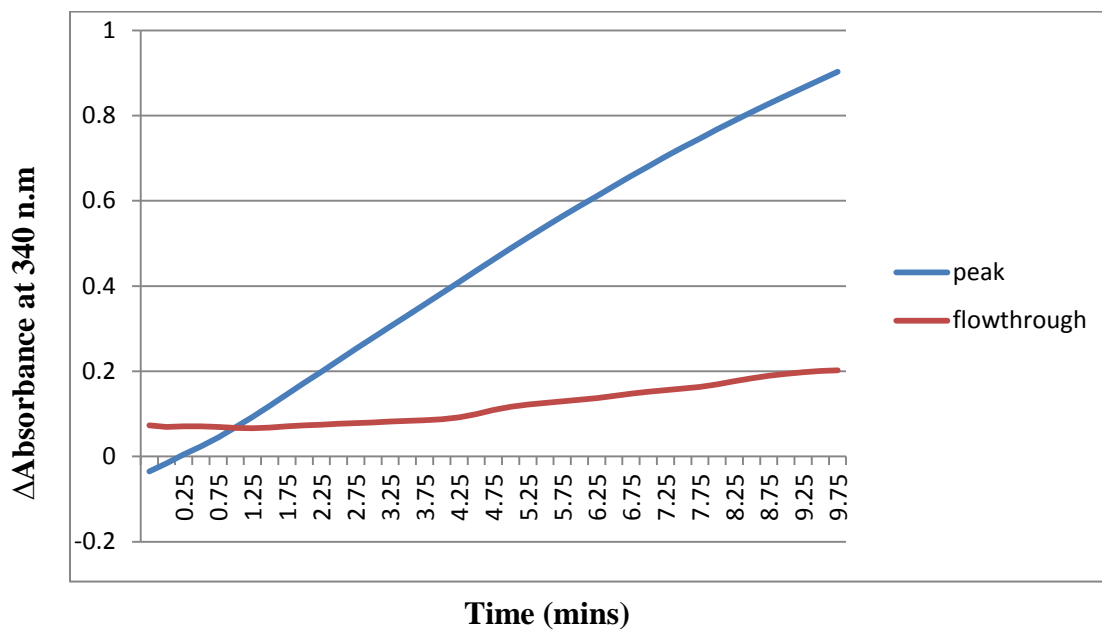
Figure 4-13: Graphical representation of absorbance at 340 n.m of HIC samples

Fraction	Volume	Total Protein (mg)	Activity (units)	Total Activity (units*ml)	Specific activity(units/mg)	Fold Purification
Lysate	3ml	15.3	2.08	6.24	0.4	1
Hic Peak	6ml	0.360	1.947	11.4	31.6	79
Hic Flowthrough	4ml	12.56	-			

Table 4-4: Purification Table of GST-MelA

GST specific activity of each sample was calculated. As per the Figure 4-13, the enzyme activity of GST-MelA and GST lysates were tested. It was found that the GST activity is not hindered due its fusion to *MelA*. GST and *E.coli* BL-21 lysate was similar to the activity of GST lysate used as positive control. These results show that *MelA* fusion to GST does not affect its biological activity, though fusion of GST to NTB, RBNT and HIC mini lost the GST activity. Specific activity of the protein load, flow-through and elutant during HIC was calculated. And as per Table 4-4, it was found that approximately 80 fold increase in the enzyme activity of the purified protein.

4.4 Discussion

From these results, we found that *MelA* can be used as the affinity tag for both 2 step purification combining HIC + IMAC and one step purification using HIC. The most desirable properties for an affinity tag are that it does not interfere with the folding of the fused protein and thus in turn the solubility and biological activity. Mis-folding of the fused protein makes it insoluble and loses its biological activity, especially with HIC this has been a big challenge. Based on our results we conclude that *MelA* when fused with a desired protein can be produced in soluble form and most of fused protein was recovered during HIC with very few contaminants. Higher specific activity was found in HIC elutant compared to the lysate loaded and the flow-through. The enzyme activity of GST is not hindered with the *MelA* fusion. From the specific activity study of the elutant from HIC experiment, it can be concluded that most of the fusion protein loaded on HIC was recovered during elution. Many of the affinity tags are required to get cleaved from the fused protein to get make it biologically active, but here the purified protein is not required to get cleaved if it is not used in vivo. Such a low salt concentration (0.25M Na_2SO_4) in the binding buffer of HIC, leads to elimination of most contaminants and only highly hydrophobic proteins regions will absorb. Elution buffer of HIC is compatible with next step i.e IMAC so there is no need of buffer exchange step required, reducing the cost and loss of product by cutting an extra step for purification. In conclusion, we report *MelA* as the first tag which can be used for purification approach using HIC and IMAC.

CHAPTER 5

Conclusions and Recommendations

5.1 Conclusions:

Here we have developed a bioseparation system comprised of a suite of affinity tags used to facilitate purification. We identified *E. coli* alpha-galactosidase as a protein which, under conditions of salt promoted adsorption, binds both Immobilized Metal Affinity Chromatography (IMAC) and Hydrophobic Interaction Chromatography (HIC) resins. The entire protein and individual domains of the protein have separately been tested as affinity tags based on a separation scheme that consists of IMAC + HIC and HIC alone. We have shown that *MelA* and NT1 (tailored *MelA*) can be successfully used as the affinity tag to purify fused proteins using Hydrophobic Interaction Chromatography at very low salt concentration in binding buffer, with very few co-eluting proteins, though entire MelA as an affinity tag is more effective in terms of recovery of fused protein. The major issues associated with HIC affinity tails i.e expression and solubility is solved by using native protein of *E.coli* as a base for the development of tag. Biological activity of fusion partner is maintained with maximum recovery when HIC is used as the single step purification with very few co-eluting contaminants.

Combining affinity tail design based on proteomic data and homologous recombination to alter genomic protein expression, within the context of improving bioseparation, has placed this work on a path towards creating a novel platform strategy with *E. coli*.

5.2 Recommendations

5.2.1 Development of tailored strain for use of HIC and HIC + IMAC based purification.

For the essential proteins, which are also the co-eluting proteins, modifications are recommended as the future goal so they are not captured during IMAC via the deletion of surface exposed histidine, and reintroduction of the mutant into genome via recombination would be a more effective strategy for important proteins, as their metabolic function would be retained and they would not carry through the bioseparation steps. In contrast, deletion of genes responsible for unnecessary enzymes like *MelA* and *GatD*, for example, do not affect glucose catabolism and likely would not be deleterious to growth and / or recombinant protein expression.

5.2.2 Fusion of *MelA* and NT1 with other proteins to check their ability as affinity tags.

We have used small peptide (FLAG) and larger protein (GST) to fuse with the *MelA* and checked its ability to work as affinity tag. But it is recommended that the developed tag should be fused and check with other proteins which do not show affinity to HIC under the given conditions naturally. Such proteins which can act as control should be fused with the tag and tested for purification conferred by the tag.

5.2.3 Comparison of developed tag (MelA and NT1) with other affinity tags.

Though *MelA* shows binding capacity to HIC and IMAC, IMAC retention can be increased by using His6, when IMAC+ HIC to be used as purification approach. Dissociation constant (Kd) is defined as half the concentration of ligand required to occupy half of the binding sites of proteins. A smaller dissociation constant value leads to higher affinity towards the ligand. So, affinity tags with lower Kd value are desired to have higher binding. Low Kd indicates that less concentration of A and is required for formation of complex [AB]. Different affinity tags can be

compared for better binding properties for both IMAC and HIC which could be quantitatively compared based on their Kd values.

5.2.4 Use of proteome approach for affinity tag development.

Here we have used a proteome approach to identify and develop affinity tag for HIC and successfully solved the problem of expression and solubility for HIC tag. Based on the success of a proteomic approach in this work, such knowledge can be used to design tags for other chromatographic techniques and other living system such as yeast.

REFERENCES

Aizawa, P., S. Winge, et al. (2008). "Large-scale preparation of thrombin from human plasma." Thrombosis Research **122**(4): 560-567.

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

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

Arnau, J., C. Lauritzen, et al. (2006). "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins." Protein Expression and Purification **48**(1): 1-13.

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

Bartlow, P., G. T. Uechi, et al. (2011). "Identification of native Escherichia coli BL21 (DE3) proteins that bind to immobilized metal affinity chromatography under high imidazole conditions and use of 2D-DIGE to evaluate contamination pools with respect to recombinant protein expression level." Protein Expression and Purification **78**(2): 216-224.

Becker, K., J. Van Alstine, et al. (2008). "Multipurpose peptide tags for protein isolation." Journal of Chromatography A **1202**(1): 40-46.

Bolanos-Garcia, V. M. and O. R. Davies (2006). "Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography." Biochim. Biophys. Acta **1760**: 1304-1313.

Brodersen, D. E. and P. Nissen (2005). "The social life of ribosomal proteins." FEBS J **272**(9): 2098-2108.

Cai, Y., M. Moore, et al. (2004). "Genomic data for alternate production strategies. I. Identification of major contaminating species for cobalt immobilized metal affinity chromatography." Biotechnol. Bioeng. **88**(1): 77-83.

Cai, Y., M. Moore, et al. (2004). "Genomic data for alternate production strategies. I. Identification of major contaminating species for cobalt(+2) immobilized metal affinity chromatography." Biotechnology and Bioengineering **88**(1): 77-83.

Carpena, X., W. Melik-Adamyan, et al. (2004). "Structure of the C-terminal domain of the catalase-peroxidase KatG from Escherichia coli." Acta Crystallographica Section D-Biological Crystallography **60**: 1824-1832.

Charcosset, C. (1998). "Purification of proteins by membrane chromatography." Journal of Chemical Technology and Biotechnology **71**(2): 95-110.

Deuerling, E., A. Schulze-Specking, et al. (1999). "Trigger factor and DnaK cooperate in folding of newly synthesized proteins." Nature **400**(6745): 693-696.

Dibb, N. J. and P. B. Wolfe (1986). "Lep operon proximal gene is not required for growth or secretion by Escherichia coli." Journal of Bacteriology **166**(1): 83-87.

Diguan, C., P. Li, et al. (1988). "VECTORS THAT FACILITATE THE EXPRESSION AND PURIFICATION OF FOREIGN PEPTIDES IN ESCHERICHIA-COLI BY FUSION TO MALTOSE-BINDING PROTEIN." Gene **67**(1): 21-30.

Douthwaite, S., R. A. Garrett, et al. (1979). "Ribonuclease-resistant region of the 5S RNA and its relation to the RNA-binding sites of proteins L18 and L25." Nucleic Acids Research **6**(7): 2453-2470.

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

Fahrner, R. L., H. L. Knudsen, et al. (2001). "Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes." Biotechnol. Genet. Eng. Rev. **18**: 301-327.

Fayet, O., T. Ziegelhoffer, et al. (1989). "THE GROES AND GROEL HEAT-SHOCK GENE-PRODUCTS OF ESCHERICHIA-COLI ARE ESSENTIAL FOR BACTERIAL-GROWTH AT ALL TEMPERATURES." Journal of Bacteriology **171**(3): 1379-1385.

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

Ferrer-Miralles, N., J. Domingo-Espin, et al. (2009). "Microbial factories for recombinant pharmaceuticals." Microbial Cell Factories **8**: 8.

Follman, D. K. and R. L. Fahrner (2004). "Factorial screening of antibody purification processes using three chromatography steps without protein A." J. Chromatogr. A **1024**: 79-85.

Frangioni, J. V. and B. G. Neel (1993). "SOLUBILIZATION AND PURIFICATION OF ENZYMATICALLY ACTIVE GLUTATHIONE-S-TRANSFERASE (PGEX) FUSION PROTEINS." Analytical Biochemistry **210**(1): 179-187.

Fry, B., T. Zhu, et al. (2000). "Characterization of growth and acid formation in a Bacillus subtilis pyruvate kinase mutant." Applied and Environmental Microbiology **66**(9): 4045-4049.

Fuchs, S. M. and R. T. Raines (2005). "Polyarginine as a multifunctional fusion tag." Protein Sci **14**: 1538-1544.

Gaberc-Porekar, V. and V. Menart (2001). "Perspectives of immobilized-metal affinity chromatography." Journal of Biochemical and Biophysical Methods **49**(1-3): 335-360.

Gassmann, M., B. Grenacher, et al. (2009). "Quantifying Western blots: Pitfalls of densitometry." Electrophoresis **30**(11): 1845-1855.

Gerdes, S. Y., M. D. Scholle, et al. (2003). "Experimental determination and system level analysis of essential genes in Escherichia coli MG1655." Journal of Bacteriology **185**(19): 5673-5684.

Gongadze, G. M., A. P. Korepanov, et al. (2005). "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. **280**(16): 16151-16156.

Gongadze, G. M., V. A. Meshcheryakov, et al. (1999). "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 **451**(1): 51-55.

Graslund, T., G. Lundin, et al. (2000). "Charge engineering of a protein domain to allow efficient ion-exchange recovery." Protein Eng **13**(10): 703-709.

Gutierrez, R., E. M. M. del Valle, et al. (2007). "Immobilized metal-ion affinity chromatography: Status and trends." Separation and Purification Reviews **36**(1): 71-111.

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

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

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

Herbst, K. L., L. M. Nichols, et al. (1994). "A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene 60 from bacteriophage T4." Proc. Natl. Acad. Sci. USA **91**(26): 12525-12529.

Hjerten, S. (1973). "SOME GENERAL ASPECTS OF HYDROPHOBIC INTERACTION CHROMATOGRAPHY." Journal of Chromatography **87**(2): 325-331.

Hjerten, S., J. Rosengren, et al. (1974). "HYDROPHOBIC INTERACTION CHROMATOGRAPHY - SYNTHESIS AND USE OF SOME ALKYL AND ARYL DERIVATIVES OF AGAROSE." Journal of Chromatography **101**(2): 281-288.

Hochuli, E., H. Dobeli, et al. (1987). "NEW METAL CHELATE ADSORBENT SELECTIVE FOR PROTEINS AND PEPTIDES CONTAINING NEIGHBORING HISTIDINE-RESIDUES." Journal of Chromatography **411**: 177-184.

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

Horne, J. R. and V. A. Erdmann (1972). "Isolation and characterization of 5S RNA-protein complexes from Bacillus stearothermophilus and Escherichia coli ribosomes." Mol. Gen. Genet. **119**(4): 337-344.

Humphreys, D. P., S. P. Heywood, et al. (2004). "Engineering of Escherichia coli to improve the purification of periplasmic Fab' fragments: chaning the pI of the chromosomally encoded PhoS/PstS protein." Protein Expression and Purification **37**: 109-118.

Jiang, C. P., J. Liu, et al. (2009). "A mechanistic study of Protein A chromatography resin lifetime." J. Chromatogr. A **1216**(31): 5849-5855.

Kalyanpur, M. (2002). "Downstream processing in the biotechnology industry." Molecular Biotechnology **22**(1): 87-98.

Kato, A., K. Maki, et al. (2007). "Mutational analysis of protein solubility enhancement using short peptide tags." Biopolymers **85**(1): 12-18.

Korepanov, A. P., G. M. Gongadze, et al. (2007). "Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in Escherichia coli." Journal of Molecular Biology **366**(4): 1199-1208.

Korobeinikova, A. V., G. M. Gongadze, et al. (2008). "5S rRNA-recognition module of CTC family proteins and its evolution." Biochemistry-Moscow **73**(2): 156-163.

Kweon, D.-H., D.-H. Lee, et al. (2002). "Characterization of polycationic amino acids fusion systems for ion-exchange purification of cyclodextrin glycosyltransferase from recombinant Escherichia coli." Biotechnol. Prog. **18**(2): 303-308.

Kyte, J. (2003). "The basis of the hydrophobic effect." Biophysical Chemistry **100**(1-3): 193-203.

Kyte, J. and R. F. Doolittle (1982). "A SIMPLE METHOD FOR DISPLAYING THE HYDROPATHIC CHARACTER OF A PROTEIN." Journal of Molecular Biology **157**(1): 105-132.

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

Lee, J., A. Goel, et al. (1994). "Flux adaptations of citrate synthase deficient Escherichia coli." Biochemical Engineering Viii **745**: 35-50.

Lengeler, J. (1975). "MUTATIONS AFFECTING TRANSPORT OF HEXITOLS D-MANNITOL, D-GLUCITOL, AND GALACTITOL IN ESCHERICHIA-COLI K-12 - ISOLATION AND MAPPING." Journal of Bacteriology **124**(1): 26-38.

Lesley, S. A. (2001). "High-throughput proteomics: Protein expression and purification in the postgenomic world." Protein Expression and Purification **22**(2): 159-164.

Li, M., P. Y. Ho, et al. (2006). "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." Biochemical Engineering Journal **30**(3): 286-296.

Li, Y. F. (2010). "Commonly used tag combinations for tandem affinity purification." Biotechnology and Applied Biochemistry **55**: 73-83.

Lightfoot, E. N. and J. S. Moscariello (2004). "Bioseparations." Biotechnol. Bioeng. **87**(3): 259-272.

Liljestrom, P. L. and P. Liljestrom (1987). "NUCLEOTIDE-SEQUENCE OF THE MELA GENE, CODING FOR ALPHA-GALACTOSIDASE IN ESCHERICHIA-COLI K-12." Nucleic Acids Research **15**(5): 2213-2220.

Litlechild, J., N. Turner, et al. (1995). "CRYSTALLIZATION AND PRELIMINARY-X-RAY CRYSTALLOGRAPHIC DATA WITH ESCHERICHIA-COLI TRANSKETOLASE." Acta Crystallographica Section D-Biological Crystallography **51**: 1074-1076.

Liu, Z., P. Bartlow, et al. (2009). "Use of proteomics for design of a tailored host cell for highly efficient protein purification." J. Chromatogr. A **1216**(12): 2433-2438.

Lu, M. and T. A. Steitz (2000). "Structure of Escherichia coli ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Angstrom resolution." Proc. Natl. Acad. Sci. U.S.A. **97**(5): 2023-2028.

Lyddiatt, A. and D. A. O'Sullivan (1998). "Biochemical recovery and purification of gene therapy vectors." Current Opinion in Biotechnology **9**(2): 177-185.

Mahn, A. and J. A. Asenjo (2005). "Prediction of protein retention in hydrophobic interaction chromatography." Biotechnology Advances **23**(5): 359-368.

Maina, C. V., P. D. Riggs, et al. (1988). "AN ESCHERICHIA-COLI VECTOR TO EXPRESS AND PURIFY FOREIGN PROTEINS BY FUSION TO AND SEPARATION FROM MALTOSE-BINDING PROTEIN." Gene **74**(2): 365-373.

Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nature Biotechnology **21**(3): 255-261.

Mattevi, A., G. Valentini, et al. (1995). "CRYSTAL-STRUCTURE OF ESCHERICHIA-COLI PYRUVATE-KINASE TYPE-I - MOLECULAR-BASIS OF THE ALLOSTERIC TRANSITION." Structure **3**(7): 729-741.

Mihara, H., R. Hidese, et al. (2008). "The iscS gene deficiency affects the expression of pyrimidine metabolism genes." Biochemical and Biophysical Research Communications **372**(3): 407-411.

Miller, C. G. and G. Schwartz (1978). "Peptidase-deficient mutants of Escherichia coli." Journal of Bacteriology **135**(2): 603-611.

Miller, D. L. and Weissbac.H (1970). "Interactions between elongation factors: displacement of GDP from Tu-GDP complex by factor Ts." Biochemical and Biophysical Research Communications **38**(6): 1016-1022.

Miller, W. T., K. A. W. Hill, et al. (1991). "EVIDENCE FOR A CYSTEINE-HISTIDINE BOX METAL-BINDING SITE IN AN ESCHERICHIA-COLI AMINOACYL-TRANSFER RNA-SYNTHETASE." Biochemistry **30**(28): 6970-6976.

Mooney, J. T., D. Fredericks, et al. (2011). "Use of phage display methods to identify heptapeptide sequences for use as affinity purification 'tags' with novel chelating ligands in immobilized metal ion affinity chromatography." Journal of Chromatography A **1218**(1): 92-99.

Nfor, B. K., T. Ahamed, et al. (2008). "Design strategies for integrated protein purification processes: challenges, progress and outlook." Journal of Chemical Technology and Biotechnology **83**(2): 124-132.

Nfor, B. K., T. Ahamed, et al. (2008). "Design strategies for integrated protein purification processes: challenges, progress and outlook." J. Chem. Technol. Biotechnol. **83**: 124-132.

Noble, M. E. M., J. P. Zeelen, et al. (1993). "STRUCTURE OF TRIOSEPHOSPHATE ISOMERASE FROM ESCHERICHIA-COLI DETERMINED AT 2.6-ANGSTROM RESOLUTION." Acta Crystallographica Section D-Biological Crystallography **49**: 403-417.

Noronha, S. B., H. J. C. Yeh, et al. (2000). "Investigation of the TCA cycle and the glyoxylate shunt in Escherichia coli BL21 and JM109 using C-13-NMR/MS." Biotechnology and Bioengineering **68**(3): 316-327.

Onesti, S., A. D. Miller, et al. (1995). "THE CRYSTAL-STRUCTURE OF THE LYSYL-TRANSFER-RNA SYNTHETASE (LYSU) FROM ESCHERICHIA-COLI." Structure **3**(2): 163-176.

Patwardhan, A. V., G. N. Goud, et al. (1997). "Selection of optimum affinity tags from a phage-displayed peptide library - Application to immobilized copper(II) affinity chromatography." Journal of Chromatography A **787**(1-2): 91-100.

Perkins, D. N., D. J. C. Pappin, et al. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-3567.

Porath, J. (1988). "HIGH-PERFORMANCE IMMOBILIZED-METAL-ION AFFINITY-CHROMATOGRAPHY OF PEPTIDES AND PROTEINS." Journal of Chromatography **443**: 3-11.

Porath, J. (1988). "IMAC - IMMOBILIZED METAL-ION AFFINITY BASED CHROMATOGRAPHY." Trac-Trends in Analytical Chemistry **7**(7): 254-259.

Porath, J. (1992). "IMMOBILIZED METAL-ION AFFINITY-CHROMATOGRAPHY." Protein Expression and Purification **3**(4): 263-281.

Przybycien, T. M., N. S. Pujar, et al. (2004). "Alternative bioseparation operations: life beyond packed-bed chromatography." Curr. Opin. Biotech. **15**(5): 469-478.

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

Queiroz, J. A., C. T. Tomaz, et al. (2001). "Hydrophobic interaction chromatography of proteins." Journal of Biotechnology **87**(2): 143-159.

Roque, A. C. A., C. R. Lowe, et al. (2004). "Antibodies and genetically engineered related molecules: production and purification." Biotechnol. Prog. **20**(3): 639-654.

Sassenfeld, H. M. and S. J. Brewer (1984). "A Polypeptide Fusion Designed for the Purification of Recombinant Proteins." Bio/Technology **2**: 76-81.

Sauer, U. and B. J. Eikmanns (2005). "The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria." FEMS Microbiology Reviews **29**(4): 765-794.

Schroeder, U., B. Henrich, et al. (1994). "Peptidase D of Escherichia coli K-12, a metallopeptidase of low substrate specificity." FEMS Microbiol. Lett. **123**(1-2): 153-159.

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

Schwartz, C. J., O. Djaman, et al. (2000). "The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli." P. Natl. Acad. Sci. USA **97**(16): 9009-9014.

Shevchenko, A., H. Tomas, et al. (2006). "In-gel digestion for mass spectrometric characterization of proteins and proteomes." Nat. Protoc. **1**(6): 2856-2860.

Shoji, S., B. D. Janssen, et al. (2010). "Translation factor LepA contributes to tellurite resistance in Escherichia coli but plays no apparent role in the fidelity of protein synthesis." Biochimie **92**(2): 157-163.

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

Shukla, A. A., B. Hubbard, et al. (2007). "Downstream processing of monoclonal antibodies-development of platform approaches." J. Chromatogr. B **848**: 28-39.

Shukla, A. A. and J. Thommes (2010). "Recent advances in large-scale production of monoclonal antibodies and related proteins." TRENDS BIOTECHNOL **28**(5): 253-261.

Smith, D. B. and K. S. Johnson (1988). "SINGLE-STEP PURIFICATION OF POLYPEPTIDES EXPRESSED IN ESCHERICHIA-COLI AS FUSIONS WITH GLUTATHIONE S-TRANSFERASE." Gene **67**(1): 31-40.

Sprenger, G. A., U. Schorken, et al. (1995). "Transaldolase B of Escherichia coli K-12 - cloning of its gene, talB, and characterization of the enzyme from recombinant strains." Journal of Bacteriology **177**(20): 5930-5936.

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

Stoldt, M., J. Wohnert, et al. (1999). "The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition." EMBO Journal **18**(22): 6508-6521.

Szymanski, M., M. Z. Barciszewska, et al. (2002). "5S ribosomal RNA database." Nucleic Acids Research **30**(1): 176-178.

Teter, S. A., W. A. Houry, et al. (1999). "Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains." Cell **97**(6): 755-765.

Thömmes, J. and M. Etzel (2007). "Alternatives to Chromatographic Separations." Biotechnol. Prog. **23**: 42-45.

Tiwari, N., L. Woods, et al. (2010). "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 Expression and Purification **70**(2): 191-195.

Triggsraine, B. L., B. W. Doble, et al. (1988). "NUCLEOTIDE-SEQUENCE OF KATG, ENCODING CATALASE HPI OF ESCHERICHIA-COLI." Journal of Bacteriology **170**(9): 4415-4419.

Tugcu, N., D. J. Roush, et al. (2008). "Maximizing productivity of chromatography steps for purification of monoclonal antibodies." Biotechnol. Bioeng. **99**(3): 599-613.

Utter, M. F. and H. M. Kolenbrander (1972). The Enzymes. New York and London, Academic Press, Inc.

van Reis, R. and A. Zydney (2001). "Membrane separations in biotechnology." Current Opinion in Biotechnology **12**(2): 208-211.

Vorackova, I., S. Suchanova, et al. (2011). "Purification of proteins containing zinc finger domains using immobilized metal ion affinity chromatography." Protein Expression and Purification **79**(1): 88-95.

Waugh, D. S. (2005). "Making the most of affinity tags." Trends in Biotechnology **23**(6): 316-320.

Yang, C., Q. Hua, et al. (2003). "Analysis of Escherichia coli anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout." Biotechnology and Bioengineering **84**(2): 129-144.

Zatloukalova, E. and Z. Kucerova (2004). "Separation of cobalt binding proteins by immobilized metal affinity chromatography." Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **808**(1): 99-103.

Zuurmond, A. M., A. K. Rundlof, et al. (1999). "Either of the chromosomal tuf genes of E-coli K-12 can be deleted without loss of cell viability." Molecular and General Genetics **260**(6): 603-607.

APPENDIX

Evaluation of Escherichia coli proteins that burden non-affinity-based chromatography as a potential strategy for improved purification performance

ABSTRACT

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. Affinity chromatography methods are unrivaled in their selectivity, easily resolving target proteins from crude lysates, but they come with a significant cost. Reported in this study are preliminary efforts to integrate downstream separation with upstream host design by evaluating co-eluting host proteins that most severely burden two different non-affinity based column processes. Phosphoenolpyruvate carboxykinase and peptidase D were significant contaminants during serial purification of GFPuv by hydrophobic interaction and anion exchange chromatography. Ribosomal protein L25 dominated non-target binding of polyarginine-tagged GFPuv on cation exchange resin. Implications for genetic knockout or site-directed mutagenesis resulting in diminished column retention are discussed for these and other identified contaminants.

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 therapeutics (Langer and Ranck 2006; Nfor, Ahamed et al. 2008; Tugcu, Roush et al. 2008). A shift has thus occurred, as the manufacturing steps limiting productivity are now downstream. Beyond initial product recovery steps, chromatographic separation remains the cornerstone in downstream processing of therapeutics in both industry and academia. In an effort to quicken time-to-clinic and market, research efforts have focused on cutting material costs, improving productivity at large-scale, and developing robust, generic separation steps (Shukla, Hubbard et al. 2007; Shukla and Hinckley 2008). Since downstream processes account for 50-80% of total manufacturing costs, these efforts to ameliorate purification of high-value, high-quality products are critical to success in the biopharmaceutical industry (Fahrner, Knudsen et al. 2001; Lightfoot and Moscariello 2004; Roque, Lowe et al. 2004; Aldridge 2006).

Affinity chromatography media is favored for its superb selectivity, particularly media with conjugated proteins for monoclonal antibody capture such as protein A. The remarkable performance comes at a cost, however, as protein A resin is over 30 times more expensive than typical ion exchange resins (Follman and Fahrner 2004). 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 (Jiang, Liu et al. 2009). To mitigate costs, investigations that seek to find alternatives to packed-bed chromatography are 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 (Przybycien, Pujar et al. 2004; Thömmes and Etzel 2007; Shukla and Thommes 2010)). 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 protein A in the purification of Chinese Hamster Ovary-derived monoclonal antibodies (Follman and Fahrner 2004). 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 (Sassenfeld and Brewer 1984). Variants of a positively charge 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 (Graslund, Lundin et al. 2000; Hedhammar and Hober 2007). Separation to high purity of cyclodextrin glycosyltransferase fused to polylysine or polyarginine has also been reported(Kweon, Lee et al. 2002).

Use of non-affinity based matrices is favored for their low cost and robustness towards harsh cleaning procedures, but lesser selectivity means multiple steps are typically needed to remove host contaminants. We seek to improve efficiency in the downstream phase by integrating rational host design upstream. 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. Co-eluting 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.These drive up operational costs and detriment overall yield. Prior studies by our group and collaborators have demonstrated this approach. Identification of contaminating *Escherichia coli* proteins and tailoring of the host cell for improved downstream processing has been applied in applications of immobilized metal affinity chromatography (Cai, Moore et al. 2004; Liu, Bartlow et al. 2009; Tiwari, Woods et al. 2010; Bartlow, Uechi et al. 2011). Similar work has also elucidated proteins that commonly diminish IMAC performance (Bolanos-Garcia and Davies 2006) .

Such methodology is the basis of the work presented in this study on non-affinity-based matrices. We used a similar approach to improve technical 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.

Materials and Methods

Strains, Cell Growth and Medium

E. 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*RI 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 *Hin*dIII 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.

Overnight cultures in 25 g/L Luria–Bertani (LB), supplemented with 100 ug/mL ampicillin, were diluted 1:25 in 500 mL fresh LB in a 2 L flask. Cells were induced with 1 mM IPTG at 2 h when A_{660} was approximately 0.4. Fermentations were carried out in a shaker at 37 °C and 200 rpm for 7 h until the A_{660} reached 2.2 – 2.5. Cell pellets were collected by centrifugation at 5000 *g* for 20 min at 4 °C and stored at -80 °C prior to lysis.

Sample Preparation

Cell pellets were resuspended in the appropriate chromatography starting buffer and supplemented with 0.5% Triton X-100, 0.4 mM $MgCl_2$, 1 mM phenylmethylsulphonyl fluoride (PMSF), as well as 0.5 mg/mL lysozyme unless otherwise noted. 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.

Purification of non-tagged GFPuv

All column purifications were performed with an ÄKTAprimePlus purification system with real-time monitoring of total protein absorbance at 280nm and prepacked 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 column volumes (CV) of start buffer, five CV of eluting high salt 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 eluting 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 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 eluting 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 spinning down 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, Fisherbrand, 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 column order was reversed. Five mL of AEX bound proteins were similarly dialyzed overnight in 0.5 L HIC eluting buffer. Sample adjustment to 1.5 M $(NH_4)_2SO_4$ preceded the optimized HIC step.

Ammonium Sulfate Precipitation

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. With exception to the 1.5 M aliquot, stock solution was again added according to the formula provided by EnCor 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 bis-Tris 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.

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 (pI 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 elutions. 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.

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, 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 aquire the density

profile was approximately one-third the lane width, according to Gassmann *et al*(Gassmann, Grenacher et al. 2009).

Peptide Mass Fingerprinting Analysis

Gel spots of interest were manually picked and digested using the methods of Shevchenko *et al* (Shevchenko, Tomas et al. 2006) with a few modifications. 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 as instructed. Trypsin Gold (Promega, Madison, WI, USA) 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 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).(Perkins, Pappin et al. 1999) 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 +/- 0.3 Da or less.

Results and Discussion

Purification of non-tagged GFPuv

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 (pI 5.7) are reasonably average in the context of the host proteome. Ideally, a group of model proteins with a spectrum of 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 most fundamental conditions for GFPuv purification via HIC and AEX was carried out. We 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 eluting 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 eluting conditions that significantly deviate from those reported here. Contaminants relevant to that specific process could then be identified.

Initial experiments on HiTrap Phenyl FF (low-substitution) columns revealed that loading soluble lysate adjusted to 1.5 M ammonium sulfate allowed adequate GFP retention and had superior resolution than found with high-substitution matrix (data not shown). GFPuv breakthrough at the end of the loading phase, measured by the fluorescence exiting the column normalized to the cleared lysate, was 6% and 30% with 1.5 M and 1.2 M ammonium sulfate, respectively. Further, 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 1, which shows sequential "cuts" of an ammonium sulfate precipitation. OmpF 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, 33% sequence coverage). Initial trials on Q columns determined that an optimal loading buffer was 20 mM Bis-Tris, 50 mM NaCl at pH 6.2. While GFPuv retention was comparable at pH 6.0 during trials with soluble lysate, peak resolution was slightly advantageous at pH 6.2. With the chosen salt concentration and pH, 0.5 units more basic than the target's p*I*, GFPuv was present in the first major peak following gradient salt elution. Further, GFPuv breakthrough was not detected during column loading.
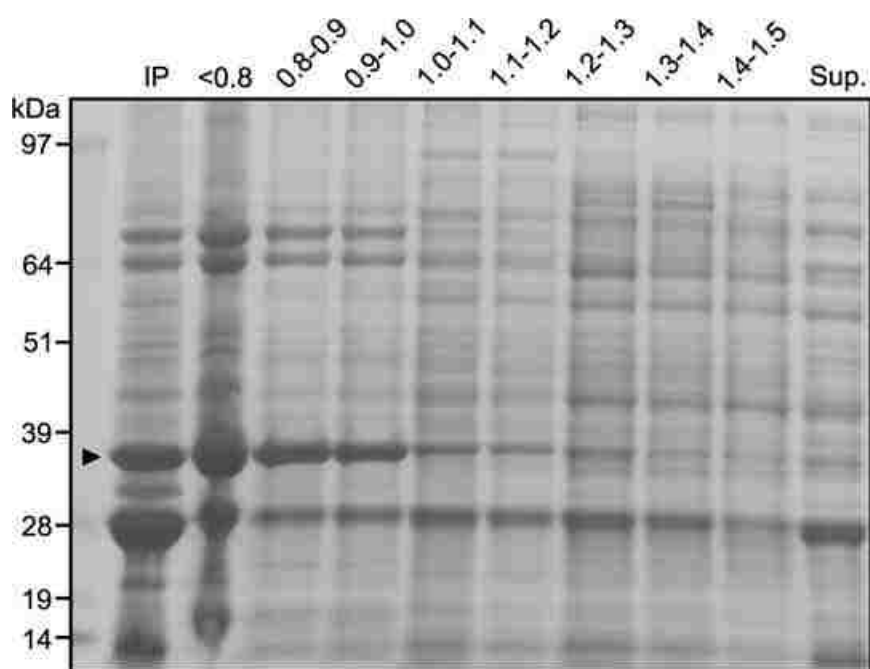
Figure -1: Ammonium sulfate precipitation of proteins expressed in BL21 (DE3) pGFPuv

Purification Steps	Total protein (mg)	GFP (mg)	yield	Purity	Purification
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	13	7.9	90	50	2.6

Table 1: Relevant Purification data for each downstream processing

Gene	Protein Name	Protein Mass (Da)	Protein p*I*	UniProtKB Accession No.
ileS	Isoleucyl-tRNA synthetase	104231	5.66	**P00956**
acnB	Aconitate hydratase 2	93439	5.24	**P36683**
pckA	Phosphoenolpyruvate carboxykinase	59606	5.46	**P22259**
pepD	Aminoacyl-histidine dipeptidase	52882	5.20	**P15288**
gltA	Citrate synthase	47984	6.21	**P0ABH7**
sucC	Succinyl-CoA ligase [ADP-forming] subunit beta	41367	5.37	**P0A836**
gcvT	Aminomethyltransferase	40121	5.36	**P27248**
talB	Transaldolase B	35197	5.11	**P0A870**
cysK	Cysteine synthase A	34468	5.83	**P0ABK5**
tsf	Elongation factor Ts	30404	5.22	**P0A6P1**
dapD	Tetrahydrodipicolinate N-succinyltransferase	29873	5.56	**P0A9D8**
gmhA	Phosphoheptose isomerase	20802	5.97	**P63224**
ahpC	Alkyl hydroperoxide reductase subunit C	20748	5.03	**P0AE08**

Table 2: Proteins co-eluting with GFPuv during purification via ammonium sulfate precipitation, HIC, and AEX
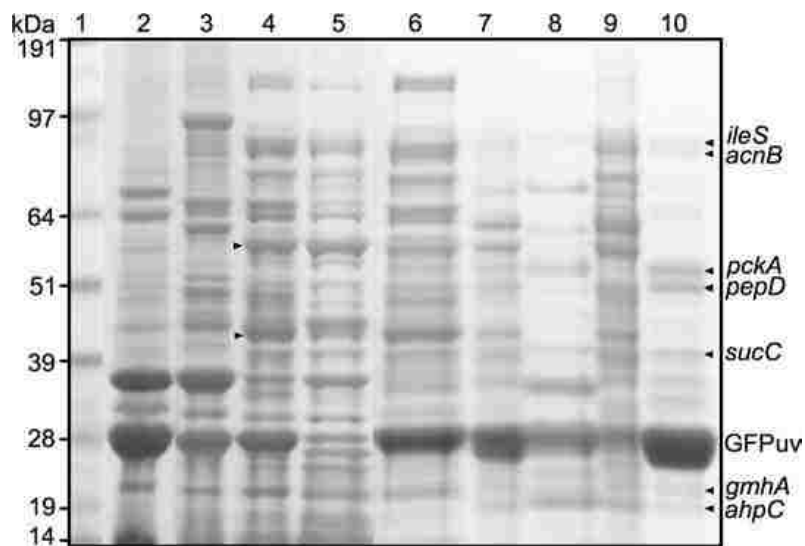
Figure 2: SDS-PAGE of GFPuv purified by HIC-AEX

Lanes: 1, molecular weight ladder; 2, insoluble pellet; 3, ammonium sulfate precipitated proteins; 4, soluble extract; 5, HIC flow through; 6, HIC bound proteins; 7, after dialysis; 8, AEX flow through; 9, small AEX elution peak; 10, GFPuv peak. Lanes 5, 8, and 9 were prepared from TCA/DOC precipitated proteins

Proteins recovered all steps of the purification are illustrated in Figure 2 and the accompanying process data is given in Table 1. 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. 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 5 ml were used for the final column step; however when accounting for actual GFP loss during dialysis, overall yield was 55%. While our aim was not centered on yield optimization, other dialysis methods could have been mitigated loss between column steps. Minor flow through in the HIC step and the GFPuv fraction pooling strategy attribute to much of the detriment to yield during chromatography, but overall it is comparable to multi-step non-affinity schemes reported elsewhere (Follman and Fahrner 2004).

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 are presented in Table 2.

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*) the beta subunit of ADP-forming succinyl-coa ligase (*sucC*) and citrate synthase (*gltA*). Mutants lacking *sucC* and *gltA* are undesirable for biotechnology applications since they have been shown to produce elevated acetate levels (Lee, Goel et al. 1994; Li, Ho et al. 2006). Enzymes involved in amino acid biosynthesis are tetrahydrodipicolinate N-succinyltransferase (*dapD*) and cysteine synthase A

(*cysK*). Others include the essential class I isoleucyl-tRNA synthetase (*ileS*);(Antonio, McFerran et al. 2002) the T subunit of the glycine cleavage complex aminomethyltransferase (*gcvT*); transaldolase B (*talB*), which has been previously purified by ammonium sulfate precipitation and AEX;(Sprenger, Schorken et al. 1995) alkyl hydroperoxide reductase subunit C (*ahpC*); phosphoheptose isomerase (*gmhA*); and elongation factor Ts (*tsf*), which associates with the EF-Tu-GDP complex at the ribosome and induces exchange of GDP to GTP.(Miller and Weissbac.H 1970) Nonessential proteins in this group total less than 8% of the final pool by densitometry; and thus knockout of such individual genes, if not detrimental to the cell, would not be expected to significantly enhance purification efficiency.

Other gene products opportune 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, these proteins hinder target binding capacity without necessarily reducing final purity. Indeed, protein 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 1 and lane 3 of Fig. 2. 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) (Duval, Nicoloff et al. 2009). 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 2 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 of *groL* are not viable and those deficient in *tufA* have a slow-growth phenotype in complex media (Fayet, Ziegelhoffer et al. 1989; Zuurmond, Rundlof et al. 1999).

A combined proteomic and host cell engineering approach can offer significant gains in downstream processing efficiency by genetically removing the most burdensome co-eluting proteins. Gene *pckA* encodes phosphoenolpyruvate carboxykinase (PEPCK) 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 (Miller and Schwartz 1978; Schroeder, Henrich et al. 1994). 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 (Henrich, Monnerjahn et al. 1990). *E. coli* posses many peptidases, however, and *pepD*-deficient mutants can still utilize a diverse array of peptides (Miller and Schwartz 1978). 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 (Utter and Kolenbrander 1972). 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 (Sauer and Eikmanns 2005). 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 (Yang, Hua et al. 2003). It was also noted that this flux pattern was similarly found in the low acetate producer BL21,(Noronha, Yeh et al. 2000) 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 (Yang, Hua et al. 2003). These reports lend credence to a plausible use for $\Delta pckA$ strains in biotechnology applications.

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

Purification of arg$_6$-GFPuv

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-eluting proteins could thus be applicable to a variety of eluting strategies. Soluble lysates were 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. 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 showed no target retention (not shown). Proteins were eluted with a linear NaCl gradient to 1 M. No proteins were recovered in the subsequent step elution to 2 M.

Proteins were sampled at each step in the process and fractionated by SDS-PAGE and presented in Figure 3, with corresponding purification data shown in Table 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 expectedly superior to either individual step in the prior strategy due to the endowed localized charge. The purity achieved may see underwhelming in comparison to other studies where multiple steps were used,(Sassenfeld and Brewer 1984) but the process yield compares very favorably to previous successful uses of this method (Kweon, Lee et al. 2002). 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 expression (lane 2), even more so than in the un-tagged strain cultured identically in spite of evidence that polyarginine tags improve solubility (Kato, Maki et al. 2007). This effect could possibly have been mitigated with the addition of NaCl during resuspension (Fuchs and Raines 2005). 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).
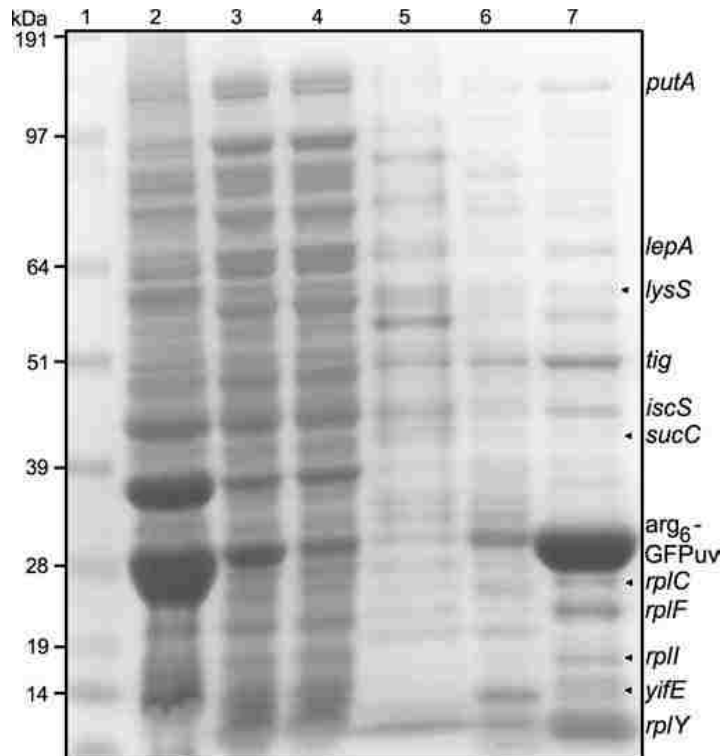
Figure 3: SDS-PAGE of arg$_6$-GFPuv purification by CEX.

Lanes: 1, molecular weight ladder; 2, insoluble pellet; 3, soluble extract; 4, flow through;

5, wash fraction; 6, first eluting peak; 7, arg$_6$-GFPuv peak

Gene	Protein Name	Protein Mass (Da)	Protein p*I*	UniProtKB Accession No.
putA	Bifunctional protein putA	143725	5.69	**P09546**
lepA	Elongation factor 4	66528	5.40	**P60785**
lysS	Lysyl-tRNA synthetase	57567	5.11	**P0A8N3**
tig	Trigger factor	48163	4.83	**P0A850**
iscS	Cysteine desulfurase	45061	5.94	**P0A6B7**
sucC	Succinyl-CoA ligase [ADP-forming] subunit beta	41367	5.37	**P0A836**
rplC	50S ribosomal protein L3	22230	9.91	**P60438**
rplF	50S ribosomal protein L6	18892	9.71	**P0AG55**
rplI	50S ribosomal protein L9	15759	6.17	**P0A7R1**
yifE	UPF0438 protein yifE	13125	6.10	**P0ADN2**
rplY	50S ribosomal protein L25	10687	9.60	**P68919**

Table 3: Proteins co-eluting with arg_6-GFPuv during purification by CEX

Eleven proteins confirmed to co-elute with arg$_6$-GFPuv are presented in Table 3. Some lesser contaminating proteins are lysyl-tRNA synthetase (*lysS*), bifunctional protein PutA (*putA*), and UPF0438 protein (*yifE*). Cysteine desulfurase (*iscS*) acts in iron-sulfur clusters after assembly with scaffold homolog IscU and possibly has a role in regulation of genes associated with pyrimidine metabolism(Mihara, Hidese et al. 2008). Mutants deficient in *iscS* have been characterized to have a slow growth phenotype (Schwartz, Djaman et al. 2000).

Trigger factor (*tig*) 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(Baneyx and Mujacic 2004). While Δ*tig* strains have no obvious growth phenotype or increased protein aggregation, it must be compensated by the DnaK/J and GroEL/S systems;(Deuerling, Schulze-Specking et al. 1999; Teter, Houry et al. 1999) therefore *tig* mutants are ill-suited for biotech applications that utilize over-expression of recombinant proteins. For further detail, readers are encouraged to see these reviews (Baneyx and Mujacic 2004; Hoffmann, Bukau et al. 2010).

Elongation factor 4 (*lepA*) is a translational GTPase with an apparent ability to catalyze reverse translocation *in vitro*(Qin, Polacek et al. 2006) and was thought to play a role in translation fidelity. Recent evidence, however, not only confirms prior reports that mutants lacking *lepA* have no deleterious effect on growth rate(Dibb and Wolfe 1986), but also suggests that its gene product does not contribute to translation fidelity (Shoji, Janssen et al. 2010).

Four out of the thirteen identified contaminants are 50S ribosomal proteins L3 (*rplC*), L6 (*rplF*), L9 (*rplI*), and the overwhelmingly present 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,(Schulze and Nierhaus 1982; Herbst, Nichols et al. 1994; Brodersen and Nissen 2005) though L25 deserves further consideration. It was the first discovered representative the CTC (catabolite controlled) family in 1972(Horne and Erdmann 1972). 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,(Douthwaite, Garrett et al. 1979; Gongadze, Meshcheryakov et al. 1999; Harms, Schluenzen et al. 2001) a domain strictly conserved in bacteria(Szymanski, Barciszewska et al. 2002) and distinctive for consecutive non-Watson-Crick base pairs. Indeed, evidence has shown that all CTC proteins possess a domain, homologous to L25, responsible for 5S binding at this site, their only known target for interaction (Fedorov, Meshcheryakov et al. 2001; Gongadze, Korepanov et al. 2005). Given the wealth of biochemical, interaction, and structural studies, questions remain concerning their cellular function unique 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 (Korepanov, Gongadze et al. 2007). In *E. coli*, cells survive knockout of L25 but exhibit inhibited growth and protein translation which could be recovered by L25 expression from plasmid;(Korepanov, Gongadze et al. 2007) nevertheless such a strategy is unsuitable for biotechnology applications.

Sequence alignment has revealed surprisingly low sequence homology among 300 CTC proteins given the nature in which they associate with the ribosome. Only five residues are strictly conserved (Stoldt, Wohnert et al. 1998) In L25, these are R9, R21, Y31, H88 and D90. It has been shown that these amino acids interact primarily with the phosphoribose backbone via hydrogen bonding, inaccessible to the solvent (Stoldt, Wohnert et al. 1999; Lu and Steitz 2000; Fedorov, Meshcheryakov et al. 2001). Subsequent 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 (Gongadze, Korepanov et al. 2005; Korobeinikova, Gongadze et al. 2008). This is attributed to their solvent accessibility. 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, and such mutation results in complex dissolution (Gongadze, Korepanov et al. 2005).

In contrast with affinity techniques, CEX interaction is based on electrostatic charge. Modifications of proteins that prompt retention loss are thus not likely limited to one or two select residues, especially with large proteins. *E. coli* L25 however, is a uniquely good candidate for residue modification due to its small size and few occasions with consecutive positively charged residues. In yet unpublished results, our group and colleagues have used this approach to modify catabolite repressor protein (CAP) and triosephosphate isomerase (TPI) to eliminate retention on immobilized metal columns. A similar approach by 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 (Humphreys, Heywood et al. 2004). Based on these and prior findings discussed, 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. And use of a host strain deficient in *lepA* and with altered L25 could have significant, intrinsic advantages for CEX separation of polyarginine tagged proteins.

Conclusions

Reported in this study are the preliminary efforts to integrate rational upstream host design and downstream efficiency using proteomic tools to evaluate co-eluting 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.