5-2013

Bio-Separation Process Improvement via Genomic Manipulation: Development of Novel Strains for Use in Immobilized Metal Affinity Chromatography (IMAC)

Ryan Curtis Haley
University of Arkansas, Fayetteville

BIOSEPARATION PROCESS IMPROVEMENT VIA GENOMIC MANIPULATION: DEVELOPMENT OF NOVEL STRAINS FOR USE IN IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY (IMAC)

BIO-SEPARATION PROCESS IMPROVEMENT VIA GENOMIC MANIPULATION: DEVELOPMENT OF NOVEL STRAINS FOR USE IN IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY (IMAC)

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

By

Ryan C. Haley
Bachelor of Science in Cellular Biology, 2004
Northeastern State University

May 2013
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dissertation Director:

_____

Dr. Robert R. Beitle

Thesis Committee:

_____

Dr. Jeannine Durdik

_____

Dr. Jin-Woo Kim

_____

Dr. Robyn Goforth

DISSERTATION DUPLICATION RELEASE

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

Agreed _____

Ryan C. Haley

Refused _____

Ryan C. Haley

ACKNOWLEDGEMENTS

Many people deserve my gratitude for helping to make this dissertation possible.
I would first and foremost like to thank the invaluable direction and guidance from my Ph.D. advisor, Dr. Robert R. Beitle, who has been not only a mentor but someone I have come to consider a close friend. Dr. Beitle has always seen the bright side of mistakes made along the way, and has been there with a calm sense of motivation to stay the course during times of struggle. It is only through his continued optimism, patience, and encouragement that I have learned to become an independent scientist.

I would like to thank the contributions of the other members of my dissertation committee: Dr. Robyn Goforth, Dr. Jeannine Durdik, and Dr. Jin Woo Kim. Dr. Robyn Goforth deserves special thanks for making her laboratory available to me and offering guidance during the course of my research. I would also like to thank Alicia Kight for her assistance and instruction during the molecular techniques. Special thanks goes to Dr. Rajaramesh Varakala, Dr. Neha Tawiri, Ellen Brune, and McKenzie Fruktel -- my lab partners, research colleagues, and friends during this project.

Most importantly, my sincerest gratitude goes to my parents and sister for their unconditional support and encouragement through the years.

Financial support from the Fulbright College of Arts and Sciences, Department of Chemical Engineering, Arkansas Biosciences Institute and National Science Foundation are gratefully acknowledged.

TABLE OF CONTENTS

Acknowledgements
Table of Contents
List of Tables
List of Figures
Scope of Work 1
1. Overview 3
2. Literature Review 6
2.1. –omics: Large Scale Applied Biological Sciences 6
2.2. Genomics 7
2.3. Transcriptomics 8
2.3.1. DNA Microarray Technology 10
2.3.2. RNA-Seq 11
2.4. Proteomics 12
2.4.1. Liquid Chromatography 14
2.4.2. IMAC 15

Part I: A description of the *E. coli* total contaminant pool encountered during immobilized metal affinity chromatography

3. Abstract 18
4. Introduction 19
5. Materials and Methods 21
5.1. Cell culture and Sample Preparation 21
5.2. IMAC and Sample Preparation 22
5.3. Protein Identification 23
5.4. Data Analysis 24
6. Results 25
6.1. IMAC Contaminant Proteins 25
7. Discussion 38
8. Conclusion 41

Part II: A redesigned Escherichia coli triosephosphate isomerase restores growth properties in a bacterial strain useful for Immobilized Metal Affinity Chromatography (IMAC)

9. Abstract 43
10. Introduction 44
11. Materials and Methods 46

11.1. Bacterial Strains and Growth 46
11.2. PCR Methodology 47
11.3. Plasmid Construction 49
11.4. GST:eTIM Expression in E.coli VR101:pGEX -6P-2_tpiA.IMAC.1 51
11.5. Purification of eTIM 52
11.6. Assay of Triosephosphate Isomerase 53
12. Results 54
13. Discussion 66

Part III: Future extension of proteomic approach toward development of a minimal genome recombinant expression host.

14. Discussion 68
15. Considerations for Future Work 71

References 76

Appendix I: Compilation of contaminant species 87
Appendix II: Essential vs. non-essential 104

LIST OF TABLES

Part I: A description of the *E. coli* total contaminant pool encountered during immobilized metal affinity chromatography

Table 1: Genes Associated with the Total Contaminant Pool of $Co^{(ii)}$ – IMAC
Table 2: Genes Associated with the Total Contaminant Pool of $Ni^{(ii)}$ – IMAC
Table 3: Genes Associated with the Total Contaminant Pool of $Zn^{(ii)}$ – IMAC

Part II: A redesigned Escherichia coli triosephosphate isomerase restores growth properties in a bacterial strain useful for Immobilized Metal Affinity Chromatography (IMAC)

Table 1: Comparison of specific activity of triosephosphate isomerase.

LIST OF FIGURES

Part I: A description of the *E. coli* total contaminant pool encountered during immobilized metal affinity chromatography

Figure 1: Distribution of a total contaminant pool (TCP) with respect to essentiality of gene product.
Figure 2: Distribution of commonality between contaminant pools with respect to Co(II), Ni(II), and Zn(II) chelates used in IMAC separations.
Figure 3: Distribution of contaminating species found in IMAC metalloproteome relative to isoelectric point.
Figure 4a: Distribution of contaminant proteins found in Co(II) IMAC purifications when an IDA chelate and pH step elution is employed.
Figure 4b: Distribution of contaminant proteins found in Ni(II) IMAC purifications when an IDA chelate and pH step elution is employed.
Figure 4c: Distribution of contaminant proteins found in Zn(II) IMAC purifications when an IDA chelate and pH step elution is employed.
Figure 5: Distribution of nonessential contaminating proteins of the total contaminant pool based on gene product functional category.

Part II: A redesigned Escherichia coli triosephosphate isomerase restores growth properties in a bacterial strain useful for Immobilized Metal Affinity Chromatography (IMAC)

Figure 1: Cloning strategy implemented to obtain p*tpiA.IMAC.1*.
Figure 2: TIM sequence alignments (63 isoforms) showing region on mutagenesis for the removal surface exposed histidine.
Figure 3: Growth Characterization of E.coli RH101: pGEX -6P-2_tpiA.IMAC.1 when compared to E.coli VR101 and E.coli BL21 (DE3).
Figure 4: 12.5% SDS-Page gel showing cleaved eTIM does not bind Co(II) IMAC column.

Part III: Future extension of proteomic approach toward development of a minimal genome recombinant expression host.

Figure 1: Circos plot of the E. coli chromosome and IMAC metalloproteome according TCP and ECP data provided in Part I

1. SCOPE OF WORK

The dissertation is comprised of three parts. **Part I** describes proteomic analysis of native bacterial proteins from *Escherichia coli (E.coli)* that bind during Immobilized Metal Affinity Chromatography (IMAC). **Part II** describes the value in exploiting proteome based data as a tool toward the design an *E. coli* expression strain that is particularly useful when Immobilized Metal Affinity Chromatography is employed as the initial capture step of a homologous protein purification process. **Part III** describes a methodology of chromosomal mapping of all contaminant gene products.

The objective of **Part I** was to identify all *E. coli* proteins that bind to $Co^{(II)}$, $Ni^{(II)}$, and $Zn^{(II)}$ IMAC columns, describing the isoelectric point, molecular weight, and metabolic essentiality of the characterized proteins were considered. Information regarding this group of proteins is presented and used to define the IMAC bioseparation-specific metalloproteome of *E. coli*. Such data concerning the potential contaminant pool is useful for the design of separation schemes, as well as designing optimized affinity tails and strains for IMAC purification. **Part II** examined proteins known to co-elute during $Co^{(II)}$, $Ni^{(II)}$, and $Zn^{(II)}$ IMAC purifications. Methods to circumvent the effects of punitive protein removal were proposed and carried out. Specifically, triosephosphate isomerase (TIM; *tpiA* gene product), a protein known to bind during IMAC, was redesigned through site directed mutagenesis to eliminate surface exposed histidine. By extension of this rational, **Part III** provides a theoretical model of using *in silico* mapping (Circos diagrams) to create a practical system of applying data described in **Part I.** Such a tool, has potential to allow future investigators the possibility of mapping large scale genomic

deletions; significantly streamlining cell line development when compared to the individual targeting methodologies seen in **Part II**.

2. OVERVIEW

Use of recombinant proteins as therapeutic agents continues to increase in medicine. An April 2010 global market report indicates that approximately 50% of all newly approved drugs in some way incorporate a bio-pharma active ingredient. Forecasts in this report also indicate the global market for biopharmaceuticals is projected to reach US$182.5 billion by 2015, a fifteen percent increase from current standings (UK Global Research, 2010). It is foreseeable, that this kind of growth will require significant streamlining of current process methodologies to facilitate sustainable global demand.

In recent years, developments in upstream technologies have dramatically improved bioprocessing scale, yield, and performance. Historically, due to a lack of available information of the internal workings of biological systems, purification research primarily focused on downstream process improvements; where developments in purification chemistries and fluid dynamics predominated (Porath, Carlsson, Olsson, & Belfrage, 1975). However, with more recent advancements in the understanding of biological systems as a whole, contemporary studies have looked more seriously into the possibility of making improvements to the recombinant host itself; *vis. genetic engineering of specialized recombinant expression hosts* (Tan, Kern, & Selleck, 2005).

For decades, the *Escherichia coli* recombinant expression system has been used for characterization of protein structure and function and it has become a fundamental tool in biochemical research practices. Simultaneously, industrial practices have implemented the *E.coli* expression platform as one of the most efficient and cost effective tools for the production of

both industrial and simple unconjugated recombinant proteins of medicinal value. During this time, the addition of an affinity tag for simplifying the targeting proteins with unknown purification characteristics has become common. Indeed, in industrial settings, it is common practice to include a well characterized immobilized metal affinity tag to all expression constructs at the onset of any new project regardless of the target protein properties (Hengen, 1995).

The research presented hypothesized that if one has prior knowledge of which contaminant proteins must be circumvented to obtain enhanced purity during a recombinant purification process; he/she could use such information to customize a recombinant strain that perimts a more simplified protein purification process. Using this novel host cell could (i.) decrease the size of the chromatography columns required, (ii.) shorten the time required to obtain the final recombinant product, and (iii.) reduce overall costs of the purification process itself.

This dissertation describes *E. coli* proteins that bind to IMAC columns under common industrial and academic conditions, the effect of modifications to such contaminant proteins at the molecular level with relation to overall cell health, and the potential for using large scale knock-outs (i.e. deleting large portions of a bacterial chromosome) to derive a minimal genome recombinant expression host. As part of a parallel study, in Part I we characterize the *E.coli* metallo-proteome; comprising the IMAC Total Contaminant Pool (TCP) and Elution Contaminant Pool (ECP). Relative concentrations of individual contaminant species, which have the greatest effect on decreasing column capacity at the initial capture step, have been defined (Varakala, 2008). The data describing contaminant proteins can be utilized to select the most

efficient combination of chromatography conditions to come to a homologous sample. Selections of secondary and polishing purification steps could then be determined based on physical properties of the remaining contaminant pool.

As a complimentary continuation of this work, Part II of this dissertation describes a methodology of simplifying large scale purification processes by using molecular techniques to knock-out or genetically modify endogenous contaminant proteins known to co-elute during IMAC purifications from *E.coli* feed-streams. Growth characterization of a representative mutant strain is analyzed relative to expression of a recombinant target.

Part III describes the future application of data produced in Parts I and II, respectively. Future research could use this data to create of a functional map of the E. coli chromosome to detail specific locations of essential and non-essential IMAC contaminant genes. Therefore, continuation of this work gives rise to the application and utility of being able to perform large scale chromosomal modifications to produce a minimal genome recombinant host. Finally, a personal recommendation to continue this work by multiplexing the production of complex biomolecules in mammalian cell lines is briefly discussed.

3. LITERATURE REVIEW

3.1 -omics: Large Scale Applied Biological Sciences

For most in the biological sciences community, the time spanning the mid-1990's to present could be coined as the "–omics era". With the advent and implementation of laser adsorption measurement, robotic automation, and microprocessor integrated platforms into daily use analytical equipment; combined with tethering such equipment to the World Wide Web (Berners-Lee, 1999), biologist and biological engineers are now able to obtain large scale data sets in near real-time. The technology of the –omics era has brought about dramatic changes in the way biological experiments are contrived and carried out (Kiechle, Zhang, & Holland-Staley, 2004). The sciences of genomics, transcriptomics, and proteomics all converge to allow researchers the ability to answer biological questions that, in the past, could merely be speculated as to the functions, interactions, regulatory mechanisms, or pathology of macromolecules within the context of cellular function.

3.2 Genomics

The study of genomics originated as a discipline within the scope of genetics where the major focus was driven by the complete sequencing of an organism's genome. Today, the field has diversified into multiple efforts that aim to determine not only the DNA sequence of organisms, but also aims to include high resolution genome mapping (O'Brien et al., 1993). In 1995, *H. influenza* (1.8 Mb) was the first free-living organism to have its genomic DNA (*gDNA*) sequenced and annotated (Fleischmann et al., 1995). Since this time, with the advent of high through-put shotgun sequencing, as well as tools such as the Illumina™ sequencing platform, the field of genomics has considerably changed the way biological science is conducted. At the time of this manuscript, the NCBI BioProject database cites archival of 2,192 Eukaryotes, 14,448 Prokaryotes, and 297 Archaea genome sequences in their database.

The first near complete sequence of an individual human genome was published in June 2007 (Levy et al., 2007; Venter et al., 2007). Modern genomics has since become a mainstay in medicine in the 21st century, as the promise of "individualized" medical treatment moves closer to becoming reality (Evans & Relling, 2004). Genomics could be considered the primary level of study in biological systems. More recent applications in field have shifted toward comparative genomics. Here deep sequencing techniques are used to discriminate short single nucleotide polymorphisms (SNPs) within the genome. Subsequently, comparisons of such an individual sequence are made to that of the population consensus to determine if abnormal or allele specific traits are present (Ajay, Parker, Abaan, Fajardo, & Margulies, 2012).

3.3 Transcriptomics

The transcriptome is considered to be the entire set of RNA molecules expressed in an organism, tissue, or cell type. Transcriptomics is considered to be a secondary level of study in biological systems. It has the overarching goal of providing researchers statistically measurable information concerning active expression of genes within a cell, tissue, or organism. Because of the dynamic nature of the transcriptome, gene expression studies have a requirement of defining the cellular or organismic physiological state at the specific moment of sample collection.

Recent publications in the field of transcriptomics predominantly include: cellular or tissue response to treatment/ medication (de Groot et al., 2007; Patterson et al., 2008), differential regulatory function attributed to specific genetic mutation or cancer phenotype (Andersson, Sulkowski, & Porath, 1987; Augenicht, Taylor, Anderson, & Lipkin, 1991; Augenlicht & Kobrin, 1982; Augenlicht, Wahrman, Halsey, & al., 1987; Chakravarchy & Pietenpol, 2003; Jain et al., 2012; Morin et al., 2008; Wu et al., 2010), stem cell or developmental expression profiles (Cloonan et al., 2008; Fujiwara et al., 2011; Hermann et al., 2009; Jain, et al., 2012; Li, Yang, Nakashima, & Rana, 2011), and alternative regulatory or metabolic pathway(s) organisms/ cells may utilize when metabolically challenged (Çakir, Kirdar, & Ülgen, 2004; Hirai et al., 2005; Hirai et al., 2004).

Early transcript expression profiling was performed by Northern blot analysis and the production of complimentary DNA (cDNA) libraries (Alwine, Kemp, & Stark, 1977; Belyavsky, Vinogradova, & Rajewsky, 1989; Schlamp et al., 2008). Although still considered to be a

principal molecular technique, more recent technological advances now allow for high throughput genome wide expression profiling through DNA microarray technology (Maskos & Sothern, 1992; Winn et al., 2011), serial analysis gene expression (SAGE) (Nystrom, Fierlbeck, Granqvist, Kulak, & Ballermann, 2009; Velculescu, Zhang, Vogelstein, & Kenzler, 1995; Wu, et al., 2010), and quantitative PCR (qPCR) (Jeanty, Longrois, Mertes, Wagner, & Devaux, 2010; Pecson, Martin, & Kohn, 2009), quantitative reverse transcription PCR (qRT-PCR) (Curtis et al., 2010; Jain, et al., 2012).

3.3.1 DNA Microarray Technology

Since first being published in 1982, DNA microarrays have revolutionized the bioinformatics field (Augenlicht & Kobrin, 1982). By allowing investigators to compare expression profiles of hundreds to even thousands of genes simultaneously, this technology has significantly contributed to the advancement of biomedical research. For example, following initial publication exploring carcinoma transplants in immuno-compromised mice, Augenicht and colleagues' published two follow-up studies using DNA Chip technology in human colonic cancer studies (Augenicht, et al., 1991; Augenlicht, et al., 1987). Since their publications', the ability to simultaneously contrast transcriptome-wide analyses of experimental groups has led to a far greater understanding of genotypic cancer expression patterns. Knowledge gained from this work has given the medical community a powerful tool for accurately predicting how aggressively to a treat cancer patients or patients facing other genetically linked diseases.

Germane to this dissertation is the work by Haddadin and Harcum (Haddadin & Harcum, 2005). Using fed-batch fermentation combined with whole genome microarrays, they were able to characterize gene expression profiles of wild-type *Escherichia coli* MG1655 to that of *E.coli* MG1655 transformed with an IPTG inducible plasmid, under varying states. Their work confirms significant departures in gene expression between the two strains under various growth states, as well as under recombinant induction.

3.3.2 RNA-Seq

Whole transcriptome shotgun sequencing or "RNA-Seq" has been touted as the next "revolutionary tool for transcriptomics" (Wang, Gerstein, & Snyder, 2009). Built on deep sequencing technology, RNA-Seq offers the researcher a quantitative, high-throughput sequence-based tool for surveying entire transcriptomes at single-base resolution (Holt & Jones, 2008; Wilhelm et al., 2008). RNA-Seq has been applied to Saccharomyces cerevisiae (Nagalakshmi et al., 2008), Saccharomyces pombe (Wilhelm, et al., 2008), Arabidopsis thaliana, Mus musculus (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008), as well as Human cancer and stem cell lines (Cloonan, et al., 2008; Morin, et al., 2008).

It should be noted that it is common in the field of transcriptomics for experimental designs to use *ex vivo* models by way of culturing the tissue or cell line of interest (Fujiwara et al. 2011, Hermann et al. 2009, Li et al. 2011). This methodology affords the researcher the ability to influence and measure expression patterns across multiple cell lines, *viz* parallel experimental groups, simultaneously. However, due to the relative instability and dynamic expression patterns inherent to RNA molecules, this 'experimental tool' could conceivably bias the outcome. To account for bias, a recurring pattern in transcriptomics studies is the inclusion of a parallel control group obtained through direct excision of the organ/ tissue of interest, followed by isolating the RNA transcripts via phenol: chloroform reagent (Chomczynski and Sacchi 2006, Hermann et al. 2009). Inclusion of this type of control specimen allows researchers the ability to resolve variations in expression across multiple treatments by normalizing transcript patterns; consequently giving a higher confidence interval to the overall study.

3.4 Proteomics

Proteomics has been defined as the study of the entire compliment of proteins, including modifications made to a particular set of proteins, produced by a cell, organism or system (Marc R. Wilkins et al., 1996; M.R. Wilkins, Sanchex, Gooley, & *al.*, 1996). Following genomics and transcriptomics, proteomics is considered the tertiary level of study of biological systems. Within the field there are notable applications, techniques and tools that directly contribute to the overarching discipline.

Most studies in this field use tools, such as two dimensional gel electrophoresis (2DGE) combined with mass spectroscopy (MS) to determine functional expression levels within an organism under defined conditions. For example, in an article comparing proteomic separation techniques, Chiou and Wu offer a comparison of common electrophoretic methods (Chiou & Wu, 1999). Gygi and colleagues give insights in to potential pitfalls associated with using 2DGE as a separation technique (Gygi, Corthals, Zhang, Rochon, & Aebersold, 2000). Fey and Larsen offer expert opinions on the application of 2DGE, suggesting that while the tool has definite drawbacks, it is the best methodology for high through-put proteome screening (Fey & Larsen, 2001).

By far, the most relevant biological application of 2DGE - MS has been in comparative expression proteomics. Areas of application are very broad, including toxicological response (Sa-Correia & Teixeira, 2010), metabolomics (Christopher Kirkpatrick, 2001), purification process development (Kumar, Tabor, & Richardson, 2004), hepatocellular binding of heavy

metals (S. D. Smith, She, Roberts, & Sarkar, 2004), metalloproteomics (S. D. Smith, et al., 2004), cancer biomarker identification for prognostic treatment strategy (Chakravarchy & Pietenpol, 2003), systems biology (Aebersold & Mann, 2003), protein-protein interactions (Kumar et al., 2004), and the integration of cellular signaling networks(Choudhary & Mann, 2010), and Complexomics (Lasserre et al., 2006) to name a few.

3.4.1 Liquid Chromatography

Liquid chromatography is traditionally defined as a process of separating a complex aqueous mixture into its composing parts. Typically, a feed solution is introduced to a packed column containing a stationary phase where fractionation is accomplished through specific chemistry inherent to each part making up the whole of the feed stream. This action is accomplished by continuously feeding solvent to the column following introduction of the solution to be fractionated, while simultaneously collecting individual peaks as they elute off the column.

3.4.2 Immobilized Metal Affinity Chromatography

Immobilized-Metal Affinity Chromatography (IMAC) is a fractionation technique that utilizes covalently bound chelating compounds on solid chromatographic supports to capture metal ions, which serve as affinity ligands for various peptides / proteins (Gaberc-Porekar & Menart, 2001b). Porath and colleagues first published on IMAC as a separation method in 1975 (Porath, et al., 1975). Advantages of IMAC, compared to other affinity techniques, include ease of column regeneration and stability of the stationary phase.

Up to now, divalent $Cu^{(II)}$, $Ni^{(II)}$, $Zn^{(II)}$, and $Co^{(II)}$ have successfully been used as affinity ligands in IMAC (Ueda, Gout, & Morganti, 2003). Under correct stereo special configuration, these divalent transition metals demonstrate favorable binding properties to macromolecules displaying N, O, and S (Gaberc-Porekar & Menart, 2001b). Amino acids such as histidine, cysteine, glutamic acid, and aspartic acid have electron donating side chains, however histidine shows the greatest affinity towards chelated metals under conditions of high salt content (Yip, Nakagawa, & Porath, 1989). Several chelates are commercially available for IMAC, including iminodiacetic acid (IDA), nitrilotriacetic acid (NTA), and carboxymethylated aspartic acid; however IDA or NTA chemistry predominate the field (Chaga, 2001).

Since the introduction of metal affinity tags by Smith and colleagues, and subsequent development by Beitle and Ataai, IMAC has developed into one of the most robust methods for the purification of recombinant proteins (Arnau, Lauritzen, Petersen, & Pedersen, 2006; Beitle & Ataai, 1993; Lichty, Malecki, Agnew, Michelson-Horowitz, & Tan, 2005; M. C. Smith, Furman,

Ingolia, & Pidgeon, 1988). IMAC is widely used as an initial capture step in recombinant protein purification, however the presence of nascent host proteins that display similar affinity toward transitional metals currently preclude IMAC as a single step purification method for recombinant proteins. Notably, loss of column capacity and a complication in gradient elution often occur when genomic proteins are retained during adsorption.

Previous studies have also shown IMAC useful in separating nucleotides, probing a protein surface amino acid residues, and studying protein – protein interactions (Andersson, et al., 1987; Hubert & Porath, 1980; Rossetto, Schiavo, Laureto, Fabbiani, & Montecucco, 1992).

Part I: A description of the *E. coli* total contaminant pool encountered during immobilized metal affinity chromatography

4.0 ABSTRACT

A description of native bacterial proteins from *Escherichia coli* that bind during Immobilized Metal Affinity Chromatography (IMAC) has been prepared by passing extracts through IMAC columns of varying metal ion, and identifying proteins that bind through proteome analysis. Specifically, $Co^{(II)}$, $Ni^{(II)}$, and $Zn^{(II)}$ IMAC columns were examined, and the isoelectric point, molecular weight, and essentiality of the characterized proteins were determined. Information regarding this group of proteins is presented and used to define the IMAC bioseparation-specific metalloproteome of *E. coli*. Such data concerning the potential contaminant pool is useful for the design of separation schemes, as well as designing optimized affinity tails and strains for IMAC purification.

5.0 INTRODUCTION

Proteomics is the analysis of gene and cellular function at the protein level. Traditionally it has been used to determine biologic events including protein expression, post-translational modifications, and protein-protein interactions. This analysis is done by combining modern techniques like mass spectrometry, high-resolution chromatography and bioinformatics (Giometti, 2003; Shi, Xiang, Horváth, & Wilkins, 2004). Other uses of proteomics include the search for biomarkers in clinical diagnosis and therapy (Drake, Cazares, Semmes, & Wadsworth, 2005). Proteomics depends on advanced and sensitive technologies which allow rapid separation and identification of proteins (Hoog & Mann, 2004). Typically the resolution of a mixture of proteins from a sample is done by 2 dimensional (2D) gel electrophoresis (Fey & Larsen, 2001). Determination of proteins is done by identifying their digested peptides by mass spectroscopy (MS) (Drake, et al., 2005). The combination of 2D gel electrophoresis, MS, and pretreatment strategies including chromatography has been established as a successful technique for protein identification and characterization (Gygi, et al., 2000; Stasyk & Huber, 2004).

Germane to this work is the use of proteome based data to characterize Immobilized Metal Affinity Chromatography (IMAC). Porath et al. introduced IMAC as a separation method in 1975 (Porath, et al., 1975). This method is based on the affinity between peptides / proteins and metal ions bound to a support matrix. IMAC can also be used to separate nucleotides (Hubert & Porath, 1980), and has proven useful in probing surface amino acids of proteins (Andersson, et al., 1987; Rossetto, et al., 1992) and interactions between proteins

(Andersson, et al., 1987). IMAC has advantages compared to other affinity techniques that include ease of regeneration and stability. Up to now, divalent Cu, Ni, Zn, and Co has been used in IMAC (Ueda, et al., 2003). These divalent transition metals have favorable binding with N, O, and S and under conditions of high salt content, favor nitrogen coordination (Gaberc-Porekar & Menart, 2001b). Within the column, transition metal ions (e.g. Ni(II)) are bound by immobilized chelating ligands. Several commercial chelating groups are available for IMAC: iminodiacetic acid (IDA), nitrilotriacetic acid (NTA), and carboxymethylated aspartic acid (Chaga, 2001). IDA is a commonly used as metal-chelating ligand in IMAC (Lindgren, 1994). Amino acids like histidine, cysteine, glutamic acid and aspartic acid have electron donating side chains, but histidine has the greatest affinity towards chelated metals under conditions of high NaCl content (Yip, et al., 1989). With the introduction of affinity tags (e.g. His_6), IMAC has developed into one of the most robust methods for the purification of recombinant proteins (Arnau, et al., 2006; Lichty, et al., 2005).

Although IMAC is widely used for single-step purification of recombinant proteins, the presence of host cell proteins having an affinity toward transition metal ions remain problematic. Notably, the loss of column capacity and the complication of gradient elution occur when genomic proteins are retained during adsorption. *A priori* selection of chromatographic steps is nearly impossible in a rational manner due to lack of knowledge regarding the contaminant proteins. To this end, this article reports on contaminant proteins of *Escherichia coli* that will bind to IMAC columns under various conditions. Identification of these metal binding proteins will aid in the design of purification strategies for recombinant proteins.

6.0 MATERIAL AND METHODS

6.1 Cell culture and Sample Preparation

Escherichia coli BL21 cells were obtained from the Stratagene (La Jolla, CA). Cells were cultured in Luria-Bertani (LB) medium in a 7-L Applikon Bioreactor (Applikon Inc., Dover, NJ) with 5 L working volume at 37° C with 200 rpm agitation rate and 500 ml/min aeration rate. Cells were harvested during late exponential phase. Cells pellets were obtained by centrifugation at 10,000 g for 10 min. Cell pellets were suspended in 100mL IMAC Loading Buffer (0.5 M NaCl, 0.05 M phosphate, pH 7.2) supplemented with non-chelating protease inhibitor from Sigma (St. Louis, MO). Cell lysis was performed by sonication on ice using twenty 1 sec pulses using a Branson Sonifier S-150 (Danbury, CT). For improved lysis, the cell lysate was frozen at -80º C overnight and sonication was repeated on following day. Cellular debris was removed by centrifugation at 30,000 rcf for 20 min, followed by filtration with a 0.2 µm syringe filter to remove any remaining particulate matter. Collected supernatant samples were applied directly to IMAC columns.

6.2 IMAC and sample preparation

Chromatography experiments were performed on an AKTA™ FPLC (GE Healthcare, Piscataway, NJ) system at room temperature. HiTrap™ Chelating Sepharose (pre-packed columns) of 5 ml volume were charged with Co(II) or Ni(II) or Zn(II). Metal ions were coupled to the resin by applying two column volumes (CV) of a 50mM solution of the corresponding metal chloride. Unbound metal ions were removed by washing with 5 CV of deionized water. Binding/wash buffer contained 0.05 M phosphate, 0.5 M NaCl, at pH 7.2. Samples were filtered with a 0.2 µm syringe filter to remove any solids and applied after the column was equilibrated with 3 CV of binding/wash buffer. The column was washed with 3 CV of binding/wash buffer to remove non-specific binding proteins from the column. Bound proteins were eluted with 50mM ethylene diamine tetraacetic acid (EDTA). Samples were desalted using a pre-packed column of 53 ml (GE Healthcare) by isocratic elution with 5mM phosphate buffer supplemented with Sigma protease inhibitor tablets to reduce protease activity to the extent possible. Samples were stored in -80^{o} C before concentration. Protein samples were concentrated using an Amicon™ stirred cell using 5000 MWCO membranes (Millipore, Billerica, MA). Total protein concentration was measured with a DC protein assay kit according to the manufacturer (Bio-Rad, Hercules, CA).

6.3 Protein Identification

Concentrated samples were sent to Midwest Biosciences (Lawrence, KS) for protein identification. The methodology for protein identification procedure may be found at the following website (http://www.proteinid.com/proteinid.html). Briefly samples were digested with trypsin and applied to a peptide trap column to remove salt and impurities. Peptides were separated by microcapillary C18 reverse-phase chromatography column and delivered to MS for identification. MS/MS spectra corresponding to peptides were identified by matching protein sequence databases. This analysis was done using TURBOSEQEST™ software.

6.4 Data Analysis

A property list of the proteins that bind to IMAC columns was developed based on literature information and *in silico* analysis. Essentiality of genes corresponding to contaminant proteins was based on Gerdes et al (Gerdes et al., 2003). This reference contains a comprehensive experimental assessment of *E. coli* MG1655. Isoelectric point (pI) and molecular weights were calculated by using a pI/MW calculation tool on the ExPASy server (http://www.expasy.ch/tools/pi_tool.html). Isoelectric values of proteins were calculated based on the amino acid pKa values and molecular weights were calculated from the isotopic masses of amino acids.

7.0 RESULTS

7.1 IMAC Contaminant Proteins

A comprehensive list of the genes responsible for the expression of *E. coli* proteins that bind to High Trap Chelating Sepharose with $Co^{(II)}$, $Ni^{(II)}$, or $Zn^{(II)}$ as the affinity ligand has been compiled (Tables 1, 2 and 3). Proteins were bound to the aforementioned metal affinity ligands by the IDA chelating group, and eluted using a step in EDTA concentration. Chromatography was performed both in the presence and absence of imidazole. For each metal ion examined, Figures 1 and 2 summarize the distribution of a total contaminant pool (TCP) with respect to essentiality of gene product. The TCP of each metal ion had a slightly larger pool of nonessential proteins, with Co(II) IDA binding both the least number of gene products deemed essential and a favorable ratio of essential: nonessential (E:N) of 0.61. Note that for each metal ion, total proteins identified in the TCP did not equal the sum of essential and nonessential proteins. The disparity was due to the fact that although the genome of *E. coli* is well documented, there still a remains significant fraction of gene products within the proteome that remain uncharacterized and are not well studied with respect to function and essentiality (Blattner et al., 1997).

Table 1: Genes associated with total contaminant pool of Co-IMAC

accC	*fusA*	*mfd*	*qseB*	*yciK*
aceE	*gapA*	*mgtA*	*rho*	*ydfG*
aceF	*gidB*	*miaB*	*rimM*	*yebU*
adhE	*glgB*	*minC*	*rne*	*yecP*
ahpF	*glk*	*mrp*	*rph*	*yegQ*
alaS	*glmS*	*mukB*	*rplC*	*yeiE*
aroH	*glnD*	*murC*	*rplN*	*yfbG*
asnB	*gltB*	*murE*	*rplY*	*yfbU*
atpC	*gltX*	*nagB*	*rpoB*	*yfcG*
atpG	*glyA*	*narL*	*rpoC*	*yfgD*
basR	*groL*	*nrdA*	*rpoD*	*yfgM*
bcp	*guaA*	*nrdD*	*rpsB*	*yfhQ*
can	*guaB*	*nuoG*	*rpsJ*	*yfiA*
carB	*gyrA*	*nusG*	*rstA*	*ygiN*
cbpA	*gyrB*	*pdxB*	*sdhA*	*ygjF*
chbR	*hemD*	*pflA*	*slyD*	*yhaJ*
clpA	*hemE*	*pflB*	*speG*	*yhbG*
clpB	*hepA*	*phoP*	*tdk*	*yhbH*
coaA	*hinT*	*prfA*	*thrS*	*yhbW*
crp	*hslR*	*prfC*	*tiaE*	*yhfT*
cueO	*infA*	*prkB*	*tpiA*	*yhhX*
cusF	*kdsA*	*proB*	*treC*	*yhiQ*
cydA	*leuC*	*prsA*	*tufA*	*yhiR*
dapB	*lpd*	*pta*	*tufB*	*yifE*
dnaE	*malP*	*purB*	*yacF*	*yjjK*
dnaK	*map*	*purH*	*yajD*	*yjtD*
dnaN	*mdoB*	*purU*	*ybfF*	*yraL*
folE	*mdoD*	*putA*	*ycbY*	*ytfP*
folX	*metH*	*pyrB*	*ycfP*	
fur	*metL*	*pyrI*	*ycgK*	

Table 2: Genes associated with total contaminant pool of Ni-IMAC

aceE	*gatA*	*miaB*	*relA*	*ybiV*
aceF	*gatD*	*minC*	*rho*	*ycfP*
adhE	*gcvP*	*mpaA*	*ribD*	*ycgK*
ahpF	*gidB*	*mpl*	*rne*	*yciK*
alaS	*glgB*	*mrp*	*rph*	*ydcY*
artP	*glmS*	*msrA*	*rplC*	*ydiI*
atpC	*glnD*	*mukB*	*rplF*	*yebU*
atpG	*gltB*	*murC*	*rplY*	*yecP*
basR	*gltX*	*murE*	*rpoB*	*yegQ*
bcp	*glyA*	*narL*	*rpoC*	*yfaY*
can	*groL*	*nrdA*	*rpsB*	*yfbU*
carB	*guaA*	*nrdD*	*rpsC*	*yfhQ*
cbpA	*guaB*	*nuoG*	*rstA*	*yfiA*
cbpM	*gyrA*	*nusG*	*sdhA*	*ygdH*
clpA	*gyrB*	*pdxB*	*speA*	*ygiN*
clpB	*hemN*	*pflA*	*sucA*	*yhaJ*
clpP	*htpG*	*pflB*	*suhB*	*yhbG*
corA	*infA*	*phoP*	*tdk*	*yhbH*
crr	*iscS*	*ppiA*	*thrS*	*yhbW*
cueO	*ispD*	*prfB*	*tpiA*	*yhhX*
cusF	*lpdA*	*prfC*	*treC*	*yhiR*
cydA	*lysS*	*proB*	*trmC*	*yieN*
cysD	*lysU*	*pta*	*trpS*	*yifE*
dnaE	*maa*	*purB*	*truD*	*yjbR*
dnaK	*malP*	*purH*	*tufA*	*yjjK*
dnaN	*mdoD*	*purL*	*tufB*	*yliG*
dps	*metH*	*putA*	*uspG*	*yncB*
edd	*metL*	*pykA*	*ybeD*	*yncE*
folE	*mfd*	*pyrB*	*ybgJ*	*yneH*
				zwf

Table 3: Genes associated with total contaminant pool of Zn-IMAC

ace	*galU*	*mdoD*	*pyrG*	*uup*
aceF	*gapA*	*metH*	*pyrI*	*yahK*
acnB	*gatD*	*metL*	*rffE*	*yccX*
adhE	*gdhA*	*mfd*	*rho*	*ycdW*
ahpF	*gidA*	*mgtA*	*ribD*	*ycfP*
arcA	*gidB*	*miaB*	*ribE*	*ycgK*
argE	*glk*	*mpl*	*rnb*	*yciK*
artP	*glmS*	*mukB*	*rnc*	*ydcF*
asnA	*gltB*	*nagB*	*rne*	*ydeA*
asnB	*gltX*	*nagD*	*rplC*	*ydfG*
basR	*glyA*	*narL*	*rplF*	*ydhF*
Bcp	*groL*	*nikR*	*rplN*	*yfbG*
Can	*gst*	*nrdA*	*rplS*	*yfbU*
carA	*guaA*	*nudC*	*rplY*	*yfcG*
carB	*guaB*	*nuoG*	*rpoB*	*yffB*
cbpM	*gyrA*	*nusG*	*rpoC*	*yfgD*
clpA	*gyrB*	*ompA*	*rpsA*	*yfiA*
clpB	*hemE*	*ompF*	*rpsF*	*yhbG*
clpP	*hepA*	*ompR*	*rstA*	*yhbH*
cpxR	*hinT*	*pdxH*	*sdhA*	*yhbW*
csrA	*hisB*	*pflA*	*slyD*	*yhiR*
cusF	*hisH*	*pflB*	*speD*	*yieN*
cydA	*hrpA*	*phoB*	*speG*	*yifE*
dapB	*hslV*	*phoP*	*tdk*	*yjjK*
Dfp	*hybC*	*prfC*	*thrA*	*yjtD*
dnaK	*ilvD*	*proB*	*thrS*	*yliG*
Edd	*infA*	*prsA*	*tiaE*	*yqcD*
elaB	*iscS*	*pta*	*tpiA*	*yrdA*
fabI	*kdsA*	*purB*	*treC*	
Fbp	*leuC*	*purH*	*tufA*	
folE	*lexA*	*purL*	*tufB*	
Fur	*lysU*	*purU*	*ugd*	
fusA	*malP*	*putA*	*uhpA*	
galF	*map*	*pyrB*	*uspG*	

Figure 1: Distribution of a total contaminant pool (TCP) with respect to essentiality of gene product.

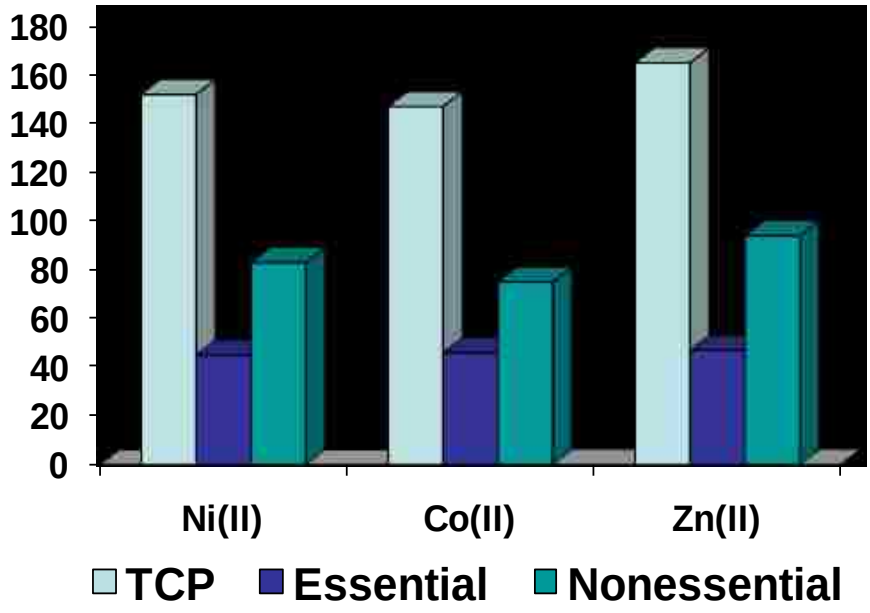
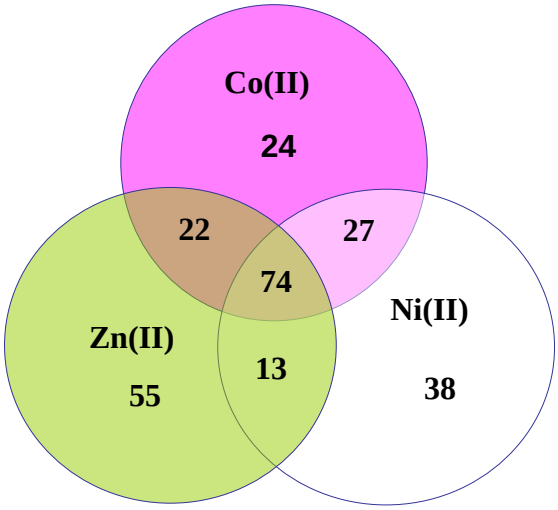


Figure 2: Distribution of commonality between contaminant pools with respect to Co(II), Ni(II), and Zn(II) chealates used in IMAC separations.



Other comparisons were made between the TCP corresponding to each metal ion. TCPs of Co(II) and Ni(II) had the greatest similarity, in agreement with prior studies (Y. Cai et al. 2004). A lack of commonality between Zn(II) and Ni(II) also occurred, as 22 identified proteins bound to both Co(II) and Zn(II). 2D representations of pI and MW for total contaminant proteins identified for three metals are presented in Figures 3 and 4, with this data representing the first report of IMAC TCPs. Virtually all of the TCP resided in the range 5.5 - 7.0, and proteins with pI values above 7.0 were not encountered to a significant extent. Indeed, there were less than ten proteins with theoretical or measured pI values between 7.0 and 10.5. Most of the TCPs resided between pI values of 5 - 7 and below 80 kDa MW. One gene product had a pI value greater than 7 and a MW greater than 60 kDa in Co(II) and Zn(II) TCPs, and no gene product was found in that range for Ni(II).

Figure 3: Distribution of contaminating species found in IMAC metalloproteome relative to isoelectric point.

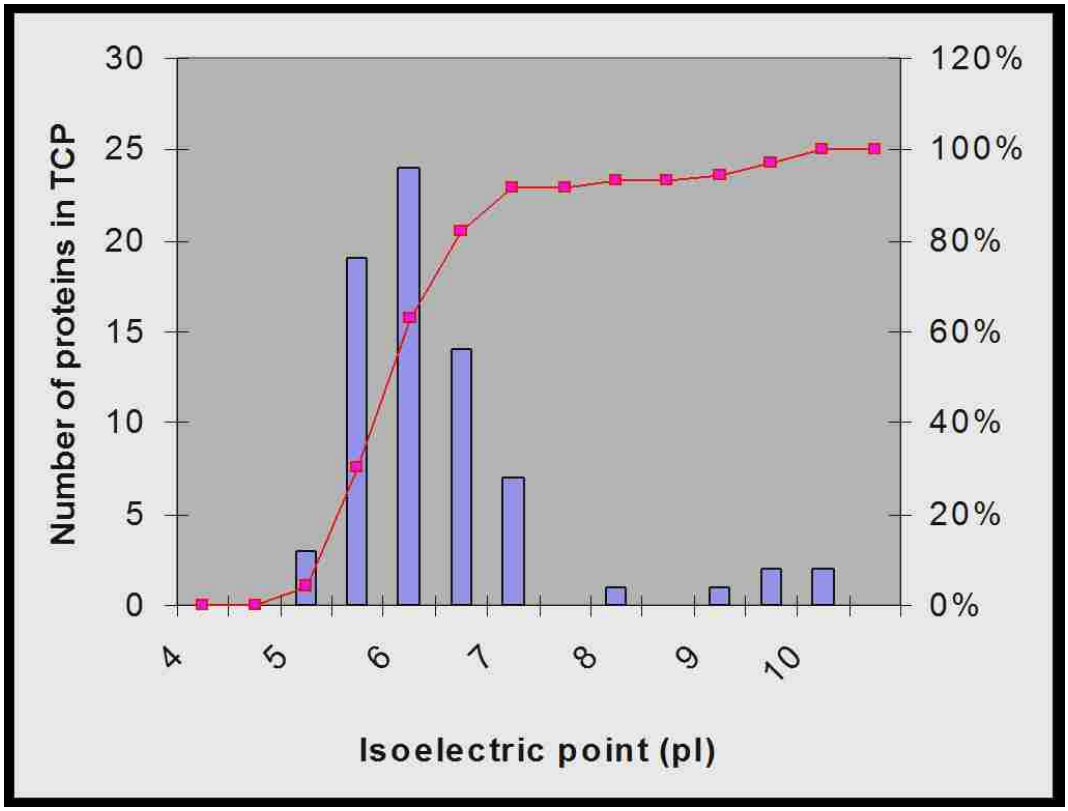


Figure 4a: Distribution of contaminant proteins found in Co(II) IMAC purifications when an IDA chelate and pH step elution is employed.

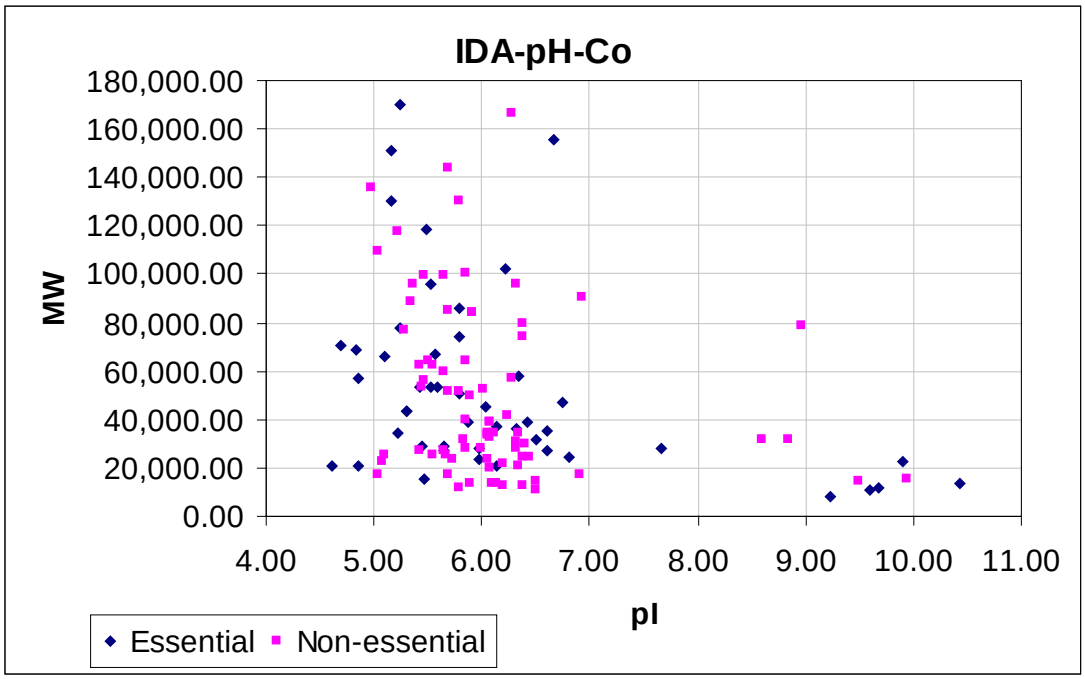


Figure 4b: Distribution of contaminant proteins found in Ni(II) IMAC purifications when an IDA chelate and pH step elution is employed.

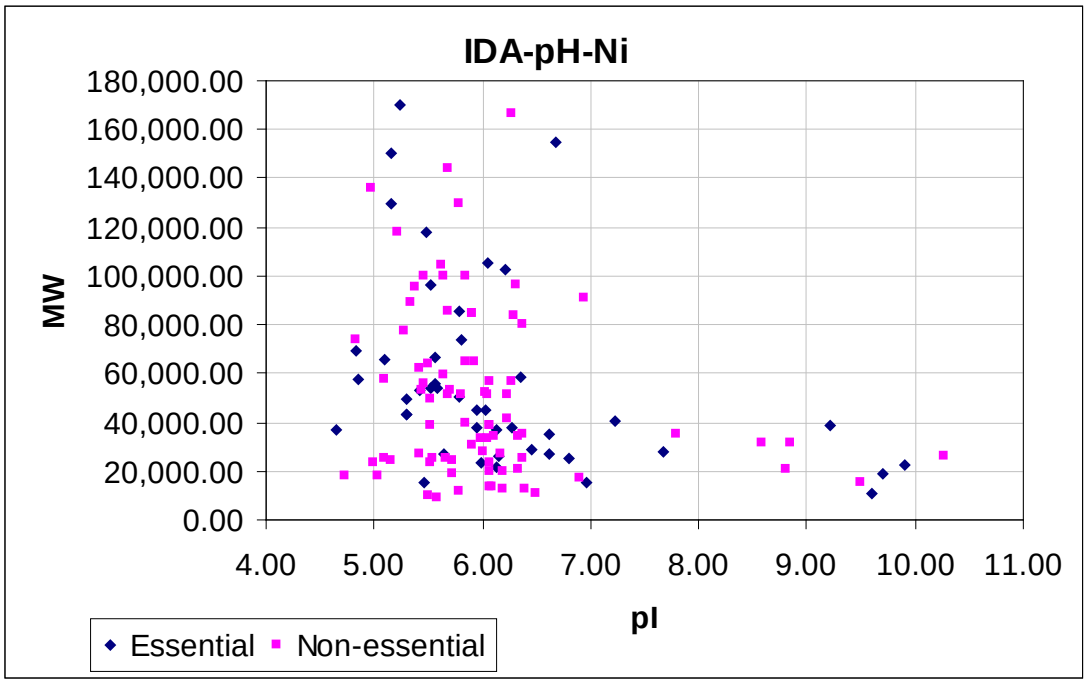
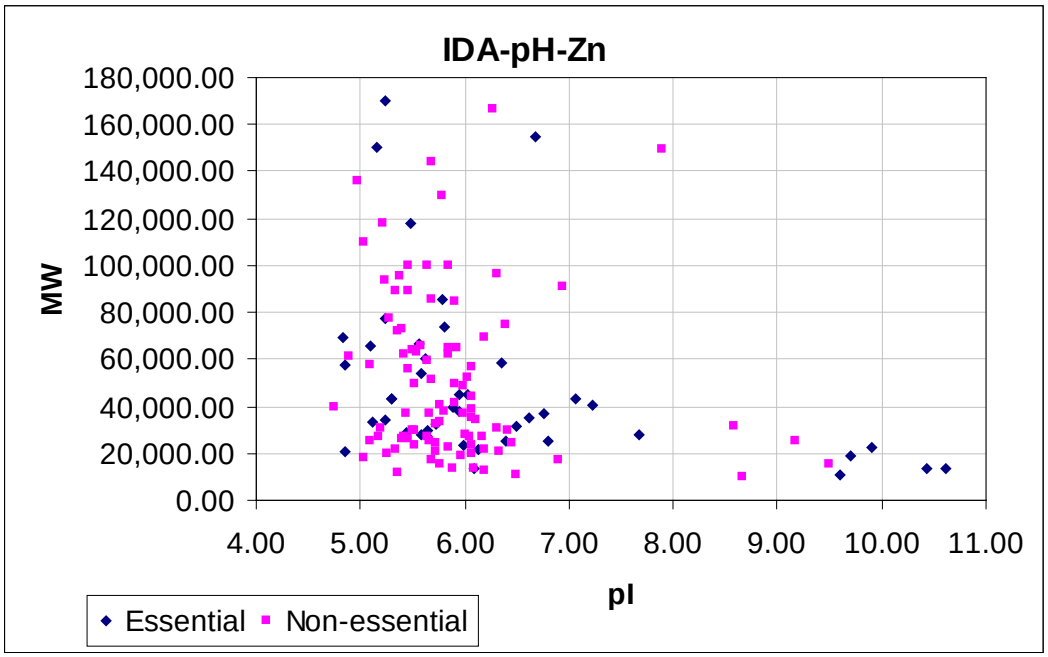
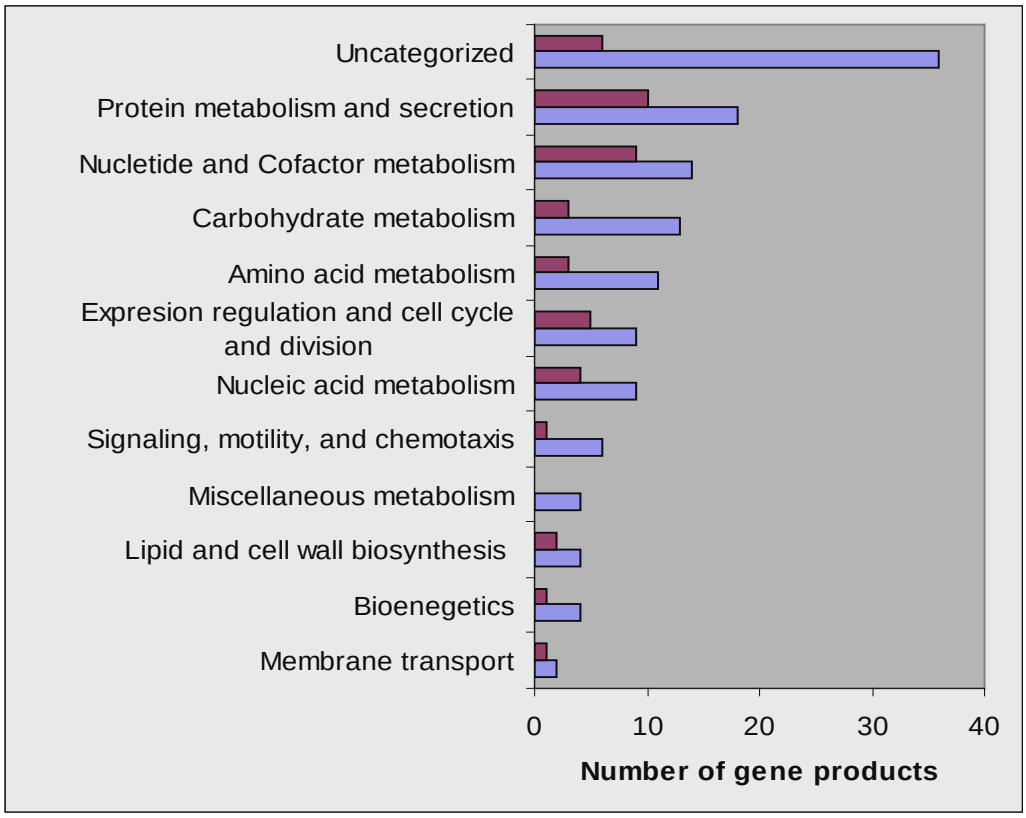


Figure 4c: Distribution of contaminant proteins found in Zn(II) IMAC purifications when an IDA chelate and pH step elution is employed.



Information regarding the metabolic function of genes was determined. Figure 5 describes the distribution of nonessential proteins for several functional categories (e.g. biosynthesis, carbohydrate metabolism) within a TCP, with the charted value defined as a percentage of the total TCP for a given condition (e.g. Co(II) in the presence of imidazole). For example, if one considers entries which describe binding in the presence of imidazole, 11, 7, and 11 percent of the TCPs for Zn(II), Ni(II), and Co(II), respectively, are considered nonessential and related to amino acid biosynthesis. The group which represents the largest qualification was *unclassified*, representing approximately 25 percent of the total populations. In rank order, the top five categories of proteins in a TCP were: *unclassified*, *protein metabolism and secretion*, *nucleotide and cofactor metabolism*, *carbohydrate metabolism*, and *amino acid metabolism*. With the exception of *protein metabolism and secretion* and *nucleotide and cofactor metabolism*, each category had less than half essential proteins.

Figure 5: Distribution of nonessential contaminating proteins of the total contaminant pool based on gene product functional category.



8.0 DISCUSSION

IMAC, an established method of bioseparation, served as the focus of a study of how modern biochemical techniques can shed light on groups of proteins. Specific to this work was the initial characterization of proteins which bind to IDA-bound chelated metal ions. To our knowledge no attempt has been made to provide a large dataset describing the proteome associated with a bioseparation method. While the converse has occurred, *viz*. IMAC has been used to fractionate a proteome, data of this nature has not been viewed with respect to bioprocesses design (Shi, et al., 2004; Stensballe, Andersen, & Jensen, 2001; Sun, Chiu, & He, 2005).

Characterizations that result from this work can guide improvements in affinity tails, as well as enhancements to the recombinant host itself. The aforementioned are deemed long term improvements, while short term gains are possible by using the dataset in a direct fashion. The latter will be discussed first. The data are helpful to find useful differences between a hypothetical recombinant target protein and impurities; appropriate physical or chemical properties (surface charge, molecular weight, hydrophobicity) may be chosen to select the separation steps that follow the IMAC affinity capture step. For example, if one chose to express a recombinant product in the range of 60 - 80 kDa with an isoelectric point between 6 - 7, Figures 3 and 4 aid in understanding the nature of the contaminating proteins that may be encountered. Theoretically, by using $Co^{(II)}$ IMAC one will encounter two host cell impurities which could be removed by ion exchange or gel filtration. More importantly, these two gene products are non-essential according to literature and could be removed from consideration via gene knockout. Similar comments may be made with respect to zones with minimal contaminant

for divalent nickel and zinc. As the MW / pI characteristics of a hypothetical target moves towards the cluster for each metal ion (cf. Figures 4a-c) the total impurities that may be encountered increase, but number of essential proteins do not dramatically change. Again, gene knockout of a potential candidate is a useful strategy as the likelihood of encountering a nonessential gene product remains favorable. Should one choose to develop an improved IMAC strain that displays a minimum TCP, deletion of nonessential gene products or modification of histidine content for proteins deemed essential is very possible based on this data set. Indeed, as more of the *E. coli* proteome is characterized, information such as this could serve as a powerful tool to develop an IMAC-optimized host strain that lacks the nonessential contaminants and / or alter their histidine content. Factoring in the space defined by MW / pI values, several zones of few impurities could be expanded via this strategy.

There is a great deal of overlap between TCPs for all three metal ions as shown in Figure 2. Further breakdown with respect to metabolic function (cf. Figure 5) shows how promising the prospect of developing an IMAC-optimized bacterial strain can be due to the lack of essentiality identified for the majority of proteins. The predicted proteome for *E. coli* has been described (Han & Lee, 2006). The predictive proteome of 4,288 ORF, which correlates well with experimentally derived subsets, can be compared to the metalloproteome of $Co^{(II)}$, $Ni^{(II)}$, or $Zn^{(II)}$. The metalloproteomes favored alkaline proteins (pI >7) over acidic proteins (pI<7), and showed a similar distribution of acidic proteins to that of the genome. However, the TCPs presented few alkaline proteins, arguably with one in the range above 60 kDa for three metals. This suggests the design of a bifunctional IMAC + ion exchange affinity tail. Bifunctional tails have been described in previous literature reports (Lichty, et al., 2005; Tan, et al., 2005). They have not

been designed *a priori* based on bioprocess data.

9.0 CONCLUSIONS

E. coli genomic proteins that bind to IMAC columns through IDA as the chelating agent for different metal ligands were identified using proteomic techniques. Commonality, essentiality and metabolic functions of these proteins were presented. Three metals showed similarity in binding, with approximately 50% of each metal protein pool common. Interestingly, 70% of each metal binding protein pool was deemed non-essential with respect to bacterial growth per literature data. These protein pools represent the total contaminant pool for IMAC capture step, with further characterization helpful to select complementary steps to IMAC, develop new affinity tails, develop simple elution methods, and ultimately build an *E. coli* host strain(s) that express a reduced contaminant pool.

Part II: A redesigned Escherichia coli triosephosphate isomerase restores growth properties in a bacterial strain useful for Immobilized Metal Affinity Chromatography (IMAC)

10.0 Abstract

The bacterium Escherichia coli is one of the most commonly used organisms in biotechnology for recombinant protein production and high-throughput development of biopharmaceuticals. The focus of this article is the utilization of proteome based data to design an E. coli expression strain that is improved for initial protein capture via Immobilized Metal Affinity Chromatography (IMAC). Proteome data was specifically applied to guide the modification of a known IMAC binding protein, triosephosphate isomerase (tpiA gene product), and the use of site directed mutagenesis eliminated binding properties. The designer tpiA gene, when reintroduced into an E. coli strain deficient in this enzyme activity, produced a functional protein lacking in surface exposed histidine and was able to restore glycolytic function.

Keywords

Escherichia coli, Immobilized metal affinity chromatography, PCR mutagenesis, Protein engineering, Triosephosphate isomerase

11.0 Introduction

Within the application of immobilized metal affinity chromatography (IMAC), protein affinity to divalent metals is principally dependent on correct spatial orientation of surface exposed histidine (Gaberc-Porekar and Menart 2001). While it is an established method to add extra histidine in the form of an affinity tag (e.g. His_6) to provide increased specificity towards IMAC, other contaminating species can easily show affinity to the chelated metal ion (Co(II), Ni(II), Zn(II) or Cu(II)) (Bolanos-Garcia and Davies 2006; Cai et al. 2004; Gaberc-Porekar and Menart 2001). These contaminating species, when adsorbed to the chromatography column can cause a reduction in column capacity, complications in gradient elution and a need for other companion bioseparation steps. For example, a survey of several IMAC papers indicate that when His_6 is used in conjunction with Co(II), (i) approximately 18 - 43% of proteins bound to the column are not the target recombinant, (ii) elution conditions of the peak of interest can vary from 20 - 500 mM imidazole, and (iii) it is necessary for chromatographic steps like ion exchange or size exclusion to accompany initial IMAC purification (Efremenko et al. 2006; Ha et al. 2008; Hutchinson and Chase 2006; Hutchinson et al. 2006; Liu et al. 2007; McCluskey et al. 2007; N. Abdullah and Chase 2005). For these reasons, elimination of contaminant species from the genomic contaminant pool is likely to improve the recovery process as a whole.

In order to develop a bacterial host for use in recombinant protein expression that is modified to ease the chromatographic burden for IMAC, a logical first step is to qualitatively establish any similarities between metalloproteomes when divalent metal ions are used as the affinity ligand (Cai et al. 2004). For example, triosephosphate isomerase (*tpiA* gene product) was identified as a

common protein during IMAC adsorption regardless of metal ion chosen for an IMAC capture step, and is of sufficiently high concentration to affect column capacity when the recombinant target protein (e.g. membrane protein) is not expressed in high quantities. To modify the *tpiA* gene such that the transcribed protein will not significantly bind IMAC resins, substitutions for three histidine codons via mutagenesis are necessary. The study described in this paper investigated alterations in the *tpiA* gene to greatly diminish IMAC adsorption. We report on the design of an engineered *tpiA* gene, enzymatic activity of the triosephosphate isomerase containing three residue mutations, and restoration of *E. coli* growth when introduced via plasmid transformation.

Retention of this glycolytic enzyme is necessary to maintain integrity of the Embden-Meyerhof pathway. Triosephosphate isomerase is a catabolic enzyme that does not require cofactors, prosthetic groups, or metal ions. Its functional role is located at the branching step of the glycolytic pathway, where it catalyzes the interconversion of dihydroxyacetone phosphate to glycerol-3-phosphate through acid-base catalysis and electrophilic stabilization, respectively (Raines et al. 1986). It is seen in nature as a homodimer, with a subunit molecular weight of approx. 26 500 Daltons (Putman et al. 1972) and large hydrophobic interdigitating loops at the subunit interface that interact at the back wall of the active site (Casal et al. 1987; Schliebs et al. 1997). Triosephosphate isomerase of *Escherichia coli* possesses three surface histidine residues near the N-terminus, with two forming a His-X_2-His motif, hypothesized to be the cause of adsorption.

12.0 Material and Methods

12.1 Bacterial Strains and Growth

E. coli VR101 (*tpiA*-), a glycolysis compromised strain, was a kind gift from the Saab-Rincon lab (Saab-Rincon et al. 2001). This strain was prepared through homologous recombination of an *E. coli* strain (JM101) designed to introduce a kanamycin resistance gene into the *tpiA* open reading frame (ORF). This insertion rendered the chromosomal gene inactive. *E. coli* BL21 (DE3) was used to provide a copy of a wild-type *tpiA* template for PCR amplification followed by ligation into the pGEX-6P-2 vector (GE Life Sciences, Uppsala, Sweden). *E.coli* BL21 Star (Invitrogen, Carlsbad, CA) was used as a comparable expression strain for production of recombinant product. *E.coli* TB-1 was used for stable cryogenic storage of DNA constructs.

12.2 PCR Methodology

E. coli BL21 (DE3) chromosomal DNA was purified using a Chromosomal DNA Extraction Kit (Qiagen, Valencia, CA) to provide a wild-type template for PCR amplification of the *tpiA* gene sequence. PCR amplification was carried out using Platinum *Taq* Polymerase Supermix™ (Invitrogen, Carlsbad, CA) with the following primers: Forward (BamHI site underlined) CGGGATCCATGCGACATCCTTTAGTGATGGGTAACTG; Reverse (EcoRI site underlined) GCAGAAGCGGCTAAACAGGCTTAAGAATTCC. Briefly, the forward primer begins at the ATG starting codon for the *tpiA* gene with an upstream BamHI site to allow for directional ligation into the pGEX-6P-2 vector. The reverse primer includes the TAA stop sequence seen within the genomic sequence allowing for amplification of the entire ORF, and an EcoRI site immediately downstream to allow for directional ligation into the vector.

Mutagenesis PCR was carried out using purified plasmid containing wild type *tpiA* as template DNA. Platinum Taq Polymerase Supermix™ (Invitrogen, Carlsbad, CA) supplemented with 2µl dimethyl sulfoxide (DMSO) and the following primers were used: 5' forward (*BamHI* site underlined; mutation points in bold; template annealing sequence in italics) GCGGATCCATGCGAGCTCCTTTAGTGATGGGTAACTGGAAACTGAACGGCAGCCGC GTCATGGTTCGCGAGCTGGTTTCTAACCTGCGTAAAGAG; 3' reverse (EcoRI site underlined) GCAGAAGCGGCTAAACAGGCTTAA GAATTCC. The forward mutagenesis primer contained the same directional ligation properties as the wild type primer. The annealing sequence was moved down stream within the ORF to allow the mutation sequence to act as an extended overhang with respect to the template sequence. The annealing sequence of the

forward primer was designed to match the physical properties of the 3’ reverse primer used in the wild type *tpiA* gene amplification. PCR product containing the mutations will henceforth be referred to as *tpiA:IMAC.1*.

12.3 Plasmid Construction

Both the *tpiAwt* and *tpiA:IMAC.1* PCR products as well as the pGEX-6P-2 vector (3µg) (GE Life Sciences, Uppsala, Sweden) were double digested using *EcoRI* (2µl at 20kU/ml) and *BamHI* (2µl at 20kU/ml) (New England Biolabs, Ipswich, MA, USA). Double digests were purified using a Qiagen Gel Extraction kit, followed by visualization by agarose gel electrophoresis. The *tpiAwt* and *tpiA:IMAC* sequences (112ng) were independently ligated into the digested plasmid (100 ng) using T4 DNA ligase at 28°C for 2 hours. Plasmids were transformed into *E. coli* TB-1 which was subsequently selected on LB agar containing ampicillin (150 µg/ml).

Resulting transformants were checked by PCR using MCS primers (GE Biosciences, Uppsala, Sweden) with parallel cultures of the transformant colonies grown in 1mL LB supplemented with 150µg/mL ampicillin. PCR checks were visualized on a 1% agarose gel, with positive amplification cultures correlating to positives grown overnight at 37°C. Qiagen Mini-Preps™ were performed to isolate plasmids. Resulting plasmids were sequenced at the DNA Sequencing Core Facility, University of Arkansas for Medical Sciences using an Applied Biosystems 3100 Genetic Analyzer.

Chemically competent *E. coli* VR101 was transformed with the *tpiA:IMAC.1* plasmid. The resulting transformants were selected on LB agar containing ampicillin (150 µg/ml) and 55 µg/ml kanamycin. Chemically competent *E. coli* BL21 Star was also transformed with plasmids, with the resulting transformants plated on LB agar containing ampicillin (150 µg/ml). Interest in the former strain stemmed from the basic premise of the work, restoration of triosephosphate

isomerase activity, whereas the latter strain was used to produce the enzyme in large quantities for enzyme characterization (kinetic activity and IMAC adsorption properties).

12.4 Expression in *E.coli* VR101:pGEX -6P-2_tpiA.IMAC.1

Seed cultures were inoculated 1:100 from cryogenic stock into Luria Bertani (LB) antibiotic supplemented media in 50mL conical tubes (ampicillin (150µg/mL); kanamycin (55 µg/ml)). Cultures were grown overnight in an orbital incubator set to 37°C and 200rpm. The following morning, one liter Erlenmeyer flasks containing 500mL working volume of LB and antibiotic were inoculated. Recombinant expression of the GST:eTIM fusion protein was induced at OD600 = 0.35 and culture growth was monitored for an additional 6 hrs. Cell pellet was recovered by centrifugation (10,000 rcf; 10 min). The pellet was resuspended in 20mL PBS supplemented with protease inhibitor without EDTA and sonicated at 0°C using 3 cycles of 15 x 1 second bursts, with a 3-5 minute resting period between rounds. Crude lysate was frozen in -80°C freezer overnight. The following morning the crude lysate was thawed on ice and the sonication step was repeated. Cellular debris was cleared by centrifugation at 25,000 rpm for 20 minutes. The supernatant was decanted into a clean 50mL conical tube and stored at -20°C. Cell debris pellet was stored for future use at -80°C.

12.5 Purification of eTIM

Cell lysate was thawed on ice and syringe filtered through a 0.45µm PES filter. Ten milliliters cleared lysate was loaded into an Atka FPLC super loop and passed over a 5mL GST column at 1.0 ml/min, allowing 3 column volumes (CV) flow through to clear any non-bound proteins from the column. A 5ml aliquot of "Cleavage Buffer" supplemented with 500U of PreScission™ Protease was loaded into the FPLC super loop and passed over the column at 1.0 ml/min to initiate GST:eTIM cleavage. Cleaved eTIM was collected as a single 10mL collection tube and immediately transferred to ice until further use. Following elution, the GST column was washed with several CVs of 50mM Tris pH-7.2 to remove residual glutathione from the column.

Cleavage fractions were quantified by BioRad DC Assay and visualized by SDS PAGE. Cleaved eTIM was prepped for buffer exchange using a G-25 26/10 desalting column into IMAC Bind/Wash Buffer (0.5M NaCl; 50mM Na2PO4; 10mM Imidazole pH = 7.47). Recovered eTIM fraction was collected in a 10mL volume and stored at -20°C for further analysis. A HiTrap HP Chelating column was charged with 3 CV of 5mg/mL $CoCl_2$ and rinsed with 10 CV of Milli-Q water. Column was equilibrated with IMAC Bind/Wash Buffer for 3 CV. Five milliliters of a lysate containing eTIM sample were loaded to the column at 1mL/min, and 5mL fractions were collected in the flow through and elution steps. As will be shown, the eTIM did not bind to the column and was collected in the flow through. Each 5mL fraction containing eTIM flow-through was combined to create a single 10mL fraction for activity studies.

12.6 Assay of Triosephosphate Isomerase

A kinetic activity assay from Sigma-Aldrich Life Sciences (St. Louis, MO) was used to measure specific activity of the engineered protein relative to wild-type kinetic properties, using purchased rabbit enzyme as a control (Sigma-Aldrich Life Sciences, St. Louis, MO). Total protein was measured using a BioRad DC assay kit (BioRad, Hercules CA). To gauge restoration of in vivo activity, cells transformed with tpiA.IMAC.1 were grown in batch cultures and cell growth was monitored. In these cultures, no IPTG was added for induction.

13.0 Results

Figure 1 describes the overall cloning strategy for the preparation of the engineered isozyme of triosephosphate isomerase that does not show binding affinity towards IMAC columns. PCR was used to amplify the chromosomal copy of tpiA and clone it into a pGEX vector, which was subcloned into the E.coli TB-1. This plasmid was in turn used as the template sequence for a mutagenic PCR reaction which resulted in a *tpiA* gene with the following substitutions: H3A, H17V, H20R[1]. These substitutions were based on a sequence analysis using the bioinformatics software application Bio-Edit™ (Ibis Biosciences, Carlsbad, CA). While the software application was originally designed for phylogenetic analysis of protein isoforms, here we implement it as a protein engineering tool to infer to the essentiality of surface exposed histidine. By comparing sequence data across multiple species, such data gives incite as to which residue substitution are most likely to result in a favorable outcome.

Sixty three sequence alignments (**Figure 2**) of various triosephosphate isomerase isoforms indicate significant amino acid variation within the region; including aliphatic, hydrophobic, and amphiphilic residues. Respective alanine, valine, and arginine residue substitutions were chosen in light of micro-environment chemistries within the region of interest, thus dictating the most favorable substitutions for each respective position. During construct design, significant consideration was given as to the location of histidine removal relative to location of the enzyme active site. It was determined that due to the significant distance between the location of mutagenesis and the enzymes active site (centered on position Glu_{167}) there was little chance of

[1] Amino acid numbering follows the convention for *E. coli* str. K-12 sub-strain MG1655. Designations use the single letter amino acid code. H3A signifies that in the engineered enzyme Histidine-3 has been changed to Alanine.

interfering with enzymatic kinetics. Of note, care was taken to preserve the integrity of Lys_{11}, due to its punitive role in stabilizing substrate intermediates (Kempf et al. 2007; Nickbarg et al. 1988; Nickbarg and Knowles 1988; Pompliano et al. 1990).

E. coli VR101 growth properties in the presence and absence of the *tpiA:IMAC.1* plasmid were examined in batch cultures (**Figure 3**). A trace of the OD for the parent strain indicated a severely compromised metabolism, as evident by the exceedingly long approach to a steady state value of 0.2 OD. In contrast, plasmid bearing VR101 responded quite well to the reintroduction of engineered *tpiA* with the aforementioned residue mutations, as evident by its growth curve (Figure 3). Batch growth resulted in a higher final OD of 1.7, which was eight and one half times that of the parent strain. The maximum specific growth rate (μ_{max}) was calculated, defined by

$$\mu_{\max} = \frac{1}{X}\frac{dX}{dt}$$

where *X* is cell concentration and *t* time. Under conditions employed in this study, the value of the maximum growth rate was 0.84 h-1. Also presented in **Figure 3** is the growth characteristics of BL21 (DE3), a strain used commercially for the production of recombinant protein. Aside from a small departure late during exponential growth, both BL21 (DE3) and *E. coli* VR101 harboring *tpiA:IMAC.1* displayed similar values of both μmax and final OD. The similarities in growth characteristics indicate that in vivo, isomerase activity has been restored to a value comparable to typical *E. coli*.

Figure 1: Cloning strategy implemented to obtain p*tpiA.IMAC.1*.

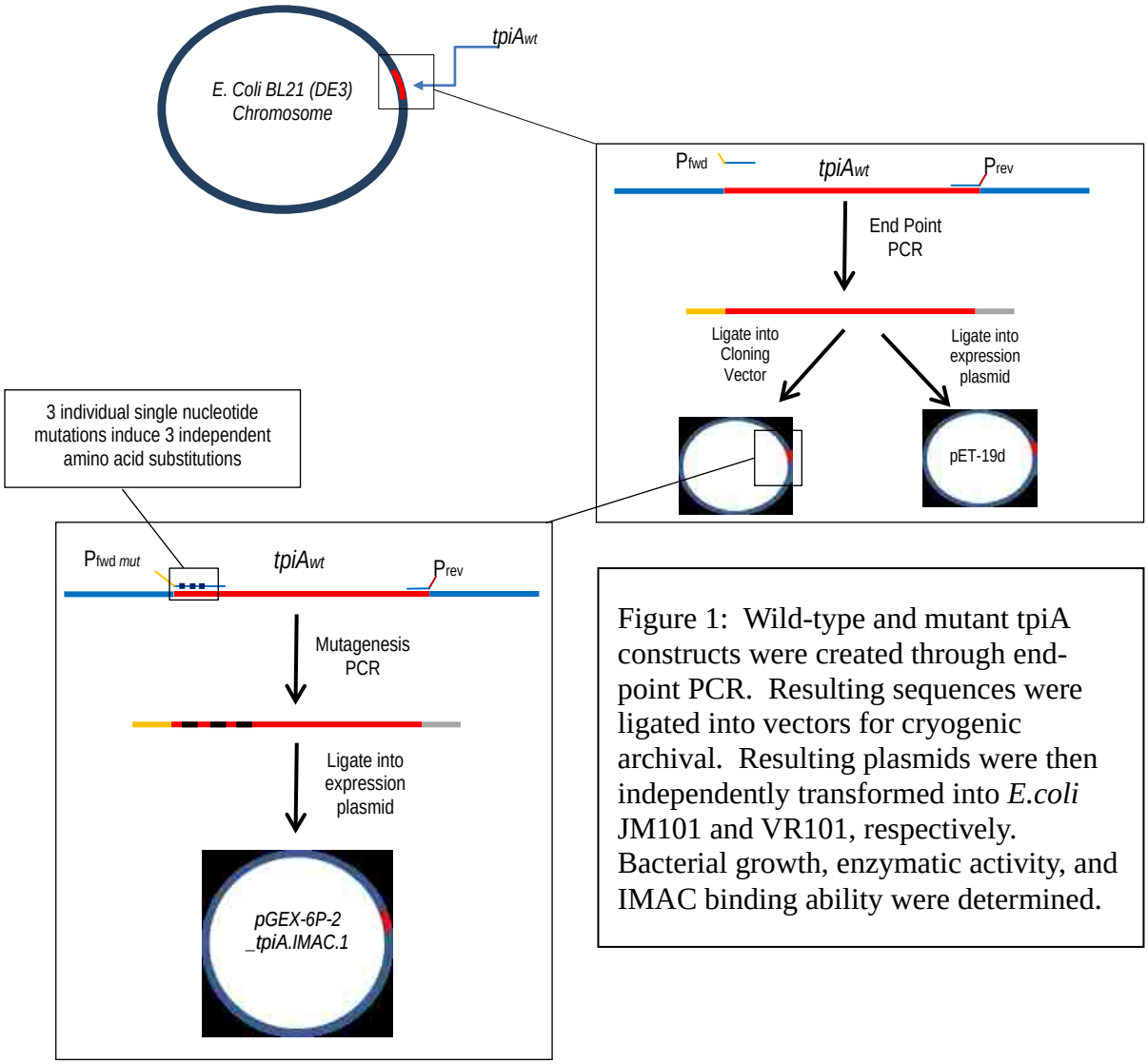


Figure 1: Wild-type and mutant tpiA constructs were created through end-point PCR. Resulting sequences were ligated into vectors for cryogenic archival. Resulting plasmids were then independently transformed into *E.coli* JM101 and VR101, respectively. Bacterial growth, enzymatic activity, and IMAC binding ability were determined.

Figure 2: TIM sequence alignments of 20 species showing region on mutagenesis for the removal surface exposed histidine.

gi|16131757 -------MRH PLVMGNWKLNGSRH-MVHELVSNLRKELAG--VAGCAVAIAPPEMYIDMAKREAEGS--HIMLGAQNVDLN-LSGAFTGE

gi|15834098 -------MRH PLVMGNWKLNGSRH-MVHELVSNLRKELAG--VAGCAVAIAPPEMYIDMAKREAEGS--HIMLGAQNVDLN-LSGAFTGE

gi|26250685 -------MRH PLVMGNWKLNGSRH-MVHELVSNLRKELAG--VAGCAVAIAPPEMYIDMAKREAEGS--HIMLGAQNVDLN-LSGAFTGE

gi|19075524 ----MA--RK FFVGGNFKMNGSLE-SMKTIIEGLNTTKLN--VGDVETVIFPQNMYLITTRQQVKK---DIGVGAQNVFDK-KNGAYTGE

gi|16767347 -------MRH PLVMGNWKLNGSRH-MVNELVANLRKELTG--VAGCDVAIAPPEMYIDLAKRAAAGS--HIMLGAQNVDLN-LSGAFTGE

gi|15599942 -------MRR PLVAGNWKMHGTHS-SVAELIKGLR-QLAL--PSGVDVAVMPPCLFISQVIQGLAGK--AIDVGAQNSAVEPMQGALTGE

gi|125972663 ------MSRK VIAAGNWKMNKTPK-EAVEFVQALKGRVA--DAD-TEVVVGVPFVCLPGVVEAAKGS--NIKVAAQNMHWE-KGAFTGE

gi|33597944 --MTTAENRARLVLGNWKMHGNLA-ENAALLAELR-AADA--AAHCEMGVCVPFPYLAQTAAALQGS--AIGWGAQDVSAGAYTGE

gi|15608576 ------MSRKPLIAGNWKMNLNHY-EAIALVQKIAFSLPDKYYDRVDVAVIPPFTDLRSVQTLVDGDKLRLTYGAQDLSPH-SGAYTGD

gi|33865357 -------MRRPVIAGNWKMHMTCA-QARDYMAAFLPQIER-APQDREIVLAPPFTALSTMAAAAEHS--VVGLASQNVHWQ-DHGAFTAE

gi|168185507 -------MRKAIIAGNWKMNNTIS-QGLKLVEELKPLVA--GAN-SDVVVCPPTLALDAVVKATEGT--NIKVGAQNMHFE-ESGAFTGE

gi|19745720 ------MSRKPIIAGNWKMNKNPQ-EAKAFVEAVASKLP--STDLVDVAVAAPAVDLVTTIEAAKDS--VLKVAAQNCYFE-NTGAFTGE

gi|6320255 ----MA--RTFFVGGNFKLNGSKQ-SIKEIVERLNTASIP--EN-VEVVICPPATYLDYSVSLVKKP--QVTVGAQNAYLK-ASGAFTGE

gi|4507645 ----MAPSRKFFVGGNWKMNGRKQ-SLGELIGTLNAAKVP--AD-TEVVCAPPTAYIDFARQKLDP---KIAVAAQNCYKV-TNGAFTGE

gi|15674692 ------MSRK PIIAGNWKMNKNPQ-EAKAFVEAVASKLP--STDLVDVAVAAPAVDLVTTIEAAKDS--VLKVAAQNCYFE-NTGAFTGE

gi|15679059 MLEDLELKDTPIVILNFKTYLESTGERALELASICGDVAD--ETGVNMAVAPQHMDLHRVSDAVEIP-----VLAQHIDAV-DAGGHTGS

gi|33591972 --MTTAENRARLVLGNWKMHGNLA-ENAALLAELR-AADA--AAHCEMGVCVPFPYLAQTAAALQGS--AIGWGAQDVSAH-AKGAYTGE

gi|29832840 -----MTSRMPLMAGNWKMNLNHL-EAIAHVQKLAFALADKDYEACEVAVLPPYTDLRSVQTLVDGDKLKIKYGAQDVSAH-DSGAYTGE

gi|58337021 ------MSRTPIIAGNWKLHMNPE-QTTEFVDAVKGKLP--DPSKVESLICAPAVDLDALRKAAEGS--NLHIGAENCYFE-DEGAYTGE

gi|29832840 -----MTSRMPLMAGNWKMNLNHL-EAIAHVQKLAFALADKDYEACEVAVLPPYTDLRSVQTLVDGDKLKIKYGAQDVSAH-DSGAYTGE

gi|30064787 -------MRH PLVMGNWKLN GSRH-MVHEL VSNLRKELAG --VAGCAVAI APPEMYIDMA KREAEGS--H IMLGAQNVDL N-LSGAFTGE

gi|16131757 TSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVLKEQGLTPVLCIGETEAENEAGKTEEVCARQIDAVLKTQGAAAFEGAVIAYE

gi|15834098 TSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVLKEQGLTPVLCIGETEAENEAGKTEEVCARQIDAVLKTQGAAAFEGAVIAYE

gi|262506855 TSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVLKEQGLTPVLCIGETEAENEAG KTEEVCARQIDAVLKTQGAA AFEGAVIAYE

gi|19075524 NSAQSLIDAGITYTLTGHSE RRTIFKESDEFVADKTKFALEQGLTVVACIGETLAEREANETINVVVRQLNAIADKVQN--WSKIVIAYE

gi|16767347 TSAEMLKDIGAQYIIIGHSERRTYHKESDE LIAKKFAVLK EQGLTPVLCI GETEAENEAG KTEEVCARQI DAVLKTQGAA AFEGAVIAYE

gi|15599942 TAPSQLADVGCSMVLVGHSERRLILGESDEVVSRKFAAAQSCGLVPVLCVGETRAEREAGKTLEVVARQLGSVIDELGVGFARAVVAYE

gi|125972663 VSGPMLAELGVDYVIIGHSERRQYFGETDETVNKKVHAAFKYGLKPIICVGESLTQREQGVTAELVRYQVKIALLGLSAEQVKEAVIAYE

gi|33597944 VAAPMLAEFGCRWVLVGHSERRTLHAESDQLVADKARAALEAGLTPVVCVGESLQEREGGNTLGVRQLEPVL-ALGRDALVRMVLAYE

gi|15608576 VSGAFLAKLGCSYVVVGHSERRTYHNEDDALVAAKAATALKHGLTPIVCIGEHLDVREAGNHVAHNIEQLRGSLAGLLAEQIGSVVIAYE

gi|33865357 ISAEMLLEHGVAYTIVGHSEPRKYFSESDEQINHRARCSQAKGLIPIVCVGESDEQRERGEAERVIRRQIEQGLEGLDAN---KLVVAYE

gi|168185507 IAPAMLEELGVKYVILGHSERRQYFGENDADLNKKMKKAFEHNLTPILC GETLEEREADVTEEVLAKQI KLDLAGLTKE QIAETVIAYE

gi|19745720 TSPKVLAEMGADYVVIGHSERRDYFHETDEDINKKAKAIFANGLTPIICCGESLETYEAGKAAEFVGAQVSAALAGLSAEQVASLVLAYE

gi|6320255 NSVDQIKDVGAKWVILGHSERRSYFHEDDKFIADKTKFALGQGVGVILCIGETLEEKKAGKTLDVVERQL NAVLEEVKD- -WTNVVVAYE

gi|4507645 ISPGMIKDCGATWVVLGHSERRHVFGESDELIGQKVAHAL AEGLGVIACI GEKLDEREAG ITEKVVFEQT KVIADNVKD- -WSKVVLAYE

gi|15674692 TSPKVLAEMGADYVVIGHSERRDYFHETDEDINKKAKAIFANGLTPIVCCGESLETYEAGKAVEFVGAQVSAALAGLSAEQVASLVLAYE

gi|15679059 ILAECARDAG AAGTLINHSE KRMQLADIEW VISR----MK ELEMMSVVCT NNVMTTAAAA ALG------- ---------- ---PDFVAVE

gi|33591972 VAAPMLAEFGCRWVLVGHSERRTLHAESDQLVADKARAALEAGLTPVVCVGESLQEREGGNTLGVIERQLEPVL-ALGRALVRMVLAYE

gi|29832840 ISGSMLAKLKCTYVAVGHSERRQYHHETDEIVNAKVKASFRHGLIPILCV GEELEVREAG NHVTHTLTQV EGGLKDVPAEQAETIVIAYE

gi|58337021 TSPKVLKEMGIDYVIIGHSERRGYFHETDEDINKKAKAIFANGMKPIICCGESLETREANKQEDWVVAQIKAALDGLTAEQVSSLVIAYE

gi|29832840 ISGSMLAKLKCTYVAVGHSERRQYHHETDEIVNAKVKASFRHGLIPILCVGEELEVREAG NHVTHTLTQV EGGLKDVPAE QAETIVIAYE

gi|30064787 TSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVLK EQGLTPVLCI GETEAENEAG KTEEVCARQI DAVLKTQGAA AFEGAVIAYE

gi|16131757 PVWAIGTGKSATPAQAQAVHKFIRDHIA-KVDANIAEQVIIQYGGSVNASNAAELFAQPDIDGALVG-GA SLKADAFAVI VKAAEAAKQA

gi|15834098 PVWAIGTGKSATPAQAQAVHKFIRDHIA-KVDANIAEQVIIQYGGSVNASNAAELFAQPDIDGALVG-GA SLKADAFAVI VKAAEAAKQA

gi|26250685 PVWAIGTGKSATPAQAQAVHKFIRDHIA-KVDANIAEQVIIQYGGSVNASNAAELFAQPD IDGALVG-GA SLKADAFAVI VKAAEAAKQA

gi|19075524 PVWAIGTGKT ATPEQAQEVH AEIRKWATNK LGASVAEGLR VIYGGSVNGG NCKEFLKFHD IDGFLVG-GA SLKP-EFHNI VNVHSL---

gi|16767347 PVWAIGTGKSATPAQAQAVHKFIRDHIA-KADAKIAEQVIIQYGGSVNASNAAELFAQPDIDGALVG-GA SLKADAFAVI VKAAEAAKQA

gi|15599942 PVWAIGTGLT ASPAQAQEVHAAIRAQLA-AENAEVAKGVRLLYGGSVKAASAAELFGMPD IDGGLVG-GA SLNADEFGAI CRAAGS----

gi|125972663 PIWAIGTGKT ATNEQAEEVC GIIRECIKEL YGQDVAEAIR IQYGGSVNAA NAAELFNMPN IDGGLVG-GA SLKLDDFEKI AKYNK-----

gi|33597944 PVWAIGTGRT ASPEQAQEVH SAIRVALD-G LQAS---QVR VLYGGSVKGA NAASLFAMPD IDGGLVG-GA SLVAEEFLRI AAA-------

gi|15608576 PVWAIGTGRVASAADAQEVCAAIRKELASLASPRIADTVRVLYGGSVNAKNVGDIVAQDDVDGGLVG-GASLDGEHFATLAAIAAGGPL

gi|33865357 PIWAIGTGKT CEAAEANRIC GLIRSWVG-- -----ATDLI IQYGGSVKPT NIDELMAMSD IDGVLVG-GA SLKPDSFARI ANYQAI----

gi|168185507 PIWAIGTGKT ATSDQAEETI AFVRKTVAGM FGAEAAEKMR IQYGGSVKPA TIKEQMAKPN IDGGLIG-GA SLKAADFAAIVNFDK-----

gi|19745720 PIWAIGTGKSATQDDAQNMCKAVRDVVAADFGQEVADKVRVQYGGSVKPENVKDYMACPDVDGALVG-GASLEADSFLAL LDFLN---

gi|6320255 PVWAIGTGLA ATPEDAQDIH ASIRKFLASK LGDKAASELR ILYGGSANGS NAVTFKDKAD VDGFLVG-GA SLKP-EFVDI INSRN-----

gi|4507645 PVWAIGTGKT ATPQQAQEVH EKLRGWLKSN VSDAVAQSTR IIYGGSVTGA TCKELASQPD VDGFLVG-GA SLKP-EFVDI INAKQ----

gi|15674692 PIWAIGTGKSATQDDAQNMCKAVRDVVAAD FGQEVADKVR VQYGGSVKPE NVKDYMACPD VDGALVG-GA SLEADSFLAL LDFLN-

gi|15679059 PPELIGSGIP VSRAEPEVIT GSVDAVKK-- ----VNPEVS VLCGAGISTG DDMKAAVDLG AEGVLLASGI ILADSPRDAL LDLVSKV---

gi|33591972 PVWAIGTGRT ASPEQAQEVH SAIRVALD-G LQAS---QVR VLYGGSVKGA NAASLFAMPD IDGGLVG-GA SLVAEEFLRI AAA-------

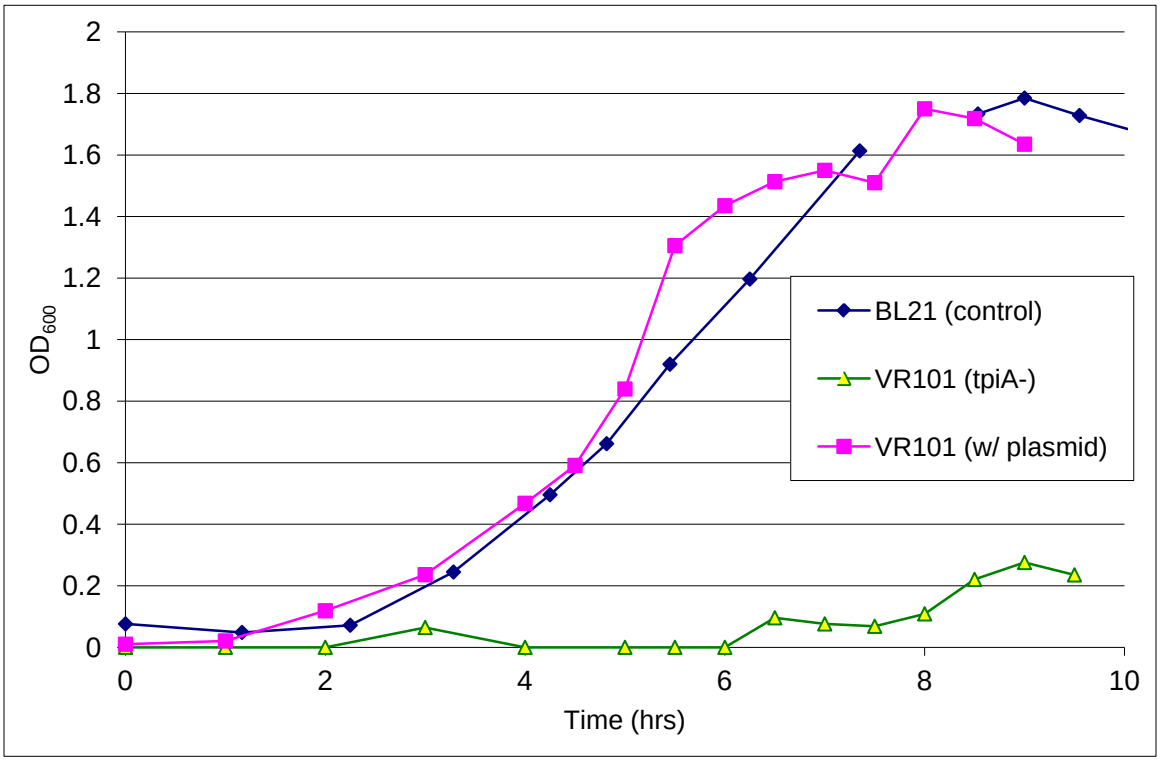
gi|29832840 PVWAIGTGKV CGADDAQEVC AAIRAKLAEL YSQELADQVR IQYGGSVKSG NVAEIMAKPD IDGALVG-GA SLDADEFVKI ARFRDQ-

gi|58337021 PIWAIGTGKT ASSDQAEEMC KTIRETVKDL YNEETAENVR IQYGGSVKPA NVKELMSKPD IDGGLVG-GA SLDPESFLAL VNYQD---

gi|29832840 PVWAIGTGKV CGADDAQEVC AAIRAKLAEL YSQELADQVR IQYGGSVKSG NVAEIMAKPD IDGALVG-GA SLDADEFVKI ARFRDQ-

gi|30064787 PVWAIGTGKSATPAQAQAVHKFIRDHIA-KVDANIAEQVIIQYGGSVNASNAAELFAQP IDGALVG-GA SLKADAFAVI VKAAEAAKQA

Figure 3: Growth Characterization of *E.coli* RH101*: pGEX -6P-2_tpiA.IMAC.1* when compared to *E.coli* VR101 and *E.coli* BL21 (DE3).



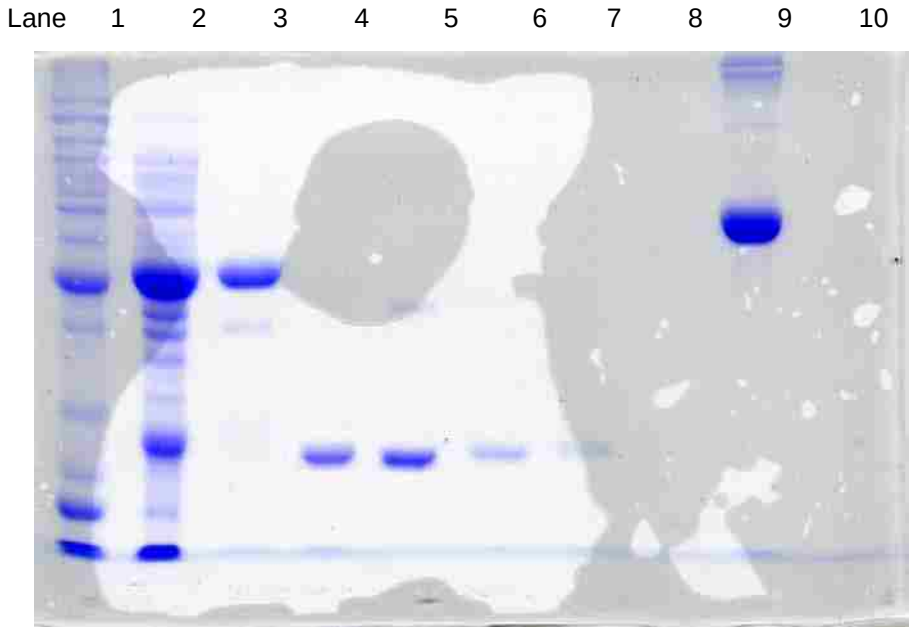
To address the possibility that differences in expression levels could, in part, explain the restoration of growth, kinetic properties of the engineered *tpiA* gene product was examined. Similar specific activity values demonstrate that enzyme from E. coli, rabbit, and E. coli *tpiA:IMAC.1* possess similar catalytic potential. As seen in **Table 1**, calculations show specific activities to be nearly identical (U/mg). This fact, when combined with growth data, leads us to believe that the described mutations have no effect on the function of the enzyme itself.

Table 1: Specific activity of the engineered TIM enzyme compared to Rabbit TIM and *E.coli wt*TIM.

Protein Sample	*Ave. U/mL*	*Standard Deviation*	*Concentration Protein*	*Units/ mg protein*
Rabbit TIM	0.579	0.1093	2.49 ng/mL	232.3
E.coli TIMwt	1221.6	24.99	5.26 µg/mL	232.3
*E.coli e*TIM	384.72	75.01	1.66 µg/mL	232.3

While eTIM restored growth conditions it was also necessary to confirm that the modified protein no longer bound to an IMAC chelating resin. **Figure 4** shows chromatography fractions following Prescission protease cleavage of eTIM. Lane 6-8 confirm that the modified eTIM does not bind to $Co^{(II)}$ IMAC column as evidenced by the presence of bands in lane 6 and 7, and its absence in lane 8.

Figure 4: 12.5% SDS-Page gel showing cleaved eTIM does not bind Co(II) IMAC column.



Lane 1: Benchmark Ladder (10µl)

Lane 2: Crude Lysate (7µL)

Lane 3: GST: TIM elution (10µL)

Lane 4: GST cleaved (10µl)

Lane 5: eTIM cleaved (10µl)

Lane 6: eTIM IMAC load (10µl)

Lane 7: eTIM IMAC flow-through

Lane 8: EDTA purge of $Co^{(II)}$ column

Lane 9: 5µg load BSA

14.0 Discussion

Comparing all three OD trends indicated that the overall strategy defined in this paper restored growth properties to levels comparable to that of commercially used strains. In general, genomic proteins that either reduce column capacity, require a gradient designed to minimize contaminant elution, and / or necessitate the addition of other purification steps can be removed from consideration by first eliminating the chromosomal copy of the essential gene product. When the protein is essential for cell growth, designing and constructing a version of the protein with a lack of binding features, and transforming the deletion strain with an appropriate plasmid harboring the designer gene can resupply wild-type activity while simplifying protein isolation. For the case of IMAC purification, removal of surface exposed histidine accomplishes this task.

Triosephosphate isomerase shows extensive conservation with respect to active site residues. However, regions that contain the three surface exposed histidine residues show few conserved features when compared across multiple species alignments. Of the conserved features, Glu_{167}, His_{95}, and possibly Lys_{11} have been described as the primary acting residues offering catalytic function (Nickbarg et al. 1988; Nickbarg and Knowles 1988). In the active site, Glu_{165} acts as a catalytic base during the enolization step of the conversion of dihydroxyacetone phosphate to glycerol-3-phosphate (Raines et al. 1986; Straus et al. 1985). This reaction is stabilized by a short α-helix region flanked by His_{95} and His_{103}, in which His_{95} acts as an acid during the collapse of the reaction to form glycerol-3-phosphate, while stabilizing hydrogen bonding during the enolization step (Lodi and Knowles 1993). These histidine residues are buried in the secondary structure, and are not capable of interacting with IMAC media (Aparicio et al. 2003; Blacklow et al. 1991).

Based on ClustalW2 alignments and a thorough search of the literature, surfaced exposed histidine at the N-terminal region shows no functional role with respect to folding or catalytic activity. As such, these residues were replaced using PCR mutagenesis. In frame residue substitutions were designed into the forward primer that allowed for residue replacement at the locations corresponding to the three histidine residues that were surface exposed. The codon substitutions resulted in Ala_3 for His_3, Val_{18} for His_{18}, and Arg_{21} for His_{21} in the engineered *E.coli tpiA* gene product. Kinetic assays confirmed the preservation of enzymatic activity, confirming to hypothesis that that N-terminal surface exposed histidine were unimportant from a catalytic standpoint.

While *in vitro* measurements were encouraging, growth characteristics provided the final test of the engineered *tpiA* gene. Leaky expression of the modified *tpiA* gene product clearly restored growth. When comparing it to the *E. coli* BL-21control strain, it is with in the approximate range of wild-type growth characteristics. The fact that the modified *tpiA* gene is expressed from plasmid is inconsequential, as it may be chromosomally integrated should one desire a plasmid free host strain. To this end, future efforts will investigate protein expression and IMAC purification to illustrate improvement in bioseparation by, in particular, recovering column capacity.

Part III: Future extension of proteomic approach toward development of a minimal genome recombinant expression host.

14.0 Discussion

To date, *a priori* selection of appropriate chromatographic steps for isolating a homologous recombinant protein has been impossible in any rational manner. This is, in part, due to two specific deficits of knowledge relating to understanding of the biochemical composition of the feed stream: i.) a lack of knowledge regarding specific contaminant proteins from which the recombinant target must be resolved, ii.) relatively little empirical data is available as to how a recombinant host could modify gene expression should mutagenesis be employed toward optimizing the crude lysate.

This dissertation has addressed these gaps by defining the *Escherichia coli* metalloproteome. Herein, I have provided a complete list of all IMAC contaminant proteins, and where possible I have annotated the genomic contaminant protein by g.i. #, associated gene name(s), protein name(s), and metabolic function. The complete IMAC contaminant protein list was assembled based on replicated empirical data ($n=3$), and using an industry standard recombinant expression host grown under batch fermentation conditions.

From an academic perspective, addressing the second knowledge deficit is a significantly more challenging prospect. As a proof of concept toward the generation of a specialized 'purification process' recombinant host, I chose characterize deletion and/ or site-directed mutagenesis as methods of circumventing IMAC contaminant proteins. Rational, for choosing two contaminant proteins was based on data indicating proportional concentration of individual contaminants relative to the overall concentration of the elution contaminant pool (ECP).

As an example of eliminating a non-essential contaminant protein, the *melA* gene product was found to be the second most concentrated protein species to make up the ECP. Melibiose dehydrogenase (*melA*) is a constitutively expressed protein which under laboratory controlled growth conditions for recombinant protein production is considered metabolically non-essential. I demonstrate that an *E.coli: ΔmelA* knock-out strain was successfully able to express a recombinant product with yields statistically similar to that of the *E.coli* BL21 (DE3) control strain. Thus, I demonstrate a simple gene knock-out to be a functional method of eliminating a non-essential genomic contaminant from both the total contaminant pool (TCP) and/or elution contaminant pool (ECP).

As an example of eliminating a metabolically essential contaminant protein, triosephosphate isomerase (TIM), *tpiA* gene product, was chosen as it propotionally makes up the highest relative concentration contaminant protein species. Triosephosphate isomerase is a high abundance constitutively expresses enzyme functionally localized to the branching step of glycolysis. While TIM is not classified as metabolically essential, it was determined that due to the punitive consequences with regard to cell growth and recombinant protein production, gene knock-out could not be considered as a viable option. Using bioinformatics sequence alignment and three dimentional protein structure tools, I was able to determine a logical alternative would be to use PCR mediated site-directed mutagenesis to replace the N-teminal IMAC binding motif, thus eliminating TIM from the ECP and TCP. Using a published enzymatic assay I was able to show that the engineered IMAC TIM (eTIM) enzyme was statistically identical with respect to specific activity. Furthermore, upon transformation into a strain lacking a function *tpiA* gene, the eTIM

plasmid was able to completely restore normal growth characteristics when compared to the *E.coli* BL21 (DE3) control.

15.0 Considerations for Future Work:

As described in Part II, the traditional approach to microbial cell line development would dictate independent mutagenesis techniques for each contaminant protein species, followed by growth and recombinant expression characterization to determine functionality. For the independent researcher this method would likely be considered both economic and time prohibitory, particularly when dealing with an extensive contaminant pool of more than 6-8 proteins. Therefore, I believe a logical extension of this project would be to implement the provided metallo-proteome data into a more functional tool.

Mapping the *E. coli* chromosome in such a way as to detail specific locations of essential vs. non-essential IMAC contaminant gene products would a productive first step. Due to the scope of information required, computer generated mapping would likely be the most practical method to comprehensively analyze my dataset included in Part I of this dissertation. Such a task is ideally suited for *in silico* mapping software platforms that allow for multiplex data output.

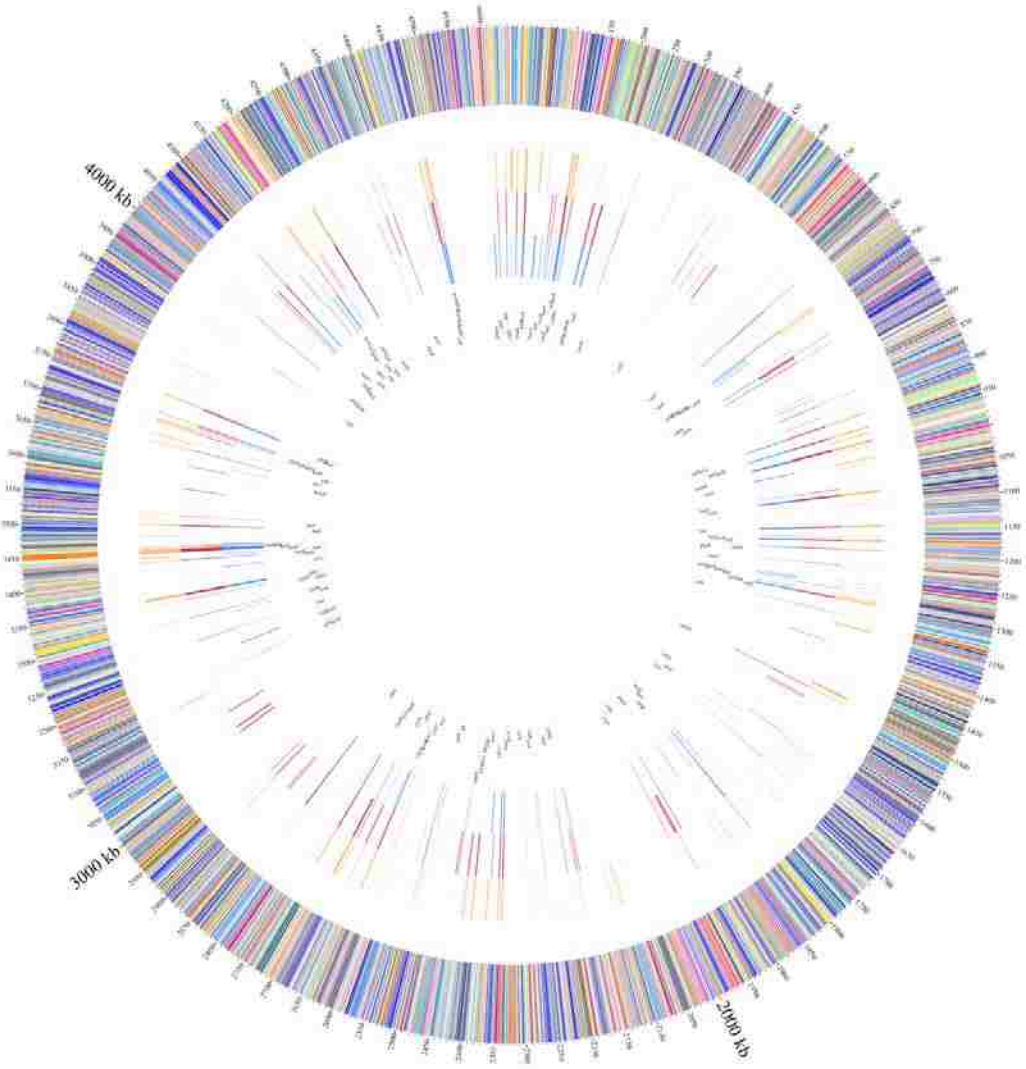
Ellen Brune, a current Ph.D. student in the Beitle Group, has recently done just this. By inputting the TCP and ECP data set from Part I of my work into the Circos software platform; she and colleagues' are currently in the process of demonstrating how it is possible to visualize each contaminant gene product relative to its respective location within the *E.coli* chromosome.

Indeed, this facet of the overall project is beginning to show the significant value of this data by allowing for a global perspective of how best to utilize proteome driven data toward recombinant cell line development.

As previously mentioned, one such tool for viewing this type data is the Circos software platform. Circos was first published in Genome Research in 2009 (Krzywinski et al., 2009). Originally designed as a comparative genomics tool, Circos was specifically intended to easily visualize genetic data across multiple species; again demonstrating the utility of using phylogenetic analysis tools toward the genetic engineering of specialized cell lines. Figure 1 shows an example of how Circos diagrams could be used to map the *E.coli* chromosome with the respect to TCP and ECP data[2]. The full *E.coli* chromosome is presented on the outer ring applying NCBI COG color coding to correlate gene function. The internal marks connect metalloproteome data for each metal ion used in this study. As an example, to illustrate the overall impact of this tool, my original data set, described in Part I of this dissertation comprised over one hundred printed pages. However, using Circos mapping future researchers will have the ability to analyze the same data set in a small fraction of the time.

[2] Adapted with permission from the Ph.D. proposal of Ellen Brune to whom full credit for this figure is due.

Figure 1: Circos plot of the *E. coli* chromosome and IMAC metalloproteome according TCP and ECP data provided in Part I.



The utility of the Circos plot lies in the streamlining of the cell line development process as a whole. As opposed to a gene specific knock-out (GSK) or the PCR derived site-directed mutagenesis techniques implemented in Part II, the Circos plot will allow future researchers to utilize large-scale knockout (LSK) techniques to oblate a larger number of contaminating protein species simultaneously (Osterman & Gerdes, 2008).

References

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. [10.1038/nature01511]. *Nature, 422*(6928), 198-207.

Ajay, S. S., Parker, S. C. J., Abaan, H. O., Fajardo, K. V. F., & Margulies, E. H. (2012). Accurate and comprehensive sequencing of personal genomes. *Genome Research*. doi: 10.1101/gr.123638.111

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci USA, 74*(12), 5350-5354.

Andersson, L., Sulkowski, E., & Porath, J. (1987). Facile Resolution of a-Fetoproteins and Serum Albumins by Immobilized Metal Affinity Chromatography. *Cancer Research, 47*, 3624-3626.

Aparicio, R., Ferreira, S. T., & Polikarpov, I. (2003). Closed Conformation of the Active Site Loop of Rabbit Muscle Triosephosphate Isomerase in the Absence of Substrate: Evidence of Conformational Heterogeneity. *J. Mol. Biol., 334*, 1023-1041. doi: 10.1016/j.jmb.2003.10.022

Arnau, J., Lauritzen, C., Petersen, G. E., & Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. [Comparative Study Review]. *Protein Expression and Purification, 48*(1), 1-13. doi: 10.1016/j.pep.2005.12.002

Augenicht, L. H., Taylor, J., Anderson, L., & Lipkin, M. (1991). Patterns of gene expression that characterize the colonic mucosa in patients at genetic risk for colonic cancer. *PNAS, 88*, 3286-3289. doi: doi:10.1073/pnas.88.8.3286

Augenlicht, L. H., & Kobrin, D. (1982). Cloning and Screening of Sequences Expressed in a Mouse Colon Tumor. *Cancer Research, 42*, 1088-1093.

Augenlicht, L. H., Wahrman, M. Z., Halsey, H., & al., e. (1987). Expression of Cloned Sequences in Biopsies of Human Colonic Tissue and in Colonic Carcinoma Cells Induced to Differentiate *In Vitro*. *Cancer Research, 47*, 6017-6021.

Beitle, R. R., & Ataai, M. M. (1993). One-Step Purification of a Model Periplasmic Protein from Inclusion Bodies by Its Fusion to an Effective Metal-Binding Peptide. *Biotechnol. Prog., 9*, 64-69.

Belyavsky, A., Vinogradova, T., & Rajewsky, K. (1989). PCR-based cDNA library construction: geneal cDNA libraries at the level of a few cells. *Nucleic Acids Research, 17*(8), 2919-2932.

Berners-Lee, T. (Ed.). (1999). *Weaving the Web: The Originalo Design and Ultimate Destiny of the World Wde Web by Its Inventor* (Vol. 1). San Francisco: Harper.

Blacklow, S. C., Liu, K. D., & Knowles, J. R. (1991). Stepwise Improvements in Catalytic Effectiveness: Independence and Interdependence in Combinations of Point Mutations of a Sluggish Triosephosphate Isomerase. *Biochemistry, 30*(34), 8470-8476.

Blattner, F. R., III, G. P., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., . . . Shao, Y. (1997). The Complete Genome Sequence of Escherichia coli K-12. *Science, 277*(5331), 1453-1462. doi: 10.1126/science.277.5331.1453

Bolanos-Garcia, V. M., & Davies, O. R. (2006). Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects, 1760*(9), 1304-1313.

Cai, Y., Moore, M., Goforth, R., Henry, R., & Beitle, R. (2004). Genomic data for alternate production strategies. I. Identification of major contaminating species for Cobalt(+2) immobilized metal affinity chromatography. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.]. *Biotechnology and bioengineering, 88*(1), 77-83. doi: 10.1002/bit.20212

Cai, Y., Moore, M., Goforth, R., Henry, R., & Beitle, R. (2004). Genomic Data for Alternate Production Strategies. I. Identification of Major Contaminating Species for $Cobalt^{+2}$ Immobilized Metal Affinity Chromatography. [Journal Article]. *Biotechnology and Bioengineering, 88*(1), 77-83. doi: 10.1002/bit.20212

Çakir, T., Kirdar, B., & Ülgen, K. Ö. (2004). Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. [10.1002/bit.20020]. *Biotechnology and bioengineering, 86*(3), 251-260.

Casal, J. I., Ahern, T. J., Davenport, R. C., Petsko, G. A., & Klibanov, A. M. (1987). Subunit Interface of Triosephosphate Isomerase: Site- Directed Mutagenesis and Characterization of the Altered Enzyme. *Biochemistry, 26*(5), 1258-1264.

Chaga, G. S. (2001). Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. *Journal of Biochemical and Biophysical Methods, 49*(1–3), 313-334. doi: 10.1016/s0165-022x(01)00206-8

Chakravarchy, B., & Pietenpol, J. A. (2003). Combined Modality Management of Breast Cancer: Development of Predictive Markers Through Proteomics. *Seminars in Oncology 30*(4), 23-36.

Chiou, S.-H., & Wu, S.-H. (1999). Evaluation of commonly used electrophoretic methods for the analysis of proteins and peptides and their application to biotechnology. *Analytica Chimica Acta, 383*, 47-60.

Choudhary, C., & Mann, M. (2010). Decoding signalling networks by mass spectrometry - based proteomics. *Nature Reviews: Molecular Cell Biology, 11*, 427-439.

Christopher Kirkpatrick, L. M. M., Nikki E. Oyelaken, Yuliya N. Yancheva, Russell Maurer, and Joan L. Slonczewski. (2001). Acetate and Formate Stress: Opposite Responses in the Proteome of Escherichia coli. *Journal of Bacteriology, 183*(21), 6466-6477.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., . . . Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. [Research Support, Non-U.S. Gov't]. *Nature methods, 5*(7), 613-619. doi: 10.1038/nmeth.1223

Curtis, K. M., Gomez, L. A., Rios, C., Garbayo, E., Raval, A. P., Perez-Pinzon, M. A., & Schiller, P. C. (2010). EF1alpha and RPL13a represent normalization genes suitable for RT-qPCR analysis of bone marrow derived mesenchymal stem cells. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Molecular Biology, 11*, 61. doi: 10.1186/1471-2199-11-61

de Groot, M. J., Daran-Lapujade, P., van Breukelen, B., Knijnenburg, T. A., de Hulster, E. A., Reinders, M. J., . . . Slijper, M. (2007). Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. [Research Support, Non-U.S. Gov't]. *Microbiology, 153*(Pt 11), 3864-3878. doi: 10.1099/mic.0.2007/009969-0

Drake, R. R., Cazares, L. H., Semmes, O. J., & Wadsworth, J. T. (2005). Serum, salivary and tissue proteomics for discovery of biomarkers for head and neck cancers. *Expert Review of Molecular Diagnostics, 5*(1), 93-100. doi: 10.1586/14737159.5.1.93

Efremenko, E., Votchitseva, Y., Plieva, F., Galaev, I., & Mattiasson, B. (2006). Purification of His6–organophosphate hydrolase using monolithic supermacroporous polyacrylamide cryogels developed for immobilized metal affinity chromatography. [Journal]. *Appl Microbiol Biotechnol, 70*, 558-563. doi: 10.1007/s00253-005-0103-x

Evans, W. E., & Relling, M. V. (2004). Moving towards individualized medicine with pharmacogenomics. *Nature, 429*, 464-468. doi: 10.1038/nature02626

Fey, S. J., & Larsen, P. M. (2001). 2D or not 2D. *Current Opinion in Chemical Biology, 5*(1), 26-33. doi: 10.1016/s1367-5931(00)00167-8

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., . . . al, e. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. . *Science, 269*(5523), 496-512.

Fujiwara, M., Yan, P., Otsuji, T. G., Narazaki, G., Uosaki, H., Fukushima, H., . . . Yamashita, J. K. (2011). Induction and enhancement of cardiac cell differentiation from mouse and human induced pluripotent stem cells with cyclosporin-A. [Evaluation Studies

Research Support, Non-U.S. Gov't]. *PloS one, 6*(2), e16734. doi: 10.1371/journal.pone.0016734

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

Gaberc-Porekar, V., & Menart, V. (2001b). Perspectives of immobilized-metal affinity chromatography. *j. Biochm. Biophys. Methods, 49*, 335-360.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., . . . Bryant, S. H. (2010). The NCBI BioSystems database. [Research Support, N.I.H., Intramural]. *Nucleic Acids Research, 38*(Database issue), D492-496. doi: 10.1093/nar/gkp858

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., . . . Osterman, A. L. (2003). Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655. *Journal of Bacteriology, 185*(19), 5673-5684. doi: 10.1128/jb.185.19.5673-5684.2003

Giometti, C. S. (2003). Proteomics and Bioinformatics. In D. S. Richard & D. V. Timothy (Eds.), *Advances in Protein Chemistry* (Vol. Volume 65, pp. 353-369): Academic Press.

Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., & Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Proceedings of the National Academy of Sciences of the United States of America, 97*(17), 9390-9395. doi: 10.1073/pnas.160270797

Ha, E.-J., Kim, Y.-J., An, S. S. A., Kim, Y.-R., Lee, J.-O., Lee, S.-G., & Paik, H.-j. (2008). Purification of His-tagged proteins using Ni^{2+} -poly(2-acetamidoacrylic acid) hydrogel. [Journal]. *Journal of Chromatography B, 876*, 8-12. doi: 10.1016/jjchromb.2008.10.020

Haddadin, F. T., & Harcum, S. W. (2005). Transcriptome profiles for high-cell-density recombinant and wild-type Escherichia coli. [Research Support, U.S. Gov't, Non-P.H.S.]. *Biotechnology and bioengineering, 90*(2), 127-153. doi: 10.1002/bit.20340

Han, M. J., & Lee, S. Y. (2006). The Escherichia coli proteome: past, present, and future prospects. [Historical Article Research Support, Non-U.S. Gov't Review]. *Microbiology and molecular biology reviews : MMBR, 70*(2), 362-439. doi: 10.1128/MMBR.00036-05

Hengen, P. N. (1995). Purification of His-Tag fusion proteins from Escherichia coli. *Trends in Biochemical Sciences, 20*(7), 285-286. doi: 10.1016/s0968-0004(00)89045-3

Hermann, B. P., Sukhwani, M., Simorangkir, D. R., Chu, T., Plant, T. M., & Orwig, K. E. (2009). Molecular dissection of the male germ cell lineage identifies putative spermatogonial stem cells in rhesus macaques. [Research Support, N.I.H., Extramural]. *Human reproduction, 24*(7), 1704-1716. doi: 0.1093/humrep/dep073

Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., . . . Saito, K. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. [Research Support, Non-U.S. Gov't]. *The Journal of Biological Chemistry, 280*(27), 25590-25595. doi: 10.1074/jbc.M502332200

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., . . . Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. [Research Support, Non-U.S. Gov't]. *Proceedings of the National Academy of Sciences of the United States of America, 101*(27), 10205-10210. doi: 10.1073/pnas.0403218101

Holt, R. A., & Jones, S. J. (2008). The new paradigm of flow cell sequencing. [Research Support, Non-U.S. Gov'tReview]. *Genome research, 18*(6), 839-846. doi: 10.1101/gr.073262.107

Hoog, C. L. d., & Mann, M. (2004). Proteomics. *Annu. Rev. Genomics Hum. Genet, 5*.

Hubert, P., & Porath, J. (1980). Metal chelate affinity chromatography : I. Influence of various parameters on the retention of nucleotides and related compounds. *Journal of Chromatography A, 198*(3), 247-255. doi: 10.1016/s0021-9673(00)84764-0

Hutchinson, M. H., & Chase, H. A. (2006). Intensified Process for the Purification of an Enzyme from Inclusion Bodies Using Integrated Expanded Bed Adsorption and Refolding. *Biotechnol. Prog., 22*(4), 1187-1193. doi: 10.1021/bp060055q

Hutchinson, M. H., Morreale, G., Middelberg, A. P. J., & Chase, H. A. (2006). Production of Enzymatically Active Ketosteroid Isomerase following Insoluble Expression in Escherichia coli. *Biotechnology and Bioengineering, 95*(4), 724- 733. doi: 10.1002/bit.21043

Jain, A. K., Allton, K., Iacovino, M., Mahen, E., Milczarek, R. J., Zwaka, T. P., . . . Barton, M. C. (2012). p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *PLoS biology, 10*(2), e1001268. doi: 10.1371/journal.pbio.1001268

Jeanty, C., Longrois, D., Mertes, P. M., Wagner, D. R., & Devaux, Y. (2010). An optimized protocol for microarray validation by quantitative PCR using amplified amino allyl labeled RNA. [Research Support, Non-U.S. Gov't Validation Studies]. *BMC genomics, 11*, 542. doi: 10.1186/1471-2164-11-542

Kempf, J. G., Jung, J.-y., Ragain, C., Sampson, N. S., & Loria, J. P. (2007). Dynamic Requirements for a Functional Protein Hinge. *J. Mol. Biol., 368*, 131-149. doi: 10.1016./j.jmb.2007.01.074

Kiechle, F. L., Zhang, Z., & Holland-Staley, C. A. (2004). The -omics Era and Its Impact. *Arch Pathol Lab Med., 128*, 1337-1345.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. [Research Support, Non-U.S. Gov't]. *Genome research, 19*(9), 1639-1645. doi: 10.1101/gr.092759.109

Kumar, J. K., Tabor, S., & Richardson, C. C. (2004). Proteomic analysis of thioredoxin-targeted proteins in Escherichia coli. [Research Support, U.S. Gov't, Non-P.H.S.

Research Support, U.S. Gov't, P.H.S.]. *Proceedings of the National Academy of Sciences of the United States of America, 101*(11), 3759-3764. doi: 10.1073/pnas.0308701101

Lasserre, J. P., Beyne, E., Pyndiah, S., Lapaillerie, D., Claverol, S., & Bonneu, M. (2006). A complexomic study of Escherichia coli using two-dimensional blue native/SDS polyacrylamide gel electrophoresis. [Research Support, Non-U.S. Gov't]. *Electrophoresis, 27*(16), 3306-3321. doi: 10.1002/elps.200500912

Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., . . . Venter, J. C. (2007). The Diploid Genome Sequence of an Individual Human. *PLoS Biol, 5*(10), e254. doi: 10.1371/journal.pbio.0050254

Li, Z., Yang, C. S., Nakashima, K., & Rana, T. M. (2011). Small RNA-mediated regulation of iPS cell generation. *The EMBO journal, 30*(5), 823-834. doi: 10.1038/emboj.2011.2

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification, 41*(1), 98-105. doi: 10.1016/j.pep.2005.01.019

Lindgren, G. E. (1994). Immobilized metal affinity chromatography (IMAC). *Am. Biotechnol. Lab., 12*(7), 36.

Liu, Q., Lin, J., Liu, M., Tao, X., Wei, D., Ma, X., & Yang, S. (2007). Large scale preparation of recombinant human parathyroid hormone 1-84 from *Escherichia coli*. *Protein Expression and Purification, 54*, 212-219. doi: 10.1016/j.pep.2007.03.009

Lodi, P. J., & Knowles, J. R. (1993). Direct Evidence for the Exploitation of an alpha- Helix in the Catalytic Mechanism of Triosephosphate Isomerase. *Biochemistry, 32*(16), 4338-4343.

Maskos, U., & Sothern, E. M. (1992). Ologonucleotide hybridisations on glass supports: a novel linker for oligonucleotide synthesis and hybridisation properties of oligonucleotides synthesised in situ. *Nucleic Acids Research, 20*(7), 1679-1684.

McCluskey, A. J., Poon, G. M. K., & Gariepy, J. (2007). A rapid and universal tandem-purification strategy for recombinant proteins. *Protein Science, 16*, 2726- 2732.

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., . . . Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. [Research Support, Non-U.S. Gov't]. *BioTechniques, 45*(1), 81-94. doi: 10.2144/000112900

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Nature methods, 5*(7), 621-628. doi: 10.1038/nmeth.1226

N. Abdullah, & Chase, H. A. (2005). Removal of Poly-Histidine Fusion Tags From Recombinant Proteins Purified by Expanded Bed Adsorption. [Journal]. *Biotechnology and Bioengineering, 92*(4), 501-513. doi: 10.1002/bit.20633

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Science, 320*(5881), 1344-1349. doi: 10.1126/science.1158441

Nickbarg, E. B., Davenport, R. C., Petsko, G. A., & Knowles, J. R. (1988). Triosephosphate Isomerase: Removal of a Putatively Electrophilic Histidine Residue Results in a Subtle Change in Catalytic Mechanism. *Biochemistry, 27*(16), 5948-5960.

Nickbarg, E. B., & Knowles, J. R. (1988). Triosphosphate Isomerase: Energetics of the Reaction Catalysed by the Yeast Enzyme Expressed in *Escherichia coli*. *Biochemistry, 27*, 5939-5947.

Nystrom, J., Fierlbeck, W., Granqvist, A., Kulak, S. C., & Ballermann, B. J. (2009). A human glomerular SAGE transcriptome database. [Research Support, Non-U.S. Gov't]. *BMC nephrology, 10*, 13. doi: 10.1186/1471-2369-10-13

O'Brien, S. J., Womack, J. E., Lyons, L. A., Moore, K. J., Jenkins, N. A., & Copeland, N. G. (1993). Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics, 3*, 103-112. doi: 10.1038/ng0293-103

Osterman, A. L., & Gerdes, S. Y. (Eds.). (2008). *Methods in Molecular Biology* (Vol. 416). Totowa, NJ: Humana Press Inc.

Patterson, A. D., Li, H., Eichler, G. S., Krausz, K. W., Weinstein, J. N., Jr., A. J. F., . . . Idle, J. R. (2008). UPLC-ESI-TOF MS-Based Metabolomics and Gene Expression Dynamics Inspector Self Organizing Metabolomic Maps as Tools for Understanding the Cellular Response to Ionizing Radiation. *Anal Chem., 80*(3), 665-674.

Pecson, B. M., Martin, L. V., & Kohn, T. (2009). Quantitative PCR for determining the infectivity of bacteriophage MS2 upon inactivation by heat, UV-B radiation, and singlet oxygen: advantages and limitations of an enzymatic treatment to reduce false-positive results. [Research Support, Non-U.S. Gov't]. *Applied and environmental microbiology, 75*(17), 5544-5554. doi: 10.1128/AEM.00425-09

Pompliano, D. L., Peyman, A., & Knowles, J. R. (1990). Stabilization of a Reaction Intermediate as a Catalytic Device: Definition of the Functional Role of the Flexible Loop in Triosephosphate Isomerase. *Biochemistry, 29*(13), 3186-3194.

Porath, J., Carlsson, J., Olsson, I., & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature, 258*(5536), 598-599.

Putman, S. J., Coulson, A. F. W., Farley, I. R. T., Riddleston, B., & Knowles, J. R. (1972). Specificity and Kinetics of Triose Phosphate Isomerase from Chicken Muscle. *Biochem. J. , 129*, 301-310.

Raines, R. T., Sutton, E. L., Straus, D. R., Gilbert, W., & Knowles, J. R. (1986). Reaction Energetics of a Mutant Triosephosphate Isomerase in Which the Active- Site Glutamate Has Been Changed to Aspartate. *Biochemistry, 25*(22), 7142-7154.

Research, U. G. (2010). Biopharmaceuticals: A Global Market Overview *Biopharmaceuticals* (Vol. 1, pp. 240). London, England.

Rossetto, O., Schiavo, G., Laureto, P. P. d., Fabbiani, S., & Montecucco, C. (1992). Surface topography of histidine residues of tetanus toxin probed by immobilized metal ion affinity chromatography. *Biochem. J. , 285*, 9-12.

Sa-Correia, I., & Teixeira, M. C. (2010). 2D electrophoresis - based expression proteomics: a mircrobiologist's perspective. *Expert Rev. Proteomics, 7*(6), 943-953.

Schlamp, K., Weinmann, A., Krupp, M., Maass, T., Galle, P., & Teufel, A. (2008). BlotBase: a northern blot database. *Gene, 427*(1-2), 47-50.

Schliebs, W., Thanki, N., Jaenicke, R., & Wierenga, R. K. (1997). A Double Mutation at the Tip of the Dimer Interface Loop of Triosephosphate Isomerase Generates Active Monomers with Reduced Stability. *Biochemistry, 36*(32), 9655-9662.

Shi, Y., Xiang, R., Horváth, C., & Wilkins, J. A. (2004). The role of liquid chromatography in proteomics. *Journal of Chromatography A, 1053*(1-2), 27-36. doi: 10.1016/j.chroma.2004.07.044

Smith, M. C., Furman, T. C., Ingolia, T. D., & Pidgeon, C. (1988). Chelating Peptide-immobiliaed Metal Ion Affinity Chromatography: A new concept in affinity chromatography for recombinant proteins. *The Journal of Biological Chemistry, 263*(15), 7211-7215.

Smith, S. D., She, Y.-M., Roberts, E. A., & Sarkar, B. (2004). Using Immobilized Metal Affinity Chromatography, Two-Dimensional Electrophoresis and Mass Spectometry to Identify Hepatocellular Proteins with Copper-Binding Ability. *Journal of Proteome Research, 3*, 824-840.

Stasyk, T., & Huber, L. A. (2004). Zooming in: Fractionation strategies in proteomics. *Proteomics, 4*(12), 3704-3716. doi: 10.1002/pmic.200401048

Stensballe, A., Andersen, S., & Jensen, O. N. (2001). Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. [10.1002/1615-9861(200102)1:2<207::AID-PROT207>3.0.CO;2-3]. *Proteomics, 1*(2), 207-222.

Straus, D., Raines, R., Kawashima, E., Knowles, J. R., & Gilbert, W. (1985). Active site of triosephosphate isomerase: *In vitro* mutagenesis and characterization of an altered enzyme. *Proc. Natl. Acad. Sci. USA, 82*, 2272-2276.

Sun, X., Chiu, J.-F., & He, Q.-Y. (2005). Application of immobilized metal affinity chromatography in proteomics. *Expert Reviews of Proteomics, 2*(5), 649-657.

Tan, S., Kern, R. C., & Selleck, W. (2005). The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expression and Purification, 40*(2), 385-395. doi: 10.1016/j.pep.2004.12.002

Ueda, E. K. M., Gout, P. W., & Morganti, L. (2003). Current and prospective applications of metal ion–protein binding. *Journal of Chromatography A, 988*(1), 1-23. doi: 10.1016/s0021-9673(02)02057-5

Varakala, R. (2008). *Proteome Based Improvements For Immobilized Metal Affinity Chromatography in Esherichia coli Expression System.* Ph.D. Dissertation, University of Arkansas, Fayetteville.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kenzler, K. W. (1995). Serial analysis of gene expression. *Science, 270*, 484-487.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Zhu, X. (2007). The sequence of the human genome. [Research Support, Non-U.S. Gov't]. *Science, 291*(5507), 1304-1351. doi: 10.1126/science.1058040

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. [Review]. *Nature reviews. Genetics, 10*(1), 57-63. doi: 10.1038/nrg2484

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., . . . Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. [10.1038/nature07002]. *Nature, 453*(7199), 1239-1243. doi: http://www.nature.com/nature/journal/v453/n7199/suppinfo/nature07002_S1.html

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., . . . Hochstrasser, D. F. (1996). From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. [10.1038/nbt0196-61]. *Nat Biotech, 14*(1), 61-65.

Wilkins, M. R., Sanchex, J. C., Gooley, A. A., & *al., e.* (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev., 13*, 19-50.

Winn, M. E., Shaw, M., April, C., Klotzle, B., Fan, J. B., Murray, S. S., & Schork, N. J. (2011). Gene expression profiling of human whole blood samples with the Illumina WG-DASL assay. [Comparative StudyResearch Support, N.I.H., ExtramuralResearch Support, Non-U.S. Gov't]. *BMC genomics, 12*, 412. doi: 10.1186/1471-2164-12-412

Wu, Z. J., Meyer, C. A., Choudhury, S., Shipitsin, M., Maruyama, R., Bessarabova, M., . . . Liu, X. S. (2010). Gene expression profiling of human breast tissue samples using SAGE-Seq. [Comparative StudyResearch Support, N.I.H., ExtramuralResearch Support, Non-U.S. Gov't]. *Genome research, 20*(12), 1730-1739. doi: 10.1101/gr.108217.110

Yip, T.-T., Nakagawa, Y., & Porath, J. (1989). Evaluation of the interaction of peptides with Cu(II), Ni(II), and Zn(II) by high-performance immobilized metal ion affinity

chromatography. *Analytical Biochemistry*, *183*(1), 159-171. doi: 10.1016/0003-2697(89)90184-x

Appendix I: Compilation of Contaminant Species

Appendix I list all elution contaminant species observed for each of the three metals investigated in this study ($Co^{(II)}$, $Ni^{(II)}$, and $Zn^{(II)}$). Contaminant identification was obtained through tandem mass spectrometry of individual column elutions. Mascot™ data mining was used to associate peptide mass fingerprints to gene products from the *E. coli* proteome. Resulting proteins showing greater than 95 percent confidence level were archived as IMAC elution contaminant species.

ABBREVIATIONS	Database Origin
ref	National Institute of Standards and Technology (NIST)
gb	National Center for Biotechnology Information (NCBI)
emb	European Bioinformatics Institute (EMBL-EBI)
pdb	Protein Data Bank
dbj	DNA Data Bank of Japan

To maximize utility of the document, redundancies in the mass spectral output were removed for clarity. Disagreements with contaminant protein numbers found in Appendix II are attributed to uncharacterized/ unnamed proteins within the dataset. Unedited mass spectrometry data has been archived in the laboratory of Dr. Robert Beitle, Ph.D., P.E.

Contaminants Observed in $Co^{(II)}$ IMAC only

Reference	gi Number	Database	File ID	Peptide
1	91213796	ref	YP_543782.1	Mg2+-importing ATPase, P-type 1 (MgtA)
2	91214109	ref	YP_544095.1	ABC transporter ATP-binding protein YjjK
3	75208804	ref	ZP_00702003.1	AICAR transformylase/IMP cyclohydrolase PurH
4	26246587	ref	NP_752626.1	Alkyl hydroperoxide reductase subunit F
5	537084	gb	AAA97139.1	alternate gene name mgt; CG Site No. 497
6	75234553	ref	ZP_00718894.1	ATPase components of ABC transporters with duplicated ATPase
7	147365	gb	AAA24422.1	ATP-dependent protease binding subunit
8	216538	dbj	BAA03143.1	BasR
9	91212788	ref	ZP_00702003.1	bifunctional AICAR formyltransferase/IMP cyclohydrolase
10	15834183	ref	NP_312956.1	bifunctional phosphoribosylaminoimidazole-carboxamide formyltransferase
11	75207853	ref	ZP_00708331.1	Biotin carboxylase
12	26245956	ref	NP_751995.1	carbamoyl-phosphate synthase; large subunit
13	75229613	ref	ZP_00716152.1	Cation transport ATPase
14	13786943	pdb	1FL2	Chain A, Catalytic Component of Alkylhydroperoxide Reductase (Ahpf F)
15	1942723	pdb	1EFU	Chain C, Elongation Factor Complex Ef-TuEF-Ts
16	91210463	ref	YP_540449.1	CoA-linked acetaldehyde dehydrogenase
17	91209194	ref	YP_539180.1	conserved hypothetical protein YadF
18	1788586	gb	AAC75312.1	conserved protein
19	1786291	gb	AAC73213.1	conserved protein
20	87082081	gb	AAC75354.2	conserved protein
21	91210106	ref	YP_540092.1|	delta-1-pyrroline-5-carboxylate dehydrogenase
22	26246130	ref	NP_752169.1	DNA polymerase III subunit alpha
23	1790552	gb	AAC77074.1	DNA-binding response regulator in two-component regulatory system
24	1788589	gb	AAC75315.1	fused UDP-L-Ara4N formyltransferase/UDP-GlcA C-4'-decarboxylase
25	75228928	ref	ZP_00715520.1	Glucan phosphorylase
26	91209331	ref	YP_539317.1	glutamate 5-kinase

27	15833345	ref	NP_312118.1	glutamate synthase large subunit
28	75255300	ref	ZP_00727133.1	Glyceraldehyde-3-phosphate dehydrogenase
29	4377499	emb	CAA26035.1	glycogen phosphorylase
30	26248643	ref	NP_754683.1	hypothetical protein c2797
31	26248681	ref	NP_754721.1	hypothetical protein c2835
32	15832397	ref	NP_311170.1	hypothetical protein ECs3143
33	15832432	ref	NP_311205.1	hypothetical protein ECs3178
34	91211549	ref	YP_541535.1	hypothetical protein YfbG
35	75515848	ref	ZP_00737983.1	IMP dehydrogenase/GMP reductase
36	15833513	ref	NP_312286.1	maltodextrin phosphorylase
37	75190291	ref	ZP_00703558.1	NAD-dependent aldehyde dehydrogenases
38	75512755	ref	ZP_00731043.1	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD sub
39	1788619	gb	ZP_00731043.1	NADH:ubiquinone oxidoreductase, chain G
40	75259718	ref	ZP_00731020.1	Nucleoside-diphosphate-sugar epimerases
41	56540641	gb	AAV92774.1	PmrA
42	15830514	ref	NP_309287.1	proline dehydrogenase
43	1788943	gb	AAC75641.1	protein disaggregation chaperone
44	91211588	ref	YP_541574.1	protein YfbU
45	15829384	ref	NP_308157.1	putative carbonic anhdrase
46	91210003	ref	YP_539989.1	pyruvate formate lyase activating enzyme 1
47	91212872	ref	YP_542858.1	ribosome-associated heat shock protein 15
48	15832671	ref	NP_311444.1	serine hydroxymethyltransferase
49	15831680	ref	NP_310453.1	threonyl-tRNA synthetase
50	91212813	ref	YP_542799.1	transcription antitermination protein
51	15830746	ref	NP_309519.1	transcription-repair coupling factor
52	75513895	ref	ZP_00736248.1	Triosephosphate isomerase
53	75189611	ref	ZP_00702878.1	Uncharacterized conserved protein
54	75227376	ref	ZP_00714154.1	Uncharacterized conserved protein
55	75187076	ref	ZP_00700343.1	Uncharacterized conserved protein
56	41964	emb	CAA29949.1	unnamed protein product
57	41627	emb	CAA26133.1	unnamed protein product
58	43066	emb	CAA23560.1	unnamed protein product
59	466691	gb	AAB18530.1	unnamed protein product
60	42497	emb	CAA25363.1	unnamed protein product
61	499181	emb	CAA25357.1	unnamed protein product
62	42821	emb	CAA23627.1	unnamed protein product

63	42773	emb	CAA38206.1	unnamed protein product
64	41676	emb	CAA68776.1	unnamed protein product
65	43112	emb	CAA25253.1	unnamed protein product
66	42284	emb	CAA23597.1	unnamed protein product
67	41039	emb	CAA23528.1	unnamed protein product
68	581130	emb	CAA23584.1	unnamed protein product

Contaminants Observed in $Ni^{(II)}$ IMAC only

Reference	gi Number	NCBI Database	File ID	Peptide
1	15832855	ref	NP_311628.1	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
2	75228118	ref	ZP_00714789.1	2-methylthioadenine synthetase
3	91209238	ref	YP_539224.1	30S ribosomal protein S2
4	75258319	ref	ZP_00729750.1	3'-Phosphoadenosine 5'-phosphosulfate (PAPS) 3'-phosphatase
5	26246419	ref	NP_752458.1	5-amino-6-(5-phosphoribosylamino) uracil reductase
6	56199744	gb	AAV84339.1	60 kDa chaperonin
7	91214109	ref	YP_544095.1	ABC transporter ATP-binding protein YjjK
8	91212678	ref	YP_542664.1	Acetyl CoA carboxylase; biotin carboxylase subunit
9	75188736	ref	ZP_00702003.1	AICAR transformylase/IMP cyclohydrolase (PurH)
10	15832808	ref	NP_311581.1	alanyl-tRNA synthetase
11	26246587	ref	NP_752626.1	Alkyl hydroperoxide reductase; subunit F
12	75189732	ref	ZP_00702999.1	Allophanate hydrolase subunit 1
13	15834446	ref	NP_313219.1	ammonium transport system structural protein
14	15834469	ref	NP_313242.1	anaerobic ribonucleoside triphosphate reductase
15	75238135	ref	ZP_00722139.1	Anthranilate phosphoribosyltransferase
16	15833068	ref	NP_311841.1	arginine decarboxylase
17	75235446	ref	ZP_00719654.1	Aspartate carbamoyltransferase
18	15832652	ref	NP_311425.1	ATP synthase; beta subunit
19	15833927	ref	NP_312700.1	ATP synthase; subunit epsilon
20	75259275	ref	ZP_00730623.1	ATPases involved in chromosome partitioning
21	83585826	ref	ZP_00924466.1	ATPases with chaperone activity; ATP-binding subunit
22	91209919	ref	YP_539905.1	ATP-dependent clp protease ATP-binding subunit clpA
23	41114	emb	CAA40846.1	ATP-dependent protease regulatory subunit
24	26250788	ref	NP_756828.1	B12-dependent methionine synthase
25	216538	dbj	BAA03143.1	BasR
26	26250778	ref	NP_756818.1	bifunctional phosphoribosyl-aminoimidazolecarboxamide formyltransferase

27	75512769	ref	ZP_00735271.1	BioD-like N-terminal domain of phosphotransacetylase
28	75254956	ref	ZP_00726900.1	Biotin carboxylase
29	26245956	ref	NP_751995.1	Carbamoyl-phosphate synthase; large subunit
30	75512001	ref	ZP_00734596.1	Carbonic anhydrase
31	75242370	ref	ZP_00726132.1	Cation transport ATPase
32	75239112	ref	ZP_00723092.1	Chaperonin GroEL (HSP60 family)
33	26248955	ref	NP_754995.1	ClpB protein
34	91210463	ref	YP_540449.1	CoA-linked Acetaldehyde dehydrogenase
35	83585801	ref	ZP_00924442.1	Coenzyme F420-dependent N5
36	26246951	ref	NP_752991.1	Condesin; subunit B
37	58176674	pdb	1RWU	Conserved Protein Ybed; A Chain A
38	26250631	ref	NP_756671.1	coproporphyrinogen III oxidase
39	91213772	ref	YP_543758.1	CysQ protein
40	91211857	ref	YP_541843.1	Cysteine desulfurase
41	87081779	gb	AAC73827.2	cytochrome d terminal oxidase; subunit I
42	91212281	ref	YP_542267.1	Decarboxylating glycine dehydrogenase
43	75230173	ref	ZP_00716674.1	Dehydrogenases with different specificities
44	26247036	ref	NP_753076.1	Delta-1-pyrroline-5-carboxylate dehydrogenase
45	15833925	ref	NP_312698.1	D-fructose-6-phosphate amidotransferase
46	15829373	ref	NP_308146.1	dihydrolipoamide acetyltransferase
47	91209184	ref	YP_539170.1	dihydrolipoamide dehydrogenase
48	75511822	ref	ZP_00734440.1	Dihydroxyacid dehydratase/phosphogluconate dehydratase
49	91213221	ref	YP_543207.1	DNA gyrase subunit B
50	91213223	ref	YP_543209.1	DNA polymerase III; beta-subunit
51	15829440	ref	NP_308213.1	DNA polymerase III; subunit alpha
52	33357282	pdb	1L8I	DNA Protection And Binding; Dps Protein; L Chain L
53	26246788	ref	NP_752828.1	DNA protection during starvation conditions
54	75230656	ref	ZP_00717127.1	DNA-binding ferritin-like protein (oxidative damage protection)
55	26248709	ref	NP_754749.1	erythronate-4-phosphate dehydrogenase
56	15832458	ref	NP_311231.1	Erythronate-4-phosphate dehydrogenase
57	91211380	ref	YP_541366.1	Galactitol-Specific Enzyme IIA of Phosphotransferase System
58	14278248	pdb	1FS0	Gamma/ EPSILON ATP SYNTHASE; E Chain E

59	75237019	ref	ZP_00721075.1	Glucan phosphorylase
60	75242483	ref	ZP_00726227.1	Glucosamine 6-phosphate synthetase
61	91209331	ref	YP_539317.1	glutamate 5-kinase
62	26249799	ref	NP_755839.1	Glutamate synthase; [NADPH] large chain precursor
63	15832532	ref	NP_311305.1	glutamyl-tRNA synthetase
64	91210996	ref	YP_540982.1	glyceraldehyde-3-phosphate dehydrogenase A
65	26249318	ref	NP_755358.1	Glycine dehydrogenase
66	75515689	ref	ZP_00737824.1	Glycine/D-amino acid oxidases (deaminating)
67	4377499	emb	CAA26035.1	glycogen phosphorylase
68	75235452	ref	ZP_00719660.1	Glycosidase
69	1943396	pdb	1GTP	Gtp Cyclohydrolase I; T Chain T
70	91211926	ref	YP_541912.1	heat shock protein
71	91210562	ref	YP_540548.1	hypothetical oxidoreductase YciK
72	26248681	ref	NP_754721.1	hypothetical protein c2835
73	26248712	ref	NP_754752.1	hypothetical protein c2870
74	75210279	ref	ZP_00710441.1	hypothetical protein EcolB_01002524
75	75255649	ref	ZP_00727421.1	hypothetical protein EcolE2_01004299
76	15829900	ref	NP_308673.1	hypothetical protein ECs0646
77	15830132	ref	NP_308905.1	hypothetical protein ECs0878
78	15830926	ref	NP_309699.1	hypothetical protein ECs1672
79	15831647	ref	NP_310420.1	hypothetical protein ECs2393
80	15832432	ref	NP_311205.1	hypothetical protein ECs3178
81	38704075	ref	NP_311235.2	hypothetical protein ECs3208
82	15833295	ref	NP_312068.1	hypothetical protein ECs4041
83	91209704	ref	YP_539690.1	hypothetical protein UTI89_C0659
84	91211373	ref	YP_541359.1	hypothetical protein UTI89_C2360
85	91212578	ref	YP_542564.1	hypothetical protein UTI89_C3587
86	91213596	ref	YP_543582.1	hypothetical protein UTI89_C4643
87	91209194	ref	YP_539180.1	hypothetical protein YadF
88	26246680	ref	NP_752720.1	Hypothetical protein ybgJ
89	91209834	ref	YP_539820.1	hypothetical protein YbiB
90	91210263	ref	YP_540249.1	hypothetical protein YcfP
91	91210899	ref	YP_540885.1	hypothetical protein YdiI
92	26248098	ref	NP_754138.1	Hypothetical protein yebU
93	91211621	ref	YP_541607.1	hypothetical protein YfcK
94	91213273	ref	YP_543259.1	hypothetical protein YieN
95	26250492	ref	NP_756532.1	Hypothetical protein yieN
96	91212533	ref	YP_542519.1	hypothetical transcriptional regulator YhaJ
97	91211831	ref	YP_541817.1	inosine-5'-monophosphate dehydrogenase

98	91211860	ref	YP_541846.1	inositol monophosphatase
99	15833016	ref	NP_311789.1	lysyl-tRNA synthetase
100	91213341	ref	YP_543327.1	Magnesium, Cobalt, and Nickel Transporter
101	15833513	ref	NP_312286.1	Maltodextrin phosphorylase
102	91213796	ref	YP_543782.1	Mg2+-importing ATPase (MgtA); P-type 1
103	75227515	ref	ZP_00714276.1	MoxR-like ATPases
104	38704071	ref	NP_311194.2	NADH dehydrogenase; gamma subunit
105	91210444	ref	YP_540430.1	NarL transcriptional dual regulator
106	48429400	gb	AAT42468.1	NrdA
107	75235453	ref	ZP_00719661.1	Oxygen-sensitive ribonucleoside-triphosphate reductase
108	147130	gb	AAA24310.1	PdxB protein
109	26249306	ref	NP_755346.1	Peptide chain release factor 2
110	91212840	ref	YP_542826.1	peptidyl-prolyl cis-trans isomerase A (rotamase A)
111	26250714	ref	NP_756754.1	Peroxidase/catalase HPI
112	15832435	ref	NP_311208.1	phosphate acetyltransferase
113	91211075	ref	YP_541061.1	phosphogluconate dehydratase
114	75187102	ref	ZP_00700369.1	Phosphoglycerate dehydrogenase
115	26248921	ref	NP_754961.1	phosphoribosylformyl-glycinamidine synthase
116	75236283	ref	ZP_00720392.1	Phosphotransferase system mannitol/fructose-specific IIA do
117	15829423	ref	NP_308196.1	PII uridylyl-transferase
118	56540641	gb	AAV92774.1	PmrA
119	75188498	ref	ZP_00701765.1	Predicted S-adenosylmethionine-dependent methyltransferase
120	15830514	ref	NP_309287.1	proline dehydrogenase
121	56553868	pdb	1TJ2	Proline Dehydrogenase Domain; A Chain A
122	75230227	ref	ZP_00716728.1	Protein chain release factor A
123	91210391	ref	YP_540377.1	protein YcgK precursor
124	91211588	ref	YP_541574.1	protein YfbU
125	91212449	ref	YP_542435.1	protein YgiN
126	38704200	ref	NP_312715.2	putative 2-component regulator
127	15829630	ref	NP_308403.1	putative ATP binding protein of a transporter system
128	91211400	ref	YP_541386.1	putative ATPase
129	15829384	ref	NP_308157.1	putative carbonic anhdrase
130	15831283	ref	NP_310056.1	putative glycoprotein
131	91213267	ref	YP_543253.1	putative methyltransferase GidB
132	13362678	dbj	BAB36631.1	putative peptidase

133	26248459	ref	NP_754499.1	Putative protease yegQ
134	15833543	ref	NP_312316.1	putative regulator
135	15831799	ref	NP_310572.1	putative rRNA methylase
136	15829721	ref	NP_308494.1	Pyrimidine deaminase
137	91209182	ref	YP_539168.1	pyruvate dehydrogenase E1 component
138	91210003	ref	YP_539989.1	pyruvate formate lyase activating enzyme 1
139	15831818	ref	NP_310591.1	pyruvate kinase
140	75511991	ref	ZP_00734586.1	Pyruvate/2-oxoglutarate dehydrogenase complex
141	75512959	ref	ZP_00735440.1	Response regulators consisting of a CheY-like receiver domain
142	15831568	ref	NP_310341.1	response transcriptional regulatory protein RstB
143	26247224	ref	NP_753264.1	Ribonuclease E
144	75512710	ref	ZP_00735212.1	Ribonucleotide reductase; alpha subunit
145	75258017	ref	ZP_00729491.1	Ribosome-associated protein Y (PSrp-1)
146	91212806	ref	YP_542792.1	RNA polymerase beta subunit
147	75240443	ref	ZP_00724377.1	rRNA methylase
148	15832671	ref	NP_311444.1	serine hydroxymethyltransferase
149	91209847	ref	YP_539833.1	stationary phase nucleoid protein Dps
150	15832860	ref	NP_311633.1	sulfate adenylyltransferase subunit 2
151	75210206	ref	ZP_00710374.1	Threonyl-tRNA synthetase
152	91212813	ref	YP_542799.1	transcription anti-termination protein
153	91213304	ref	YP_543290.1	transcription termination factor Rho
154	91210819	ref	YP_540805.1	transcriptional regulatory protein RstA
155	15830746	ref	NP_309519.1	Transcription-repair coupling factor (superfamily II helicase)
156	91212815	ref	YP_542801.1	translation elongation factor EF-Tu
157	2392136	pdb	1AH9	Translational Initiation Factor (If1)
158	15834470	ref	NP_313243.1	Trehalase 6-P hydrolase
159	75513895	ref	ZP_00736248.1	Triosephosphate isomerase
160	26249151	ref	NP_755191.1	tRNA pseudouridine synthase D
161	91209155	ref	YP_539141.1	UDP-N-acetylmuramate-L-alanine ligase
162	75187076	ref	ZP_00700343.1	Uncharacterized conserved protein
163	75512344	ref	ZP_00734896.1	Uncharacterized conserved protein
164	75514816	ref	ZP_00737045.1	Uncharacterized protein conserved in bacteria
165	43112	emb	CAA25253.1	unnamed protein product
166	42886	emb	CAA23639.1	unnamed protein product
167	41964	emb	CAA29949.1	unnamed protein product
168	43066	emb	CAA23560.1	unnamed protein product
169	42773	emb	CAA38206.1	unnamed protein product

170	42497	emb	CAA25363.1	unnamed protein product
171	41039	emb	CAA23528.1	unnamed protein product
172	434012	emb	CAA24742.1	unnamed protein product
173	91210668	ref	YP_540654.1	YdcG precursor

Contaminants Observed in $Zn^{(II)}$ IMAC

Reference	gi Number	NCBI Database	File ID	Peptide
1	15830974	ref	NP_309747.1	2-dehydro-3-deoxyphosphooctonate aldolase
2	91211354	ref	YP_541340.1	2'-deoxycytidine 5'-triphosphate deaminase
3	26250194	ref	NP_756234.1	2-hydroxyacid dehydrogenase
4	87082289	gb	AAC76577.2	2-keto-D-gluconate reductase (glyoxalate reductase)
5	75228118	ref	ZP_00714789.1	2-methylthioadenine synthetase
6	75511704	ref	ZP_00734342.1	3-isopropylmalate dehydratase; large subunit
7	91209134	ref	YP_539120.1	3-isopropylmalate isomerase (dehydratase) subunit
8	15832331	ref	NP_311104.1	50S ribosomal protein L25
9	26246970	ref	NP_753010.1	ABC transporter ATP-binding protein Uup
10	91214109	ref	YP_544095.1	ABC transporter ATP-binding protein YjjK
11	75188007	ref	ZP_00701274.1	ABC-type (unclassified) transport system; ATPase component
12	15834140	ref	NP_312913.1	acetylornithine deacetylase
13	15829376	ref	NP_308149.1	aconitate hydratase
14	91210287	ref	YP_540273.1	adenylosuccinate lyase
15	75511729	ref	ZP_00734358.1	Alcohol dehydrogenase; class IV
16	26246587	ref	NP_752626.1	Alkyl hydroperoxide reductase; subunit F
17	75258283	ref	ZP_00729731.1	Asparagine synthetase A
18	15829958	ref	NP_308731.1	asparagine synthetase B
19	26250490	ref	NP_756530.1	asparagine synthetase; (AsnA)
20	75235446	ref	ZP_00719654.1	Aspartate carbamoyltransferase; catalytic chain
21	537087	gb	AAA97142.1	aspartate carbomoyltransferase; catalytic subunit
22	91213799	ref	YP_543785.1	aspartate-carbamoyltransferase (PyrI)
23	1333747	emb	CAA23585.1	aspartokinase II-homoserine dehydrogenase II
24	83586215	ref	ZP_00924851.1	Aspartokinases
25	14278248	pdb	1FS0	ATP SYNTHASE; E Chain E
26	15831269	ref	NP_310042.1	ATP-dependent helicase
27	91209919	ref	YP_539905.1	ATP-dependent protease; ATP-binding subunit (clpA)
28	26250788	ref	NP_756828.1	B12-dependent methionine synthase

29	15834123	ref	NP_312896.1	bifunctional aspartate kinase II/homoserine dehydrogenase II
30	15834183	ref	NP_312956.1	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase
31	91209091	ref	YP_539077.1	carbamoyl-phosphate synthase small chain
32	75515621	ref	ZP_00737778.1	Carbamoylphosphate synthase; large subunit
33	75515620	ref	ZP_00737777.1	Carbamoylphosphate synthase; small subunit
34	75188955	ref	ZP_00702222.1	Cation transport ATPase
35	91210026	ref	YP_540012.1	cell division protein MukB
36	24054822	gb	AAN45715.1	chaperone Hsp60; GroEL
37	91209071	ref	YP_539057.1	chaperone Hsp70; DNA biosynthesis; auto-regulated heat shock proteins
38	26248955	ref	NP_754995.1	ClpB protein
39	91210463	ref	YP_540449.1	CoA-linked acetaldehyde dehydrogenase
40	15830261	ref	NP_309034.1	Condesin subunit B
41	91209194	ref	YP_539180.1	conserved hypothetical protein YadF
42	87081995	gb	AAC74944.2	copper homeostasis protein
43	91211857	ref	YP_541843.1	cysteine desulfurase
44	75211496	ref	ZP_00711586.1	Cysteine sulfinate desulfinase/cysteine desulfurase
45	87081779	gb	AAC73827.2	cytochrome d terminal oxidase; subunit I
46	26249982	ref	NP_756022.1	DamX protein
47	26247036	ref	NP_753076.1	Delta-1-pyrroline-5-carboxylate dehydrogenase
48	26250472	ref	NP_756512.1	D-fructose-6-phosphate amidotransferase
49	83584870	ref	ZP_00923528.1	Diadenosine tetraphosphate hydrolase (Ap4A)
50	15829288	ref	NP_308061.1	dihydrodipicolinate reductase
51	15829373	ref	NP_308146.1	dihydrolipoamide acetyltransferase
52	15832368	ref	NP_311141.1	DNA gyrase; Topoisomerase type II; subunit A
53	26250759	ref	NP_756799.1	DNA-directed RNA polymerase; beta subunit
54	147896	gb	AAA50992.1	elongation factor G
55	26247623	ref	NP_753663.1	enoyl-(acyl carrier protein) reductase
56	75240068	ref	ZP_00724029.1	Enoyl-[acyl-carrier-protein] reductase (NADH) [Escherichia
57	26248709	ref	NP_754749.1	erythronate-4-phosphate dehydrogenase
58	15831113	ref	NP_309886.1	exoribonuclease II

59	15830240	ref	NP_309013.1	formate acetyltransferase 1
60	15830988	ref	NP_309761.1	formyltetrahydrofolate deformylase
61	75230210	ref	ZP_00716711.1	Formyltetrahydrofolate hydrolase [Escherichia coli B7A]
62	147895	gb	AAA50991.1	FusA
63	1790376	gb	AAC76922.1	fused aspartokinase II/homoserine dehydrogenase II
64	87082027	gb	AAC75083.2	fused histidinol-phosphatase/imidazoleglycerol-phosphate dehydratase
65	1788589	gb	AAC75315.1	fused UDP-L-Ara4N formyltransferase; UDP-GlcA C-4'-decarboxylase
66	5739461	gb	AAD50483.1	GalF
67	1786437	gb	AAC73346.1	gamma-glutamate kinase
68	75237019	ref	ZP_00721075.1	Glucan phosphorylase
69	75228928	ref	ZP_00715520.1	Glucan phosphorylase
70	24417729	gb	AAN60453.1	glucose-1-phosphate uridylyltransferase
71	15833937	ref	NP_312710.1	glucose-inhibited division protein A
72	91209331	ref	YP_539317.1	glutamate 5-kinase
73	15831721	ref	NP_310494.1	glutamate dehydrogenase
74	551808	gb	AAA23908.1	glutamate synthase
75	15833345	ref	NP_312118.1	glutamate synthase large subunit
76	75186830	ref	ZP_00700097.1	Glutamine amidotransferase
77	91211830	ref	YP_541816.1	glutamine-hydrolyzing amonia dependent GMP synthetase
78	26248778	ref	NP_754818.1	glutamyl-tRNA synthetase
79	26247882	ref	NP_753922.1	Glutathione S-transferase
80	75257054	ref	ZP_00728614.1	Glutathione S-transferase
81	75228875	ref	ZP_00715476.1	Glycine/serine hydroxymethyltransferase
82	4377499	emb	CAA26035.1	glycogen phosphorylase
83	75212206	ref	ZP_00712246.1	Glycosidases
84	75214531	ref	ZP_00713099.1	GTPases - translation elongation factors
85	91211926	ref	YP_541912.1	heat shock protein
86	26247246	ref	NP_753286.1	HIT-like protein ycfF
87	75231824	ref	ZP_00718165.1	HrpA-like helicases
88	91210562	ref	YP_540548.1	hypothetical oxidoreductase (YciK)
89	26248141	ref	NP_754181.1	hypothetical protein c2289
90	26248643	ref	NP_754683.1	hypothetical protein c2797
91	26248681	ref	NP_754721.1	hypothetical protein c2835
92	75255649	ref	ZP_00727421.1	hypothetical protein EcolE2_01004299
93	15829865	ref	NP_308638.1	hypothetical protein ECs0611
94	15829900	ref	NP_308673.1	hypothetical protein ECs0646

95	38703933	ref	NP_309508.2	hypothetical protein ECs1481
96	15830926	ref	NP_309699.1	hypothetical protein ECs1672
97	15831274	ref	NP_310047.1	hypothetical protein ECs2020
98	15831647	ref	NP_310420.1	hypothetical protein ECs2393
99	15832397	ref	NP_311170.1	hypothetical protein ECs3143
100	15832432	ref	NP_311205.1	hypothetical protein ECs3178
101	91212860	ref	YP_542846.1	hypothetical protein UTI89_C3886
102	91210263	ref	YP_540249.1	hypothetical protein YcfP
103	91210659	ref	YP_540645.1	hypothetical protein YdcF
104	91210899	ref	YP_540885.1	hypothetical protein YdiI
105	26248556	ref	NP_754596.1	Hypothetical protein yeiR
106	91211549	ref	YP_541535.1	hypothetical protein YfbG
107	91212992	ref	YP_542978.1	hypothetical protein YhiR
108	15832078	ref	NP_310851.1	imidazole glycerol phosphate synthase subunit HisH
109	75229117	ref	ZP_00715692.1	Imidazoleglycerol-phosphate dehydratase
110	91211831	ref	YP_541817.1	inosine-5'-monophosphate dehydrogenase
111	1786882	gb	AAC73762.1	isopentenyl-adenosine A37 tRNA methylthiolase
112	15829330	ref	NP_308103.1	isopropyl malate isomerase; large subunit
113	75242656	ref	ZP_00726400.1	Lactate dehydrogenase
114	1787820	gb	AAC74612.1	L-allo-threonine dehydrogenase; NAD(P)-binding
115	15833513	ref	NP_312286.1	maltodextrin phosphorylase
116	9257169	pdb	3MAT	Methionine Aminopeptidase Transition-State Inhibitor Complex; A Chain A
117	91213796	ref	YP_543782.1	Mg2+-importing ATPase; MgtA
118	91209707	ref	YP_539693.1	N-acetyl glucosamine metabolism
119	83588394	ref	ZP_00927018.1	NAD/FAD-utilizing enzyme apparently involved in cell division
120	91211575	ref	YP_541561.1	NADH dehydrogenase I; chain G
121	26246640	ref	NP_752680.1	NagD protein
122	91210444	ref	YP_540430.1	NarL transcriptional dual regulator
123	48429400	gb	AAT42468.1	NrdA
124	75259718	ref	ZP_00731020.1	Nucleoside-diphosphate-sugar epimerases
125	147130	gb	AAA24310.1	PdxB protein
126	91214089	ref	YP_544075.1	peptide chain release factor RF-3
127	75256286	ref	ZP_00727958.1	Peptide chain release factor RF-3
128	75242168	ref	ZP_00725962.1	Periplasmic glucans biosynthesis protein
129	15832435	ref	NP_311208.1	phosphate acetyltransferase

130	26246403	ref	NP_752442.1	Phosphate regulon transcriptional Regulatory protein (phoB)
131	91211075	ref	YP_541061.1	phosphogluconate dehydratase
132	75243131	ref	ZP_00726837.1	Phosphoglycerate dehydrogenase and related dehydrogenases
133	15832677	ref	NP_311450.1	phosphoribosylformylglycinamidine synthase
134	1787458	gb	AAC74291.1	phosphoribosylpyrophosphate synthase
135	75228259	ref	ZP_00714914.1	Phosphoribosylpyrophosphate synthetase
136	12514920	gb	AAG56065.1	phosphoribosylpyrophosphate synthetase
137	56540577	gb	AAV92742.1	PmrA
138	91209466	ref	YP_539452.1	positive response regulator for pho regulon
139	75188498	ref	ZP_00701765.1	Predicted S-adenosylmethionine-dependent methyltransferase
140	75242722	ref	ZP_00726466.1	Predicted transcriptional regulators containing the CopG/Ar
141	75210026	ref	ZP_00710209.1	Predicted UDP-glucose 6-dehydrogenase
142	91210785	ref	YP_540771.1	probable oxidoreductase YdfG
143	26247820	ref	NP_753860.1	Probable oxidoreductase ydfG
144	15830514	ref	NP_309287.1	proline dehydrogenase
145	56553868	pdb	1TJ2	Proline Dehydrogenase Domain; A Chain A
146	1788943	gb	AAC75641.1	protein disaggregation chaperone
147	75188266	ref	ZP_00701533.1	Protein involved in catabolism of external DNA
148	86517026	gb	ABC98205.1	PrsA
149	86516774	gb	ABC98079.1	PurB
150	26247168	ref	NP_753208.1	Putative 2-hydroxyacid dehydrogenase ycdW
151	15829384	ref	NP_308157.1	putative carbonic anhdrase
152	15830664	ref	NP_309437.1	putative dehydrogenase
153	91209729	ref	YP_539715.1	putative esterase/lipase YbfF
154	72003790	gb	AAZ65830.1	putative GalF
155	15831283	ref	NP_310056.1	putative glycoprotein
156	91213267	ref	YP_543253.1	putative methyltransferase GidB
157	15832611	ref	NP_311384.1	putative oxidoreductase
158	91209183	ref	YP_539169.1	pyruvate dehydrogenase
159	91210003	ref	YP_539989.1	pyruvate formate lyase activating enzyme 1
160	75234414	ref	ZP_00718775.1	Pyruvate-formate lyase

161	75212945	ref	ZP_00712943.1	Response regulators consisting of a CheY-like receiver domain
162	26246420	ref	NP_752459.1	riboflavin synthase; subunit beta
163	26247224	ref	NP_753264.1	Ribonuclease E
164	91211527	ref	YP_541513.1	Ribonucleoside diphosphate reductase 1; alpha subunit B1
165	91210428	ref	YP_540414.1	ribose-phosphate pyrophosphokinase
166	43017	emb	CAA23653.1	ribosomal protein L14
167	83584673	ref	ZP_00923339.1	Ribosomal protein L3
168	75214891	ref	ZP_00713370.1	Ribosomal protein L6P/L9E
169	75258017	ref	ZP_00729491.1	Ribosome-associated protein Y (PSrp-1)
170	91212806	ref	YP_542792.1	RNA polymerase beta subunit
171	1073424	pir		S53980 hypothetical protein A - Escherichia coli
172	15832671	ref	NP_311444.1	serine hydroxymethyltransferase
173	75231698	ref	ZP_00718039.1	Short-chain alcohol dehydrogenase
174	75195640	ref	ZP_00705710.1	Short-chain alcohol dehydrogenase of unknown specificity
175	688004	gb	AAB31771.1	spirosin; 95 kDa spirosome subunit (N-terminal)
176	15831680	ref	NP_310453.1	threonyl-tRNA synthetase
177	687792	gb	AAC43729.1	threonyl-tRNA synthetase
178	42685	emb	CAA27600.1	thyA (C-terminal)
179	24022343	gb	AAN41256.1	thymidilate synthetase
180	83584784	ref	ZP_00923446.1	Thymidylate synthase
181	91212223	ref	YP_542209.1	thymidylate synthetase
182	14916721	sp	P58220	TKRA_ECO57 2-ketogluconate reductase (2KR) (2-ketoaldonate reductase)
183	91212813	ref	YP_542799.1	transcription antitermination protein
184	91213304	ref	YP_543290.1	transcription termination factor rho
185	15830856	ref	NP_309629.1	transcriptional regulatory protein
186	26250678	ref	NP_756718.1	Transcriptional Regulatory protein cpxR
187	91210819	ref	YP_540805.1	transcriptional regulatory protein RstA
188	15831568	ref	NP_310341.1	transcriptional regulatory protein RstB
189	75255629	ref	ZP_00727408.1	Transcription-repair coupling factor (superfamily II helicase)
190	91212815	ref	YP_542801.1	translation elongation factor EF-Tu
191	2392136	pdb	1AH9	Translational Initiation Factor If1
192	26251145	ref	NP_757185.1	Trehalose-6-phosphate hydrolase
193	75513895	ref	ZP_00736248.1	Triosephosphate isomerase
194	15832083	ref	NP_310856.1	UDP-glucose 6-dehydrogenase
195	18266397	gb	AAL67551.1	UDP-glucose pyrophosphorylase (GalF)

196	75259180	ref	ZP_00730528.1	UDP-glucose pyrophosphorylase [Escherichia coli E22]
197	18266410	gb	AAL67564.1	UDP-glucose-6-dehydrogenase Ugd
198	15833322	ref	NP_312095.1	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
199	75187076	ref	ZP_00700343.1	Uncharacterized conserved protein
200	75512080	ref	ZP_00734665.1	Uncharacterized protein involved in chromosome partitioning
201	75513021	ref	ZP_00735483.1	Uncharacterized protein involved in copper resistance
202	75514702	ref	ZP_00736957.1	Universal stress protein UspA and related nucleotide-binding
203	41964	emb	CAA29949.1	unnamed protein product
204	581130	emb	CAA23584.1	unnamed protein product
205	43257	emb	CAA25773.1	unnamed protein product
206	42886	emb	CAA23639.1	unnamed protein product
207	42497	emb	CAA25363.1	unnamed protein product
208	42984	emb	CAA25720.1	unnamed protein product
209	466691	gb	AAB18530.1	unnamed protein product
210	43066	emb	CAA23560.1	unnamed protein product
211	434010	emb	CAA24740.1	unnamed protein product
212	42821	emb	CAA23627.1	unnamed protein product
213	43112	emb	CAA25253.1	unnamed protein product
214	41627	emb	CAA26133.1	unnamed protein product
215	43268	emb	CAA25785.1	unnamed protein product
216	15834174	ref	NP_312947.1	uroporphyrinogen decarboxylase
217	91211327	ref	YP_541313.1	UTP-glucose-1-phosphate uridylyltransferase
218	26248415	ref	NP_754455.1	UTP--glucose-1-phosphate uridylyltransferase
219	91210391	ref	YP_540377.1	YcgK precursor
220	26247702	ref	NP_753742.1	ydcF
221	91210668	ref	YP_540654.1	YdcG precursor
222	91211588	ref	YP_541574.1	YfbU
223	58177208	pdb	1WPB	Yfbu Gene Product; P Chain P
224	48429440	gb	AAT42488.1	YleA

Appendix II: Contaminant protein metabolic function & essentiality for growth under recombinant production conditions.

Appendix II compiles Elution Contaminant Pool (ECP) proteins by metabolic function and essentiality. Essentiality was determined by biological activity of each contaminant protein relative to the cell's ability to utilize an alternate metabolic pathway(s) to drive growth. Metabolic function was determined using the EcoCyc® metabolomics database. Due to the analogous nature of the protein characterization work, where available, I have provided the Blattner ID number for each contaminant protein.

Abbreviations	Definition
E	Essential for growth or biological function
N	Non-essential for growth or biological function
X	Non-essential; knock-out significantly inhibits growth
?	Uncharacterized protein or metabolic pathway

An essentiality value of "X" was given to distinguish a scenario intrinsic to recombinant protein production. While traditional gene knock-out of this contaminant gene is not considered lethal under laboratory growth conditions; the effect of such a mutation would likely have an unfavorable impact on recombinant production or process through-put.

IMAC $Zn^{(II)}$ Contaminant Gene Products and Determined Essentiality for Growth

All Genes #	Gene	E/N	Length	Protein Description & E.C. #	Blattner #	Pathway
1	*accC*	X	449	Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA carboxylase (EC 6.4.1.2)	b3256	LPC
2	*aceE*	N	887	Pyruvate dehydrogenase E1 component (EC 1.2.4.1)	b0114	CHM
3	*aceF*	E	630	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	b0115	NCM
4	*acnB*	N	865	Aconitate hydratase 2 (EC 4.2.1.3)	b0118	CHM
5	*adhE*	N	891	Aldehyde-alcohol dehydrogenase	b1241	MSM
6	*ahpF*	N	531	Alkyl hydroperoxide reductase subunit F (EC 1.6.4.-)	b0606	MSM
7	*alaS*	E	876	Alanyl-tRNA synthetase (EC 6.1.1.7)	b2697	PMS
8	*argE*	X	383	Acetylornithine deacetylase (EC 3.5.1.16)	b3957	AAM
9	*aroH*	E	348	Phospho-2-dehydro-3-deoxyheptonate aldolase, Trp-sensitive (EC 4.1.2.15)	b1704	AAM
10	*artP*	N	242	Arginine transport ATP-binding protein artP	b0864	MTR
11	*asnA*	N	330	Aspartate--ammonia ligase (EC 6.3.1.1)	b3744	AAM
12	*asnB*	N	554	Asparagine synthetase B [glutamine-hydrolyzing] (EC 6.3.5.4)	b0674	AAM
13	*atpC*	E	139	ATP synthase epsilon chain (EC 3.6.3.14)	b3731	BEN
14	*atpG*	N	287	ATP synthase gamma chain (EC 3.6.3.14)	b3733	BEN
15	*basR*	N	222	Transcriptional regulatory protein basR/pmrA	b4113	SMC
16	*bcp*	N	156	Bacterioferritin co-migratory protein	b2480	UNC
17	*carA*	N	382	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5)	b0032	AAM
18	*carB*	N	1073	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	b0033	AAM
19	*cbpA*	N	306	Curved DNA-binding protein	b1000	UNC
20	*clpA*	N	758	ATP-dependent clp protease ATP-binding subunit clpA	b0882	PMS
21	*clpB*	N	857	ClpB protein	b2592	UNC
22	*clpP*	N	207	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92)	b0437	PMS
23	*coaA*	E	316	Pantothenate kinase (EC 2.7.1.33)	b3974	NCM
24	*corA*	E	316	Magnesium and cobalt transport protein corA	b3816	MTR
25	*cpxR*	N	232	Transcriptional regulatory protein cpxR	b3912	SMC

26	*crp*	X	210	Catabolite gene activator	b3357	RCD
27	*crr*	N	169	PTS system, glucose-specific IIA component (EC 2.7.1.69)	b2417	MTR
28	*csrA*	X	61	Carbon storage regulator	b2696	RCD
29	*cueO*	N	516	Blue copper oxidase cueO precursor	b0123	UNC
30	*cydA*	E	523	Cytochrome D ubiquinol oxidase subunit I (EC 1.10.3.-)	b0733	BEN
31	*cysD*	N	302	Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4)	b2752	MSM
32	*dapB*	E	273	Dihydrodipicolinate reductase (EC 1.3.1.26)	b0031	AAM
33	*dfp*	E	430	DNA/pantothenate metabolism flavoprotein	b3639	NCM
34	*dnaE*	E	1160	DNA polymerase III alpha subunit (EC 2.7.7.7)	b0184	NAM
35	*dnaK*	E	638	Chaperone protein dnaK	b0014	PMS
36	*dnaN*	X	366	DNA polymerase III, beta chain (EC 2.7.7.7)	b3701	NAM
37	*dps*	N	167	DNA protection during starvation protein	b0812	UNC
38	*edd*	N	603	Phosphogluconate dehydratase (EC 4.2.1.12)	b1851	CHM
39	*elaB*	N	101	ElaB protein	b2266	UNC
40	*fabI*	E	262	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9)	b1288	LPC
41	*fbp*	N	332	Fructose-1,6-bisphosphatase (EC 3.1.3.11)	b4232	CHM
42	*folE*	E	222	GTP cyclohydrolase I (EC 3.5.4.16)	b2153	NCM
43	*folX*	N	120	D-erythro-7,8-dihydroneopterin triphosphate epimerase (EC 5.-.-.-)	b2303	NCM
44	*fur*	N	148	Ferric uptake regulation protein	b0683	RCD
45	*fusA*	E	704	Elongation factor G	b3340	PMS
46	*galF*	E	297	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	b2042	CHM
47	*galU*	E	302	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	b1236	CHM
48	*gapA*	E	331	Glyceraldehyde 3-phosphate dehydrogenase A (EC 1.2.1.12)	b1779	CHM
49	*gatA*	N	150	PTS system, galactitol-specific IIA component (EC 2.7.1.69)	b2094	MTR
50	*gatD*	E	346	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	b2091	CHM
51	*gcvP*	N	957	Glycine dehydrogenase [decarboxylating] (EC 1.4.4.2)	b2903	AAM
52	*gdhA*	N	447	NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	b1761	MSM

53	*gidA*	N	629	Glucose inhibited division protein A	b3741	UNC
54	*gidB*	N	207	Glucose inhibited division protein B	b3740	UNC
55	*glgB*	N	728	1,4-alpha-glucan branching enzyme (EC 2.4.1.18)	b3432	CHM
56	*glk*	N	321	Glucokinase (EC 2.7.1.2)	b2388	CHM
57	*glmS*	E	609	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	b3729	CHM
58	*glnD*	E	890	[Protein-PII] uridylyltransferase (EC 2.7.7.59)	b0167	SMC
59	*gltB*	N	1517	Glutamate synthase [NADPH] large chain precursor (EC 1.4.1.13)	b3212	MSM
60	*gltX*	E	471	Glutamyl-tRNA synthetase (EC 6.1.1.17)	b2400	PMS
61	*glyA*	E	417	Serine hydroxymethyltransferase (EC 2.1.2.1)	b2551	AAM
62	*groL*	E	548	60 kDa chaperonin	b4143	PMS
63	*gst*	N	201	Glutathione S-transferase (EC 2.5.1.18)	b1635	BEN
64	*guaA*	N	525	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2)	b2507	NCM
65	*guaB*	N	488	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	b2508	NCM
66	*gyrA*	E	875	DNA gyrase subunit A (EC 5.99.1.3)	b2231	NAM
67	*gyrB*	X	804	DNA gyrase subunit B (EC 5.99.1.3)	b3699	NAM
68	*hemD*	E	246	Uroporphyrinogen-III synthase (EC 4.2.1.75)	b3804	NCM
69	*hemE*	E	354	Uroporphyrinogen decarboxylase (EC 4.1.1.37)	b3997	NCM
70	*hemN*	N	459	Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-)	b3867	NCM
71	*hepA*	N	968	RNA polymerase associated protein	b0059	UNC
72	*hisB*	N	356	Histidine biosynthesis bifunctional protein hisB	b2022	AAM
73	*hisH*	N	196	Imidazole glycerol phosphate synthase subunit hisH (EC 2.4.2.-)	b2023	AAM
74	*hrpA*	N	1281	ATP-dependent helicase hrpA	b1413	UNC
75	*hslR*	N	133	Heat shock protein 15	b3400	SMC
76	*hslV*	N	176	ATP-dependent protease hslV (EC 3.4.25.-)	b3932	SMC
77	*hybC*	N	567	Hydrogenase-2 large chain precursor (EC 1.18.99.1)	b2994	BEN
78	*ilvD*	N	605	Dihydroxy-acid dehydratase (EC 4.2.1.9)	b3771	AAM
79	*infA*	?	72	Translation initiation factor IF-1	b0884	UNC
80	*iscS*	E	412	Cysteine desulfurase (EC 4.4.1.-)	b2530	NCM
81	*ispD*	E	236	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC	b2747	LPC

				2.7.7.60)		
82	*kdsA*	N	284	2-dehydro-3-deoxyphosphooctonate aldolase (EC 4.1.2.16)	b1215	CHM
83	*leuC*	N	466	3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)	b0072	AAM
84	*lexA*	X	202	LexA repressor (EC 3.4.21.88)	b4043	NAM
85	*lpdA*	E	474	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	b0116	CHM
86	*lysS*	X	505	Lysyl-tRNA synthetase (EC 6.1.1.6)	b2890	NAM
87	*lysU*	N	505	Lysyl-tRNA synthetase, heat inducible (EC 6.1.1.6)	b4129	NAM
88	*maa*	N	183	Maltose O-acetyltransferase (EC 2.3.1.79)	b0459	CHM
89	*malP*	N	797	Maltodextrin phosphorylase (EC 2.4.1.1)	b3417	CHM
90	*map*	E	264	Methionine aminopeptidase (EC 3.4.11.18)	b0168	PMS
91	*mdoB*	X	750	Phosphoglycerol transferase I (EC 2.7.8.20)	b4359	LPC
92	*mdoH*	N	847	Periplasmic glucans biosynthesis protein mdoH	b1049	CHM
93	*metH*	N	1227	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	b4019	AAM
94	*metL*	N	810	Bifunctional aspartokinase/homoserine dehydrogenase II	b3940	AAM
95	*mfd*	N	1148	Transcription-repair coupling factor	b1114	NAM
96	*mgtA*	N	898	Mg(2+) transport ATPase, P-type 1 (EC 3.6.3.2)	b4242	MTR
97	*minC*	N	231	Septum site-determining protein minC	b1176	RCD
98	*mpl*	N	457	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (EC 6.3.2.-)	b4233	LPC
99	*mrp*	N	379	Mrp protein	b2113	UNC
100	*msrA*	N	212	Peptide methionine sulphoxide reductase (EC 1.8.4.6)	b4219	PMS
101	*mukB*	E	1486	Cell division protein	b0924	RCD
102	*murA*	N	419	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	b3189	LPC
103	*murC*	E	491	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8)	b0091	LPC
104	*murE*	E	495	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (EC 6.3.2.13)	b0085	LPC
105	*nagB*	N	266	Glucosamine-6-phosphate isomerase (EC 3.5.99.6)	b0678	CHM
106	*nagD*	N	250	NagD protein	b0675	UNC

107	*narL*	N	216	Nitrate/nitrite response regulator protein narL	b1221	SMC
108	*nikR*	N	133	Nickel responsive regulator	b3481	RCD
109	*nrdA*	E	761	Ribonucleoside-diphosphate reductase 1 alpha chain (EC 1.17.4.1)	b2234	NCM
110	*nrdD*	N	712	Anaerobic ribonucleoside-triphosphate reductase (EC 1.17.4.2)	b4238	NCM
111	*nudC*	N	257	NADH pyrophosphatase (EC 3.6.1.-)	b3996	MSM
112	*nuoG*	N	910	NADH dehydrogenase I chain G (EC 1.6.5.3)	b2283	BEN
113	*nusG*	N	181	Transcription antitermination protein nusG	b3982	RCD
114	*ompA*	N	346	Outer membrane protein A precursor	b0957	SMC
115	*ompF*	N	362	Outer membrane protein F precursor	b0929	SMC
116	*ompR*	N	239	Transcriptional regulatory protein ompR	b3405	SMC
117	*pdxB*	N	378	Erythronate-4-phosphate dehydrogenase (EC 1.1.1.-)	b2320	NCM
118	*pdxH*	N	218	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	b1638	NCM
119	*pflA*	N	246	Pyruvate formate-lyase 1 activating enzyme (EC 1.97.1.4)	b0902	CHM
120	*pflB*	N	760	Formate acetyltransferase 1 (EC 2.3.1.54)	b0903	CHM
121	*phoB*	N	229	Phosphate regulon transcriptional regulatory protein phoB	b0399	RCD
122	*phoP*	N	223	Transcriptional regulatory protein phoP	b1130	SMC
123	*ppiA*	N	190	Peptidyl-prolyl cis-trans isomerase A precursor (EC 5.2.1.8)	b3363	PMS
124	*prfA*	X	360	Peptide chain release factor 1	b1211	PMS
125	*prfB*	?	365	Peptide chain release factor 2	b2891	PMS
126	*prfC*	N	529	Peptide chain release factor 3	b4375	PMS
127	*proB*	N	367	Glutamate 5-kinase (EC 2.7.2.11)	b0242	AAM
128	*prsA*	E	315	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	b1207	CHM
129	*pta*	N	714	Phosphate acetyltransferase (EC 2.3.1.8)	b2297	MSM
130	*purB*	N	456	Adenylosuccinate lyase (EC 4.3.2.2)	b1131	NCM
131	*purL*	N	1295	Phosphoribosylformylglycinamidine synthase (EC 6.3.5.3)	b2557	NCM
132	*purU*	E	280	Formyltetrahydrofolate deformylase (EC 3.5.1.10)	b1232	NCM
133	*putA*	N	1320	Bifunctional putA protein	b1014	AAM
134	*pykA*	N	480	Pyruvate kinase II (EC 2.7.1.40)	b1854	CHM
135	*pyrB*	N	311	Aspartate carbamoyltransferase catalytic chain (EC 2.1.3.2)	b4245	NCM

136	*pyrG*	E	545	CTP synthase (EC 6.3.4.2)	b2780	NCM
137	*pyrI*	N	153	Aspartate carbamoyltransferase regulatory chain	b4244	RCD
138	*relA*	N	744	GTP pyrophosphokinase (EC 2.7.6.5)	b2784	MSM
139	*rho*	?	419	Transcription termination factor rho	b3783	RCD
140	*ribD*	E	367	Riboflavin biosynthesis protein ribD	b0414	NCM
141	*ribE*	E	213	Riboflavin synthase alpha chain (EC 2.5.1.9)	b1662	NCM
142	*rimM*	E	185	16S rRNA processing protein rimM	b2608	NAM
143	*rnb*	N	644	Exoribonuclease II (EC 3.1.13.1)	b1286	NAM
144	*rnc*	E	226	Ribonuclease III (EC 3.1.26.3)	b2567	NAM
145	*rnd*	N	375	Ribonuclease D (EC 3.1.26.3)	b1804	NAM
146	*rne*	E	1061	Ribonuclease E (EC 3.1.4.-)	b1084	NAM
147	*rph*	N	228	Ribonuclease PH (EC 2.7.7.56)	b3643	NAM
148	*rplC*	E	209	50S ribosomal protein L3	b3320	PMS
149	*rplF*	E	177	50S ribosomal protein L6	b3305	PMS
150	*rplN*	E	123	50S ribosomal protein L14	b3310	PMS
151	*rplS*	E	115	50S ribosomal protein L19	b2606	PMS
152	*rplY*	E	94	50S ribosomal protein L25	b2185	PMS
153	*rpoB*	E	1342	DNA-directed RNA polymerase beta chain (EC 2.7.7.6)	b3987	NAM
154	*rpoC*	E	1407	DNA-directed RNA polymerase beta' chain (EC 2.7.7.6)	b3988	NAM
155	*rpoD*	E	613	RNA polymerase sigma factor rpoD	b3067	RCD
156	*rpsA*	N	557	30S ribosomal protein S1	b0911	PMS
157	*rpsB*	E	241	30S ribosomal protein S2	b0169	PMS
158	*rpsC*	N	233	30S ribosomal protein S3	b3314	PMS
159	*rpsF*	E	131	30S ribosomal protein S6	b4200	PMS
160	*rpsJ*	E	103	30S ribosomal protein S10	b3321	PMS
161	*rstA*	N	242	Transcriptional regulatory protein rstA	b1608	RCD
162	*sdhA*	N	588	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	b0723	SMC
163	*slyD*	E	196	FKBP-type peptidyl-prolyl cis-trans isomerase slyD (EC 5.2.1.8)	b3349	PMS
164	*speA*	N	658	Biosynthetic arginine decarboxylase (EC 4.1.1.19)	b2938	AAM
165	*speD*	N	264	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50)	b0120	MSM
166	*speG*	N	186	Spermidine N(1)-acetyltransferase (EC 2.3.1.57)	b1584	AAM

167	*sucA*	E	933	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	b0726	CHM
168	*suhB*	E	267	Inositol-1-monophosphatase (EC 3.1.3.25)	b2533	LPC
169	*tdk*	E	205	Thymidine kinase (EC 2.7.1.21)	b1238	NCM
170	*thrA*	N	820	Bifunctional aspartokinase/homoserine dehydrogenase I	b0002	AAM
171	*thrS*	E	642	Threonyl-tRNA synthetase (EC 6.1.1.3)	b1719	PMS
172	*tpiA*	X	255	Triosephosphate isomerase (EC 5.3.1.1)	b3919	CHM
173	*treC*	N	551	Trehalose-6-phosphate hydrolase (EC 3.2.1.93)	b4239	CHM
174	*trpS*	E	334	Tryptophanyl-tRNA synthetase (EC 6.1.1.2)	b3384	PMS
175	*tufA*	E	394	Elongation factor Tu	b3339	PMS
176	*tufB*	E	394	Elongation factor Tu	b3980	PMS
177	*ugd*	N	388	UDP-glucose 6-dehydrogenase (EC 1.1.1.22)	b2028	SMC
178	*uhpA*	N	196	Transcriptional regulatory protein uhpA	b3669	SMC
179	*uup*	N	635	ABC transporter ATP-binding protein uup	b0949	MTR
180	*yacF*	N	247	Hypothetical protein yacF	b0102	UNC
181	*yahK*	N	349	Hypothetical zinc-type alcohol dehydrogenase-like protein yahK	b0325	UNC
182	*yajD*	N	115	Hypothetical protein yajD	b0410	UNC
183	*ybeD*	N	87	Hypothetical protein ybeD	b0631	UNC
184	*ybfF*	N	254	Putative esterase/lipase ybfF (EC 3.1.-.-)	b0686	UNC
185	*ybgJ*	N	218	Hypothetical protein ybgJ	b0711	UNC
186	*ybiB*	N	320	Hypothetical protein ybiB	b0800	UNC
187	*ybiV*	N	271	Hypothetical protein ybiV	b0822	UNC
188	*ycbY*	N	702	Hypothetical protein ycbY	b0948	UNC
189	*yccX*	N	92	Putative acylphosphatase (EC 3.6.1.7)	b0968	UNC
190	*ycdW*	E	325	Putative 2-hydroxyacid dehydrogenase ycdW	b1033	UNC
191	*ycfP*	E	199	Hypothetical protein ycfP	b1108	UNC
192	*ycgK*	N	133	Protein ycgK precursor	b1178	UNC
193	*yciK*	E	252	Hypothetical oxidoreductase yciK (EC 1.-.-.-)	b1271	UNC
194	*ydcF*	N	266	Protein ydcF	b1414	UNC
195	*ydcY*	N	77	Hypothetical protein ydcY	b1446	UNC
196	*ydfG*	N	248	Probable oxidoreductase ydfG (EC 1.-.-.-)	b1539	UNC
197	*ydhF*	N	298	Hypothetical oxidoreductase ydhF (EC 1.-.-.-)	b1647	UNC
198	*ydiI*	E	136	Hypothetical protein ydiI	b1686	UNC

199	*yebU*	N	481	Hypothetical protein yebU	b1835	UNC
200	*yecP*	E	323	Hypothetical protein yecP	b1871	UNC
201	*yegQ*	N	453	Putative protease yegQ (EC 3.4.-.-)	b2081	UNC
202	*yeiE*	N	293	Hypothetical transcriptional regulator yeiE	b2157	UNC
203	*yfbG*	N	660	Hypothetical protein yfbG	b2255	UNC
204	*yfbU*	N	170	Protein yfbU	b2294	UNC
205	*yfcG*	N	215	Hypothetical GST-like protein yccG	b2302	UNC
206	*yffB*	E	118	Protein yffB	b2471	UNC
207	*yfgD*	N	119	Protein yfgD	b2495	UNC
208	*yfgM*	N	206	Hypothetical protein yfgM	b2513	UNC
209	*yfiA*	N	113	Protein yfiA	b2597	UNC
210	*ygdH*	N	454	Hypothetical protein ygdH	b2795	UNC
211	*ygiN*	N	104	Protein ygiN	b3029	UNC
212	*yhaJ*	N	298	Hypothetical transcriptional regulator yhaJ	b3105	UNC
213	*yhbG*	E	241	Probable ABC transporter ATP-binding protein yhbG	b3201	UNC
214	*yhbH*	N	95	Probable sigma(54) modulation protein	b3203	UNC
215	*yhbW*	N	335	Hypothetical protein yhbW	b3160	UNC
216	*yhfT*	X	434	Hypothetical protein yhfT	b3377	UNC
217	*yhhX*	N	345	Putative oxidoreductase yhhX (EC 1.-.-.-)	b3440	UNC
218	*yhiR*	N	280	Hypothetical protein yhiR	b3499	UNC
219	*yieN*	N	506	Hypothetical protein yieN	b3746	UNC
220	*yifE*	N	112	Protein yifE	b3764	UNC
221	*yjbR*	N	118	Protein yjbR	b4057	UNC
222	*yjjK*	N	555	ABC transporter ATP-binding protein yjjK	b4391	MTR
223	*yliG*	E	441	Hypothetical protein yliG	b0835	UNC
224	*yncB*	N	376	Putative NADP-dependent oxidoreductase yncB (EC 1.-.-.-)	b1449	UNC
225	*yncE*	E	353	Hypothetical protein yncE precursor	b1452	UNC
226	*yneH*	N	308	Probable glutaminase yneH (EC 3.5.1.2)	b1524	UNC
227	*yqcD*	N	282	Hypothetical protein yqcD	b2794	UNC
228	*yraL*	N	286	Hypothetical protein yraL	b3146	UNC
229	*yrdA*	N	256	Protein yrdA	b3279	UNC
230	*ytfP*	N	113	Hypothetical protein ytfP	b4222	UNC
231	*zwf*	E	491	Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	b1852	CHM

IMAC $Co^{(II)}$ Contaminant Essentiality

#	gene	E/N	Alt. Gene Symbol	Protein Description & E.C. #	Pathway	length	MW	pI (calc.)	Transcription Start	Blattner ID
1	*aroH*	E	--	Phospho-2-dehydro-3-deoxyheptonate aldolase, Trp-sensitive (EC 4.1.2.15)	AAM	348	38,735	6.42	1786459	b1704
2	*chbR*	?	*celD*	Unknown Protein	--	280	32,968	8.41	--	--
3	*coaA*	E	*panK, rts, ts-9*	Pantothenate kinase (EC 2.7.1.33)	NCM	316	36,360	6.32	4172605	b3974
4	*crp*	X	*cap, csm*	Catabolite gene activator, cAMP receptor protein	RCD	210	23,640	8.38	3483757	b3357
5	*folX*	N	--	D-erythro-7,8-dihydroneopterin triphosphate epimerase (EC 5.-.-.-)	NCM	120	14,082	6.51	2419345	b2303
6	*glk*	N	--	Glucokinase (EC 2.7.1.2)	CHM	321	34,723	6.06	2507446	b2388
7	*hemD*	E	--	Uroporphyrinogen-III synthase (EC 4.2.1.75)	NCM	246	27,798	5.98	3987441	b3804
8	*hslR*	N	*yrfH*	Heat shock protein 15	SMC	133	15,496	9.94	3526986	b3400
9	*mdoB*	X	*yjjO*	Phosphoglycerol transferase I (EC 2.7.8.20)	LPC	750	85,494	6.10	4596971	b4359
10	*prfA*	X	*sueB, uar, ups?*	Peptide chain release factor 1	PMS	360	40,517	5.15	1264235	b1211
11	*prkB*	E	*prk?, yhfF*	Putative phosphoribulokinase	--	289	32,344	6.18	3482512	--
12	*qseB*	N	*ygiX*	Quorum sensing E.coli	--	219	2,468	6.54	--	--
13	*rimM*	E	*yfiA*	16S rRNA processing protein rimM	NAM	185	20,605	4.61	2743947	b2608

14	*rpoD*	E	--	RNA polymerase sigma factor rpoD, Sigma 70, initiates transcription	RCD	613	70,263	4.69	3210688	b3067
15	*rpsJ*	E	*nusE*	30S ribosomal protein S10	PMS	103	11,736	9.68	3450907	b3321
16	*yacF*	N	--	Hypothetical protein yacF	UNC	247	28,292	6.31	112599	b0102
17	*yajD*	N	--	Hypothetical protein yajD	UNC	115	13,364	6.14	429829	b0410
18	*ybfF*	N	--	Putative esterase/lipase ybfF (EC 3.1.-.-)	UNC	254	28,437	5.86	712025	b0686
19	*ycbY*	N	*rlmL*	Hypothetical protein ycbY	UNC	702	78,854	8.96	1007067	b0948
20	*yeiE*	N	--	Hypothetical transcriptional regulator; probable positive regulator of *lysP* transcription	UNC	293	32,724	6.07	2247638	b2157
21	*yfgM*	N	--	Hypothetical protein yfgM	UNC	206	22,176	5.07	2637303	b2513
22	*ygjF*	X	*mug, dug*	Uracil DNA-glycosylase, G:U mismatch-specific, dsDNA-specific; excises uracil from DNA by base flipping mechanism; required for mutation avoidance in stationary phase; up-regulated in stationary phase	NAM	168	18,673	9.17	--	--
23	*yhfT*	X	--	Hypothetical protein yhfT	UNC	434	46,516	6.53	3504974	b3377

24	*yhiQ*	?	--	Function unknown; in Salmonella, yhiQ is the second gene in prlC operon and in the heat shock regulon; putative SAM-dependent methyltransferase		250	26,949	6.60		--
25	*yraL*	N	--	Putative methyltransferase, function unknown	UNC	286	31,348	5.83	3290976	b3146

Common to $Co^{(II)}$ & $Zn^{(II)}$

#	gene	E/N	Alt. Gene Symbol	Protein Description & E.C. #	Pathway	AA length	MW (calc.)	pI (calc.)	Transcription start	Blattner ID
1	*asnB*	N	--	Asparagine synthetase B [glutamine-hydrolyzing] (EC 6.3.5.4)	AAM	554	62,659.01	5.55	698400	b0674
2	*dapB*	E	--	Dihydrodipicolinate reductase (EC 1.3.1.26)	AAM	273	28,756.61	5.45	28374	b0031
3	*fur*	N	--	Ferric uptake regulation protein	RCD	148	16,794.85	5.68	709869	b0683
4	*fusA*	E	*far, fus*	Elongation Factor G	PMS	704	77,581.31	5.24	3471151	b3340
5	*hemE*	E	*hemC*	Uroporphyrinogen decarboxylase (EC 4.1.1.37)	NCM	354	39,248.12	5.88	4195294	b3997
6	*hepA*	N	*rapA, yabA*	RNA polymerase associated protein	UNC	968	109,769.06	5.04	63264	b0059
7	*hinT*		*ycfF*	Purine nucleoside phosphoramidase; physiological role unknown	NAM	119	13,241.28	5.73	--	--
8	*kdsA*	N*	--	2-dehydro-3-deoxyphosphooctonate aldolase (EC 4.1.2.16)	CHM	284	30,832.69	6.32	1267388	b1215
9	*leuC*	N	--	3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)	AAM	466	49,881.83	5.90	80864	b0072
10	*map*	E	*pepM (S.t.)*	Methionine aminopeptidase (EC 3.4.11.18)	PMS	264	29,330.80	5.64	189506	b0168
11	*nagB*	N	*glmD*	Glucosamine-6-phosphate isomerase (EC 3.5.99.6)	CHM	266	29,774.20	6.41	702834	b0678
12	*prsA*	E	*prs,dnaR*	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	CHM	315	34,218.27	5.23	1261098	b1207

13	*purU*	E	*tgs, ychI*	Formyltetrahydrofolate deformylase (EC 3.5.1.10)	NCM	280	31,934.72	6.50	1287847	b1232
14	*rplN*	E	--	50S ribosomal protein L14	PMS	123	13,541.02	10.43	3445786	b3310
15	*slyD*	E	--	FKBP-type peptidyl-prolyl cis-trans isomerase(EC 5.2.1.8)	PMS	196	20,852.83	4.86	3476134	b3349
16	*speG*	N	--	Spermidine N(1)-acetyltransferase (EC 2.3.1.57)	AAM	186	21,887.01	6.20	1654208	b1584
17	*tiaE*		--	Unknown Protein *tiaE*	--	--	--	--	--	--
18	*ydfG*	N	--	Probable oxidoreductase *ydfG* (EC 1.-.-.-)	UNC	248	27,248.93	5.65	1625541	b1539
19	*yfbG*	N	*arnA, pbgP3 (S.t.), pmrI, SAF*	Hypothetical protein *yfbG*	UNC	660	74,288.89	6.39	2366059	b2255
21	*yfcG*	N	--	Hypothetical GST-like protein yccG	UNC	215	24,515.90	6.45	2418641	b2302
22	*yfgD*	N	--	Protein YfgD	UNC	119	13,398.58	5.89	2615598	b2495
23	*yjtD*	N	*lasT*	Predicted methyltransferase, function unknown	UNC	228	25,259.04	5.58	--	--

Common to $Co^{(II)}$ & $Ni^{(II)}$

#	gene	E/N	Alt. Gene Symbol	Protein Description & E.C. #	Pathway	AA Length	MW (calc.)	pI (calc.)	Transcription Start	Blattner ID
1	*accC*	X	*fabG*	Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA carboxylase (EC 6.4.1.2))	LPC	449	49,321	6.65	3403554	b3256
2	*alaS*	E	*act, ala-act, lovB*	Alanyl-tRNA synthetase (EC 6.1.1.7)	PMS	876	96,032	5.53	2820033	b2697
3	*atpC*	E	*papG, uncC*	ATP synthase epsilon chain (EC 3.6.3.14)	BEN	139	15,068	5.46	3913600	b3731
4	*atpG*	N	*papC, uncG*	ATP synthase gamma chain (EC 3.6.3.14)	BEN	287	31,577	8.84	3915893	b3733
5	*cbpA*	N	--	Curved DNA-binding protein	UNC	306	34,455	6.33	1062998	b1000
6	*cueO*	N	*yacK, cuiD (S.t.)*	Blue copper oxidase	UNC	516	56,556	6.28	137083	b0123
7	*dnaE*	E	*polC, sdgC*	DNA polymerase III alpha subunit (EC 2.7.7.7)	NAM	1160	129,905	5.16	205126	b0184
8	*dnaN*	X	--	DNA polymerase III, beta chain (EC 2.7.7.7)	NAM	366	40,587	5.25	3879949	b3701
9	*glgB*	N	--	1,4-alpha-glucan branching enzyme (EC 2.4.1.18)	CHM	728	84,337	5.91	3571135	b3432
10	*glnD*	E	--	[Protein-PII] uridylyltransferase (EC 2.7.7.59)	SMC	890	102,390	6.22	188650	b0167
11	*lpdA*	E	*lpd, dhl*	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	CHM	474	50,688	5.79	127912	b0116
12	*minC*	N	*minB*	Septum site-determining protein minC	RCD	231	24,776	6.38	1225303	b1176
13	*mrp*	N	*apbC (S.t.)*	Mrp protein	UNC	379	39,938	5.85	2192218	b2113

14	*murC*	E	--	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8)	LPC	491	53,626	5.53	100765	b0091
15	*murE*	E	--	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (EC 6.3.2.13)	LPC	495	53,344	5.42	93166	b0085
16	*nrdD*	N	--	Anaerobic ribonucleoside-triphosphate reductase (EC 1.17.4.2)	NCM	712	80,023	6.38	4460234	b4238
17	*pdxB*	N	--	Erythronate-4-phosphate dehydrogenase (EC 1.1.1.-)	NCM	378	41,368	6.23	2435871	b2320
18	*rph*	N	--	Ribonuclease PH (EC 2.7.7.56)	NAM	228	25,352	5.54	3814176	b3643
19	*rpsB*	E	--	30S ribosomal protein S2	PMS	241	26,744	6.61	189874	b0169
20	*yebU*	N	*rsmF*	Hypothetical protein yebU	UNC	481	53,228	5.44	1918241	b1835
21	*yecP*	E	*cmoB*	Hypothetical protein yecP	UNC	323	37,007	6.13	1951466	b1871
22	*yegQ*	N	--	Putative protease yegQ (EC 3.4.-.-)	UNC	453	51,193	5.80	2163690	b2081
23	*yfhQ*		*trmJ*	tRNA Cm32/Um32 methyltransferase, SAM-dependent; low abundance protein	PMS	246	27,048	5.69	--	--
24	*ygiN*	N	--	Protein ygiN	UNC	104	11,532	5.79	3171158	b3029
25	*yhaJ*	N	--	Hypothetical transcriptional regulator yhaJ	UNC	298	33,256	6.05	3251854	b3105
26	*yhhX*	N	--	Putative oxidoreductase yhhX (EC 1.-.-.-)	UNC	345	38,765	6.07	3578436	b3440
27	*ytfP*	N	--	Hypothetical protein ytfP	UNC	113	12,866	6.39	4445471	b4222

Elution Contaminants Common to $Co^{(II)}$, $Zn^{(II)}$, & $Ni^{(II)}$

All Genes #	Co(II) Only #	Gene	E/N	Alt. Gene Symbol	Enzyme Discription	Pathway	MW	Length	pI	Blattner #
2	26	*aceE*	N	--	Pyruvate dehydrogenase E1 component (EC 1.2.4.1)	UNC	99,668.49	887	5.46	b0884
3	3	*aceF*	E	--	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	RCD	66,096.07	630	5.09	b3783
5	27	*adhE*	N	*adhC, ana*	Aldehyde-alcohol dehydrogenase	NCM	96,127.24	891	6.32	b0115
6	28	*ahpF*	N	--	Alkyl hydroperoxide reductase subunit F (EC 1.6.4.-)	BEN	56,177.11	531	5.47	b0733
15	29	*basR*	N	*pmrA*	Transcriptional regulatory protein basR/pmrA	PMS	25,030.81	222	5.66	b0014
16	30	*bcp*	N	--	Bacterioferritin co-migratory protein	NCM	17,633.94	156	5.03	b2153
	67	*can*	N	--	--	CHM	--	--	--	b1779
18	31	*carB*	N	*cap, pyrA*	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	CHM	117,841.72	1073	5.22	b3729
20	32	*clpA*	N	*lopD*	ATP-dependent clp protease ATP-binding subunit clpA	PMS	84,206.93	758	5.91	b2400
21	33	*clpB*	N	*htpM*	ClpB protein	AAM	95,585.02	857	5.37	b2551
34	68	*cusF*	N	*cusX, silF (S.t.), ylcC*	Silver and copper efflux, periplasmic binding protein; confers copper and silver resistance; CusF binds both Cu(I) and Ag(I) with high affinity, but does not bind Cu(II)	PMS	12,251.16	110	6.97	b4143
30	4	*cydA*	E	--	Cytochrome D ubiquinol oxidase subunit I (EC 1.10.3.-)	RCD	58,205.08	523	6.35	b0924
35	5	*dnaK*	E	*groPAB, groPC, groPF, grpC,*	Chaperone protein	NCM	69,114.96	638	4.83	b2234

				grpF, seg						
42	6	*folE*	E	--	GTP cyclohydrolase I (EC 3.5.4.16)	NAM	24,830.62	222	6.80	b1084
48	7	*gapA*	E	--	Glyceraldehyde 3-phosphate dehydrogenase A (EC 1.2.1.12)	PMS	35,532.49	331	6.61	b3320
54	34	*gidB*	N	--	Glucose inhibited division protein B	PMS	23,431.12	207	6.06	b2185
57	8	*glmS*	E	--	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	NAM	66,894.34	609	5.56	b3987
59	35	*gltB*	N	*aspB, ossB, psiQ*	Glutamate synthase [NADPH] large chain precursor (EC 1.4.1.13)	NAM	166,709.87	1517	6.27	b3988
60	9	*gltX*	E	--	Glutamyl-tRNA synthetase (EC 6.1.1.17)	NCM	53,815.73	471	5.59	b1238
61	10	*glyA*	E	--	Serine hydroxymethyltransferase (EC 2.1.2.1)	PMS	45,316.59	417	6.03	b1719
62	11	*groL*	E	*groEL, mopA*	60 kDa chaperonin	PMS	57,328.85	548	4.85	b3339
	69	*guaA*	E	--	GMP synthase	PMS	58,679.22	525	5.24	b3980
65	36	*guaB*	N	*guaR*	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	UNC	52,022.45	488	6.02	b1108
	70	*gyrA*	E	*hisW, nalA, nfxA, norA, parD acrB, Cou, himB, hisU, hopA, nalC, parA, pcbA,*	DNA gyrase, subunit A; nalidixic acid resistance; cold shock regulon	UNC	96,963.51	875	5.09	b1271
67	65	*gyrB*	E		DNA gyrase subunit B (EC 5.99.1.3)	UNC	89,949.91	804	5.72	b3201

				pcpA						
79	1	*infA*	X	*bypA1*	Translation initiation factor IF-1	CHM	8,249.58	72	9.22	b0114
89	37	*malP*	N	*blu, malA*	Maltodextrin phosphorylase (EC 2.4.1.1)	MSM	90,522.40	797	6.94	b1241
	71	*mdoD*		*ydcG, yzzZ*	Glucans biosynthesis protein, periplasmic	MSM	62,757.95	551	5.89	b0606
93	38	*metH*	N	--	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	SMC	135,997.04	1227	4.97	b4113
94	39	*metL*	N	*metM*	Bifunctional aspartokinase/homoserine dehydrogenase II	UNC	88,887.69	810	5.34	b2480
95	40	*mfd*	N	--	Transcription-repair coupling factor	AAM	129,982.73	1148	5.79	b0033
96	41	*mgtA*	N	*corB, mgt*	Mg(2+) transport ATPase, P-type 1 (EC 3.6.3.2)	PMS	99,466.49	898	5.64	b0882
	72	*miaB*	X	*yleA*	Required for methylthiolation step of the modified tRNA nucleoside N6-(4-hydroxyisopentenyl)-2-methylthioadenosine; contains an iron-sulfur center, SAM-dependent	UNC	53,662.96	474	5.20	b2592
101	12	*mukB*	E	--	Cell division protein mukB	UNC	170,230.18	1486	5.24	b3740
107	42	*narL*	N	*frdR*	Nitrate/nitrite response regulator protein narL	MSM	23,926.79	216	5.73	b3212
109	13	*nrdA*	E	*dnaF*	Ribonucleoside-diphosphate reductase 1 alpha chain (EC 1.17.4.1)	NCM	85,775.28	761	5.79	b2508
112	43	*nuoG*	N	--	NADH dehydrogenase I chain G (EC 1.6.5.3)	CHM	100,299.20	910	5.85	b3417
113	44	*nusG*	N	--	Transcription antitermination protein nusG	AAM	20,531.51	181	6.34	b4019
119	45	*pflA*	N	*act*	Pyruvate formate-lyase 1 activating enzyme (EC 1.97.1.4)	AAM	28,204.29	246	6.00	b3940
120	46	*pflB*	N	*pfl*	Formate acetyltransferase 1 (EC 2.3.1.54)	NAM	85,357.21	760	5.69	b1114
122	47	*phoP*	N	--	Transcriptional regulatory protein phoP	MTR	25,535.22	223	5.10	b4242
126	48	*prfC*	N	*tos*	Peptide chain release factor 3	SMC	59,574.08	529	5.65	b1221
127	49	*proB*	N	*pro(2)*	Glutamate 5-kinase (EC 2.7.2.11)	BEN	39,056.52	367	6.07	b2283

129	50	*pta*	N	--	Phosphate acetyltransferase (EC 2.3.1.8)	RCD	77,172.09	714	5.28	b3982
130	51	*purB*	N	*ade(h)*	Adenylosuccinate lyase (EC 4.3.2.2)	CHM	51,542.81	456	5.68	b0902
	73	*purH*	X	--	Phosphoribosylaminoimidazolecarboxamide formyltransferase; purine synthesis	CHM	57,329.21	529	5.53	b0903
133	52	*putA*	N	*poaA, putC*	Bifunctional putA protein	SMC	143,815.16	1320	5.69	b1130
135	53	*pyrB*	N	--	Aspartate carbamoyltransferase catalytic chain (EC 2.1.3.2)	PMS	34,427.36	311	6.12	b4375
137	54	*pyrI*	N	--	Aspartate carbamoyltransferase regulatory chain	AAM	17,120.63	153	6.90	b0242
139	2	*rho*	E	*nitA, nusD, psuA, rnsC, sbaA, sun, tsu ams,*	Transcription termination factor rho	MSM	47,004.21	419	6.75	b2297
146	14	*rne*	E	*hmp1, smbB*	Ribonuclease E (EC 3.1.4.-)	NCM	118,196.73	1061	5.48	b1131
148	15	*rplC*	E	--	50S ribosomal protein L3	AAM	22,243.52	209	9.90	b1014
152	16	*rplY*	E	--	50S ribosomal protein L25	NCM	10,693.44	94	9.60	b4245
153	17	*rpoB*	E	*ftsR, groN, nitB, rif, ron, sdgB, stl, stv, tabD*	DNA-directed RNA polymerase beta chain (EC 2.7.7.6)	RCD	150,632.35	1342	5.15	b4244
154	18	*rpoC*	E	*tabB*	DNA-directed RNA polymerase beta' chain (EC 2.7.7.6)	RCD	155,160.25	1407	6.67	b1608
161	55	*rstA*	N	*urpT*	Transcriptional regulatory protein rstA	SMC	26,703.82	242	5.42	b0723
162	56	*sdhA*	N	--	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	CHM	64,421.84	588	5.85	b4239

169	19	*tdk*	E	--	Thymidine kinase (EC 2.7.1.21)	UNC	23,456.58	205	5.98	b1178
171	20	*thrS*	E	--	Threonyl-tRNA synthetase (EC 6.1.1.3)	UNC	74,014.30	642	5.80	b2294
172	66	*tpiA*	X	*tpi*	Triosephosphate isomerase (EC 5.3.1.1)	UNC	26,971.81	255	5.64	b2597
173	57	*treC*	N	*olgH, treE*	Trehalose-6-phosphate hydrolase (EC 3.2.1.93)	UNC	63,837.67	551	5.51	b3203
175	21	*tufA*	E	*kirT, pulT*	Elongation factor Tu	UNC	43,283.55	394	5.30	b3499
176	22	*tufB*	E	*kirT, pulT*	Elongation factor Tu	UNC	43,313.58	394	5.30	b3764
191	23	*ycfP*	E	--	Hypothetical protein ycfP	MTR	21,226.18	199	6.13	b4391
192	58	*ycgK*	N	--	Protein ycgK precursor	NAM	14,905.82	133	9.50	b3699
193	24	*yciK*	E	--	Hypothetical oxidoreductase yciK (EC 1.-.-.-)	CHM	27,932.91	252	7.67	b3919
204	59	*yfbU*	N	--	Protein yfbU	--	19,536.20	170	6.07	--
209	60	*yfiA*	N	*raiA*	Protein yfiA	--	12,784.59	113	6.20	--
213	25	*yhbG*	E	*lptA*	Probable ABC transporter ATP-binding protein yhbG	--	26,800.65	241	5.64	--
214	61	*yhbH*	N	*hpf*	Probable sigma(54) modulation protein	--	10,750.25	95	6.50	--
	74	*yhbW*	N	--	--	--	37,129.34	335	5.99	--
218	62	*yhiR*	N	--	Hypothetical protein yhiR	--	31,941.71	280	8.59	--
220	63	*yifE*	N	--	Protein yifE	--	13,133.58	112	6.10	--
222	64	*yjjK*	N	--	ABC transporter ATP-binding protein yjjK	--	62,442.88	555	5.43	--

Elution Contaminant Pool: All Gene Products Shown to Bind IMAC Columns if $Co^{(II)}$, $Zn^{(II)}$, or $Ni^{(II)}$ are used:

#	Gene	E/N	Alt. Gene Symbol	Protein Description & E.C. #	Pathway	length	MW (calc)	pI (calc)	Transcription Start	Blattner ID
1	*accC*	X	*fabG*	Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA carboxylase (EC 6.4.1.2))	LPC	449	49,321	6.65	3403554	b3256
2	*aceE*	N	--	Pyruvate dehydrogenase E1 component (EC 1.2.4.1)	CHM	887	99,668	5.46	123017	b0114
3	*aceF*	E	--	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	NCM	630	66,096	5.09	125695	b0115
4	*acnB*	N	*yacI, yacJ*	Aconitate hydratase 2 (EC 4.2.1.3)	CHM	865	93,498	5.24	131615	b0118
5	*adhE*	N	*adhC, ana*	Aldehyde-alcohol dehydrogenase	MSM	891	96,127	6.32	1297344	b1241
6	*ahpF*	N	--	Alkyl hydroperoxide reductase subunit F (EC 1.6.4.-)	MSM	531	56,177	5.47	638946	b0606
7	*alaS*	E	*act, ala-act, lovB*	Alanyl-tRNA synthetase (EC 6.1.1.7)	PMS	876	96,032	5.53	2820033	b2697
8	*arcA*		*cpxC, dye, fexA, msp, seg, sfrA*	Aerobic respiration control protein	SMC	238	27,292	5.21	--	b4401
9	*argE*	X	--	Acetylornithine deacetylase (EC 3.5.1.16), MULTIPLE COPIES ON THE C'SOME	AAM	383	--	--	4152426	b3957

10	*aroH*	E	--	Phospho-2-dehydro-3-deoxyheptonate aldolase, Trp-sensitive (EC 4.1.2.15)	AAM	348	38,735	6.42	1786459	b1704
11	*artP*	N	--	Arginine transport ATP-binding protein artP	MTR	242	27,022	6.17	902957	b0864
12	*asnA*	N	--	Aspartate--ammonia ligase (EC 6.3.1.1)	AAM	330	36,651	5.45	3924783	b3744
13	*asnB*	N	--	Asparagine synthetase B [glutamine-hydrolyzing] (EC 6.3.5.4)	AAM	554	62,659	5.55	698400	b0674
14	*atpC*	E	*papG, uncC*	ATP synthase epsilon chain (EC 3.6.3.14)	BEN	139	15,068	5.46	3913600	b3731
15	*atpG*	N	*papC, uncG*	ATP synthase gamma chain (EC 3.6.3.14)	BEN	287	31,577	8.84	3915893	b3733
16	*basR*	N	*pmrA*	Transcriptional regulatory protein basR/pmrA	SMC	222	25,031	5.66	4331528	b4113
17	*bcp*	N	--	Bacterioferritin comigratory protein	UNC	156	17,634	5.03	2598498	b2480
18	*can*		--	MULTIPLE COPIES ON THE C'SOME						
19	*carA*	N	*arg, cap, pyrA*	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5)	AAM	382	41,431	5.91	29651	b0032
20	*carB*	N	*cap, pyrA*	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	AAM	1073	117,842	5.22	30817	b0033
21	*cbpA*	N	--	DnaK co-chaperone; also binds curved DNA	UNC	306	34,455	6.33	1062998	b1000
22	*cbpM*		*yccD*	Modulator of CbpA co-chaperone activity	--	101	11,512	5.23	--	--
23	*chbR*		*celD*	--	--	280	32,968	8.41	--	--
24	*clpA*	N	*lopD*	ATP-dependent clp protease ATP-binding subunit clpA	PMS	758	84,207	5.91	922487	b0882
25	*clpB*	N	*htpM*	ATP-dependent protease and chaperone; protein disaggregation	UNC	857	95,585	5.37	2732193	b2592

				chaperone; role in de novo protein folding under mild stress conditions						
26	*clpP*	N	*lopP, wseA*	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92)	PMS	207	23,187	5.52	455901	b0437
27	*coaA*	E	*panK, rts, ts-9*	Pantothenate kinase (EC 2.7.1.33)	NCM	316	36,360	6.32	4172605	b3974
28	*corA*	E	--	Magnesium and cobalt uptake transporter protein corA	MTR	316	36,590	4.64	3999038	b3816
29	*cpxR*	N	*yiiA*	Transcriptional regulatory protein cpxR	SMC	232	26,312	5.39	4103251	b3912
30	*crp*	X	*cap, csm*	Catabolite gene activator, cAMP receptor protein	RCD	210	23,640	8.38	3483757	--
31	*crr*	N	*gsr, iex tgs, treD*	PTS system, glucose-specific IIA component (EC 2.7.1.69)	MTR	169	18,251	4.73	2533854	b2417
32	*csrA*	X	*zfiA*	Carbon storage regulator	RCD	61	6,856	8.16	2817168	b2696
33	*cueO*	N	*yacK, cuiD (S.t.)*	Blue copper oxidase cueO precursor	UNC	516	56,556	6.28	137083	b0123
34	*cusF*		*cusX, silF (S.t.), ylcC*	Silver and copper efflux, periplasmic binding protein; confers copper and silver resistance; CusF binds both Cu(I) and Ag(I) with high affinity to, but does not bind Cu(II)	--	110	12,251	6.97	--	--
35	*cydA*	E	--	Cytochrome d (bd-I) terminal oxidase subunit I; up regulated in biofilms and microaerobic conditions; aerobically repressed by H-NS; anaerobically repressed by Fnr	BEN	523	58,205	6.35	770678	b0733
36	*cysD*	N	--	Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4)	MSM	302	35,188	7.80	2874352	b2752

37	*dapB*	E	--	Dihydrodipicolinate reductase (EC 1.3.1.26)	AAM	273	28,757	5.45	28374	b0031
38	*dfp*	E	*coaBC*	DNA/pantothenate metabolism flavoprotein	NCM	430	43,438	7.06	3810286	b3639
39	*dnaE*	E	*polC, sdgC*	DNA polymerase III alpha subunit (EC 2.7.7.7)	NAM	1160	129,905	5.16	205126	b0184
40	*dnaK*	E	*groPAB, groPC, groPF, grpC, grpF, seg*	Chaperone protein dnaK	PMS	638	69,115	4.83	12163	b0014
41	*dnaN*	X	--	DNA polymerase III, beta chain (EC 2.7.7.7)	NAM	366	40,587	5.25	3879949	b3701
42	*dps*	N	*pexB, vtm*	DNA protection during starvation protein	UNC	167	18,695	5.72	848134	b0812
43	*edd*	N	--	Phosphogluconate dehydratase (EC 4.2.1.12)	CHM	603	64,639	5.93	1932628	b1851
44	*elaB*	N	*yfbD*	ElaB protein, Function unknown	UNC	101	11,306	5.35	2379047	b2266
45	*fabI*	E	--	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9)	LPC	262	--	--	1349063	b1288
46	*fbp*	N	--	Fructose-1,6-bisphosphatase (EC 3.1.3.11)	CHM	332	--	--	4453183	b4232
47	*folE*	E	--	GTP cyclohydrolase I (EC 3.5.4.16)	NCM	222	24,831	6.80	2241672	b2153
48	*folX*	N	--	D-erythro-7,8-dihydroneopterin triphosphate epimerase (EC 5.-.-.-)	NCM	120	14,082	6.51	2419345	b2303
49	*fur*	N	--	Ferric uptake regulation protein	RCD	148	16,795	5.68	709869	b0683
50	*fusA*	E	*far, fus*	Elongation factor G	PMS	704	77,581	5.24	3471151	b3340
51	*galF*	E	--	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	CHM	297	--	--	2112349	b2042
52	*galU*	E	--	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	CHM	302	--	--	1290680	b1236
53	*gapA*	E	--	Glyceraldehyde 3-phosphate	CHM	331	35,532	6.61	1860795	b1779

				dehydrogenase A (EC 1.2.1.12)						
54	*gatA*	N	--	PTS system, galactitol-specific IIA component (EC 2.7.1.69)	MTR	150	--	--	2173069	b2094
55	*gatD*	E	--	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	CHM	346	--	--	2170895	b2091
56	*gcvP*	N	--	Glycine dehydrogenase [decarboxylating] (EC 1.4.4.2)	AAM	957	--	--	3047061	b2903
57	*gdhA*	N	--	NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	MSM	447	--	--	1840395	b1761
58	*gidA*	N	--	Glucose inhibited division protein A	UNC	629	--	--	3923261	b3741
59	*gidB*	N	--	Glucose inhibited division protein B	UNC	207	23,431	6.06	3921308	b3740
60	*glgB*	N	--	1,4-alpha-glucan branching enzyme (EC 2.4.1.18)	CHM	728	84,337	5.91	3571135	b3432
61	*glk*	N	--	Glucokinase (EC 2.7.1.2)	CHM	321	34,723	6.06	2507446	b2388
62	*glmS*	E	--	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	CHM	609	66,894	5.56	3911296	b3729
63	*glnD*	E	--	[Protein-PII] uridylyltransferase (EC 2.7.7.59)	SMC	890	102,390	6.22	188650	b0167
64	*gltB*	N	--	Glutamate synthase [NADPH] large chain precursor (EC 1.4.1.13)	MSM	1517			3352267	b3212
65	*gltX*	E	--	Glutamyl-tRNA synthetase (EC 6.1.1.17)	PMS	471			2518692	b2400
66	*glyA*	E	--	Serine hydroxymethyltransferase (EC 2.1.2.1)	AAM	417			2683527	b2551
67	*groL*	E	--	60 kDa chaperonin	PMS	548			4368603	b4143
68	*gst*	N	--	Glutathione S-transferase (EC 2.5.1.18)	BEN	201			1712401	b1635
69	*guaA*	N	--	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2)	NCM	525			2630555	b2507
70	*guaB*	N	--	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	NCM	488			2632090	b2508
71	*gyrA*	E	--	DNA gyrase subunit A (EC	NAM	875			2337440	b2231

				5.99.1.3)						
72	*gyrB*	X	--	DNA gyrase subunit B (EC 5.99.1.3)	NAM	804			3877747	b3699
73	*hemD*	E	--	Uroporphyrinogen-III synthase (EC 4.2.1.75)	NCM	246			3987441	b3804
74	*hemE*	E	--	Uroporphyrinogen decarboxylase (EC 4.1.1.37)	NCM	354			4195294	b3997
75	*hemN*	N	--	Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-)	NCM	459			4049619	b3867
76	*hepA*	N	--	RNA polymerase associated protein	UNC	968			63264	b0059
77	*hinT*		--	--	--	--	--	--	--	--
78	*hisB*	N	--	Histidine biosynthesis bifunctional protein hisB	AAM	356			2091487	b2022
79	*hisH*	N	--	Imidazole glycerol phosphate synthase subunit hisH (EC 2.4.2.-)	AAM	196			2092557	b2023
80	*hrpA*	N	--	ATP-dependent helicase hrpA	UNC	1281			1481142	b1413
81	*hslR*	N	--	Heat shock protein 15	SMC	133			3526986	b3400
82	*hslV*	N	--	ATP-dependent protease hslV (EC 3.4.25.-)	SMC	176			4119867	b3932
83	*htpG*		--	Chaperone protein htpG	SMC	624				b0473
84	*hybC*	N	--	Hydrogenase-2 large chain precursor (EC 1.18.99.1)	BEN	567			3141004	b2994
85	*ilvD*	N	--	Dihydroxy-acid dehydratase (EC 4.2.1.9)	AAM	605			3951132	b3771
86	*infA*	?	--	Translation initiation factor IF-1	UNC	72			925666	b0884
87	*iscS*	E	--	Cysteine desulfurase (EC 4.4.1.-)	NCM	412			2659575	b2530
88	*ispD*	E	--	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)	LPC	236			2870513	b2747
89	*kdsA*	N*	--	2-dehydro-3-deoxyphosphooctonate aldolase (EC 4.1.2.16)	CHM	284			1267388	b1215
90	*leuC*	N	--	3-isopropylmalate dehydratase large	AAM	466			80864	b0072

				subunit (EC 4.2.1.33)						
91	*lexA*	X	--	LexA repressor (EC 3.4.21.88)	NAM	202			4254694	b4043
92	*lpdA*	E	--	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	CHM	474			127912	b0116
93	*lysS*	X	--	Lysyl-tRNA synthetase (EC 6.1.1.6)	NAM	505			3033194	b2890
94	*lysU*	N	--	Lysyl-tRNA synthetase, heat inducible (EC 6.1.1.6)	NAM	505			4352295	b4129
95	*maa*	N	--	Maltose O-acetyltransferase (EC 2.3.1.79)	CHM	183			479142	b0459
96	*malP*	N	--	Maltodextrin phosphorylase (EC 2.4.1.1)	CHM	797			3550106	b3417
97	*map*	E	--	Methionine aminopeptidase (EC 3.4.11.18)	PMS	264			189506	b0168
98	*mdoB*	X	--	Phosphoglycerol transferase I (EC 2.7.8.20)	LPC	750			4596971	b4359
99	*mdoH*	N	--	Periplasmic glucans biosynthesis protein mdoH	CHM	847			1110086	b1049
100	*metH*	N	--	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	AAM	1227			4221407	b4019
101	*metL*	N	--	Bifunctional aspartokinase/homoserine dehydrogenase II	AAM	810			4127415	b3940
102	*mfd*	N	--	Transcription-repair coupling factor	NAM	1148			1173187	b1114
103	*mgtA*	N	--	Mg(2+) transport ATPase, P-type 1 (EC 3.6.3.2)	MTR	898			4465199	b4242
104	*miaB*		--	--	--	--	--	--	--	--
105	*minC*	N	--	Septum site-determining protein minC	RCD	231			1225303	b1176
106	*mpaA*		--	--	--	--	--	--	--	--
107	*mpl*	N	--	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (EC 6.3.2.-)	LPC	457			4453359	b4233

108	*mrp*	N	--	Mrp protein	UNC	379	2192218	b2113
109	*msrA*	N	--	Peptide methionine sulphoxide reductase msrA (EC 1.8.4.6)	PMS	212	4439753	b4219
110	*mukB*	E	--	Cell division protein mukB	RCD	1486	975549	b0924
111	*murA*	N	--	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	LPC	419	3334135	b3189
112	*murC*	E	--	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8)	LPC	491	100765	b0091
113	*murE*	E	--	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (EC 6.3.2.13)	LPC	495	93166	b0085
114	*nagB*	N	--	Glucosamine-6-phosphate isomerase (EC 3.5.99.6)	CHM	266	702834	b0678
115	*nagD*	N	--	NagD protein	UNC	250	699549	b0675
116	*narL*	N	--	Nitrate/nitrite response regulator protein narL	SMC	216	1275052	b1221
117	*nikR*	N	--	Nickel responsive regulator	RCD	133	3616219	b3481
118	*nrdA*	E	--	Ribonucleoside-diphosphate reductase 1 alpha chain (EC 1.17.4.1)	NCM	761	2342885	b2234
119	*nrdD*	N	--	Anaerobic ribonucleoside-triphosphate reductase (EC 1.17.4.2)	NCM	712	4460234	b4238
120	*nudC*	N	--	NADH pyrophosphatase (EC 3.6.1.-)	MSM	257	4194481	b3996
121	*nuoG*	N	--	NADH dehydrogenase I chain G (EC 1.6.5.3)	BEN	910	2398191	b2283
122	*nusG*	N	--	Transcription antitermination protein nusG	RCD	181	4175322	b3982
123	*ompA*	N	--	Outer membrane protein A precursor	SMC	346	1019276	b0957
124	*ompF*	N	--	Outer membrane protein F precursor	SMC	362	986205	b0929
125	*ompR*	N	--	Transcriptional regulatory protein	SMC	239	3534222	b3405

				ompR						
126	*pdxB*	N	--	Erythronate-4-phosphate dehydrogenase (EC 1.1.1.-)	NCM	378			2435871	b2320
127	*pdxH*	N	--	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	NCM	218			1716031	b1638
128	*pflA*	N	--	Pyruvate formate-lyase 1 activating enzyme (EC 1.97.1.4)	CHM	246			950303	b0902
129	*pflB*	N	--	Formate acetyltransferase 1 (EC 2.3.1.54)	CHM	760			952777	b0903
130	*phoB*	N	--	Phosphate regulon transcriptional regulatory protein phoB	RCD	229			416366	b0399
131	*phoP*	N	--	Transcriptional regulatory protein phoP	SMC	223			1189670	b1130
132	*ppiA*	N	--	Peptidyl-prolyl cis-trans isomerase A precursor (EC 5.2.1.8)	PMS	190			3489934	b3363
133	*prfA*	X	--	Peptide chain release factor 1	PMS	360			1264235	b1211
134	*prfB*	?	--	Peptide chain release factor 2	PMS	365			3034302	b2891
135	*prfC*	N	--	Peptide chain release factor 3	PMS	529			4606983	b4375
136	*prkB*		*prk?, yhfF*	Putative phosphoribulokinase		289	32,344	6.18	--	--
137	*proB*	N	--	Glutamate 5-kinase (EC 2.7.2.11)	AAM	367			259612	b0242
138	*prsA*	E	--	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	CHM	315			1261098	b1207
139	*pta*	N	--	Phosphate acetyltransferase (EC 2.3.1.8)	MSM	714			2412767	b2297
140	*purB*	N	--	Adenylosuccinate lyase (EC 4.3.2.2)	NCM	456			1191209	b1131
141	*purH*		--	--	--	--	--	--	--	--
142	*purL*	N	--	Phosphoribosylformylglycinamidine synthase (EC 6.3.5.3)	NCM	1295			2693563	b2557
143	*purU*	E	--	Formyltetrahydrofolate deformylase (EC 3.5.1.10)	NCM	280			1287847	b1232
144	*putA*	N	--	Bifunctional putA protein	AAM	1320			1078105	b1014
145	*pykA*	N	--	Pyruvate kinase II (EC 2.7.1.40)	CHM	480			1935673	b1854

146	*pyrB*	N	--	Aspartate carbamoyltransferase catalytic chain (EC 2.1.3.2)	NCM	311			4469969	b4245
147	*pyrG*	E	--	CTP synthase (EC 6.3.4.2)	NCM	545			2907688	b2780
148	*pyrI*	N	--	Aspartate carbamoyltransferase regulatory chain	RCD	153			4469021	b4244
149	*qseB*		*ygiX*	Quorum sensing E.coli		219	2,468	6.54	--	--
150	*relA*	N	--	GTP pyrophosphokinase (EC 2.7.6.5)	MSM	744			2911673	b2784
151	*rffE*		--	--	--	--	--	--	--	--
152	*rho*	?	--	Transcription termination factor rho	RCD	419			3964032	b3783
153	*ribD*	E	--	Riboflavin biosynthesis protein ribD	NCM	367			432679	b0414
154	*ribE*	E	--	Riboflavin synthase alpha chain (EC 2.5.1.9)	NCM	213			1741266	b1662
155	*rimM*	E	*yfiA*	16S rRNA processing protein rimM	NAM	185	20,605	4.61	2743947	b2608
156	*rnb*	N	--	Exoribonuclease II (EC 3.1.13.1)	NAM	644			1346936	b1286
157	*rnc*	E	--	Ribonuclease III (EC 3.1.26.3)	NAM	226			2702083	b2567
158	*rne*	E	--	Ribonuclease E (EC 3.1.4.-)	NAM	1061			1143590	b1084
159	*rph*	N	--	Ribonuclease PH (EC 2.7.7.56)	NAM	228			3814176	b3643
160	*rplC*	E	--	50S ribosomal protein L3	PMS	209			3450563	b3320
161	*rplF*	E	--	50S ribosomal protein L6	PMS	177			3443777	b3305
162	*rplN*	E	--	50S ribosomal protein L14	PMS	123			3445786	b3310
163	*rplS*	E	--	50S ribosomal protein L19	PMS	115			2742550	b2606
164	*rplY*	E	--	50S ribosomal protein L25	PMS	94			2280537	b2185
165	*rpoB*	E	--	DNA-directed RNA polymerase beta chain (EC 2.7.7.6)	NAM	1342			4178823	b3987
166	*rpoC*	E	--	DNA-directed RNA polymerase beta' chain (EC 2.7.7.6)	NAM	1407			4182928	b3988
167	*rpoD*	E	--	RNA polymerase sigma factor rpoD	RCD	613	70,263	4.69	3210688	b3067
168	*rpsA*	N*	--	30S ribosomal protein S1	PMS	557			961218	b0911
169	*rpsB*	E	--	30S ribosomal protein S2	PMS	241			189874	b0169
170	*rpsC*	N	--	30S ribosomal protein S3	PMS	233			3447520	b3314
171	*rpsF*	E	--	30S ribosomal protein S6	PMS	131			4422696	b4200
172	*rpsJ*	E	*nusE*	30S ribosomal protein S10	PMS	103	11,736	9.68	3450907	b3321

173	*rstA*	N	--	Transcriptional regulatory protein rstA	RCD	242			1680174	b1608
174	*sdhA*	N	--	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	SMC	588			755130	b0723
175	*slyD*	E	--	FKBP-type peptidyl-prolyl cis-trans isomerase slyD (EC 5.2.1.8)	PMS	196			3476134	b3349
176	*speA*	N	--	Biosynthetic arginine decarboxylase (EC 4.1.1.19)	AAM	658			3083930	b2938
177	*speD*	N	--	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50)	MSM	264			135582	b0120
178	*speG*	N	--	Spermidine N(1)-acetyltransferase (EC 2.3.1.57)	AAM	186			1654208	b1584
179	*sucA*	E	--	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	CHM	933			757929	b0726
180	*suhB*	E	--	Inositol-1-monophosphatase (EC 3.1.3.25)	LPC	267			2661462	b2533
181	*tdk*	E	--	Thymidine kinase (EC 2.7.1.21)	NCM	205			1292750	b1238
182	*thrA*	N	--	Bifunctional aspartokinase/homoserine dehydrogenase I	AAM	820			337	b0002
183	*thrS*	E	--	Threonyl-tRNA synthetase (EC 6.1.1.3)	PMS	642			1800594	b1719
184	*tiaE*		--	--	--	--	--	--	--	--
185	*tpiA*	X	--	Triosephosphate isomerase (EC 5.3.1.1)	CHM	255			4109087	b3919
186	*treC*	N	--	Trehalose-6-phosphate hydrolase (EC 3.2.1.93)	CHM	551			4462283	b4239
187	*trmC*		--	--	--	--	--	--	--	--
188	*trpS*	E	--	Tryptophanyl-tRNA synthetase (EC 6.1.1.2)	PMS	334			3511276	b3384
189	*truD*		--	--	--	--	--	--	--	--
190	*tufA*	E	--	Elongation factor Tu	PMS	394			3468966	b3339

191	*tufB*	E	--	Elongation factor Tu	PMS	394			4173523	b3980
192	*ugd*	N	--	UDP-glucose 6-dehydrogenase (EC 1.1.1.22)	SMC	388			2097635	b2028
193	*uhpA*	N	--	Transcriptional regulatory protein uhpA	SMC	196			3848353	b3669
194	*uspG*		--	--	--	--	--	--	--	--
195	*uup*	N	--	ABC transporter ATP-binding protein uup	MTR	635			1009187	b0949
196	*yacF*	N	--	Hypothetical protein yacF	UNC	247	28,292	6.31	112599	b0102
197	*yahK*	N	--	Hypothetical zinc-type alcohol dehydrogenase-like protein yahK	UNC	349			342108	b0325
198	*yajD*	N	--	Hypothetical protein yajD	UNC	115	13,364	6.14	429829	b0410
199	*ybeD*	N	--	Hypothetical protein ybeD	UNC	87			661865	b0631
200	*ybfF*	N	--	Putative esterase/lipase ybfF (EC 3.1.-.-)	UNC	254	28,437	5.86	712025	b0686
201	*ybgJ*	N	--	Hypothetical protein ybgJ	UNC	218			742816	b0711
202	*ybiB*	N	--	Hypothetical protein ybiB	UNC	320			834471	b0800
203	*ybiV*	N	--	Hypothetical protein ybiV	UNC	271			859251	b0822
204	*ycbY*	N	*rlmL*	Hypothetical protein ycbY	UNC	702	78,854	8.96	1007067	b0948
205	*yccX*	N	--	Putative acylphosphatase (EC 3.6.1.7)	UNC	92			1029287	b0968
206	*ycdW*	E	--	Putative 2-hydroxyacid dehydrogenase ycdW	UNC	325			1097070	b1033
207	*ycfP*	E	--	Hypothetical protein ycfP	UNC	199			1164309	b1108
208	*ycgK*	N	--	Protein ycgK precursor	UNC	133			1226695	b1178
209	*yciK*	E	--	Hypothetical oxidoreductase yciK (EC 1.-.-.-)	UNC	252			1327136	b1271
210	*ydcF*	N	--	Protein ydcF	UNC	266			1485259	b1414
211	*ydcY*	N	--	Hypothetical protein ydcY	UNC	77			1515672	b1446
212	*ydeA*		--	--	--	--	--	--	--	--
213	*ydfG*	N	--	Probable oxidoreductase ydfG (EC 1.-.-.-)	UNC	248			1625541	b1539
214	*ydhF*	N	--	Hypothetical oxidoreductase ydhF	UNC	298			1723656	b1647

				(EC 1.-.-.-)						
215	*ydiI*	E	--	Hypothetical protein ydiI	UNC	136			1763656	b1686
216	*yebU*	N	--	Hypothetical protein yebU	UNC	481			1918241	b1835
217	*yecP*	E	--	Hypothetical protein yecP	UNC	323			1951466	b1871
218	*yegQ*	N	--	Putative protease yegQ (EC 3.4.-.-)	UNC	453			2163690	b2081
219	*yeiE*	N	--	Hypothetical transcriptional regulator ; probable positive regulator of lysP transcription	UNC	293	32,724	6.07	2247638	b2157
220	*yfaY*		--	--	--	--	--	--	--	--
221	*yfbG*	N	--	Hypothetical protein yfbG	UNC	660			2366059	b2255
222	*yfbU*	N	--	Protein yfbU	UNC	170			2410632	b2294
223	*yfcG*	N	--	Hypothetical GST-like protein yccG	UNC	215			2418641	b2302
224	*yffB*	E	--	Protein yffB	UNC	118			2589267	b2471
225	*yfgD*	N	--	Protein yfgD	UNC	119			2615598	b2495
226	*yfgM*	N	--	Hypothetical protein yfgM	UNC	206	22,176	5.07	2637303	b2513
227	*yfhQ*		--	--	--	--	--	--	--	--
228	*yfiA*	N	--	Protein yfiA	UNC	113			2735174	b2597
229	*ygdH*	N	--	Hypothetical protein ygdH	UNC	454			2924330	b2795
230	*ygiN*	N	--	Protein ygiN	UNC	104			3171158	b3029
231	*ygjF*		*mug, dug*	Uracil DNA-glycosylase, G:U mismatch-specific, dsDNA-specific; excises uracil from DNA by base flipping mechanism; required for mutation avoidance in stationary phase; up-regulated in stationary phase	--	168	18,673	9.17	--	--
232	*yhaJ*	N	--	Hypothetical transcriptional regulator yhaJ	UNC	298			3251854	b3105
233	*yhbG*	E	--	Probable ABC transporter ATP-binding protein yhbG	UNC	241			3341585	b3201
234	*yhbH*	N	--	Probable sigma(54) modulation protein	UNC	95			3343814	b3203
235	*yhbW*	N	--	Hypothetical protein yhbW	UNC	335			3301089	b3160

236	*yhfT*	X	--	Hypothetical protein yhfT	UNC	434	46,516	6.53	3504974	b3377
237	*yhhX*	N	--	Putative oxidoreductase yhhX (EC 1.-.-.-)	UNC	345			3578436	b3440
238	*yhiQ*		--	Function unknown; in Salmonella, yhiQ is the second gene in prlC operon and in the heat shock regulon; putative SAM-dependent methyltransferase	--	250	26,949	6.60	--	--
239	*yhiR*	N	--	Hypothetical protein yhiR	UNC	280			3643015	b3499
240	*yieN*	N	--	Hypothetical protein yieN	UNC	506			3928744	b3746
241	*yifE*	N	--	Protein yifE	UNC	112			3945709	b3764
242	*yjbR*	N	--	Protein yjbR	UNC	118			4268237	b4057
243	*yjjK*	N	--	ABC transporter ATP-binding protein yjjK	MTR	555			4628091	b4391
244	*yjtD*		--	--	--	--	--	--	--	--
245	*yliG*	E	--	Hypothetical protein yliG	UNC	441			877258	b0835
246	*yncB*	N	--	Putative NADP-dependent oxidoreductase yncB (EC 1.-.-.-)	UNC	376			1516958	b1449
247	*yncE*	E	--	Hypothetical protein yncE precursor	UNC	353			1521331	b1452
248	*yneH*	N	--	Probable glutaminase yneH (EC 3.5.1.2)	UNC	308			1611275	b1524
249	*yqcD*	N	--	Hypothetical protein yqcD	UNC	282			2923370	b2794
250	*yraL*	N	--	Putative methyltransferase, function unknown	UNC	286	31,348	5.83	3290976	b3146
251	*yrdA*	N	--	Protein yrdA	UNC	256			3426657	b3279
252	*ytfP*	N	--	Hypothetical protein ytfP	UNC	113			4445471	b4222
253	*zwf*	E	--	Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	CHM	491			1934338	b1852