COMPARISON OF THE MERCURY INDUCED PROTEOMES OF *Escherichia coli* MG1655 WITH AND WITHOUT THE NR1 PLASMID

By

ERIKA MICHELLE ZINK

A thesis submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE IN CHEMISTRY

WASHINGTON STATE UNIVERSITY
Department of Chemistry

AUGUST 2009

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of ERIKA MICHELLE ZINK find it satisfactory and recommend that it be accepted.

_____

Mary S. Lipton, Ph.D, Chair

_____

Karen E. Grant, M.S.

_____

Kathleen McAteer, Ph.D

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Mary S. Lipton for her initiation of this project and her support throughout. I would also like to extend my thanks to Anne Summers and Stephen Lavoie from the University of Georgia for the cell cultures and Susan Miller and Ben Pollaco from the University of California, San Francisco for their input on method development. I would like to thank Sam Purvine, from the Pacific Northwest National Laboratory, for his patience and guidance with the computational analyses.

This work was funded by the Department of Energy (Biological Sciences Division). The research was performed at the Environmental Molecular Sciences Laboratory (a national scientific user facility sponsored by the U.S. DOE Office of Biological and Environmental Research) located at the Pacific Northwest National Laboratory, operated by Battelle for the DOE under Contract DE-AC06-76RLO 1830.

COMPARISON OF THE MERCURY INDUCED PROTEOMES OF

Escherichia coli MG1655 WITH AND WITHOUT THE NR1 PLASMID

Abstract

By Erika Michelle Zink, M.S.
Washington State University
August 2009

Chair: Mary S. Lipton

Mercury, a naturally occurring metal, is found in varying amounts throughout the world. Exposure can result in neurotoxicity, nephrotoxicity, and gastrointestinal toxicity with the severity depending on the form and quantity of mercury involved [1]. Increased levels of environmental mercury exist due to anthropogenic productions [2-4], and although it may be impossible to avoid contact with mercury, localized areas of high concentrations can be reduced to tolerable levels through remediation techniques. These processes group into physical, chemical and biological techniques; however, current research is moving toward use of already existing biological mechanisms for remediation of heavy metals. Resistance has been shown in some bacteria to be associated with the NR1 plasmid and the *mer* operon [5].

Here we describe a proteomics based strategy to understand the complex systems of *Escherichia coli* with or without the NR1 plasmid [6, 7]. This is the first large scale, all inclusive *E. coli* mercury-induced proteome identification using high-resolution mass spectrometric proteomic techniques to determine both the qualitative and quantitative protein. We were able to identify proteins associated with the *mer* operon of the NR1 plasmid, and several proteins that contained a bound phenyl mercury, with a high level of certainty. Also identified were proteins affected by the mercury which are involved in protein synthesis. Identification of proteins targeted directly and indirectly by mercury in *E. coli* is still in the beginning stages, but the progress we have shown here is enlightening. Future studies can be expected that will improve upon our techniques, and we can also start to expand into other organisms and cell types.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Dedication

This is dedicated to my husband and best friend, Chad Zink, and to my parents,

Larry Olsen and Dawn and Budd Donelson.

Thank you all for your unwavering encouragement and support.

CHAPTER ONE

INTRODUCTION

Mercury, a naturally occurring metal, can be found in varying concentrations in nearly all corners of the world. Exposure to mercury can result in neurotoxicity, nephrotoxicity, and gastrointestinal toxicity with the severity of each depending on the form and quantity of mercury involved [1]. There are multiple processes for mercury to enter the human body. The most common means include the ingestion of food products that have been treated with mercury containing fungicides or pesticides, and the consumption of seafood or piscivorous birds from contaminated aquatic environments. Two cases of high mercury toxicity, specifically due to methyl mercury, occurred in Minimata, Japan and Iraq [8, 9]. These two incidents provided the current list of neurological symptoms related directly to mercury exposure: parasthesia (loss of sensation in the hands and feet and in areas around the mouth), ataxia (loss of coordination in gait), constriction of visual field, dysarthia (slurred speech), and hearing loss. Other neurological disorders may be linked with exposure to mercury, such as autism, Alzheimer's and Parkinson's disease; however, current research remains contradictory on these findings. Risks may also exist from dental amalgams or from vaccines containing mercury preservatives, yet existing research shows that any relation between these conditions and exposure to mercury is ambiguous [10-12].

Elemental mercury exists at room temperature as a liquid with a high vapor pressure. Alternatively, it can also be in the form of a cation with oxidation states of either +1 (mercurous) or +2 (mercuric), with the latter being more prevalent in the environment [4]. Increased levels of environmental mercury exist due to anthropogenic productions including but not limited to solid waste incineration, mining and mining drainage, past use of pesticides and fungicides that

contained mercury, industrial waste and nuclear weapons manufacture and testing [2-4]. Monatomic mercury in the atmosphere, Hg(0), can be oxidized to Hg(II) by reacting with ozone in the presence of water [13] or by conversion within organisms using a catalase and possibly other peroxidases [14]. Mercuric ions are the least mobile form of mercury and interact well with smaller soil particulates allowing them to remain as a source for conversion to more mobile forms through microbial transformations. Many other abiotic and biotic reactions can lead to the development of inorganic or organic mercury becoming mobilized through the geosphere, leading to contamination of water supplies. Infiltration of the biosphere can then occur and accumulation of high concentrations are found in both aquatic and terrestrial organisms.

The most toxic and mobile form of this metal is methyl mercury (MeHg), which is formed through microbial transformations of Hg(II). The specific targets of mercury within biological systems are not known, however it is believed that the symptoms result from the ability of mercuric ions to bond with and block the functions of sulfur-containing proteins [15-17]. The most vulnerable proteins to Hg(II) appear to be those of the electron transport chain and of the reactive oxygen species (ROS) defense system; moreover, at higher levels of Hg(II), it is believed protein adducts (RS-Hg-SR`) can also be formed. To understand fully which proteins are expressed or repressed and to identify novel proteins involved in mercury toxicity of *E. coli*, it is necessary to study the entire genetic output under contrasting culture conditions.

Although it may be impossible to avoid contact with mercury, it is possible to reduce localized concentrations to tolerable levels through remediation techniques. These processes are grouped into physical, chemical and biological means. Physical remediation can occur *in situ* or *ex situ*, but eventually relies on some form of containment of the contaminated source; whereas chemical treatment is dependent on the conversion of mercury to a less mobile form, usually

Hg(0) which escapes into the atmosphere. The largest drawbacks to either method are the cost and time required [18]. A potentially more cost effective method of remediation is through the use of biological mechanisms already in place for exposure to heavy metals such as mercury. In some bacteria, resistance has been associated with the NR1 plasmid, which contains the *mer* operon that encodes proteins involved in the reduction of Hg(II) to Hg(0). These proteins work to prevent Hg(II) from freely traveling within the cell and binding to thiol-containing proteins by restricting its activity to a few select proteins and subsequent transfer to the mercuric reductase protein, merA [5, 18-21]. The resulting Hg(0) is then released from the cell and into the atmosphere.

Techniques that have been used to study protein expression include measurements of the change in abundance between two conditions in the transcriptome or the proteome [22]. The limitation of studying transcripts is they suggest only potential protein targets. By directly analyzing the proteome we are able to reveal the actual protein targets and the types of damage produced. Early proteomics analyses were conducted by examining the protein spots on 2-dimensional electrophoresis gels (2DE); furthermore, the coupling of this technique with tandem mass spectrometry (MS/MS) has lead to the identification of hundreds of soluble proteins from many organisms [23-25]. Given that 2DE analyses work best with soluble proteins, any membrane bound proteins that may be vulnerable to Hg(II) shock have yet to be analyzed fully. It is also important to note that mercury has not been used in metal stressor studies despite its high concentrations in several environments that directly affect microbes, plants, and higher organisms including humans. This lack of a global proteomic analysis of the response of organisms to mercury is one of the issues we are beginning to resolve with this study.

In order to examine the proteomic responses caused by mercury exposure, we utilized a combination of technologies that provide the benefits of 2DE with none of the shortcomings. The accurate mass and time (AMT) tag approach [6, 7] enabled us to identify relevant proteins and the changes in their abundance in relation to the specific culture conditions. This method, which utilizes liquid chromatography coupled with mass spectrometry, also allowed us to detect lower abundance proteins because it emphasizes those proteins with significant abundance changes as opposed to showing direct comparisons of protein abundances. It is this aspect that could facilitate the potential to analyze prokaryotic and eukaryotic systems without the added expense, complexity, and time consuming methods that are associated with isotopic and derivatization labeling schemes or with 2DE comparisons.

Liquid chromatography is the separation of compounds within a liquid solution under high pressure. The sample is passed through a closed column and the solutes (in this case a mixture of peptides) are separated via chemical interactions with the stationary phase and eluted through the use of a variety of mobile phases. There are two types of liquid chromatography, both determined by the polarity differences in the stationary (packing) material and the solvents used for elution. In normal-phase chromatography, a polar stationary phase is used with a less polar mobile phase; whereas, in reverse-phase chromatography, the packing material is less polar than the elution solvent. For peptides, the method of choice is the latter.

The efficiency, or resolution, of the packing material is determined by the number of theoretical plates, which can be calculated using the van Deemter equation

$$H = A + B/u + C*u \qquad (1)$$

where A is the term used for the multiple flow paths that can be taken through the column, B is the longitudinal diffusion or broadening that occurs away from the concentrated center of a band

on the column, and C is the equilibration time between the solute and the packing material. A smaller particle size (generally 3-10 μm) results in increased resolution because it provides a more even flow rate (reducing A in Equation 1) and by providing shorter equilibration times for the samples with the column by reducing the distance between particles and thus the distance for solute diffusion (Reducing C in Equation 1). The main downfall to using smaller particles, however, is a resistance to solvent flow but this can be overcome with the use of high pressure.

Although there are many detectors available for high performance liquid chromatography (HPLC), I will focus on the mass spectrometer as that was the one used in this study. Mass spectrometry is the process of ionizing species within a mixture and then using an electric or magnetic field to determine their molecular weight. Samples enter the mass spectrometer through the sample inlet and are ionized before they are separated by their mass-to-charge (*m/z*) ratios. They are then detected by transferring their energies to electrical signals in the ion detector where they can then be visualized using computer software. Coupling a liquid chromatograph to a mass spectrometer requires a change from the liquid to gaseous state and must go from ambient pressure to a vacuum. A vacuum is used to prevent collisions with other gaseous particles (which would reduce resolution and sensitivity) and can also help in preventing damage from electrical grounding.

Samples must be introduced into the mass spectrometer without disrupting the internal vacuum. This can be achieved by direct insertion or by direct infusion. In the case of direct insertion, the sample is inserted through the sample inlet and ionized within the mass spectrometer before entering the mass analyzer. With direct infusion, the sample is first ionized then passed through the sample inlet of the mass spectrometer. Prior to the 1980s, direct infusion was not conceivable because liquid chromatography dealt with sample concentrations that were

too low and flow rates that were too high to couple to this type of detector, thus mass spectrometers were limited to use with gas chromatographs [26]. In the late 1980s, electrospray ionization was utilized for the conversion of the liquid eluent to a fine mist of charged particles suitable for analysis with a mass spectrometer [26]. The solution is passed from the column through an emitter tip that is maintained at a strong electric field. This creates a fine mist of highly charged particles from a liquid solution with ever decreasing droplet size. As the droplet size decreases, coulombic repulsion increases due to increased charge density on the surface and eventually the repulsion forces become greater than those of the solvent's surface tension and the ions are released, resulting in a spray that is known as a Taylor Cone [26]. This process can also produce multiply charged species, enabling larger molecules, such as proteins or peptides, to be analyzed via mass spectrometry.

Once the samples have been ionized and enter the mass spectrometer, they pass through the mass analyzer, where the ions are separated based on their mass-to-charge (*m/z*) ratios. Mass analyzers are characterized by their accuracy, resolution, mass range, scan speed, and ability to use tandem analysis. A quadrupole mass analyzer is made from four circular rods set in parallel, with each opposing rod connected electrically. A radio frequency (RF) is applied to each pair of rods and a current voltage is superimposed on the rods. Only those ions of a specific *m/z* will be able to pass through the electrical field to the detector, while all others will follow unstable trajectories and collide with the rods. By varying the electrical current on the rods, the ions can be separated by *m/z*. An ion trap is used to selectively trap a selected ion in the mass analyzer by means of an alternating current voltage, rather than allowing it to pass freely through. Specific ions with an *m/z* less than the prescribed value can then be ejected from the trap and either passed on to another mass analyzer (as in the case of tandem mass spectrometry) or analyzed by

an ion detector. A time-of-flight mass analyzer accelerates the ions through a field-free flight region towards a detector set at a specific electrical current. This results in the separation of the ions by their *m/z* over time. For a Fourier transform-ion cyclotron resonance mass analyzer (FTICR), magnetic fields are used to rotate ions of a specific *m/z* at a unique orbit. The ions are excited to these specific orbits by RF pulses, and it is these excited states that are detectable. The individual frequencies are isolated using a Fourier transform, providing a measure of the abundances of the individual ions present. As the ions are ejected from the mass analyzer, they come into contact with an ion detector, which transfers the kinetic energy of the ions to electrical energy that produces a usable output. The electron multiplier, as an example, is a series of dynodes, each at an increased potential. When an ion hits the detector it causes the first dynode to emit electrons that then travel to the next dynode, and so on. The resulting effect is an approximate 10^6 increase in the current.

The AMT tag strategy utilizes a high resolution capillary LC peptide separation combined with tandem mass spectrometry (MS/MS) and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) where an initial peptide identification is achieved from a MS/MS analysis that is validated and quantitated from a separate FTICR-MS analysis using in-house software [6, 7]. The combination of the two mass spectrometric approaches maintains confidence in peptide identification and enables high-throughput operation, improved dynamic range, and relevant protein abundance change comparisons.

Here we describe a strategy to understand the complex systems of *Escherichia coli* under four growth conditions, either with or without the NR1 plasmid, using AMT tags [6, 7].This is the first large scale, all inclusive *Escherichia coli* mercury induced proteome identification using high-resolution mass spectrometric proteomic techniques to determine both the qualitative and

quantitative protein identification and significant relative abundance determined from absolute peak intensities.

CHAPTER TWO

RESEARCH DESIGN AND METHODOLOGY

2.1 Culture conditions and cell lysis

All cell cultures were prepared by Stephen LaVoie at the University of Georgia as described here. Starter cultures were first created by growing either *E. coli* MG1655(seq) or MG1655(NR1) cells in 20 ml Neidhart-MOPS minimal media with 2 mg/ml of 1% uracil and 1 mg/ml 0.05% thiamine HCl. These cultures were grown in 125-ml Erlenmeyer flasks for 17 hours at 37°C and 250 rpm. Each strain of *E. coli* was grown under the following conditions: untreated, *in vitro* addition of 40 mM iodoacetamide (IAM), *in vivo* addition of 20 μM phenylmercuric acetate (PMA), and the combination of the last two conditions. A total of four conditions for each strain were prepared, resulting in eight total samples. These cultures were then transferred to 2800-ml Fern Bach Flasks containing 600 ml of the same media and incubated for 5.25 hours at 37°C and 250 rpm. For the final 15 minutes, the cultures were divided into half and 20 μM of PMA was added to one of each sample. These were then incubated under the same conditions in 1 L Erlenmeyer flasks to a final OD at 600 of 0.6455 (late log/early stationary phase). The cells were harvested by centrifugation at 10,000 rpm for 10 minutes at 4°C and then resuspended in 50 mM ammonium bicarbonate, pH 7.8. These samples were then split in half and 40 mM IAM was added to one of each before being transferred to

Pacific Northwest National Laboratory (PNNL) where we conducted further processing and analyses.

During all phases of processing, from cellular lysis to tryptic digestion, the cells were kept under non-oxygenating conditions by having all buffers bubbled with nitrogen gas prior to use and nitrogen gas blown over the surface of exposed protein solutions. All chemicals used were obtained from Sigma-Aldrich Company (St. Louis, MO), unless otherwise stated. The cells were first rinsed by centrifuging at 10,000 rpm for 10 minutes at 4°C then resuspended in 100 mM ammonium bicarbonate, pH 8.4, and split into two even samples, one of which was used for analysis and the other was stored for future use. Each sample was then processed through a PBI Barocycler NEP 3229 for ten cycles, holding at 3500 psi for 20 seconds each time. The resulting lysates were then collected and placed on ice to prevent proteolysis. Half of each protein solution was directly processed for global digests by the addition of powdered urea and thiourea to achieve respective concentrations of 7 M and 2 M.

2.2 Sub-cellular fractionation

Enrichment of the soluble and insoluble proteins on the remaining half of each solution was achieved via sub-cellular fractionation using ultracentrifugation. After removing any whole cells via centrifugation at 4000 rpm at 4°C for 2 minutes, the lysates were transferred to polycarbonate centrifuge tubes and spun at 100,000 rpm for 10 minutes at 4°C. The process was repeated once more to ensure maximum enrichment. The pellet (insoluble fraction) was solubilized into a buffer of 7 M urea, 2 M thiourea, 1% 3-[3-(Cholamidopropyl)dimethyl-ammonio]-1-proanesulfonate (CHAPS) in 50 mM ammonium bicarbonate, pH 7.8; while dry

components of urea and thiourea were added to the supernatant (soluble fraction) at concentrations of 7 M and 2 M, respectively.

2.3 Trypsin digestion

The global and sub-cellular fractions were reduced by the addition of dithiothreitol (DTT) to a final concentration of 5 mM for the global and soluble fractions and 10 mM for the insoluble fractions. They were all then incubated for 30 minutes at 60°C, after which the global and soluble fractions were diluted 10-fold with 100 mM ammonium bicarbonate, and the insoluble fractions were diluted with 50 mM ammonium bicarbonate. Enough calcium chloride was then added to obtain a final concentration of 1 mM. The samples were then tryptically digested for 3 hours at 37°C using sequencing grade modified trypsin (Promega, Madison WI) in an approximate ratio of 1:50 (wt/wt) trypsin-to-protein.

2.4 Peptide concentration and cleanup

The global and soluble fractions were desalted by using Supelco (St. Louis, MO) Supelclean C-18 SPE columns with a Supelco vacuum manifold. Three tube volumes of methanol and two tube volumes of 0.1% trifluoroacetic acid (TFA) were used to condition the column resin. After the protein solution had been added to the column, four tube volumes of 5% acetonitrile in 0.1% TFA were used to wash the peptides and the resin was allowed to go to dryness. Elution of the peptides was achieved with one tube volume of an 80% acetonitrile in 0.1% TFA solution. Because CHAPS will bind to C-18 and elute with the peptides, the membrane proteins were desalted using a Supelco Supelclean SCX tubes (St. Louis, MO). The columns were conditioned with two tube volumes each of methanol, 10 mM ammonium formate

in 25% acetonitrile, pH 3.0, 500 mM ammonium formate in 25% acetonitrile, pH 6.8, 10 mM ammonium formate in 25% acetonitrile, then nanopure water, and finally with four tube volumes of 10 mM ammonium formate in 25% acetonitrile, pH 3.0. The peptides were then added to the column after they had been acidified with 20% formic acid to a pH lower than 4.0 and rinsed with four tube volumes of 10 mM ammonium formate in 25% acetonitrile, pH 3.0 prior to allowing the column to go to dryness. The peptides were then eluted with one-and-a-half tube volumes of 80% methanol in 5% ammonium hydroxide. All elutions were concentrated by SpeedVac (Thermosavant, San Jose CA) to 100 ul and peptide concentrations were determined by BCA assay (Pierce, Rockford IL).

2.5 Strong cation exchange and fractionation

A portion of each global trypsin digestion from each of the eight growth conditions was separated with strong cation exchange (SCX) fractionation. The SCX was performed on an Agilent 1100 series HPLC (Santa Clara, CA) equipped with a Unicam 4225 (Genetic Technologies, Miami, FL) variable wavelength detector set at 280 nm. The separations were performed on a 2.1 x 200 mm (5 μm, 300 Å) polysulfoethyl A column (PolyLC, Columbia, MD) preceded by a 2.1 x 10 mm guard column for 70 minutes with the following two mobile phases: (A) 10 mM ammonium formate in 25% acetonitrile, pH 3.0 and (B) 500 mM ammonium formate in 25% acetonitrile, pH 6.8. The fractionation began isocratically for 10 minutes with mobile phase A, switching to 50% mobile phase B for 50 minutes, and then holding at 100% mobile phase B for 10 minutes. A total of twenty-five 0.56-ml fractions were collected for each of the samples, which were subsequently dried in a Savant (ThermoSavant, Marietta, OH) Speed-Vac centrifugal concentrator prior to MS analysis.

2.6 Capillary LC separations

All the peptides from the global, soluble and insoluble fractions, as well as those from the SCX fractionation, were then separated by an automated in-house LC-MS system. The capillary system consists of 100-ml Isco Model 100DM syringe pumps (Isco, Inc., Lincoln, NE), 2-position Valco Valves (Valco Instruments Co., Houston, TX) and a PAL autosampler (Leap Technologies, Carrboro, NC). Separations were accomplished using 150 μm i.d. x 360 μm o.d x 60 cm C-18 reverse-phase packed capillaries (3 μM diameter, 300 Å; Polymico Technologies, Pheonix, AZ) utilizing a two mobile-phase solvent system: (A) 0.2% acetic acid and 0.05% TFA in water and (B) 90% acetonitrile in 0.1% TFA in water. The mobile phases were degassed using an in-line Degassex Model DG4400 vacuum degasser (Phenomenex, Torrence, CA). Flow through the capillary column when equilibrated with 100% mobile phase A was approximately 2 μl/min and had a pressure of 5 kpsi. After the system was equilibrated with 100% of mobile phase A for 20 minutes, the mobile phase selection valve was switched to allow a near exponential gradient of mobile phase B.

The peptides from the global preparation were run in triplicate on an in-house manufactured LC-MS/MS system consisting of 65-ml Isco Model 65D syringe pumps (Isco, Inc., Lincoln, NE), 2-position Vaclo valves (Valco Instruments Co., Houston, TX) and a PAL autosampler (Leap Technologies, Carrboro, NC). This fully automated system utilizes four C-18 reverse-phase fused silica capillary HPLC columns that were also manufactured in-house (60-cm length of 360 um o.d. x 75 um i.d., Polymicro Technologies Inc., Phoenix, AZ). The mobile phases consisted of (A) 0.2% acetic acid and 0.05% TFA in water and (B) 0.1% TFA in 90% acetonitrile/10% water. Both solutions were degassed through utilization of an in-line Alltech

vacuum degasser (Alltech Associates, Inc., Deerfield, IL). The HPLC system was first equilibrated with 100% mobile phase A at a pressure of 10 kpsi; then a mobile phase selection valve was switched 50 min after injection to create a near-exponential gradient as mobile phase B displaced A in a 2.5 ml active mixer. The approximately 20 ul/min of flow was split prior to reaching the injection valve (5 ul sample loop) by a 30-cm length of 360 um o.d. x 15 um i.d. fused silica tubing, which enabled the maintenance of a constant pressure (10 kpsi) during the gradient. Flow through the HPLC column, when equilibrated with 100% mobile phase A was approximately 400 nL/min.

2.7 Potential mass and time tag (PMT) acquisition

The eluate from the LC-MS/MS was directly transferred into a Thermofinnigan LTQ ion trap mass spectrometer (Thermo Scientific, San Jose, CA) via electrospray ionization (ESI). The HPLC column was coupled to the mass spectrometer with an in-house interface consisting of chemically etched electrospray emitters (150 μm o.d. x 20 μm i.d.) with a spray voltage of 2.2 kV and a heated capillary temperature of 200°C [27]. Data was collected for 85 minutes, beginning 30 minutes after the sample was injected (10 minutes into gradient). Full spectra (AGC $3x10^4$) were collected from a 400-2000 *m/z* scan range followed by the data dependent ion trap MS/MS spectra (AGC $1x10^4$) of the three most abundant ions using a collisional energy of 35%. Previously analyzed ions were excluded using a dynamic exclusion time of one minute. Analysis of the MS/MS spectra was conducted using the peptide identification software SEQUEST [28] concurrently with the annotated peptide translations from the *Escherichia coli* genome sequence, strain K-12, obtained from Genbank (http://www.ncbi.nlm.nih.gov/Genbank/index.html). Initial identifications were based on a minimum cross correlation (Xcorr) score of 1.5 for all peptides that were seen at least twice in

the MS/MS experiments. For those peptides that were only seen once, the Xcorr values used were 1.9, 2.2, and 3.2 for charge states of +1, +2, and +3, respectively. All peptides had a tryptic cleavage on at least the N- or the C- terminus of the peptide. These peptides were used to create a Potential Mass and Time (PMT) Tag database to be further validated by Fourier transform ion cyclotron resonance (FTICR-MS) measurements.

2.8 Accurate mass and time tag (AMT) validation

The eluate from the LC-MS was directly transferred into a Thermofinnigan LTQ-orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) with electrospray ionization (ESI), using the in-house interface previously described [27]. Data was collected for 100 minutes, beginning 65 minutes after the sample was injected (15 minutes into gradient). Orbitrap spectra (AGC $1x10^6$) were collected using a full range scan of 400-2000 *m/z* at a resolution of 100k followed by the data dependent ion trap MS/MS spectra (AGC $1x10^4$) of the three most abundant ions using a collisional energy of 35%. Previously analyzed ions were excluded using a dynamic exclusion time of one minute, and a mass calibration mixture was introduced at the end of each analysis to calibrate all the spectra in the analysis.

The unique peptides (trypsin cleaved peptide sequence unique to one protein) detected by the LTQ ion trap mass spectrometer were compared to those identified as PMT tags. The data was subsequently processed using the PRISM Data Analysis System, a series of software tools developed in-house. The data was initially de-isotoped to give a monoisotopic mass, charge, and intensity of the major peaks in each spectrum. The data was then analyzed in a two-dimensional fashion to determine the groups of peaks that were observed in sequential spectra.

Each group, identified as a unique mass class (UMC), was characterized as having a distinctive median mass, central normalized elution time (NET), and abundance estimate, calculated as the sum of the intensities for the MS peaks of that UMC. Each UMC was determined by comparing the mass and NET to those of the 45,171 PMT tags in the *Escherichia coli* PMT tag database. Search tolerances were set to ±6 ppm for the mass and ±5% for the elution time. Those UMCs that most closely matched the PMT tags were validated as Accurate Mass and Time (AMT) tags; moreover, this technique provided a list of peptides observed and abundance estimates for each sample.

CHAPTER THREE

RESULTS AND DATA ANALYSIS

3.1 Observed proteome of *Escherichia coli*

There are 4464 predicted proteins for *E. coli* MG1655 and 123 proteins for the NR1 plasmid [5, 29]. Our analyses resulted in the identification of 1820 chromosomal proteins and 24 NR1 plasmid proteins, resulting in a proteome coverage of 41% for chromosomal proteins and 20% for plasmid proteins (Table 1). For the cells that contained no plasmid, 1590 chromosomal proteins were identified (36% coverage); in contrast when the plasmid was present, 1707 proteins were identified (38% coverage).

Of the 1820 proteins identified, 705 were identified based on one unique peptide identification in one or more analysis and 1115 were identified with a unique peptide identification count of 2 or more in one or more analysis. The proteins identified with one unique peptide identification were included in the total proteome coverage due to the high

Table 1. *Escherichia coli* proteome coverage

Category	Predicted Proteins	Observed Proteins	Percent Coverage
Chromosomal	4464	1820	41%
Non-Plasmid Chromosomal	4464	1590	36%
Plasmid Chromosomal	4464	1707	38%
Plasmid NR1	123	24	20%
Total	4483	1844	41%

filtering controls of the MS/MS data as well as the added levels of validation due to confirming an MS/MS identification with an autonomous high mass accuracy MS detection of the parent ion by FTICR-MS.

3.2 TIGR functional role category expression

The identified chromosomal proteins were organized into functional role categories based on the categorization created from the primary annotation for the *E. coli* chromosome (K-12 MG1655) obtained from the Institute of Genomic Research (TIGR) (www.tigr.org/tigr-scripts/CMR2/batch_download.dbi). Figure 1 shows the number of proteins for each functional role category as well as the number of proteins identified for the two types of cells used in this study. With an average percent identification coverage of 41% for our analyses, the non-plasmid containing cells showed a higher than average expression for proteins in the following categories: protein synthesis, 89%; purines, pyrimidines, nucleosides and nucleotides, 90%; amino acid biosynthesis,86%; transcription, 76%; protein fate, 74%; fatty acid and phospholipid metabolism, 67%; cellular processes, 60%; cell envelope, 54%; central intermediary metabolism, 61%; biosynthesis of cofactors, prosthetic groups and carriers, 61%; DNA metabolism, 50%; signal transduction, 57%; and regulatory processes, 46%. These proteins are predicted to be involved in the housekeeping functions of the cells. For those analyses that did contain the NR1 plasmid, similar results were found except for the proteins in the energy metabolism category which showed a higher than average expression (44%) for these cell samples.

Proteins that were identified from the plasmid were also parsed into functional role categories based on the roles provided by primary annotation from A CLAssification of Mobile genetic Elements (ACLAME) (http://aclame.ulb.ac.be/). The number of proteins identified in

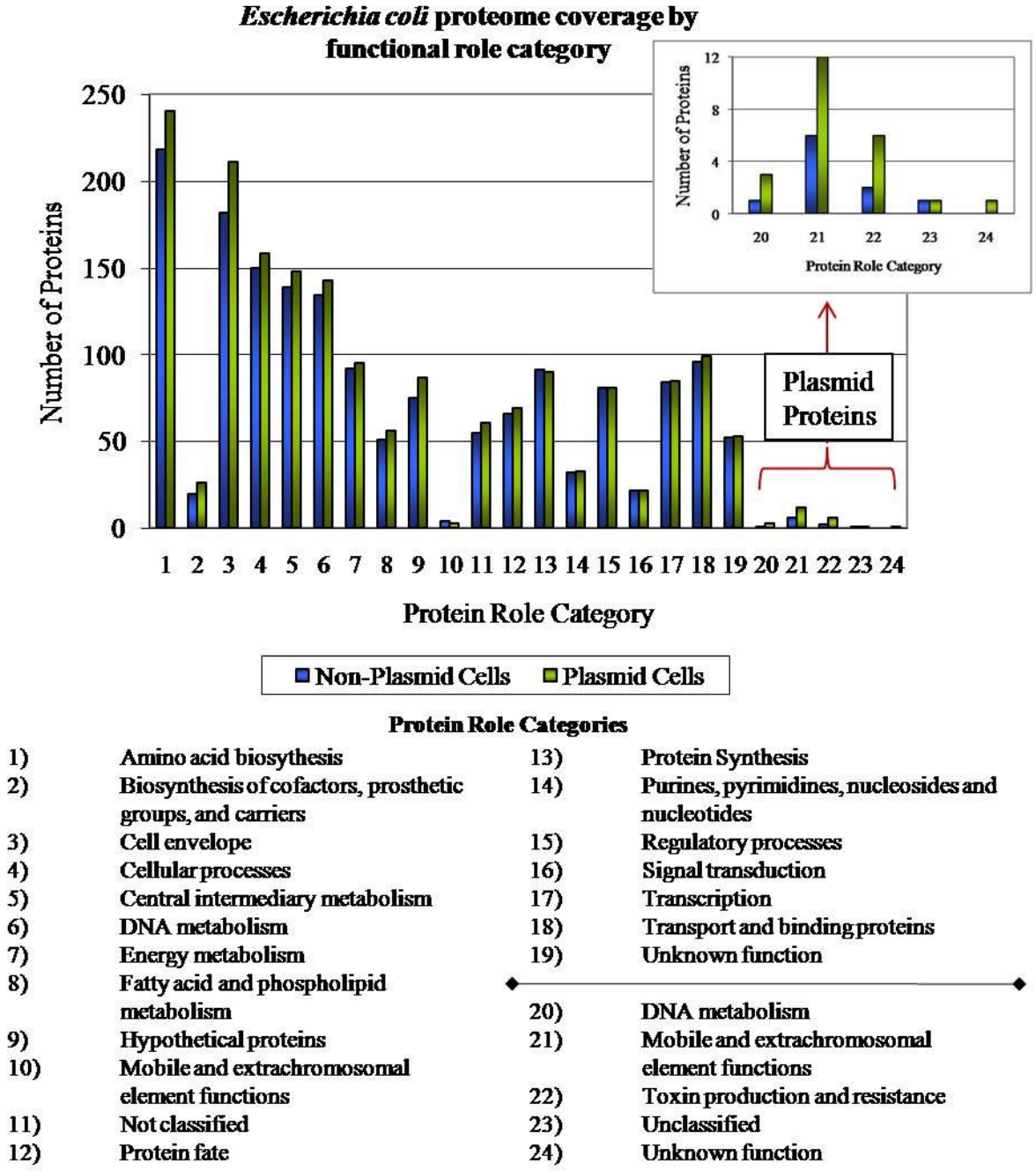


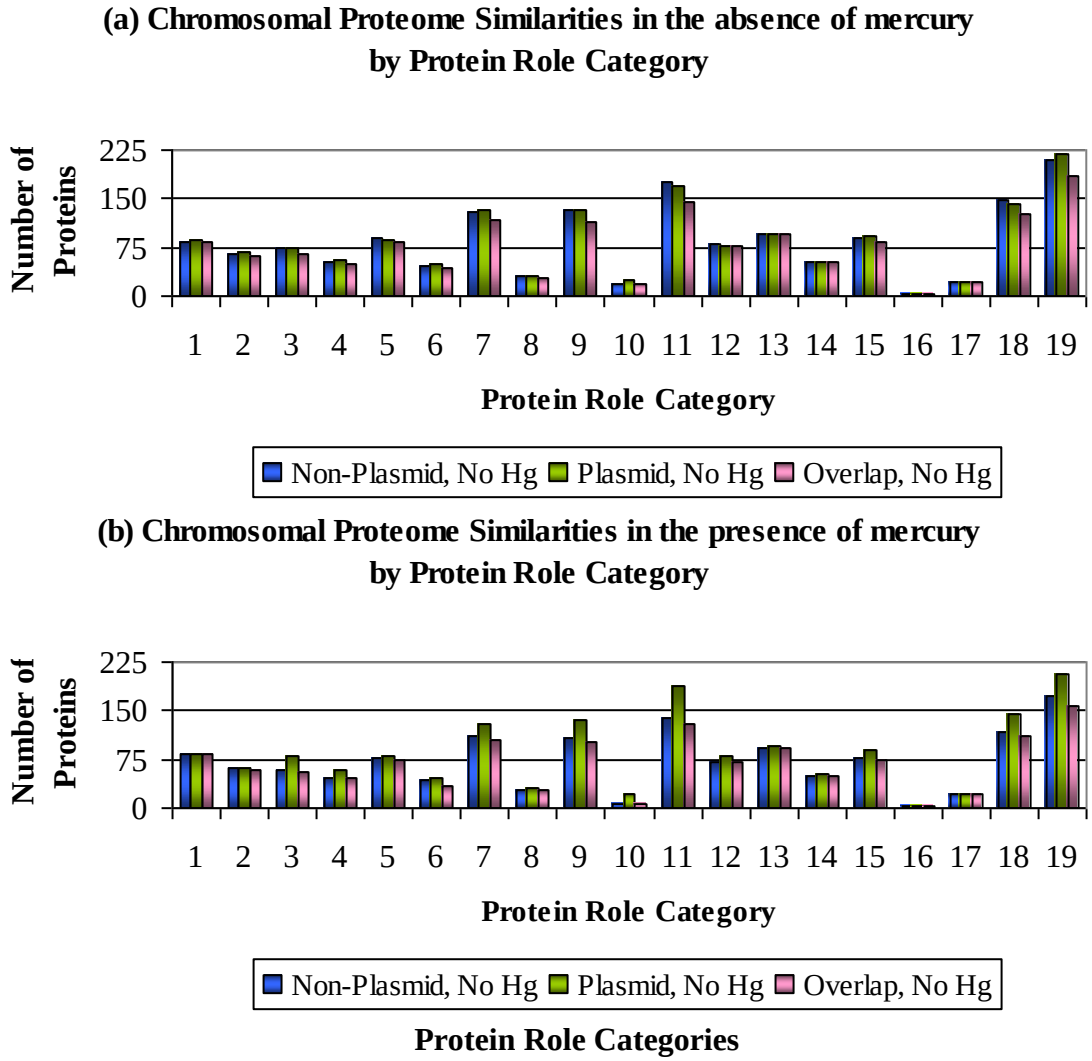
Figure 1. Number and percent coverage of proteins identified in the chromosomal proteome analysis of *Escherichia coli*. Protein coverage is divided into functional role categories as defined by TIGR primary annotation for chromosomal proteins and by ACLAME primary annotation for plasmid proteins.

this study can also be seen in Figure 1. Based on an average of 20% proteome identification coverage overall, the plasmid-containing cells showed a higher-than-average expression for the categories of DNA metabolism (21%), mobile and extrachromosomal element functions (21%), and toxin production and resistance (33%). As for the plasmid proteins that were identified in the non-plasmid containing cells, the actual numbers were well below average and are believed to be false positives.

3.3 Protein differences between cell types

Of the 1823 chromosomal proteins identified, 113 were only found in the cells without the NR1 plasmid and 235 were only found in the cells with the plasmid. This results in an overlap of 1477 proteins between the two cell cultures. The protein role categories that showed the lowest levels of variance between the cell types (5 proteins or less) in identified proteins included amino acid biosynthesis; biosynthesis of cofactors, prosthetic groups and carriers; fatty acid and phospholipid metabolism; protein synthesis; purines, pyrimidines, nucleosides and nucleotides; signal transduction; and transcription. Those categories that showed a moderate variance (6 to 20 proteins) included the cell envelope, cellular processes, central intermediary metabolism, DNA metabolism and regulatory processes. The greatest differences were seen for the following protein role categories: energy metabolism, hypothetical proteins, non-classified proteins, transport and binding proteins, and those proteins with unknown functions.

When the cell cultures are separated by whether they were exposed to phenyl mercuric acetate (PMA), interesting results appear. For those with no mercury exposure both cell types show somewhat equivalent variances (Figure 2a). However, when mercury was present, the



1) Amino acid biosynthesis
2) Biosynthesis of cofactors, prosthetic groups, and carriers
3) Cell envelope
4) Cellular processes
5) Central intermediary metabolism
6) DNA metabolism
7) Energy metabolism
8) Fatty acid and phospholipid metabolism
9) Hypothetical proteins
10) Mobile and extrachromosomal element functions
11) Not classified
12) Protein fate
13) Protein Synthesis
14) Purines, pyrimidines, nucleosides and nucleotides
15) Regulatory processes
16) Signal transduction
17) Transcription
18) Transport and binding proteins
19) Unknown function

Figure 2. (a) Chromosomal proteins in the absence of mercury (b) Chromosomal proteins in the presence of mercury.

number of unique proteins for the non-plasmid cell cultures seems to be reduced, while the cell cultures with the plasmid show an equivalent number of unique chromosomal proteins (Figure 2b). Also, there were a few protein role categories where the presence of mercury did not seem to cause a change in their overall variances: amino acid biosynthesis; fatty acid and phospholipid biosynthesis; purines, pyrimidines, nucleosides, and nucleotides; signal transduction; and transcription. It is likely that these non-changes occurred because these proteins are not affected by the presence of mercury.

3.4 Relative protein abundance estimates

In Figure 3, the protein abundance changes for 1844 proteins (all identified with at least 2 unique peptides) can be seen for the different treatments based on whether the cells had the NR1 plasmid (NR) or not (MG). The abundance values were calculated from the average abundance of all peptides for a given protein in each high resolution LC-MS experiment. This value was determined by the sum of the highest peak intensity for each scan of that eluting peptide in the total ion chromatogram for each high resolution LC-MS analysis. Figure 3 shows the abundance value comparison using a z-score for each protein abundance value, determined using Microsoft Excel, using the following equation:

$$z = (x - \mu) / \sigma \qquad (2)$$

This form of standardization begins with resetting the mean value for the entire dataset, x, to zero by subtracting the mean abundance from each data point, μ, then dividing by their standard deviation, σ. The resulting value, z, is in standardized units that correspond to the number of standard deviations each point lies above or below the mean. In short, the z-score for each protein in each experiment shows the degree of standard deviation for that protein from the

***Escherichia coli* proteome abundance changes by treatment condition**

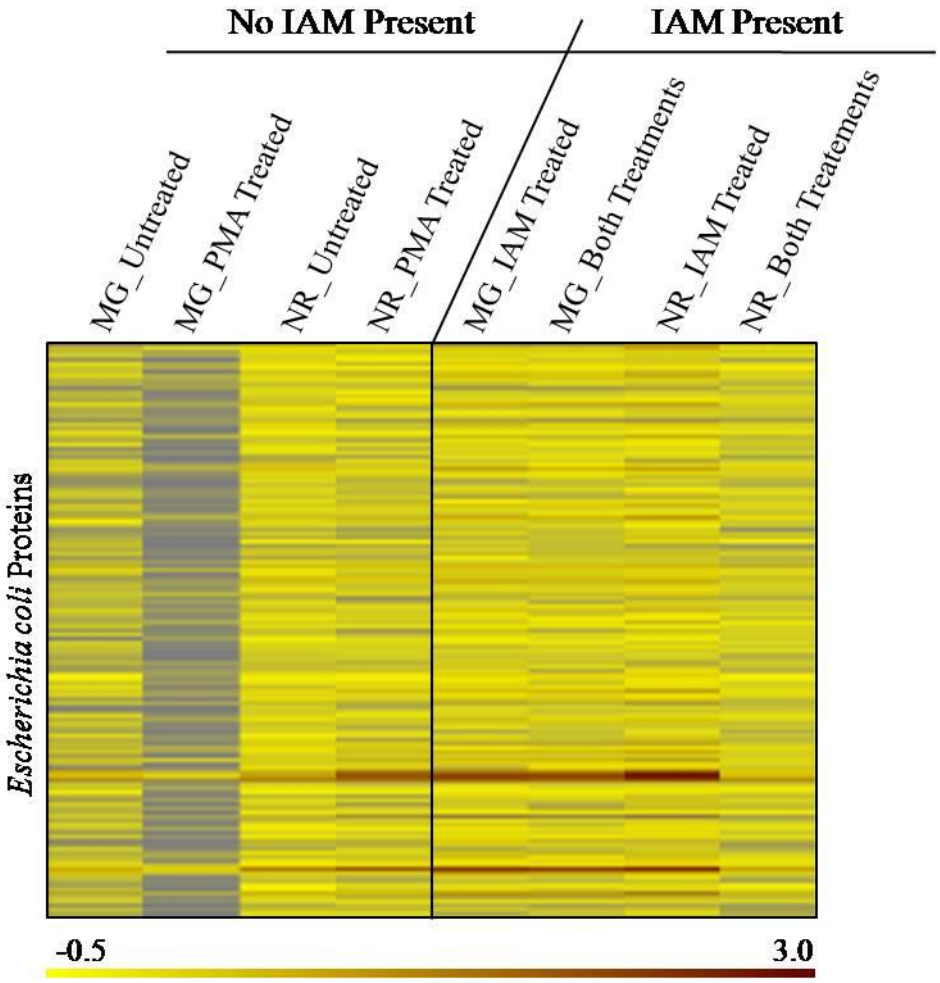


Figure 3. (a) Relative abundance data plot for 1844 proteins from Escherichia coli versus culture condition and cell fractionation. The proteins represented here were identified with at least 2 unique peptide identifications in a single LC-MS analysis.

average. A high positive z-score (shown here in red) corresponds to a significant increase in abundance; whereas, those shown as yellow depict a significant decrease in abundance from that of the standard deviation of the average. If a protein has a z-score of 0, this equates to no significant change to the average protein abundance over the different conditions of the experiment.

Due to the large dynamic range of protein concentrations found in a given cell, use of the z-score is crucial for determining protein abundance changes between experiments for both low and high abundance proteins. For each LC-MS analysis the same amount of total peptide is used, yet a level of variability will still exist in the abundances of a protein due to the dynamic nature of the sample, even if the total amount of protein is the same between two cells grown under different conditions. Sample to sample variation is accounted for by using the z-score because it shows the changes invoked by the sample for the proteins instead of a direct comparison of the protein concentrations between two samples.

3.5 Pairwise Pearson's correlation

Pearson's correlation shows the strength of the linear relationship between two variables. In this analysis, the z-scores were compared for the protein abundances that were calculated from the average abundance of all the peptides identified from every protein in each high resolution LC-MS analysis. The z-scores from one analysis were correlated to the z-scores of the other analyses resulting in the pairwise Pearson's correlation plot in Figure 4 (created using DAnTE [30]). The values on the plot vary from 0 to +1, with 0 corresponding to no correlation (red) and 1.0 corresponding to a perfectly positive or negative correlation (white).

From this plot we can see that that there is considerable correlation between the two cell

Pairwise Pearson's Correlation Plot of *Escherichia coli* proteins

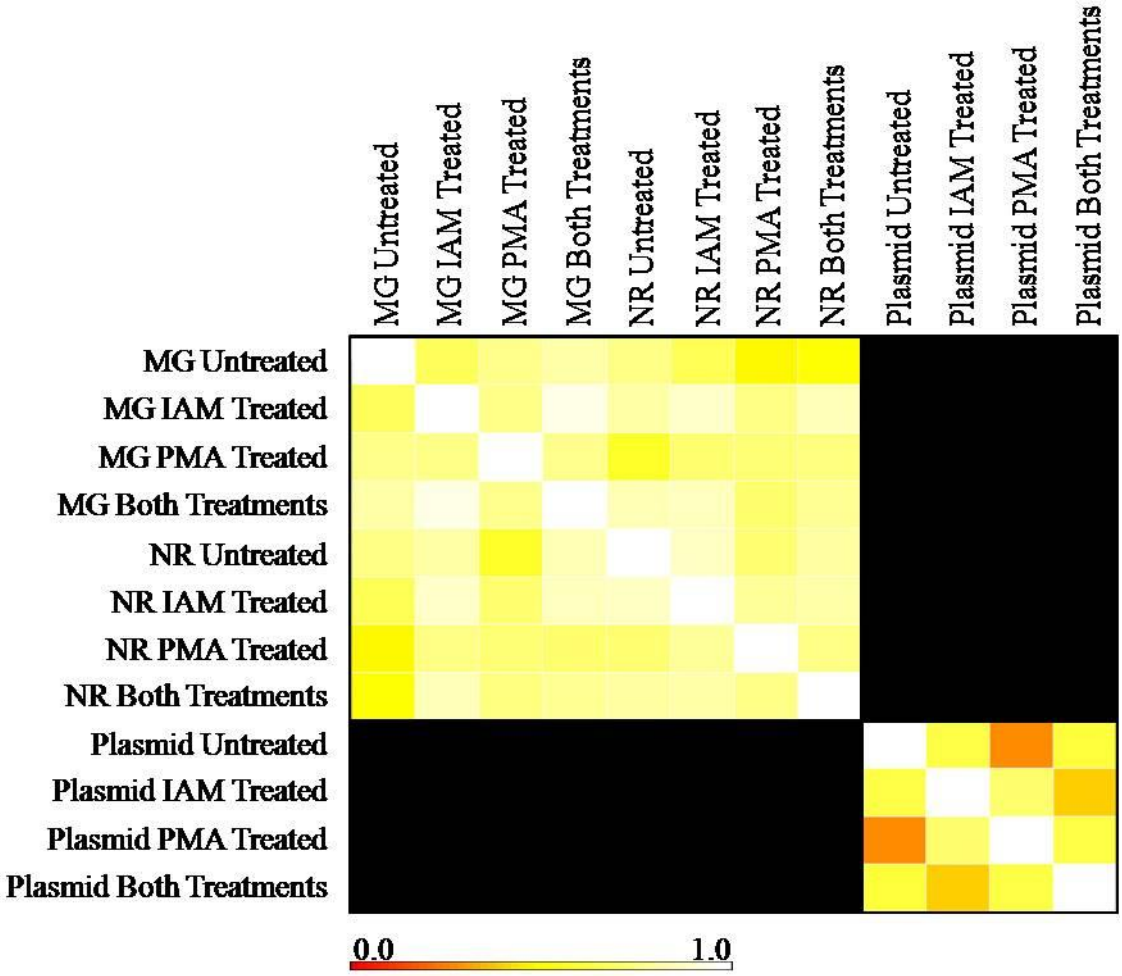


Figure 4. (a) Pairwise Pearson's Correlation plot of all *Escherichia coli* chromosomal peptide identifications from each triplicate analyses correlated for similarity to itself and every other analyses obtained from the three cell preparations. The darker the box, the less correlated the data is between two analyses. White boxes represent a perfect correlation. (b) Pairwise Pearson's Correlation plot of all NR1 Plasmid peptide identifications from each triplicate analyses correlated for similarity to itself and every other analyses obtained from the three cell preparations. The darker the box, the less correlated the data is between two analyses. White boxes represent a perfect correlation.

cultures and the different treatments for the chromosomal proteins. This is to be expected since they are both the same species of bacteria. As for the plasmid proteins, those cell cultures that were exposed to mercury and those that were not show respectable correlation values, but the interrelationship of the two treatment types show a lower correlation value, as would be expected.

3.6 Potentially mercurylated proteins

Of the proteins identified in this study, 130 peptides were identified as having a phenyl mercury adduct based on SEQUEST analyses (26 unique to the non-plasmid cell cultures (MG) and 81 unique to the plasmid containing cell cultures (NR)). These peptides were found to be associated with 104 proteins, of which only 75 passed the extensive filters used for the AMT tag process. Of these 75 proteins, when the equivalent protein without the adduct bound were looked at, 16 were found to show no significant abundance change when analyzed using the z-score no matter the cell culture or the treatment type (Figure 7a). Considering the proteins used for abundance calculations could not be identified with an adduct, there should be a decrease in abundance if the peptide gains the phenyl mercury adduct. From the remaining 59, only 13 resulted from experimentation that would allow us to fully analyze the peptide for the phenyl mercury adduct (i.e. high resolution mass spectrometry). The 15 peptides that were associated with these 13 proteins were then evaluated for their isotopic profile and compared to a theoretical profile with the phenyl mercury adduct attached using ICR-2LS, an in-house program. Figure 7b shows the observed isotopic profile for a phenyl mercurylated peptide from glutamine synthetase and the theoretical isotopic profile. Another 7 peptides were found to have the expected isotopic profile with a phenyl mercury adduct and these corresponded to the following 7 proteins: alkyl

Peptide/Protein Analysis for the Presence of a Phenyl Mercury Adduct

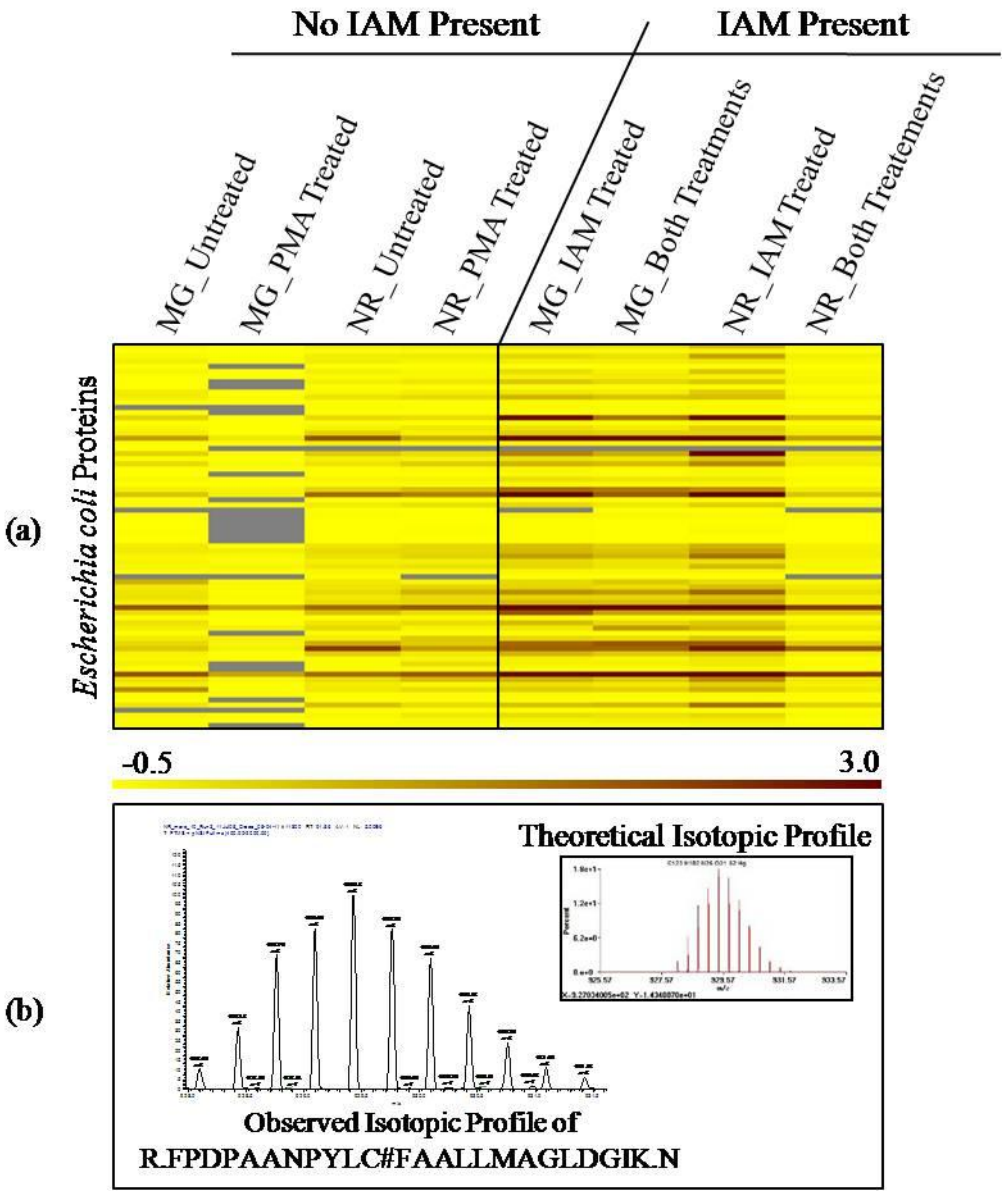


Figure 5. (a) Relative abundance data plot of 59 proteins that potentially have a phenyl mercury adduct. (b) Isotopic profile of a peptide from Glutamine Synthetase with a phenyl mercury adduct, represented by the symbol #. The isotopic profile strongly resembles the expected isotopic profile of the peptide with the phenyl mercury adduct.

hydroperoxide reductase, C22 subunit (ahpC); citrate synthase (gltA); global DNA-binding transcriptional dual regulator H-NS (hns); fused DNA-binding transcriptional dual regulator/O6-methylguanine-DNA methyltransferase (ada); ketol-acid reductoisomerase, NAD(P)-binding (ilvC); 5-methyltetrahydropteroyltriglutamate-homocysteine-S-methyltransferase (metE); and glutamine synthetase (glnA).

3.7 Mercury affected proteins

The remaining proteins that were not identified as having a phenyl mercury adduct, were analyzed for changes in abundance based on whether the cells were or were not exposed to mercury. This was done by evaluating the differences in the z-scores for each cell type (plasmid or non-plasmid containing) and whether they were exposed to phenyl mercury acetate. By choosing the proteins that showed a difference in z-score of greater than +/- 1, we were able to identify if the abundance of the protein increased or decreased by more than one standard deviation from the mean abundance. Thirteen proteins were identified as meeting this criteria and are shown in Table 2, with the majority involved in protein synthesis.

For the cell cultures without the NR1 plasmid, three proteins were shown to decrease when mercury was introduced to the cells and nine were found to increase in abundance. Of the proteins that decreased in abundance, the presence of the NR1 plasmid seemed to make no significant difference except to hold two of the proteins at a constant abundance despite the treatment conditions (metH and betB). For the proteins that increased in abundance, the presence of the NR1 plasmid caused a decrease in abundance for four of the proteins, a stability in abundance for three of the proteins and an increase in abundance for the remaining protein. The NR1 plasmid also showed an increase in abundance for the protein, rpsA, that remained at a

Table 2. *Escherichia coli* proteins affected by phenyl mercuric acetate (PMA)

Protein ID	Protein Name	Non-Plasmid Cell Cultures	Plasmid Cell Cultures	Protein Role Category
Ecoli_b4019	metH	↓	–	Amino Acid Biosynthesis
Ecoli_b0312	betB	↓	–	Cellular Processes
Ecoli_b1687	ydiJ	↓	↓	Not Classified
Ecoli_b3398	yrfF	↑	↓	Not Classified
Ecoli_b0440	hupB	↑	↓	Protein Synthesis
Ecoli_b0911	rpsA	↑	–	Protein Synthesis
Ecoli_b3294	rplQ	–	↑	Protein Synthesis
Ecoli_b3319	rplD	↑	–	Protein Synthesis
Ecoli_b3986	rplL	↑	–	Protein Synthesis
Ecoli_b4000	hupA	↑	↑	Protein Synthesis
Ecoli_b4200	rpsF	↑	↓	Protein Synthesis
Ecoli_b0854	potF	↑	↓	Transport and Binding Proteins
Ecoli_b3529	yhjK	↑	–	Unknown Function

fairly constant abundance without the plasmid.

CHAPTER FOUR

DISCUSSION

The remediation of mercury is a very current and real problem. Even though this metal is naturally occurring, anthropogenic manipulations have created areas with toxic levels of contamination. Previous exposures to animals and humans have shown us the repercussions of misuse of this metal. In this study we exposed *Escherichia coli* cells to phenyl mercury acetate in order to deduce some of the mystery behind the toxicity of mercury to living organisms. In all, we were able to identify 41% of the *E. coli* proteome, and the proteins identified showed strong correlation with those proteins expected to be involved in house-keeping and unknown proteins, which are often associated with biomarker identification. Further analysis of these proteins could eventually lead to the ability to identify levels of exposure for other organisms, including humans.

When the two different cell types were compared, it was found that the plasmid-containing cells yielded a larger number of proteins overall. Considering that the NR1 plasmid provides resistance to mercury-induced damage, it is believed that this outcome would be expected, and here we verify this theory. Interesting is the lack of protein data for the non-plasmid cells that were exposed to PMA only, which could be due to errors in sampling handling, but could also result from exposure to PMA. Further analysis is required to determine the actual cause of this lack of information.

As expected, we were able to identify two *mer* proteins from the NR1 plasmid when these cells were exposed to mercury. The merP, periplasmic mercuric ion binding protein, and the merR, activator/repressor for the mer operon, were identified. Why we did not see other mer proteins is a question that could be answered in further studies. However, identification of these proteins strengthens the reasoning for finding more proteins overall in the NR1 containing cells, especially when exposed to mercury. That is, the NR1 plasmid was able to provide protection from PMA.

It is known that mercury compounds bind to thiol-containing proteins. We were able to identify with a high level of certainty seven proteins that contained a phenyl mercury bound to a cysteine residue. In this study, there were 946 proteins containing thiols identified, resulting in 0.7% having a phenyl mercury adduct. Previous studies have shown that the proteins vulnerable to mercury compounds are from the electron transport chain and of the reactive oxygen species (ROS) defense system [15-17]. The proteins identified in this study agree with this. Three of the identified proteins are involved in methyl transferase, one of which is citrate synthase (gltA) – a protein involved in carbon fixation which provides NADH for the electron transport chain, and also provides products that can be used in glutamate metabolism. Glutamate is linked to glutathione metabolism, a compound that is involved in the ROS defense system. Other proteins involved in glutathione metabolism that were identified as having a phenyl mercury adduct include alkyl hydroperoxide reductase, C22 subunit (ahpC), an oxidoreductase, directly involved in glutathione metabolism; and glutamine synthetase (glnA), involved in the interconversion between glutamate and glutamine, the former of which feeds into glutathione metabolism.

The proteins that were not bound with mercury, but were found to be affected by the presence of mercury appear to mostly be proteins involved in protein synthesis. Five of the

thirteen proteins are ribosomal subunit proteins, two 30S subunits (rpsA and rpsF) and three 50S subunits (rplQ, rplD and rplL). The other two proteins from this role category are hupA and hupB, the alpha and beta subunits of the HU DNA-binding transcriptional regulator. Proteins involved indirectly or directly to the production of amino acids would be expected to decrease if the proteins of neighboring pathways are disrupted. The disruption of citrate synthase and glutamine synthase could be linked to the decrease in methionine synthase (metH) and of NAD-dependent betaine aldehyde dehyrogenase (betB), because these proteins are linked to amino acid biosynthesis or cellular processes related to the biosynthesis of amino acids, respectively.

Considering that thiol compounds are used for the removal of heavy metals, such as mercury, from organisms, the use of DTT and thiourea for denaturing and reducing the proteins is questionable. There have been studies that show that DTT can remove mercury from thiol-containing proteins, but whether it is bound permanently to the chelating agent depends on the type of thiol. In the case of DTT it has been found to remove mercury from proteins, but then release the metal so that it is able to rebind with another thiol-containing protein [16]. The implications of this could be that the proteins we identified as having a bound phenyl mercury may not have been the original protein to which the mercury was attached before we began treatment of the cells. Further analysis could determine if this is the case.

The identification of proteins targeted directly and indirectly by mercury in *Escherichia coli* is still in its beginning stages, but the progress we have shown here is very promising. Future studies can be expected that will improve upon our techniques, and we can also start to expand into other organisms and cell types.

1. Valko, M., H. Morris, and M.T. Cronin, *Metals, toxicity and oxidative stress.* Curr Med Chem, 2005. **12**(10): p. 1161-208.
2. Chang, T.C. and J.H. Yen, *On-site mercury-contaminated soils remediation by using thermal desorption technology.* J Hazard Mater, 2006. **128**(2-3): p. 208-17.
3. Melgar, M.J., J. Alonso, and M.A. Garcia, *Removal of toxic metals from aqueous solutions by fungal biomass of Agaricus macrosporus.* Sci Total Environ, 2007. **385**(1-3): p. 12-9.
4. Wang, Q., et al., *Sources and remediation for mercury contamination in aquatic systems--a literature review.* Environ Pollut, 2004. **131**(2): p. 323-36.
5. Liebert, C.A., R.M. Hall, and A.O. Summers, *Transposon Tn21, flagship of the floating genome.* Microbiol Mol Biol Rev, 1999. **63**(3): p. 507-22.
6. Lipton, M.S., et al., *Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags.* Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11049-54.
7. Pasa-Tolic, L., et al., *Proteomic analyses using an accurate mass and time tag strategy.* Biotechniques, 2004. **37**(4): p. 621-4, 626-33, 636 passim.
8. Bakir, F., et al., *Methylmercury poisoning in Iraq.* Science, 1973. **181**(96): p. 230-41.
9. Yorifuji, T., et al., *Long-term exposure to methylmercury and neurologic signs in Minamata and neighboring communities.* Epidemiology, 2008. **19**(1): p. 3-9.
10. *Preliminary report on the safety of dental amalgam and alternative dental restoration materials for patients and users*. 2007, Scientific Committee on Emerging and Newly Identified Health Risks.
11. Ball, L.K., R. Ball, and R.D. Pratt, *An assessment of thimerosal use in childhood vaccines.* Pediatrics, 2001. **107**(5): p. 1147-54.
12. Bernard, S., et al., *The role of mercury in the pathogenesis of autism.* Mol Psychiatry, 2002. **7 Suppl 2**: p. S42-3.
13. de Magalhaes, M.E. and M. Tubino, *A possible path for mercury in biological systems: the oxidation of metallic mercury by molecular oxygen in aqueous solutions.* Sci Total Environ, 1995. **170**(3): p. 229-39.
14. Smith, T., et al., *Bacterial oxidation of mercury metal vapor, Hg(0).* Appl Environ Microbiol, 1998. **64**(4): p. 1328-32.
15. Hansen, J.M., H. Zhang, and D.P. Jones, *Differential oxidation of thioredoxin-1, thioredoxin-2, and glutathione by metal ions.* Free Radic Biol Med, 2006. **40**(1): p. 138-45.
16. Hultberg, B., A. Andersson, and A. Isaksson, *Interaction of metals and thiols in cell damage and glutathione distribution: potentiation of mercury toxicity by dithiothreitol.* Toxicology, 2001. **156**(2-3): p. 93-100.
17. Messer, R.L., et al., *Mercury (II) alters mitochondrial activity of monocytes at sublethal doses via oxidative stress mechanisms.* J Biomed Mater Res B Appl Biomater, 2005. **75**(2): p. 257-63.
18. Mulligan, C.N., R.N. Yong, and B.F. Gibbs, *An evaluation of technologies for the heavy metal remediation of dredged sediments.* J Hazard Mater, 2001. **85**(1-2): p. 145-63.
19. Barkay, T., S.M. Miller, and A.O. Summers, *Bacterial mercury resistance from atoms to ecosystems.* FEMS Microbiol Rev, 2003. **27**(2-3): p. 355-84.
20. Westwater, J., et al., *The adaptive response of Saccharomyces cerevisiae to mercury exposure.* Yeast, 2002. **19**(3): p. 233-9.

21. Hussein, H.S., et al., *Phytoremediation of mercury and organomercurials in chloroplast transgenic plants: enhanced root uptake, translocation to shoots, and volatilization.* Environ Sci Technol, 2007. **41**(24): p. 8439-46.
22. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.
23. Chen, J.W., et al., *Expression Analysis of Up-Regulated Genes Responding to Plumbagin in Escherichia coli.* J Bacteriol, 2006. **188**(2): p. 456-63.
24. Leichert, L.I. and U. Jakob, *Protein thiol modifications visualized in vivo.* PLoS Biol, 2004. **2**(11): p. e333.
25. Rabilloud, T., et al., *Proteomics analysis of cellular response to oxidative stress. Evidence for in vivo overoxidation of peroxiredoxins at their active site.* J Biol Chem, 2002. **277**(22): p. 19396-401.
26. Siuzdak, G., *The Expanding Role of Mass Spectrometry in Biotechnology*. 2003, San Diego: MCC Press.
27. Kelly, R.T., et al., *Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry.* Anal Chem, 2006. **78**(22): p. 7796-801.
28. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.
29. Riley, M., et al., *Escherichia coli K-12: a cooperatively developed annotation snapshot--2005.* Nucleic Acids Res, 2006. **34**(1): p. 1-9.
30. Polpitiya, A.D., et al., *DAnTE: a statistical tool for quantitative analysis of -omics data.* Bioinformatics, 2008. **24**(13): p. 1556-8.