

6-26-2017

Arsenic Biotransformations in Microbes and Humans, and Catalytic Properties of Human AS3MT Variants

Jiaojiao Li

Florida International University Herbert Wertheim College of Medicine, jli041@fiu.edu

DOI: 10.25148/etd.FIDC001934

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

Recommended Citation

Li, Jiaojiao, "Arsenic Biotransformations in Microbes and Humans, and Catalytic Properties of Human AS3MT Variants" (2017). *FIU Electronic Theses and Dissertations*. 3460.
<https://digitalcommons.fiu.edu/etd/3460>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ARSENIC BIOTRANSFORMATIONS IN MICROBES AND HUMANS, AND CATALYTIC
PROPERTIES OF HUMAN AS3MT VARIANTS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMEDICAL SCIENCES

by

Jiaojiao Li

2017

To: Dean John A. Rock
Herbert Wertheim College of Medicine

This dissertation, written by Jiaojiao Li, and entitled Arsenic Biotransformations in Microbes and Humans, and Catalytic Properties of Human AS3MT Variants, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Jeffrey Boyd

Toby G. Rossman

Yuan Liu

Masafumi Yoshinaga

Barry P. Rosen, Major Professor

Date of Defense: June 26, 2017

The dissertation of Jiaojiao Li is approved.

Dean John A. Rock
Herbert Wertheim College of Medicine

Andres G. Gil
Vice President for Research and Economic Development
And Dean of University Graduate School

Florida International University, 2017

DEDICATION

This dissertation is dedicated to my husband, Zhili Peng, my daughter, Linuo Peng,
and my parents, for their love, support and understanding.

ACKNOWLEDGMENTS

I sincerely thank my advisor, Dr. Barry P. Rosen, for his guidance, friendship, helpful discussions and financial assistance throughout my graduate research at the Florida International University. I owe the deepest gratitude to his encouragement, guidance and support from the beginning to the end has enabled me to develop my project. I respect him as an excellent teacher and a rigorous scientist.

I wish to thanks to the members of my thesis committee, Dr. Jeffrey Boyd, Dr. Toby G. Rossman, Dr. Yuan Liu, and Dr. Masafumi Yoshinaga, for this work benefited greatly from their teaching, insights and advice. Thanks Dr. Donald Bowden for his advice on the EHS manuscript. He told us how to find the SNPs.

Specials thanks to Dr. El-Hage, Dr. Kun Lu, Dr. Rita Mukhopadhyay, Dr. Hiranmoy Bhattachajee and Dr. Yuk-Ching Tse-Dinh, for their assistance and critical review for my work. I wish to thank Dr. Jian Chen for his help from the initial stage of my PhD with many techniques to my graduation stage with more experiment design strategies. Thanks to Dr. Hui Dong for help me starting human AS3MT variants project. I would also like to thank all of my present and past laboratory colleagues: Dr. Dharmendra S. Dheeman, Dr. Masafumi Yoshinaga, Dr. Charles Packianathan, Dr. Venkadesh Sarkarai Nadar, Mr. Shashank Pawitwar, and Ms. Yu Yan for many stimulating discussions in the lab and their continuous help.

ABSTRACT OF THE DISSERTATION
ARSENIC BIOTRANSFORMATIONS IN MICROBES AND HUMANS, AND CATALYTIC
PROPERTIES OF HUMAN AS3MT VARIANTS

by

Jiaojiao Li

Florida International University, 2017

Miami, Florida

Professor Barry P. Rosen, Major Professor

Arsenic is the most pervasive environmental toxic substance. As a consequence of its ubiquity, nearly every organism has genes for resistance to inorganic arsenic, which are found largely in bacterial arsenic resistance (*ars*) operons. In one project I examined the role of glutaredoxin 2 (Grx2) in reduction of arsenate to arsenite. I demonstrated that Grx2 has both glutaredoxin thiol transfer activity and glutathione S-transferase (GST) activity. In a second project investigated arsenic resistance in a microbiome organism. I discovered that the human gut microflora *B. vulgatus* has eight continuous genes in its genome and these genes form an arsenical-inducible transcriptional unit. In two other projects I investigated the properties of two As(III) S-adenosylmethionine (SAM) methyltransferase (ArsM in microbes and AS3MT in animals). Methylation of inorganic arsenic is a central process in the organoarsenical biogeochemical cycle. Members of every kingdom have ArsM or AS3MT enzymes that methylate inorganic As(III) into mono- (MAs(III)), di- (DMAs(III)) and tri- (TMAs(III)) methylarsenicals. In most organisms, four conserved cysteine residues are required for methylation of As(III) to MAs(III), but methylation of MAs(III) to DMAs(III) requires only the two cysteines closest to the C-terminus. In this project we demonstrate that most fungal species have ArsM orthologs with only three conserved cysteine residues, and AfArsM from *Aspergillus*

fumigatus methylates only MAs(III) and not As(III). For human, arsenic methylation process is thought to be protective from acute high-level arsenic exposure. However, with long term low-level exposure, hAS3MT is thought to produce intracellular methylarsenite (MAs(III)) and dimethylarsenite (DMAs(III)), which are considerably more toxic than inorganic As(III) and may contribute to arsenic-related diseases. Several single nucleotide polymorphisms (SNPs) in putative regulatory elements of the hAS3MT gene have been shown to be protective. In contrast, three previously identified exonic SNPs (R173W, M287T and T306I) may be deleterious. I identified five additional intragenic variants in hAS3MT (H51R, C61W, I136T, W203C and R251H). I purified the eight polymorphic hAS3MT proteins and characterized their enzymatic properties. Each enzyme had low methylation activity through decreased affinity for substrate, lower overall rates of catalysis and/or lower stability. I propose that amino acid substitutions in hAS3MT with decreased catalytic activity lead to detrimental responses to environmental arsenic and may increase the risk of arsenic-related diseases.

TABLE OF CONTENTS

CHAPTER	PAGE
Introduction to the Dissertation	1
CHAPTER 1. Introduction.....	3
1.1 Arsenic: Occurrence and uses.....	3
1.2 Arsenic Resistance System.....	5
1.3 Arsenic As(III) S-adenosylmethionine (SAM) methyltransferases.....	9
1.4 Phylogeny of Three-cysteine ArsMs.....	16
CHAPTER 2. Statement of the Problem and Objectives	23
2.1 Statement of the Problem.....	23
2.2 Objectives.....	24
CHAPTER 3. <i>Escherichia coli</i> Grx2: A dual-function Hybrid of Glutaredoxin and Glutathione S-transferase.....	25
3.1 Abstract.....	25
3.2 Introduction.....	26
3.3 Material and Methods.....	28
3.4 Results	30
3.5 Discussion.....	39
CHAPTER 4. Expression of Arsenic Resistance genes in the Obligate Anaerobe <i>Bacteroides vulgatus</i> ATCC 8482, a Gut Microbiome Bacterium.....	41
4.1 Abstract.....	41
4.2 Introduction.....	41
4.3 Material and Methods.	44
4.4 Results.....	48
4.5 Discussion.....	54
CHAPTER 5. A novel As(III) S-adenosylmethionine Methyltransferase from <i>Aspergillus fumigatus</i>	56
5.1 Abstract.....	56
5.2 Introduction.....	57
5.3 Material and Methods.	58
5.4 Results.....	62
5.5 Discussion.....	73
CHAPTER 6. The Biological Consequences of Polymorphisms in the Human AS3MT Arsenic Methylation Gene	76
6.1 Abstract.....	76
6.2 Introduction.....	76
6.3 Materials and Methods.....	80
6.4 Results.....	84
6.5 Discussion.....	97
Future Perspectives.....	102

References.....	103
CURRICULUM VITAE	116

LIST OF TABLES

TABLE	PAGE
Table 3.1 GST activity of Grx2 and sjGST.....	38
Table 4.1 Oligonucleotide primers for RT-PCR.....	46
Table 4.2 Oligonucleotide primers for Q-RT-PCR.....	47
Table 5.1 Methylation of MAs(III) by AfArsM1 mutants in vivo and in vitro.....	70
Table 6.1. Frequencies of the eight missense hAS3MT polymorphisms in the 1000 Genomes Project, ExAC, NHLBI Exome Sequencing Project or BUSHMAN population.....	85
Table 6.2 PolyPhen2 scores of missense variants.....	88
Table 6.3 Kinetic parameters for arsenic methylation catalyzed by hAS3MT.....	94

LIST OF FIGURES

FIGURE	PAGE
Figure 1.1 Arsenic resistance system.....	6
Figure 1.2 Multiple sequence alignment of As(III) SAM methyltransferases.....	14
Figure 1.3 Proposed hAS3MT reaction scheme.....	16
Figure 1.4 A neighbor-joining phylogenetic tree showing the evolutionary relationships of fungi ArsM protein with arsenic methyltransferase proteins from members of other kingdom.....	22
Figure 3.1 Cartoon diagram of the Grx2–GSH. Ribbon diagram of the Grx2 –GSH complex (PDB entry 4kx4). Secondary-structural elements: -helix, tan; -sheet, brown; coil, cyan;310-helix, blue.	31
Figure 3.2 Superimposition of helices helices 5 and 6 of the Grx2–GSH complex (PDB entry 4kx4; tan) with the structure of the Grx2 C9S/C12S mutant (cyan; PDB entry 4ksm) and the NMR model (orchid; PDB entry 1g7o) of Grx2.....	32
Figure 3.3 Comparison of the Grx2 structure with nine other GSTs. Grx2 (4KX4) was compared with represented members of GST classes alpha (1GSE, 3.8 Å), beta (2PMT, 3.1 Å), delta (2WJU, 3.9 Å), phi (1GNW, 3.2 Å), mu (1HNA, 3.2 Å), omega(1EEM, 3.3 Å), pi (1GLP, 3.4 Å), sigma (1GSQ, 3.5 Å) and theta (1LJR, 3.4 Å). The PD B IDs and RMSD relative to Grx2 are given parentheses.....	34
Figure 3.4 Interactions of GSH with Grx2. The blue spheres represent water molecules. A potential hydrogen-bond network around the GSH molecule is indicated with dashed lines (the distances are in Å) and residues interacting with GSH are labeled.....	35
Figure 3.5 In silico analysis of the interaction of ArsC and Grx2 by molecular Docking.....	38
Figure 4.1 Resistance to inorganic and organic arsenicals in anaerobic growth <i>B. vulgatus</i> strain ATCC 8482.....	48
Figure 4.2 The <i>ars</i> operon of <i>B. vulgatus</i> strain ATCC 8482.....	50
Figure 4.3 Transcriptional analysis of the <i>arsA</i> , <i>acr3</i> , <i>arsC</i> and <i>arsD</i> genes.....	52
Figure 4.4 Transcriptional analysis of <i>orf1</i> , <i>orf2</i> and <i>orf3</i>	54
Figure 5.1 Phylogeny of arsenite S-adenosylmethyltransferases (ArsM).....	64

Figure 5.2 Multiple alignment of AfArsM orthologs.....	65
Figure 5.3 The arsenic resistance clusters in the chromosome <i>A. fumigatus</i> Af293.....	66
Figure 5.4 AfArsM1 confer MAs(III) resistance, not As(III).....	67
Figure 5.5 Methylation of MAs(III) or As(III) by AfArsM1 in vivo and in vitro.....	68
Figure 5.6 Competition of As(III) with MAs(III) for binding to AfArsM1.....	71
Figure 5.7 Methylation of MAs(V) by <i>E. coli</i> expressing AfarsM1 co-cultured with <i>Burkholderia sp. MR1</i>	72
Figure 6.1 Homology model of hAS3MT and polymorphisms.....	87
Figure 6.2 Arsenic methylation by wild type hAS3MT and polymorphic variants	89
Figure 6.3 Efficiency of arsenic methylation.....	91
Figure 6.4 Kinetics of wild type hAS3MT and polymorphic variant.....	93
Figure 6.5 Temperature stability of wild type hAS3MT and polymorphic variants	96
Figure 6.6 Molecular surface model of hAS3MT.....	101

ABBREVIATIONS AND ACRONYMS

ABBREVIATION	FULL NAME
AQP	Aquaporin
ArsA	Arsenical ATPase
ArsB	Arsenical Efflux Pump ACR3 or Related Permease
ArsC	Arsenate Reductase
ArsD	As(III) Chaperone
ArsM	As(III) S-adenosylmethionine Methyltransferase
ArsP	Methylarsenite Permease
ArsH	NADPH-dependent FMN Reductase
ArsI	MAs(III) Demethylase
ArsR	As(III)-Responsive Transcription Factor
As(III)	Trivalent Arsenite

As(V)	Pentavalent Arsenate
AS3MT	As(III) SAM Methyltransferase (in animals)
ArsM	As(III) SAM Methyltransferase (in microbes)
ATP	Adenosine Triphosphate
ATSDR	Agency for Toxic Substances and Disease Registry
DMAs(III)	Trivalent Dimethyl arsenite
DMAs(V)	Dimethylarsenate
EDTA	Ethylenediaminetetraacetic Acid
EPA	Environmental Protection Agency
ESP	Exome Sequencing Project
EXAC	Exome Aggregation Consortium
GOE	Great Oxidation Event
Grx 2	Glutaredoxin 2

GlpF	Glycerol Facilitator
GLUT	Glucose Permease
GSH	Reduced Glutathione
GST	Glutathione S-transferase
HGT	Horizontal Gene Transfer
HILIC	Hydrophilic Interaction Chromatography
HPLC	High Pressure Liquid Chromatography
IDT	Integrated DNA Technologies
MAF	Minor Allele Frequency
MAs(III)	Trivalent Methyl Arsenite
MAs(V)	Pentavalent Methylarsenate
MBP	Maltose Binding Protein

MOPS	3-(N-morpholino)propanesulfonic Acid
Nitarsone	4-nitrophenyl Arsonic Acid
PMI	Primary Methylation Index
Q-RT-PCR	Quantitative Real-time Polymerase Chain Reaction
SAM	S-adenosylmethionine
SAH	S-adenosylhomocysteine
SDS	Sodium Dodecyl Sulfate
PAGE	Polyacrylamide Gel Electrophoresis
SNPs	Single Nucleotide Polymorphisms
SMI	Secondary Methylation Index
Roxarsone	4-hydroxy-3-nitrobenzenearsonic acid
TMAs(III)	Trivalent Methyl Arsenic
TMAs(V)O	Trimethylarsine Oxide

TR-FRET	Time-resolved Förster Resonance Energy Transfer
Trx	Thioredoxin
TR	Thioredoxin Reductase
TCEP	Tris(2-carboxyethyl)phosphine

Introduction to the Dissertation

My dissertation research includes several diverse studies that are related through their interest in biological transformations of arsenic, both inorganic and organic. These are divided into major projects: 1) the role of glutaredoxin 2 (Grx2) in reduction of arsenate to arsenite; 2) arsenic resistance in a gut microbiome organism, the obligate anaerobe *Bacteroides vulgatus* ATCC 8482; 3) methylation of inorganic arsenite by the As(III) S-adenosylmethione (SAM) methyltransferase (ArsM in microbes and AS3MT in animals) from the human pathogenic fungus *Asperillus fumigatus*; and 4) catalytic properties of polymorphic variants of the human hAS3MT, which is the major focus of my dissertation. Each project is described in a separate chapter.

Arsenic is the most pervasive environmental toxic substance. As a consequence of its ubiquity, nearly every organism has genes for resistance to inorganic arsenic, which are found largely in bacterial arsenic resistance (*ars*) operons. These operons contain diverse sets of genes that vary from one organism to another. The core operon usually includes an As(III)-responsive transcription factor, ArsR, that regulates expression of the other *ars* genes, an efflux protein for As(III) that confers resistance to inorganic arsenite, either ArsB or Acr3,, and a reductase, ArsC, that reduces As(V) to As(III). The first chapter describes the ArsC arsenate reductase of *Escherichai coli* and its requirement for glutaredoxin 2 (Grx2) as an accessory protein required for ArsC. In my first project I characterized the activity of Grx2 and demonstrated that it has both redox activity and glutathione S-transferase (GST) activity. This study was published as Ye. J., Nadar. S.V., Li, J. and Rosen. B.P. Structure of *Escherichia coli* Grx2 in complex with

glutathione: a dual-function hybrid of glutaredoxin and glutathione S-transferase *Acta Cryst. D*, 70, 614-761 (2014).

In my second project, I analyzed the arsenic resistance genes in *Bacteroides vulgatus*, which is an obligate anaerobic human gut microbiota. This common microbiome organisms has an *arsRDABC* operon. In my second chapter, I demonstrated that *B. vulgatus* has arsenic-responsive genes that confer resistance to inorganic and organic arsenicals and may be responsible for its ability to maintain the organism's prevalence in the gut following dietary arsenic exposure. This study was published as Li, J., Mandal, G. and Rosen, B.P. Expression of arsenic resistance genes in the obligate anaerobe *Bacteroides vulgatus* ATCC 8482, a gut microbiome bacterium. *Anaerobes*, 39,117-123(2016).

My third and fourth projects both involve As(III) methyltransferases. The third project demonstrates that most fungal species have ArsM orthologs with only three conserved cysteine residues. ArsM from *Aspergillus fumigatus* is a functional MAs(III) SAM methyltransferase with little ability to methylate inorganic As(III), which support our hypothesis that conserved cysteine determine substrate specificity and help elucidate the catalytic mechanism of ArsM. The phylogeny of three-cysteine AfarsM also suggests horizontal transfer of gene (HGT) for a three-cysteine ArsM from bacteria to fungi. This study was published as Chen, J., Li, J., Jiang, X. and Rosen B.P. Conserved cysteine residues determine substrate specificity in a novel As(III) S-adenosylmethionine methyltransferase from *Aspergillus fumigatus*. *Mol Microbiol.* 104, 250-259 (2017).

My final and major project focused on the human arsenic biotransformation process, which is thought to be protective from acute high-level arsenic exposure. There are large individual variations for arsenic methylation in human populations, which show up as different urine DMA:MAAs arsenic metabolites ratios. Some variations are associated with single nucleotide polymorphisms (SNPs) in intronic region of the hAS3MT gene that appear to increase expression of the gene and hence increased arsenic tolerance (Schlawicke Engstrom et al. 2007). I identified SNPs within the coding region of the hAS3MT and examine the effect of single amino acid substitutions in hAS3MT on the activity of the enzyme that might explain their contributions to adverse health effects of environmental arsenic. This study is under review in *Chemical Research in Toxicology* as Li, J., Packianathan, C., Rossman, T.G. and Rosen, B.P. The biological consequences of polymorphisms in the human AS3MT arsenic methylation gene. EHP submitted (2017).

CHAPTER 1 Introduction

1.1 Arsenic Occurrence and Uses

Arsenic is a toxic metalloid that is the 20th most abundant element in the Earth's crust (Mandal and Suzuki 2002). Arsenic in bedrock is released into groundwater that is used to grow food and as drinking water, causing a world-wide health hazard that affects tens of millions of people (Naujokas et al. 2013). Arsenic is the most ubiquitous toxic substance in the environment. It contaminates soil, drinking water and our food supply. Tens of millions of people world-wide are exposed to arsenic by ingestion, inhalation, and dermal absorption routes. In the United States, as many as 25 million people are estimated to drink water with an arsenic level of >10 ppb (parts per billion) (Kozul *et al.*, 2009), the upper limit recommended by the World Health Organization and the maximal

containment level allowable by the U.S. Environmental Protection Agency (EPA). The EPA and the Agency for Toxic Substances and Disease Registry (ATSDR) rank arsenic first on the U.S. Priority List of Hazardous Substances (<http://www.atsdr.cdc.gov/SPL/index.html>). The U.S. Food and Drug Administration (FDA) considers that arsenic endangers the safety of our food supply (<http://www.fda.gov/Food/FoodborneIllnessContaminants/Metals/ucm319870.htm>). In countries such as China, arsenic in rice has been associated with increased risk of cancer in exposed populations (Li *et al.*, 2011). Consequently, the United Nations food standards body Codex Alimentarius Commission recommends that the level of arsenic in rice should not exceed 0.2 mg/kg (<http://www.fao.org/news/story/en/item/238558/icode/>). Environmental arsenic is also highly correlated with cardiovascular disease (States *et al.*, 2009), diabetes (Drobna *et al.*, 2012; Kuo *et al.*, 2015), and other human disorders (Jomova *et al.*, 2011).

In addition, humans are exposed to organoarsenicals, which can end up in water and food supplies. Both methylated and aromatic arsenicals have been used for decades for agriculture and animal husbandry. Pentavalent methylarsenate (MAs(V)) has been used as the herbicide monosodium methylarsenate (MSMA) (Akkari *et al.*, 1986); and is still used in the United States for the treatment of cotton fields, golf courses, turf management and highway medians. Roxarsone (4-hydroxy-3-nitrobenzenearsonic), nitarsonic acid (4-nitrophenyl)arsonic acid) and p-ASA(V) (p-aminophenyl arsonic acid or Atoxyl) are produced and used today in many countries as antimicrobial growth promoters for poultry and swine to control *Coccidioides* infections (Garbarino *et al.*, 2003). Even though Pfizer has voluntarily suspended U.S. production of roxarsone and atoxyl, they still produce the chemically-related compound nitarsonic acid, the only treatment for blackhead disease, or histomoniasis, in turkeys. These aromatic

arsenicals are introduced into the environment when chicken litter is used as fertilizer (Garbarino *et al.*, 2003). Pentavalent organoarsenicals are relatively nontoxic, but the trivalent forms are exceptionally toxic. Aromatic and methylarsenicals are activated to their toxin forms by microbial reduction (Chen *et al.*, 2014), which can contaminate water supplies through run-offs from farms (Fisher *et al.*, 2015), golf courses (Feng *et al.*, 2005; Matteson *et al.*, 2014) and other treated soils. Moreover, food crops such as rice accumulate methylated arsenicals, providing an additional source of dietary organoarsenicals (Lomax *et al.*, 2011; Zhao *et al.*, 2013).

1.2 Arsenic Resistance System

Life arose when the atmosphere was neutral, and toxic metal ion concentrations were higher than in present-day oceans (Zhu *et al.* 2014). To survive in the presence of arsenic, the most ubiquitous of all environmental toxins, the earliest cells must have had mechanisms for detoxifying arsenic. This intense selective pressure was the driving force for the evolution of bacterial arsenic resistance (*ars*) genes and pathways that are found in nearly every extant organism (Figure 1.1). Without arsenic detoxification systems, life could not have arisen and would not exist today.

Expression of arsenic resistance operons are usually controlled by the As(III)-responsive transcription factor ArsR: The ArsR regulatory protein binds to a specific DNA element and thus represses expression of the genes coding for the arsenic defense mechanism (San Francisco *et al.*, 1990). When As(III) enters the cell, it binds to cysteine residues in ArsR, which induces ArsR dissociation from the DNA, allowing RNA polymerase to access the operon sequence.

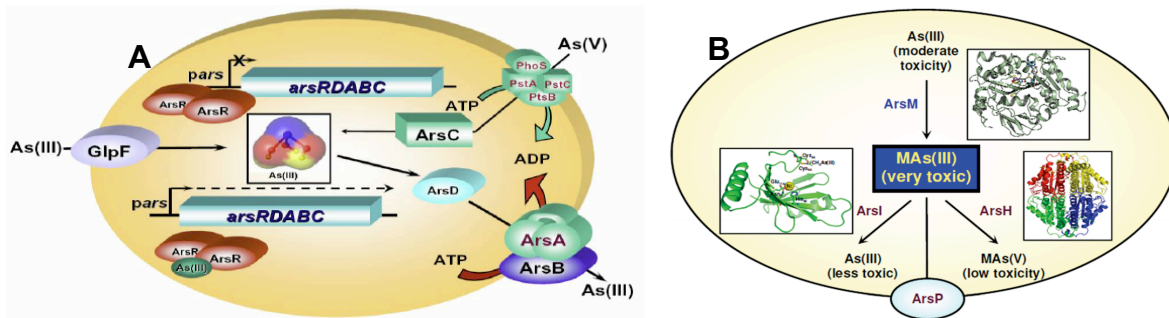


Figure. 1.1 Arsenic resistance system. (A) Arsenate (As(V)) is taken up by phosphate transporters, and arsenite (As(III)) is taken up by glycerol facilitator. As(III) binds to the ArsR repressor, inducing expression of the *arsRDABC* genes, which results in arsenate reduction and arsenite extrusion from the cells. (B) In communities of soil microbes some bacteria such as *Rhodopseudomonas palustris* carry the *arsM* gene for the As(III) SAM methyltransferase, producing highly toxic MAs(III). This trivalent organoarsenical has antibiotic-like properties. Other soil bacteria carry genes for MAs(III) resistance. Some, such as *Bacillus MD1*, have the *arsI* gene for the ArsI C-As lyase genes that confers resistance to MAs(III) by degrading it into As(III) and formaldehyde. Yet other soil bacteria such as *Pseudomonas putida* have a gene encoding ArsH, a flavoprotein that uses NADP⁺ to oxidize MAs(III) to MAs(V), thus conferring resistance. Finally, other bacteria such as *Campylobacter jejuni*, which inhabits the intestinal track of poultry and other farm animals, carry the *arsP* gene. ArsP is a MAs(III) efflux permease that extrudes trivalent organoarsenicals from cells, conferring resistance. The crystal structures of the relevant enzymes are shown next to their reactions. ArsA, arsenite pump-driving ATPase; ArsB, arsenite efflux pump ACR3 or related permease; ArsC, arsenate reductase; ArsD, As(III) chaperone; ArsR, As(III)-responsive transcription factor; ArsI, MAs(III) demethylase; ArsH, NADPH-dependent FMN reductase; ArsP, Methylarsenite permease; ArsM, As(III) S-adenosylmethyltransferase

The *ars* gene products include either ArsB or (San Francisco et al., 1988) or Acr3 (Wysocki et al., 1997; Ghosh et al., 1999) , both of which are As(III) efflux permeases. These keep the cellular concentration of As(III) at low levels, conferring resistance (Liu et al., 2013). ArsA is the catalytic subunit of the pump, which associates with ArsB, the As(III) conducting pathway. ArsD is an arsenite metallochaperone that binds arsenite and transfers it to ArsA. The interaction of ArsA and ArsD allows for resistance to As(III) at environmental concentrations.

The *arsDAB* genes confer resistance to trivalent As(III), not pentavalent As(V). As the atmosphere gradually became oxygenated 2.4 BY ago, As(III) was oxidized to As(V), and organisms faced a different challenge of arsenic toxicity in the form of arsenate. To detoxify As(V), organisms evolved reductases that reduce As(V) to As(III) (Mukhopadhyay and Rosen, 2002); which is the substrate of ArsB and can then be extruded from the cells. Alternatively, As(III) can be methylated to produce volatile arsenicals (Qin et al., 2006). These As(V) detoxification mechanisms have the benefit of getting rid of arsenic without losing phosphate, a nutrient essential to life.

There are at least three independent and unrelated families of arsenate reductases. (Mukhopadhyay and Rosen, 2002) One family of arsenate reductase that includes the chromosomal *E. coli* and plasmid R773 ArsCs uses glutaredoxin (Grx) and GSH as reductants. The second family includes the *Staphylococcus. aureus* plasmid p1258 ArsC and the *B. subtilis* chromosomal ArsC, both of which use thioredoxin (Trx) as reductant. This bacterial family of arsenate reductases is related to low molecular weight protein tyrosine phosphatases (LMW PTPase). The third family of arsenate reductases is found primarily in eukaryotic microorganisms and includes *S. cerevisiae* Acr2p and LmACR2 from the protozoan *Leishmania. major*. These eukaryotic arsenate reductases are related to the catalytic domain of the Cdc25 cell cycle protein tyrosine phosphatase (Bhattacharjee et al., 2010). From a comparison of the three structures, which show no similarity to each other, it is clear that these arsenate reductases arose multiple times by convergent evolution.

E.coli has three *grx* genes in its chromosome. The R773 ArsC has a strong preference for Grx2. To explore this preference, we solved the crystal structure of Grx2 and

determined its catalytic properties. I showed that Grx2 is a multifunctional reducing agent. It has two catalytic activities, as a glutathione (GSH) thiol transfer protein and as a GSH S-transferase (GST). Its role as a GSH thiol transferase had been demonstrated, but we postulated that GST activity would assist in forming an As(III)-SG complex that could be the substrate for ArsC, ArsM and AS3MT. My results support this hypothesis.

In addition to pathways of the above described resistance to inorganic arsenic, our laboratory recently identified parallel pathways of organoarsenical biotransformations. The *arsM* gene product, which encodes the As(III) S-adenosylmethionine (SAM) methyltransferase (EC 2.1.1.137), which is called ArsM in microbes and AS3MT in animals, methylates arsenite in three sequential steps to the trivalent species methylarsenite (MAs(III)), dimethylarsenite (DMAs(III)) and volatile trimethylarsine (TMAs(III)) (Qin *et al.*, 2009; Qin *et al.*, 2006; Dheeman *et al.*, 2014; Marapakala *et al.*, 2015). MAs(III) is considerably more toxic than As(III), which has led us to propose that MAs(III) was a primordial antibiotic. Under aerobic conditions these products are oxidized to nontoxic pentavalent arsenicals, so that methylation became a detoxifying pathway after the atmosphere became oxidizing. Other microbes have acquired the ability to re-reduce MAs(V), transforming it again into toxic MAs(III). This has led to the evolution of the *arsI*, *arsH* and *arsP* genes for MAs(III) resistance. *ArsI* is a C-As bond lyase that demethylates MAs(III) back to less toxic As(III), conferring resistance to this primitive antibiotic. *ArsH* oxidizes MAs(III) to MAs(V) (Chet *et al.*, 2015a) . *ArsP* actively extrudes MAs(III) from cells (Chet *et al.*, 2015b). This oscillation between MAs(III) synthesis and detoxification is a consequential component of the arsenic biogeochemical cycle.

1.3 ArsM and AS3MT: As(III) S-adenosylmethionine (SAM) Methyltransferases

Methyltransferases constitute a class of enzymes found in every living cell. They transfer a methyl group most frequently from S-adenosyl L-methionine (SAM or AdoMet) to a nucleophilic acceptor, producing S-adenosyl-L-homocysteine (SAH or AdoHcy) and a methylated molecule. Classification of this large superfamily is based on substrate specificity (e.g. DNA, RNA, protein, lipid and small molecules such as arsenic (Sanders et al., 2004; Schluckebier et al., 1995) and on the nucleophilic atom targeted for methylation (e.g. N, O, C or S). SAM dependent methyltransferases are involved in processes including biosynthesis, signal transduction, protein repair, chromatin regulation and gene silencing (Kozbial and Mushegian 2005; Wlodarski et al., 2011). Variation in the activity of those enzymes in humans is partially controlled by genetic polymorphisms (Weinshilboum and Raymond 1977; Weinshilboum and Sladek 1980; Rini et al., 1990), which may contribute to a more precise understanding of possible correlations between genotypes and disease-susceptibility phenotypes. “Pharmacogenetic” variation in these enzyme activities can contribute to individual differences in the metabolism, therapeutic effect, and toxicity of drugs that undergo methylation.

The gene for arsenic methylation is widespread in nature, found in many but not all members of every kingdom, from bacteria to humans. What differentiates ArsM/AS3MT from other members of the methyltransferase superfamily is the necessity to bind trivalent arsenicals and to maintain them in the reduced form, for which they have evolved and utilize four conserved cysteine residues absent in other methyltransferases. (Although, as described elsewhere, most fungi have only three of these four and have

lost the ability to methylate inorganic As(III)). ArsM or AS3MT is the primary enzyme for arsenic methylation that produces mono-, di-, and trimethylated arsenic metabolites, which is then excreted. Bacteria and archaea often evolve through horizontal gene transfer (HGT) (Fitzpatrick DA., 2012), and HGT has been suggested as a mechanism to explain the widespread occurrence of *arsM* in bacteria. The function of the enzyme is controversial. If the gene arose before the atmosphere became oxidizing, then the products would have remained trivalent, so the enzyme would activate inorganic arsenic into more toxic methylated species such as MAs(III). In extant aerobic microbes, this is clearly a detoxification process because the trivalent products are rapidly oxidized to nontoxic pentavalent species in air. But the ability to reduce MAs(V) is widespread in soil microbes. (Yoshinaga, Cai, & Rosen, 2011) These bacteria regenerate MAs(III), which, we postulate, kills off competitors. Waksman defined an antibiotic as “*a chemical substance, produced by micro-organisms, which has the capacity to inhibit the growth of and even to destroy bacteria and other micro-organisms*”. (Waksman, 1947) By this definition, MAs(III) is an antibiotic. In response to synthesis of this biological weapon, other environmental microbes evolved the mechanisms described below to detoxify MAs(III), the very definition of antibiotic resistance.

The first microbial *arsM* gene identified was from the soil bacterium *Rhodopseudomonas palustris*, (Qin et al., 2006) encoding the 283-residue ArsM protein (29,656 Da) (accession number NP_948900.1). Heterologous expression of the *arsM* from *Rhodopseudomonas palustris* conferred As(III) resistance to an arsenic-sensitive strain of *Escherichia coli*, demonstrating that the present-day microbial gene detoxifies As(III). Photosynthetic organisms play a significant role in arsenic geochemical cycling by methylation. Methylated arsenic species have been found in many photosynthetic

organisms, and several *arsM* genes and their protein products have been characterized in cyanobacteria, prokaryotic photosynthetic organisms that are ubiquitous in environments ranging from soil and water to deserts. As the dominant species in toxic algal blooms, *Microcystis* sp. PCC7806, *Nostoc* sp. PCC7120, and *Synechocystis* sp. PCC6803 are typical freshwater cyanobacteria that methylate arsenic. (Ye, Rensing, Rosen, & Zhu, 2012) Two cyanobacteria ArsM enzymes were purified and were shown to methylate arsenite *in vitro*, with trimethylarsine gas as the end product.

Why do animals methylate arsenic? Is the presence of gene for AS3MT a evolutionary fossil that inadvertently produces a toxin? Or does it play a role in arsenic detoxification? Paradoxically both are true, but the truth is more complicated. It is likely that the gene moved from prokaryotes to eukaryotes by horizontal gene transfer after the atmosphere become oxidizing. In eukaryotic microbes the gene confers arsenic resistance, so horizontal gene transfer probably conferred the ability to detoxify arsenic in lower eukaryotes. Eukaryotic algae such as the thermophilic red alga *Cyanidioschyzon merolae* isolated from hot springs such as found in Yellowstone National Park methylate arsenic rapidly, producing trimethylarsine gas. (Qin et al., 2009) The *arsM* gene from this alga confers resistance when heterologously expressed in *E. coli*. Similarly, the *arsM* gene from the mesophilic alga *Chlamydomonas reinhardtii* confers resistance in *E. coli*. (Chen, Qin, Zhu, de Lorenzo, & Rosen, 2013) It also rapidly transforms arsenic into nontoxic arsenosugars. So clearly arsenic methylation in eukaryotic microbes is a detoxification process.

Rapid and high level arsenic biomethylation and biovolatilization has also been observed in the aquatic protozoa *Tetrahymena thermophile*, which converted 66% of 40 μM

inorganic arsenic into MAs(V) and DMAs(V) within 48 hr (Yin et al., 2011). The invertebrate sea squirt *Ciona intestinalis*, one of the closest living relative of vertebrates, has an AS3MT gene, and this primitive chordate methylates inorganic arsenic (Thomas et al., 2009). Fungi are also prolific arsenic methylators (Bentley & Chasteen, 2002), and the genomes of some fungi such as *Aspergillus* have been shown to have novel AS3MT genes (accession number GAO81309). In contrast, unlike lower plants such as algae, higher plants do not have an AS3MT gene and do not methylate arsenic. Instead, methylated arsenicals in plants originate from rhizosphere microbes (Lomax et al., 2011).

In higher eukaryotes the situation is less clear. Higher eukaryotes methylate arsenic, and the gene for AS3MT was originally identified in rat (Lin et al., 2002). In zebrafish the AS3MT gene is expressed in liver and intestine, where inorganic arsenic is transformed into the mono- and dimethylated species (Hamdi et al., 2009). In mammals the products of liver methylation by AS3MT were thought to be the relatively nontoxic pentavalent species because those species found in urine (Challenger., 1974). In support of this concept, AS3MT knockout mice are more susceptible to arsenic in drinking water. (Yokohira et al. 2010; Dodmane et al. 2010). These results suggest that mammalian AS3MT is protective against acute arsenic exposure. However, more careful studies have shown that the urinary products are actually trivalent that oxidize when left in air (Hughes et al., 2009; Del Razo et al., 2011; Le et al., 2000). Purified arsenic methyltransferases, both human AS3MT and algal ArSM, produce MAs(III), not MAs(V) (Dheeman et al., 2009; Maraphakala et al., 2015). AS3MT produces sufficient intracellular MAs(III) to cause arsenic-related diseases such as bladder and other forms of cancer, diabetes and neurological disorders. (Abernathy et al., 1999; Abernathy et al.,

2003) In regions of the world with high arsenic in drinking water and food, these diseases are often found in exposed individuals. But most of the population of the world is exposed continually to low levels of arsenic that do not produce obvious symptoms but increase the probability of arsenic-related diseases later in life.

Since evolution favors survival of the young and is little concerned with disease after the organism's reproductive period is over, the AS3MT gene would be maintained in humans even if it led to long-term increases in arsenic-related diseases. A number of polymorphisms have been identified in the human AS3MT gene, both intronic and in the coding region of the gene (Agusa et al., 2011; Wood et al., 2006). and the distribution of polymorphisms in arsenic endemic regions of the world are associated with increased risk of skin lesions (Pierce et al., 2012). One such intronic polymorphism was identified in inhabitants of the northern Argentinean Andes, an arid region where elevated arsenic concentrations in drinking water is common (Schlebusch et al., 2012). In a genome-wide association study (GWAS) there was a strong association between an extragenic polymorphism on chromosome 10q24.32 that increases expression of the AS3MT and the amount of methylated arsenicals in urine. This is the first evidence that adaptation to environmental arsenic has likely driven an increase in the frequencies of protective AS3MT polymorphisms.

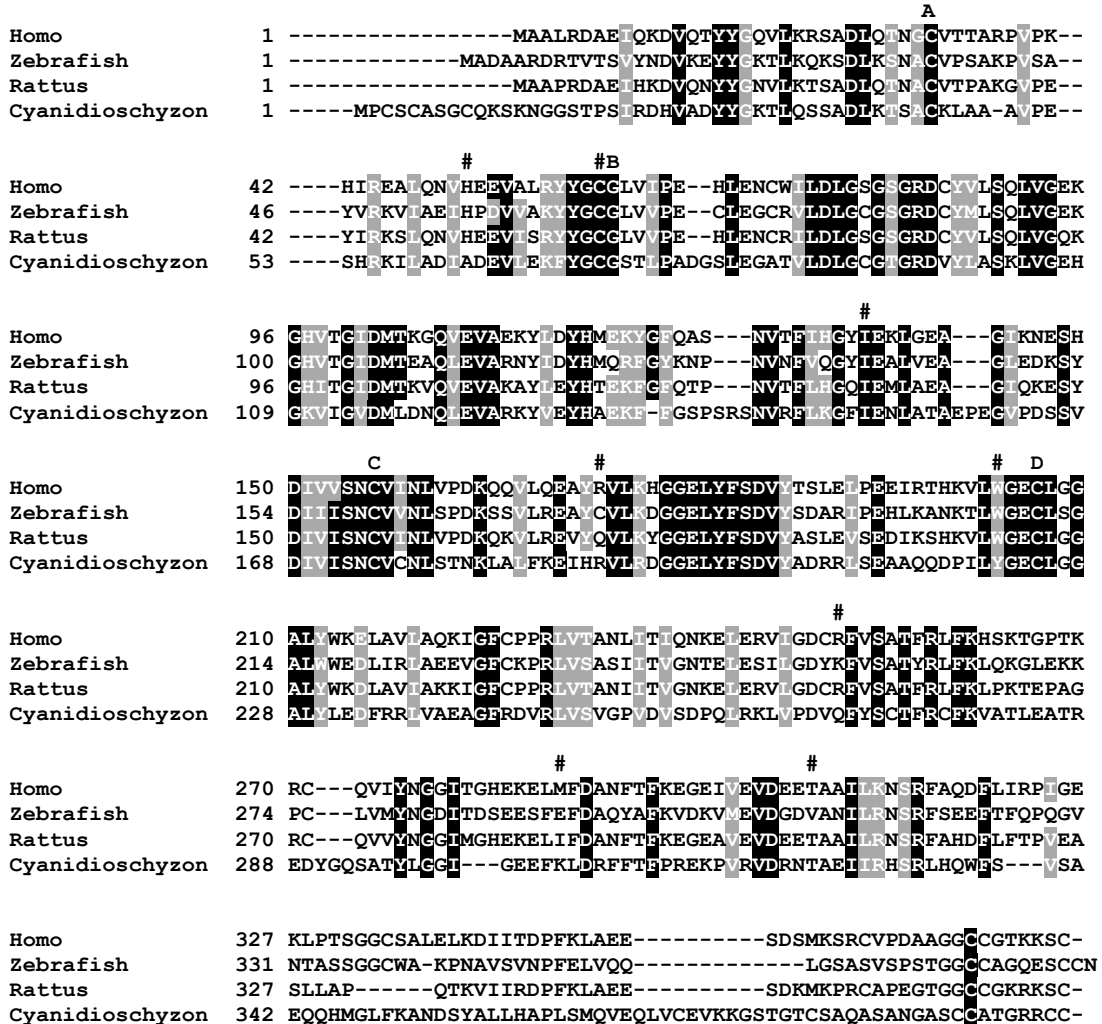


Fig. 1.2. Multiple sequence alignment of As(III) SAM methyltransferases. *Homo sapiens* As(III) S-adenosylmethionine methyltransferase (AAI19639) was aligned with eukaryotic orthologues from *Danio rerio* (zAS3MT, NP_001034928), *Rattus norvegicus* (rAS3MT, NP_543166) and *Cyanidioschyzon merolae* (CmArsM, FJ476310). Black shading: conserved residues; Grey shading: conservative replacements; conserved cysteine residues are labeled as A, B, C, D; #, polymorphic residues.

There has been recent progress on the mechanism of algal ArsM and human AS3MT. Alignment of their sequences shows four conserved cysteine residues (Fig. 1.2). The role of the cysteine residues in the catalytic cycle has been explored through a combination of structural, biochemical and molecular biological analyses. The crystal structure of the *C. merolae* CmArsM was solved at 1.6 Å (Ajees, Marapakala, Packianathan, Sankaran, & Rosen, 2012). The enzyme has three domains, an N-

terminal SAM binding domain, a middle As(III) binding domain, and a C-terminal domain of unknown function. A structure with bound phenylarsenite (PhAs(III)) shows that the two N-terminal conserved cysteines are present in a disulfide bond, and the other two conserved cysteines bind the arsenic atom (Fig. 1.3) (Dheeman et al., 2014). Each of the four cysteine residues was mutagenized in both the algal and human enzymes (Dheeman et al., 2014; Marapakala et al., 2012). The results were unexpected: all four single mutants lost all methylation activity, but, when either of the first two cysteines was changed, the enzyme lost the ability to methylate inorganic As(III) to MAs(III) (the first methylation step) but could still methylate MAs(III) to DMAs(III) (the second methylation step). These results are consistent with the crystallographic results that the second two cysteines form the arsenic binding site. In contrast, the first cysteine pair has a different role in catalysis that is only required during the first methylation (As(III)→MAs(III)) but not the second (MAs(III)→DMAs(III)). We proposed that disulfide bond formation between the first two cysteines is a required step. Transfer of the S-methyl group from SAM would oxidize the As(III) to form an enzyme-bound MAs(V) moiety. In order to proceed to the second methylation step, the arsenic atom must be re-reduced to trivalency. We suggest that one of first two cysteines reduces the pentavalent methylarsenic intermediate during disulfide bond formation. The disulfide bond is then reduced by thioredoxin, which has been shown to be a cofactor in AS3MT methylation (Ding et al., 2012; Thomas et al., 2004; Water et al., 2004), allowing the second methylation step to proceed. Since the first cysteine pair is not required for the second methylation step, that reaction may use a different combination of cysteine residues for reduction of an enzyme-bound DMAs(V) intermediate. Interestingly, a disulfide bond between the second and third cysteine residues is observed in a crystal structure with

bound SAM (Ajees et al., 2012). Thus a cascade of two successive disulfide bonds may be a feature of the methylation catalytic cycle.

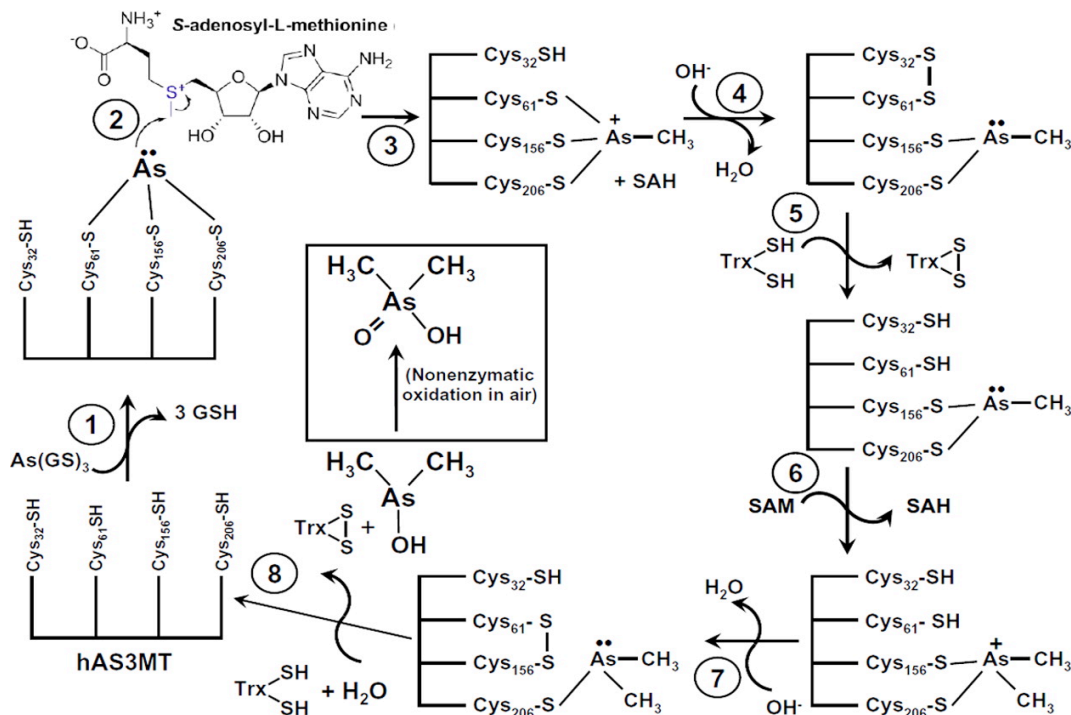


Figure 1.3. Proposed hAS3MT reaction scheme: (1) In the first round of methylation, hAS3MT binds As(III) in a series of three thiol transfer reactions from As(GS)₃; (2) the methyl group of SAM is attacked by the arsenic lone pair; (3) a pentavalent MAs(V) intermediate is formed and (4) reduced to an enzyme-bound MAs(III) intermediate by Cys₃₂ with formation of a Cys₃₂-Cys₆₁ disulfide; (5) the disulfide is reduced with Trx, and the enzyme undergoes the next round of methylation, (6) forming a pentavalent DMAs(V) intermediate, (7) which is reduced to DMAs(III) by Cys₆₁. This form is a Cys₆₁-Cys₁₅₆ disulfide, which (8) is reduced by Trx, regenerating the enzyme and releasing the major soluble product, DMAs(III). Finally, in air, trivalent DMAs(III) is nonenzymatically oxidized to DMAs(V). (Dheeman et al., 2014)

1.4 Phylogeny of Three-cysteine ArsMs

Both cluster mapping and phylogenetic analysis suggest that most horizontal gene transfer (HGT) events were of bacterial origin, and independently transferred to other kingdoms of life hundreds of millions years ago. Thus bacteria represent a rich source of

new functions potentially available for other kingdoms. HGT of the As(III) S-adenosylmethionine (SAM) methyltransferase gene (*arsM* in microbes and AS3MT in animals), which provides organisms the capacity to tolerate nonessential arsenic, occurred several times among species of different kingdoms (Chen et al., 2017).

According to the fossil record and molecular dating, the last cyanobacterial common ancestor (LCCA) is believed to have existed as early as 3 billion years ago (Bya) (Schirmer et al., 2015), considerably before the accumulation of oxygen in the atmosphere (2.45-2.32 Bya) during the Great Oxidation Event (GOE) (Bekker et al., 2004). Arsenic metabolism and cycling by microorganisms evolved at least 2.7 billion years ago (Sforna et al., 2014). Overall, the finding of an *arsM* gene in the LCCA implies that ArsM arose in an anoxic world before the GOE. This result calls the primordial role of ArsM as one of detoxification into question: methylation of arsenite increases toxicity by producing the more toxic MAs(III) and DMAs(III), which are thermodynamically stable in the oxygen-free environment that existed before the GOE (Canfiel et al., 2006). Thus, ArsM would not be a detoxification enzyme unless there is oxygen to convert the products to less toxic MAs(V) and DMAs(V). We postulate that the enzyme originally evolved to produce high toxic trivalent methylarsenicals that functioned as antibiotics to kill competitors in the primordial environment; after the GOE the trivalent species were oxidized to pentavalent forms, transforming ArsM into a detoxifying enzyme (Chen et al., 2015b).

As described above, bacterial, algal and mammalian ArsMs studied to date all have four conserved cysteine residues.(Fig. 1.2) Functionally, the two closest to the N-terminus form one pair, and the other two form a second pair (Marapakala et al., 2012, 2015;

Dheeman et al., 2014). For clarity, they are labeled CysA, CysB, CysC and CysD according to their distance from the N-terminus. All four cysteines are required for the first methylation step (As(III)→MAs(III)), but the second methylation step (MAs(III)→DMAs(III)) requires only the second pair, CysC and CysD. CysC and CysD form the binding site for trivalent arsenicals, explaining why the enzyme is inactive in their absence. Mechanistically CysA and CysB are thought to be involved reduction of an enzyme-bound MAs(V) intermediate, producing a disulfide bond between CysA and CysB and an enzyme-bound MAs(III) complex (Marapakala et al., 2015; Dheeman et al., 2014). The disulfide bond is reduced by thioredoxin, regenerating the active form of the enzyme. Thus the two conserved cysteine pairs are considered essential for ArsM function.

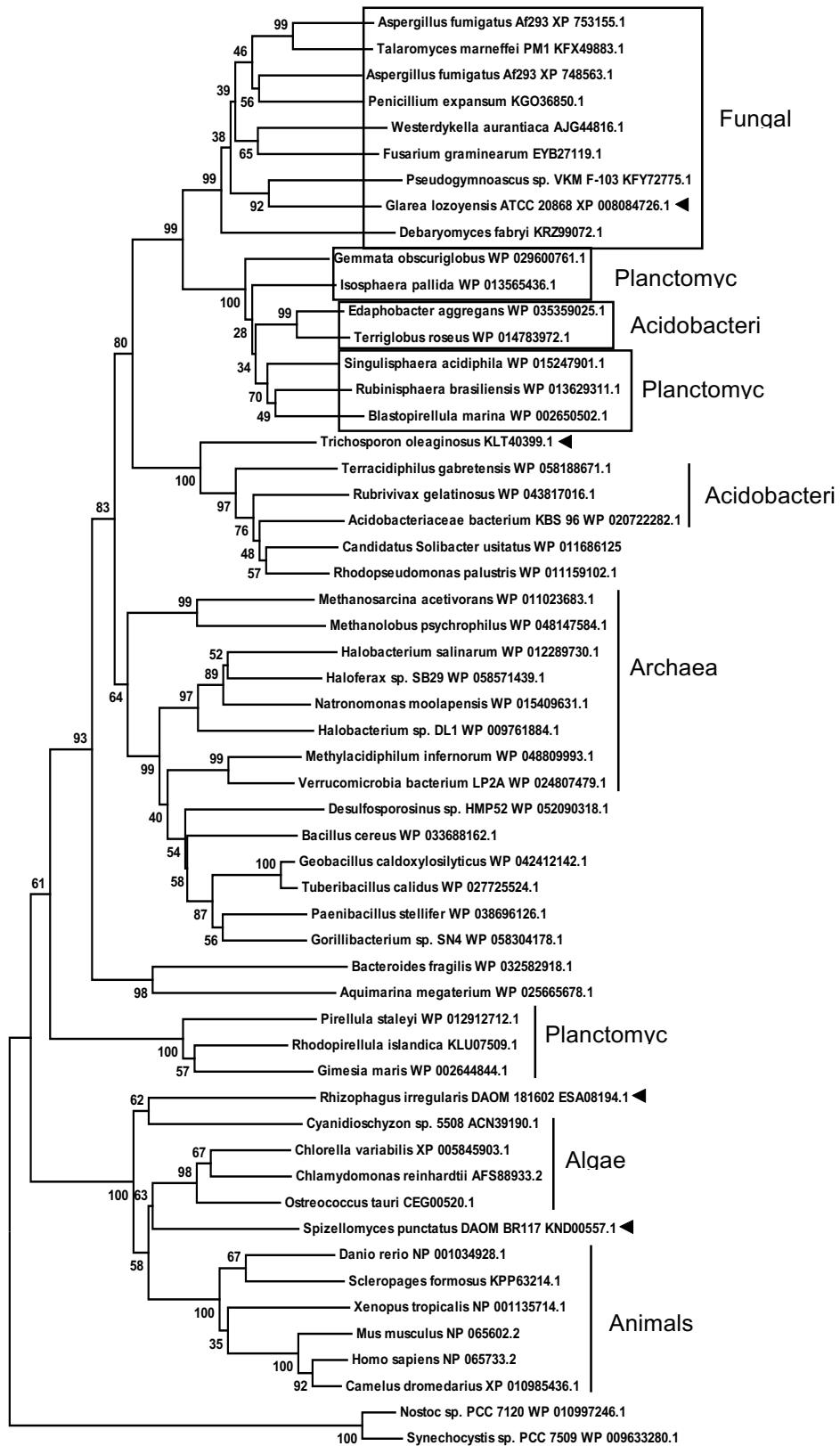
An emerging pattern is that most inter-kingdom HGT of *arsM* gene involves acquisition of genes from bacteria. Bacteria are abundant in nearly all environments, and many eukaryotes survive by eating and digesting them in large numbers (Huang JL., 2013), so bacterial genes, including *arsM*, are a source of foreign genetic material for eukaryotes. Secondly, bacteria are metabolically more diverse than eukaryotes, and, as such, have much more to offer in the way of new and beneficial functions, such as arsenic methylation. In addition, many of the bacteria have mechanisms, such as the type IV secretion system, to transfer DNA to the environment and to other organisms (Wallden et al., 2010). These ecological links between donor and recipient lineages further implies that bacteria provide a rich source of genetic material that allows rapid adaptations to new ecological niches, in this case, in the context of arsenic stress. Although fungi represent the most recalcitrant of all organisms to gene transfer due to its robust cell walls and lost phagotrophic capacities, conjugation-like transfer and ecological

association with other microorganisms may have promoted HGT between prokaryotes and fungi (Fitzpatrick DA., 2012). We conducted a phylogenetic analysis of ArsM sequences in 55 fungal species, which demonstrated that the majority of *arsM* gene products appear to have only three of the four conserved cysteine residues (loss the first conserved cysteine). These fungal ArsMs have low similarity to most bacterial ArsMs and animal AS3MTs. A representative fungus is *Aspergillus fumigatus*, a common soil organism. *A. fumigatus* Af293 has four ArsM orthologs, AfArsM1, 2 and 3 cluster in one group, while AfArsM4 is in a more distant group. Comparing AfArsM1 with animal and plant orthologs, it shares only 26% identity with human AS3MT and 29% with algal CmArsM (*Cyanidioschyzon* sp). This low similarity indicates a more distant evolutionary relationship of fungal ArsMs with other eukaryotic As(III) SAM methyltransferases. Two other fungal species, *A. lentulus* and *Neo-sartorya fischeri* NRRL 181, have multiple ArsMs that fall into the same two separate groups as the AfArsMs. This distribution in two groups in multiple organisms suggests that the ancestral genes may be been acquired by separate horizontal gene transfers from bacterial sources. The ancestor of AfArsM1 may have given rise to AfArsM1, 2 and 3 by gene duplications, leading to multiple copies in *A. fumigatus* Af293.

In addition, the fungal species *A. fumigatus* has more arsenic resistance genes than any other eukaryote, and they are all resemble bacterial arsenic resistance determinants (Chen et al., 2017). In the eight chromosomes of *A. fumigatus* Af293 there are six clusters of *ars* genes that are similar to their bacterial counterparts. Included in the *ars* clusters on chromosomes 1, 3, 5 and 8 are four AfarsM genes. The original HGT must have been an entire bacterial *ars* operon. This includes ArsH, which gives MAs(III) resistance, implying that *A. fumigatus* comes into contact with both inorganic and

organic arsenicals. The acquisition of bacterial arsenite reductase genes as well as other arsenic resistance genes by fungi reflects the arsenic dilemma faced by fungi during their evolution trajectory. Thus, HGT of multiple *ars* genes from bacteria to fungi, including those involved in arsenate reduction and As(III) methylation, provides synergistic arsenic tolerance. Why does *Aspergillus fumigatus* have so many clusters of *ars* genes, including four *arsM* genes? Multiple *ars* clusters may confer additive tolerance to arsenic. On the other hand, multiple clusters may make it possible to express a cellular activity in a wider range of environmental conditions. For example, *Pseudomonas putida* KT2440 bears two virtually identical *arsRBCH* operons, but they are expressed at different temperatures (Paez-Espino et al., 2017). Thus, the number and expression of arsenic-related genes, and their acquisition, may depend on the selective pressure of varying environmental stresses.

To trace the bacterial origin of AfArsM, a phylogenetic analysis of ArsM sequences of 55 representative prokaryotes and eukaryotes organisms was conducted. (Fig. 1.4) On the whole, the fungal sequences were most closely related to bacterial sources from Planctomycetes and Acidobacteria, which also have three conserved cysteines. These two bacteria sequences are quite distant from other prokaryotic ArsMs with four conserved cysteines. The fungal ArsM of *Glarea lozoyensis* ATCC 20868 (XP_008084726.1) has four conserved cysteines but clustered with other fungi. On the other hand, the fungal ArsMs from *T. oleaginous*, *R. irregularis* and *S. punctatus*, which each have four conserved cysteines, did not cluster with other fungi, suggesting that those three ArsMs were acquired from bacterial sources more recently than the initial horizontal gene transfer to fungi. This analysis suggests that the ArsM of most fungi originated from Planctomycetes and Acidobacteria (Chen et al., 2017).



0.1

Figure 1.4. A neighbor-joining phylogenetic tree showing the evolutionary relationships of fungi ArsM protein with arsenic methyltransferase proteins from members of other kingdoms. ArsM sequences with three conserved cysteines are highlighted by closed squares. Fungal ArsM sequences with four conserved cysteines are indicated by black triangles.

The genes for ArsM, which were transferred from various bacteria to other kingdoms of life, meet these qualifications. Firstly, ArsMs contain both SAM binding and catalytic domains, and require no other proteins for their activity (Ajees et al., 2012). Secondly, the ability to methylate arsenic into innocuous organic forms is of immediate advantages for organisms suffering from nonessential arsenic stress, such as archaeal extremophiles which live in extreme environments such as hot springs, where the concentration of arsenic is often very high (Zhu et al., 2014). Acquiring an arsenite methylation pathway might therefore be critical for adaptation to extreme environments. This applies to humans living in high arsenic environments as well, where rapid methylation of arsenic is thought to be protective from acute high-level arsenic exposure. Recent epidemiological research data show that adaptation to tolerate the environmental stressor arsenic has likely driven an increase in the frequencies of protective variants of AS3MT, providing the first evidence of human adaptation to a toxic chemical. Single nucleotide polymorphisms (SNPs) in the hAS3MT gene which lead to higher arsenic metabolization capacity have been selected as an adaptive mechanism for populations to survive in an arsenic-laden environment. For example, one protective hAS3MT polymorphism that produces low urinary excretion of MAs (7.5%) and a higher percentage of DMAs (78%) is found in indigenous populations in the Argentinean Andes exposed for generations to elevated arsenic in their water supply (0.8 mg/L) (Schlawicke Engström et al., 2007). The increased ratio of DMAs:MA in urine in this population may reflect natural selection for SNPs in the non-coding region of the human hAS3MT gene that increases expression of the gene, leading to increased arsenic tolerance. Recently

inhabitants of Camarones in the Arica y Parinacota Region in Chile who are exposed to >1mg/L of arsenic in their drinking water were found to have four protective genetic variants of the hAS3MT gene (G12390C, C14215T, T14458C, and G35991A) (Apata et al. 2017). These variants were all associated with efficient arsenic metabolism, providing convincing evidence of human adaptation to persistent high levels of arsenic (Apata et al. 2017; Schlebusch et al. 2013; Schlebusch et al. 2015).

CHAPTER 2. Statement of the Problem and Objectives

2.1 Statement of the Problem

Arsenic methylation, the primary biotransformation in the human body, is catalyzed by hAS3MT. This process is thought to be protective from acute high-level arsenic exposure. However, with long term low-level exposure, hAS3MT produces intracellular methylarsenite (MAs(III)) and dimethylarsenite (DMAs(III)), which are considerably more toxic than inorganic As(III) and may contribute to arsenic-related diseases. Several single nucleotide polymorphisms (SNPs) in putative regulatory elements of the hAS3MT gene have been shown to be protective. Until this study, only three SNPs that encode single amino acid changes in the hAS3MT protein, R173W, M287T and T306I, have been identified (Wood et al. 2006). What other hAS3MT SNPs exist in human populations, and how do they affect arsenic methylation activity and disease susceptibility?

The human gut contains 100-fold more genes than our own genome, and more than 90% are anaerobes. In the rat gut microbiome arsenic affects the diversity of microbial populations (Lu *et al.*, 2013) and their metabolism (Lu *et al.*, 2014). Exposure to arsenite

disturbs the mouse gut microbiome and substantially alters its metabolomics profile (Lu *et al.*, 2014). The prevalence of four families of *Firmicutes* bacteria significantly decreased, whereas *Bacteroidetes* families remained relatively unchanged following arsenic exposure, suggesting that *Bacteroides* may be intrinsically resistance to arsenic. In spite of the impact of arsenic on the microbiome, nothing is known about the effect of arsenic on microbiome gene expression. Our gut microbiome is exposed to dietary arsenic before it enters our blood stream; in effect, it is potentially our first-line for sensing environmental arsenic. There are only a few studies of anaerobic arsenic-inducible gene expression in environmental microbes such as the facultative anaerobe *Shewanella putrefaciens* (Murphy and Saltikov, 2009; Stolz *et al.*, 2006) and, to our knowledge, none in the obligate anaerobes of the gut microbiome. This lack of information limits our understanding of how the microbiome responds to the dietary arsenic.

2.2 Objectives

The goal of this study was to examine the effect of single amino acid substitutions in hAS3MT on the activity of the enzyme that might explain their contributions to adverse health effects of environmental arsenic. I identified five additional intragenic variants in hAS3MT (H51R, C61W, I136T, W203C and R251H). I purified the eight polymorphic hAS3MT proteins and characterized their enzymatic properties. Each enzyme had low methylation activity through decreased affinity for substrate, lower overall rates of catalysis and/or lower stability. I propose that amino acid substitutions in hAS3MT with decreased catalytic activity lead to detrimental responses to environmental arsenic and may increase the risk of arsenic-related diseases.

In human gut microbiome *B.vulgatus* study, expression of each of the *ars* genes was

determined by quantitative real time polymerase chain reaction (Q-RT-PCR) in response to exposure to trivalent arsenite, pentavalent arsenate and to the methylated arsenicals MAs(V) and MAs(III). The *ars* genes were induced most effectively by trivalent arsenicals. The *arsD* gene was the most highly expressed of the *ars* operon genes. Since these genes might be expressed even before arsenic enters our blood stream, I propose that expression of the *ars* genes in a gut microbe has to potential to be adapted for use as a biomarker of dietary exposure to arsenic. Knowledge of the response of a gut microbe to arsenic exposure increases understanding of how microbiome diversity is maintained.

CHAPTER 3. *Escherichia coli* Grx2: A dual-function Hybrid of Glutaredoxin and Glutathione S-transferase

3.1 Abstract

We solved the structure of glutaredoxin 2 (Grx2) from *Escherichia coli* co-crystallized with glutathione (GSH) to 1.60 Å. The structure of a mutant with active site residues Cys9 and Cys12 changed to serine crystallized in the absence of glutathione was solved to 2.4 Å. Grx2 has an N-terminal domain characteristic of glutaredoxins, and the overall structure is congruent with the structure of glutathione S-transferases (GSTs). Purified Grx2 exhibited GST activity. Grx2, which is the physiological electron donor for arsenate reduction by *E. coli* ArsC, was docked with ArsC. The docked structure could be fitted with GSH bridging the active sites of the two proteins. We propose that Grx2 is a novel Grx/GST hybrid that functions in two steps of the ArsC catalytic cycle: as a GST, it catalyzes glutathionylation of the ArsC-As(V) intermediate and, as a glutaredoxin, it catalyzes deglutathionylation of the ArsC-As(III)-SG intermediate.

3.2 Introduction

The objective of this project was to elucidate the properties of Grx2 that allows it serve as the primary electron donor for arsenate reduction. My hypothesis is that the true substrate of ArsC, ArsM and perhaps other arsenic enzymes and transporters is the glutathionylated species, As(GS)₃. Although As(III) will react slowly with GSH, I postulate that the reaction is catalyzed by a GST enzyme. Here I demonstrate that Grx2, unlike other glutaredoxins, has bifunctional activity as both as Grx and GST. Glutaredoxins are oxidoreductases (Fernandes & Holmgren, 2004) that belong to thioredoxin (Trx) superfamily (Martin, 1995). They are ubiquitous in most living organisms, from bacteria to human, and some viruses. Their primary function is to maintain the reduced state of cysteine residues in intracellular proteins, and they are involved in the reduction of proteins during oxidative stress (Stroher & Millar, 2012). *Escherichia coli* Grx1 was the first identified glutaredoxin, and it functions as a GSH-dependent electron donor for ribonucleotide reductase (Holmgren, 1976). Grxs can reduce mixed disulfides between protein and GSH, a process termed deglutathionylation. There are two mechanisms for deglutathionylation. One is a monothiol reaction in which only the N-terminal cysteine residue is required for reducing the mixed disulfide. The other is a dithiol reaction in which both active site cysteine residues form an intramolecular disulfide bond during reduction of protein disulfides (Stroher & Millar, 2012).

Most organisms have several glutaredoxin isoforms. *E. coli* has four, Grx1, Grx2, Grx3 and Grx4, which fall into three categories in terms of structural and catalytic properties. Grx1 and Grx3 are classical 10 kDa glutaredoxins with two conserved active site cysteine residues in a CPYC motif that fall into the first category. Grx2 also has a CPYC

motif but is larger at 215 residues (24,532 Da) and belongs to second category. Grx4 is in the third category with a monothiol active site CGFS (Holmgren, 1976, Åslund *et al.*, 1994, Åslund *et al.*, 1996, Fernandes *et al.*, 2005).

Grx2 distinguishes itself from the other three glutaredoxins by its abundance in the cell (Vlamiš-Gardikas *et al.*, 1997), relatively larger molecular weight, low sequence similarity and unique kinetic properties. Unlike the other Grxs, it is not an electron donor to ribonucleotide reductase (Vlamiš-Gardikas *et al.*, 1997). Grx2 reduces protein cysteine glutathione mixed disulfides but it is not efficient like other glutaredoxins (Lundström-Ljung *et al.*, 1999). The only primary function identified for Grx2 is that is the most effective hydrogen donor for reduction of arsenate by plasmid R773 ArsC arsenate reductase, and its catalytic efficiency is considerably higher than the other three *E. coli* glutaredoxins (Shi *et al.*, 1999).

The NMR structure of reduced Grx2 from *E. coli* shows that it has the N-terminal domain with a conserved Trx fold motif (Xia *et al.*, 2001) and the C-terminal domain, which is not present in the other three glutaredoxins, is structurally similar to glutathione S-transferases (GSTs), in particular the human ω class GST (Board *et al.*, 2000). However the lack of structural information on the *E. coli* Grx2-GSH complex or on the oxidized form of Grx2 and its catalytic properties, the mechanism of reduction of mixed disulfides by Grx2 remains unclear. We described the crystal structures of the glutathione-bound complex and a C9S/C12S double mutant Grx2, which mimics the reduced form. I showed that purified Grx2 exhibited GST activity with the artificial substrate 1-chloro-2,4-dinitrobenzene (CDNB), as do authentic GSTs. Rigid-body docking of Grx2 with ArsC using the ZDOCK server suggests that the C-terminal domain is involved in interaction

with ArsC, which has been proposed to undergo a catalytic cycle involving formation of an ArsC-As(V)-SG intermediate, reduction to ArsC-As(III)-SG intermediate, which is deglutathionylated with release of the product, As(III). We propose that Grx2 is a novel Grx/GST hybrid that participates functions as a GST in glutathionylation of the ArsC-As(V) intermediate and as a Grx in deglutathionylation of the ArsC-As(III)-SG intermediate.

3.3 Material and Methods

3.31 Structure Determination and Refinement

Crystallization and X-ray data collection of the Grx2-GSH complex and the Grx2 C9S/C12S mutant have been reported (Sheng *et al.*, 2007). The Grx2-GSH complex structure was solved by molecular replacement with *AMoRe* (Navaza, 2001) using the first conformer of the NMR structure of reduced Grx2 (PDB ID: 1G7O) as an initial model (Xia *et al.*, 1992). The Matthews coefficient of 2.3 suggests that one protein molecule is in the asymmetric unit. The best solution with R-factor of 46.3% and correlation coefficient of 49.3% was taken for further model building. Twenty cycles of rigid body and twenty cycles of restrained refinements in *REFMAC5* (Vagin *et al.*, 2004) improved the correlation coefficient to 76% and reduced the R-factor to 37%. The side chains and GSH molecule were fitted into the electron density map using *COOT* (Emsley & Cowtan, 2004). The electron density of GSH and Cys9 residue clearly shows that there is no disulfide bond between them. Water molecules were added at various stages of refinement. One cycle of simulated annealing refinement was carried out in *PHENIX* (Adams *et al.*, 2002). 5% randomly selected reflections were used as the test for calculation of R-free. All side chains are put for the refinement. The electron density

(2Fo-Fc map) of the backbone atoms and all side chain atoms are complete at 1 σ contour level. The final R-factor and R-free values were 16.4% and 19.4%, respectively. Models were generated using *Chimera* (Pettersen *et al.*, 2004). Models with electron density map were generated using *PyMol* (DeLano, 2001).

The Grx2 mutant structure was also solved by molecular replacement with AMoRe using the partially refined Grx2-GSH structure as an initial model. There is one molecule in the asymmetric unit. The refinement was carried out as above. The final R-factor and R-free values were 20.6% and 28.0%, respectively. The electron density (contoured at 1 σ level) of all backbone atoms and side chain are complete except the side chains Arg57 and Lys81. Coordinates and structure factors have been deposited for the Grx2-GSH complex (PDB ID: 4KX4) and the Grx2 C9S/C12S mutant (PDB ID: 4KSM) (Bernstein *et al.*, 1977).

3.32 GST activity

Grx2 was purified as described (Shi *et al.*, 1999). *Schistosoma japonicum* mu-class sjGST was expressed from *E. coli* bearing plasmid pET41a+ and purified by Ni-NTA affinity chromatography using the same conditions as for Grx2. GST activity was measured using 0.3 μ M of either purified Grx2 or sjGST in an assay buffer consisting of 0.1 M potassium phosphate (pH 6.5), 2.5 mM reduced GSH and the indicated concentrations of CDNB (Habig *et al.*, 1974, Mannervik & Danielson, 1988, Wilce & Parker, 1994). After preincubation at 37°C for a few minutes, Grx2 was added to initiate the reaction, which was followed from the change in absorption at 340 nm. The concentration dependence for GSH was assayed with 4 mM CDNB, and the concentration dependence for CDNB was assayed with 1 mM GSH.

3.33 Docking of ArsC-Grx2

The *E. coli* ArsC crystal structure (PDB ID. 1I9D) was used for analyzing ArsC-Grx2 interactions. Two sets of coordinates without water molecules, the ArsC-Grx2-GSH (full structure) and the ArsC-Grx2-GSH (C-terminal domain eliminated), were submitted to the ZDOCK (version ZD3.02) server (Pierce *et al.*, 2011), and the five top-ranked predictions were minimized using NAMD2 (Phillips *et al.*, 2005) and VMD (Humphrey *et al.*, 1996). GSH is a tripeptide with γ -carboxylate of glutamic acid attached to the amino-terminus of the dipeptide cysteine-glycine. The ZDOCK server does not accept GSH directly as a ligand, so the neutral form of the tripeptide Glu-Cys-Gly was used to approximate GSH for docking analysis. The coordinates of minimized model (ArsC-Grx2-GSH) is available in Supplementary section.

3.4 Results

3.41 Overall structure

As predicted, the crystal structure of Grx2-GSH complex shows an N-terminal domain, a C-terminal domain and one bound GSH molecule. The N-terminal domain (residues 1 to 72) adopts a topology similar to that of a thioredoxin fold, with four β sheets flanked by three α helices (Fig. 3.1). From the N-terminus the structure begins with a β strand (β 1) followed by an α helix (α 1) and a second β strand (β 2) that is parallel to β 1. Helix α 2 connects strands β 2 and β 3. Strand β 3 is antiparallel to strands β 1 and β 2. Strand β 4 is antiparallel to β 3, and its end leads to the third α helix. The four β sheets are almost in the same plane, with helices α 1 and α 3 below this plane, and helix α 2 above it and facing solvent. The C-terminal domain consists of eight α helices and two 3_{10} helices.

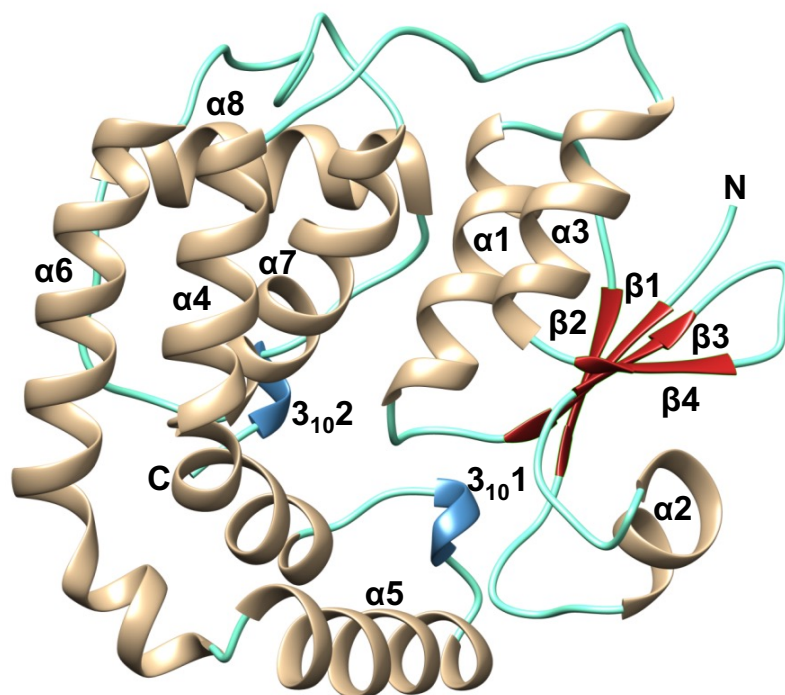


Figure 3.1. Cartoon diagram of Grx2-GSH. Ribbon diagram of the Grx2-GSH complex (PDB entry 4kx4). Secondary-structural elements: -helix, tan; -sheet, brown; coil, cyan; 3₁₀-helix, blue.

structure is similar to glutathione-bound Grx2 from *Salmonella typhimurium* (PDB ID. 3IR4). The root mean square deviation (RMSD) of the main chain atoms of both structure is 0.45 Å and sequence identity is 82.6 %.

3.42 Comparison of the Grx2-GSH and Grx2 C9S/C12S structures

The structure of the reduced form of Grx2 did not crystallize, but the Grx2 C9S/C12S mutant did, so that structure as used to approximate that of the reduced form. The RMSD between the main chain atoms of Grx2-GSH and that of the C9S/C12S structure is 1.13 Å. All of the residues of the mutant structure superimposed with the Grx2-GSH structure except Ala128, Ser129, Ala130 and Gly131. If these residues are excluded, the

One 3₁₀ helix is between helices α4 and α5. The other 3₁₀ helix is located just before the C terminus. Helices α4, α6 and α7 form a three-helix bundle, with helix α8 almost perpendicular to it. An 11-residue linker connects the N- and C-terminal domains. The Grx2-GSH

RMSD becomes 0.57 Å, indicating that this region contributes most of the differences between the two structures. $\alpha 5$ extends to Ser129 in Grx2-GSH, but in C9S/C12S structure Ala128 and Ser129 break the hydrogen bonds necessary for forming an α -helix (Fig. 3.2). The backbone of these residues in the Grx2-GSH structure superimposed with that of Grx2 from *S. typhimurium*, which also has GSH at its binding site. The *S. typhimurium* backbone atoms matched well with the *E. coli* Grx2-GSH structure but not with the C9S/C12S structure. When the twenty first conformer (best representative conformer) from the NMR

structure of reduced *E. coli* glutaredoxin (Xia *et al.*, 2001) was superimposed on our Grx2-GSH and C9S/C12S crystal structures, the trace of the backbone atoms nearly matched with the C9S/C12S structure (Fig. 3.2) and the linker connecting helices $\alpha 5$ and $\alpha 6$ is extended. This conformational change may explain how GSH interacts with Lys125 in

helix $\alpha 5$. In the Grx2-GSH structure, helix $\alpha 5$ is more extended and its compactness is

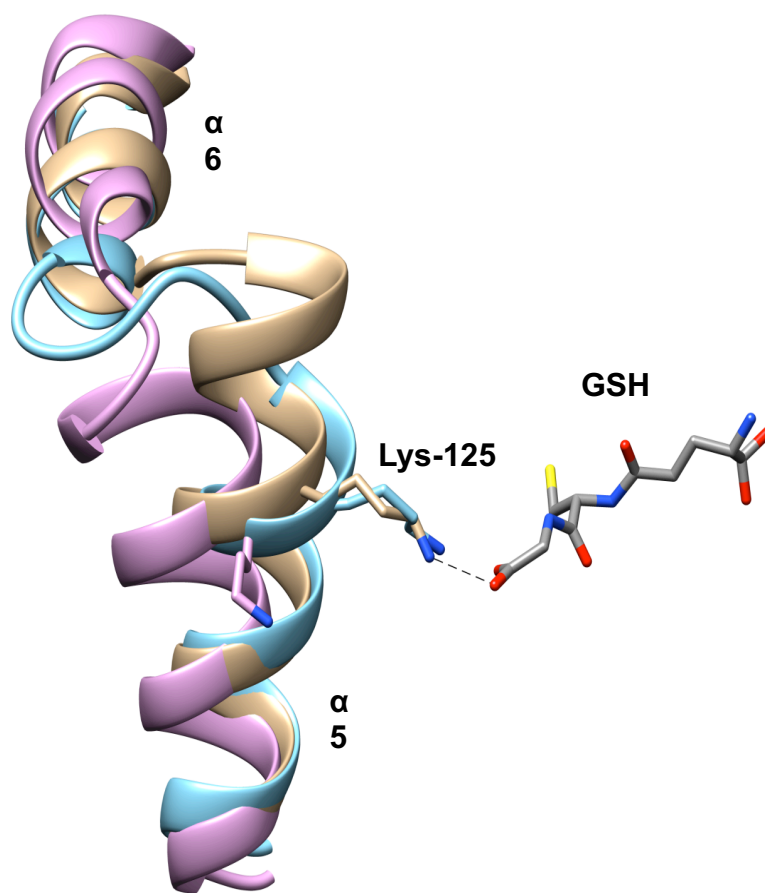


Figure 3.2 Superimposition of helices 5 and 6 of the Grx2-GSH complex (PDB entry 4kx4; tan) with the structure of the Grx2 C9S/C12S mutant (cyan; PDB entry 4ksm) and the NMR model (orchid; PDB entry 1g7o) of Grx2.

stabilized by a salt bridge formed between Lys125 and the carboxylate oxygen atom of the glyceryl residue of GSH. In contrast, in the C9S/C12S structure and NMR structure there is no GSH molecule, the helix $\alpha 5$ is more flexible by the extension of the linker.

3.43 Comparison with GST Structures

Structural homology searches using the DALI online server (Holm *et al.*, 2006) showed that Grx2 is similar to the structures of GSTs even though there is little sequence similarity. The RMSD between Grx2 and different types of GSTs are in the range of 3.0 to 4.0 Å (Fig. 3.3). The thioredoxin-like fold of N-terminal domain of Grx2 corresponded to that region of all types of GSTs. In C-terminal domain, the three helix bundle ($\alpha 4$, $\alpha 6$ and $\alpha 7$) and helix $\alpha 8$ are common for Grx2 and all type of GSTs. However these helices are arranged in different manner or topology in Grx2 and different GSTs. In the Grx2 structure, the C-terminus ends with a 3_{10} helix, but in GSTs there are one or more α helices at the C-terminus (Fig. 3.3). Helix $\alpha 5$ in Grx2 is distinguished from all type of GSTs except human theta-class GST (PDB ID. 1LJR) (Rossjohn, McKinsty *et al.*, 1998). The orientation of the C-terminal α helix of 1LJR coincides with helix $\alpha 5$ of Grx2, but the topology of these helices in both structures is different. In Grx2 the helix is placed between helices $\alpha 4$ and $\alpha 6$, while in 1LJR, it is placed at the end of the domain.

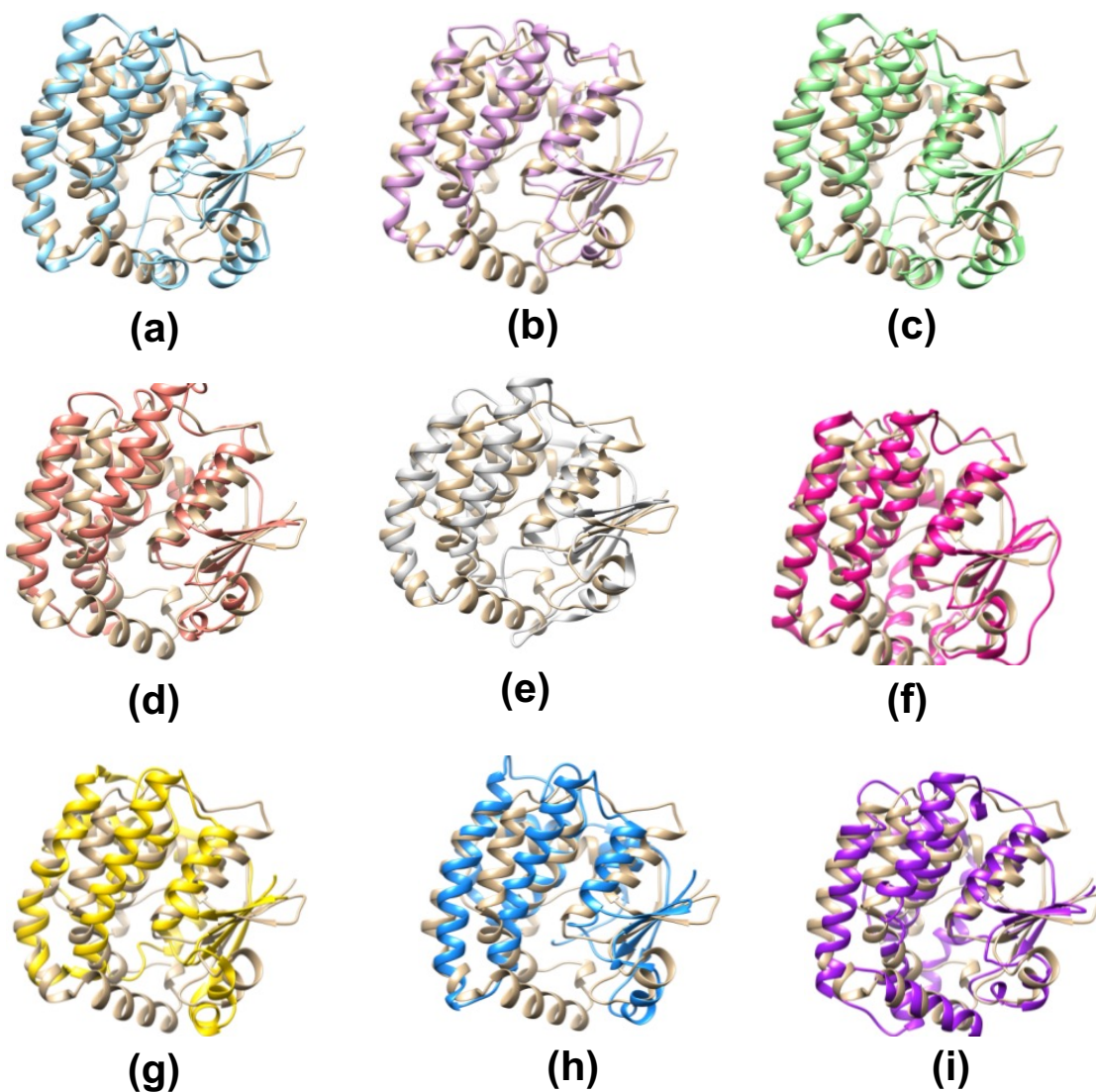


Figure 3.3 Comparison of Grx2 structure with nine other GSTs. Grx2 (4KX4) was compared with represented members of GST classes alpha (1GSE, 3.8 Å), beta (2PMT, 3.1 Å), delta (2WJU, 3.9 Å), phi (1GNW, 3.2 Å), mu (1HNA, 3.2 Å), omega (1EEM, 3.3 Å), pi (1GLP, 3.4 Å), sigma (1GSQ, 3.5 Å) and theta (1LJR, 3.4 Å). The PDB IDs and RMSD relative to Grx2 are given in parentheses.

3.44 GSH Binding Site (G-site)

GSH makes extensive interactions with residues from the N-terminal domain of Grx2 (Fig. 3.4). The oxygen atoms of the α -carboxylate group of the GSH glutamyl residue are

hydrogen bonded to the main chain amino group and the side chain hydroxyl group of Ser62. The N-terminal amino group of GSH makes a hydrogen bond to the side chain of Glu61. The main chain amino and carboxyl groups of the cysteine residue make hydrogen bonds with the backbone of Val48. The carboxylate atom of the GSH glycyl residue interacts with both Lys46 from C-terminal domain and Lys125 from the C-terminal domain. The distance between sulfur atoms of the GSH cysteinyl residue and Cys9 is 2.98 Å, which is too long for a typical disulfide bond (2.05 Å). Six water molecules also interact with different regions of the GSH molecule. The distance between the sulfur atom of Cys9 and that of Cys12 is 3.6 Å, which is also too long for a disulfide bond. However, the possibility that synchrotron radiation broke a disulfide bond between Cys12 and either Cys9 or GSH cannot be excluded. In above mentioned Grx2-GSH interactions, only two interactions with the equivalents of Grx2 residues Cys9-GSH and Val48-GSH are conserved in *E. coli* Grx1 (Xia *et al.*, 1992) and Grx3

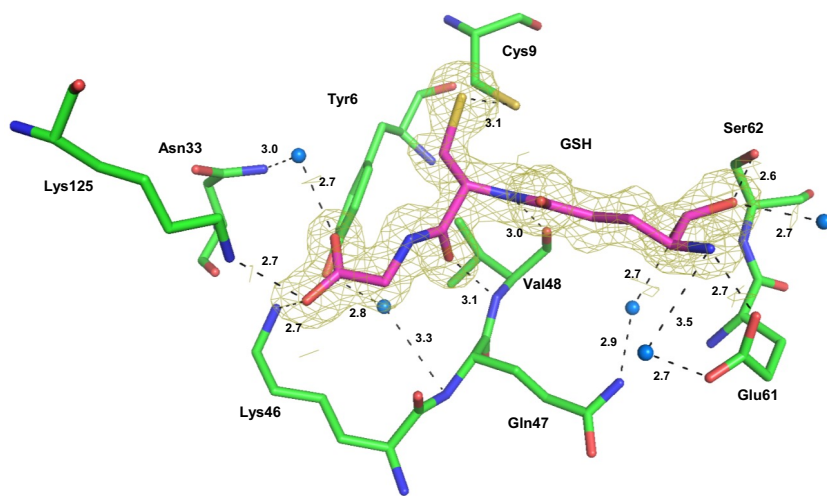


Figure. 3.4. Interaction of GSH with Grx2. The blue sphere represents water molecule. A potential hydrogen-bond network around the GSH molecule is indicated with dashed lines (the distances are in Å) and residues interacting with GSH are labeled.

(Nordstrand *et al.*, 1999). Structural alignment of Grx2 with members of different classes of GSTs using STRAP (Gille & Frommel, 2001) shows that the GSH binding residues are similar in each. In six Grx2-

GSH interactions, three interactions are structurally conserved in all GSTs. First,

interaction of carboxylate oxygen atoms of GSH glutamyl residue with either a serine or threonine residue (Grx2 residue Ser62) is conserved. Second, the N-terminal amino group of GSH interacts with a glutamine or glutamate residue in all GSTs. Third, the main chain atoms of cysteinyl residues interact with hydrophobic residues such as valine, leucine and methionine. The carboxylate oxygen atom of the glycyl residue of GSH interacts with a lysine in several GSTs (1EEM (Board *et al.*, 2000), 2PMT (Rosjohn, Polekhina *et al.*, 1998), 1GLP (Garcia-Saez *et al.*, 1994), 1GSQ (Ji *et al.*, 1995), 1HNA (Raghunathan *et al.*, 1994) and 2WJU (Tars *et al.* 2010) and an arginine residue in others (1LJR (Rosjohn, McKinstry *et al.*, 1998), 1GSE (Cameron *et al.*, 1995). However this interaction is not structurally conserved. Overall the residues in Grx2 that interact with GSH are conserved or partially conserved in GSTs. GSH-binding residues Cys9, Lys46, Val48, Glu61 and Ser62 in Grx2 are conserved in only human omega-class GST.

3.45 Hydrophobic Binding Site (H-site)

In general, GSTs have a binding site for hydrophobic substrates (H-site), and their topologies are accessible by wide range of substrates (Rosjohn, McKinstry *et al.*, 1998). For Grx2, the only substrate identified to date is the ArsC-S-As(V) complex. The Grx2-GSH and Grx2 C9S/C12S structures have an acetate anion and a 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris) cation, respectively, in their H-site. Residues from $\alpha 7$ (Gln174, Pro177, Leu178, Arg180, Leu182, Asn181 and Leu 184), $\alpha 4$ (Leu90, Asn94, Ala97, Asn98, Leu101, Leu102 and Phe105), $\alpha 5$ (Phe122 and Lys126) and $\alpha 1$ (Pro10, Tyr11, Cys12 and Lys14) define the H-site. These hydrophobic and charged residues form a narrow cleft. In the Grx2-GSH structure, the acetate ion was placed on the cleft,

interacting with residue Asn98 and making van der Waals interaction with glutathione. In the Grx2 C9S/C12S structure, the Tris molecule interacts with Lys126 and Ser12.

3.46 Conformational Changes upon GSH Binding

Comparison of the crystal structures of the Grx2-GSH complex and C9S/C12S indicates how binding of GSH triggers a conformational change from coil to α helix. In the absence of GSH, the loop that connects helices α 4 and α 5 is extended and provides flexibility for movement of helix α 5. In this conformation, the core of the structure is now easily accessible to GSH. In the GSH-bound conformation, there is a salt bridge between the carboxylate oxygen atom of the glycyl residue of GSH and Lys125 in helix α 5.

3.47 Grx2 has GST Activity

I examined the ability of Grx2 to function as a GST by conjugation of GSH with the model substrate CDNB (Table 3.1). Purified Grx2 exhibited a K_m value of 2 mM for CDNB and 0.31 for GSH. Grx2 was compared with an authentic GST, the *S. japonicum* mu-class sjGST. The two had nearly the same affinity for GSH. The Grx2 turnover number as a function of GSH concentration was approximately 20% of that of sjGST, although somewhat less for the artificial substrate CDNB. These results demonstrate that Grx2 has reasonable GST activity.

Table 3.1 GST activity of Grx2 and sjGST.

Protein	Substrate					
	GSH			CDNB		
	K_m mM	k_{cat} sec ⁻¹	k_{cat}/K_m sec ⁻¹ mM ⁻¹	K_m mM	k_{cat} sec ⁻¹	k_{cat}/K_m sec ⁻¹ mM ⁻¹
Grx2	0.31	0.58	1.9	2.0	0.41	0.21
sjGST	0.28	2.6	9.3	1.1	3.4	3.1

3.48 *In Silico* Docking Analysis of the ArsC-Grx2 Interaction

In *E. coli*, the ArsC arsenate reductase transforms arsenate to arsenite using GSH as the source of reducing potential (Shi *et al.*, 1999). In the catalytic cycle, glutaredoxins participate by deglutathionylation of the predicted ArsC Cys12-S-As(III)-SG complex (Martin *et al.*, 2001). Grx2 has considerably higher affinity for ArsC than Grx1 or Grx3 and is

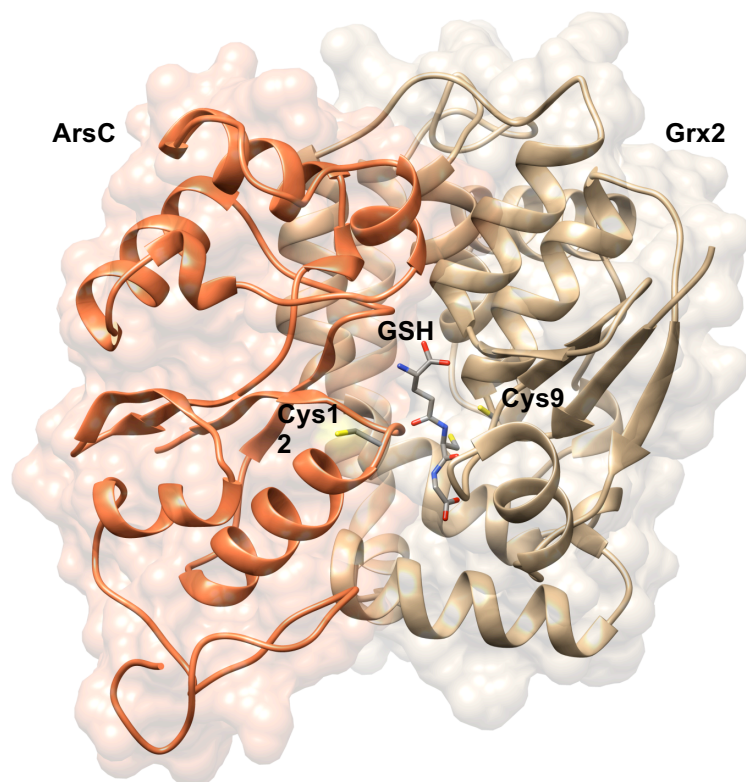


Figure 3.5. *In silico* analysis of the interaction of ArsC and Grx2 by molecular docking.

the likely *in vivo* source of electrons for ArsC-catalyzed arsenate reduction (Shi *et al.*, 1999). It should be emphasized that, to date, no other physiological function has been identified for Grx2. To predict how Grx2 might interact with ArsC, the two proteins were blindly docked (rigid body docking), with the highest ranking solution shown (Fig. 3.5). In

this solution the catalytic residues of both proteins were found to face each other. The distance between catalytic residues Cys12 in ArsC and Cys9 in Grx2 is 13.7 Å, and the distance between Cys12 in ArsC and Cys residue from the tripeptide Glu-Cys-Gly, which was used to represent GSH, is 12.4 Å. When the GSH analogue molecule is rotated (in COOT) with respect to its long axis, its thiolate is directed towards the catalytic residue Cys12 of ArsC, and the distance between them is reduced to 6.6 Å. The GSH binding residues Glu61, Ser62 and the active site residue Tyr11 of Grx2 are present at the interface, and these residues may assist in the predicted exchange of GSH from ArsC to Grx2. Other Grx2 residues in the C-terminal domain, including Glu170, Lys99, Asn98, Asn94 and Arg91, are also predicted to interact with ArsC. Docking of ArsC with only the N-terminal thioredoxin-like domain of Grx2 did not yield reasonable models. This may explain why Grx2 is the preferred glutaredoxin for ArsC rather than the smaller Grx1 or Grx3 proteins, which lack a GST-like C-terminal domain.

3.5 Discussion

In *E. coli* ArsC reduces pentavalent arsenate to trivalent arsenite with GSH as the source of reducing potential (Oden *et al.*, 1994). The reaction requires a glutaredoxin as an electron donor to regenerate active enzyme, and *E. coli* Grx2 is highly preferred over Grx1 or Grx3. The proposed ArsC reaction scheme includes four steps and three intermediates (Martin *et al.*, 2001). In the first step, arsenate directly attacks the thiolate of active site residue Cys12. The second step involves glutathionylation to form the active site intermediate ArsC-Cys12-S-As(V)-SG. The third step is deglutathionylation of the complex by Grx2 and formation of the ArsC-Cys12-S-As(III) intermediate, with concomitant formation of the mixed Grx2-S-SG disulfide. Interestingly, the Grx2-GSH complex crystallized only in the presence of sodium arsenate, although arsenate was not

observed in the final structure. This may reflect a specific interaction with As(V), but another possibility is that arsenate serves as an electron acceptor during disulfide bond formation. Is there direct interaction of ArsC and Grx2 during the catalytic cycle? *In silico* docking analysis of ArsC with Grx2 suggests that that two form a tight complex in which the ArsC active site residue Cys12 and the glutathione binding residues of Grx2. A Glu-Cys-Gly tripeptide representing the glutathione moiety is located between the two proteins, and an arsenic atom would be expected to bridge the gap.

Structurally Grx2 is related to both glutaredoxins and GSTs. Overall, Grx2 is structurally similar to GSTs in spite of a lack of sequence similarity, and its activity with the model substrate CDNB demonstrates that Grx2 is a GST. The GSH binding residues of Grx2 are best conserved with the structure of human omega-type GST, and the orientation of the C-terminal α helices are best matched with the structure of human theta type GST. Thus Grx2 does not appear to belong to any known type of GST, suggesting that it may form a new class of GST with a unique function as both glutaredoxin and glutathione S-transferase. A novel feature of Grx2 is that it is a hybrid of the small N-terminal glutaredoxin domain and the C-terminal GST domain. In *E. coli* Grx2 is the major glutaredoxin, both in terms of intracellular concentration, where it comprises 2/3 of the total cytosolic glutaredoxin and in terms of activity, where it provides 81% of total glutaredoxin activity (Åslund *et al.*, 1994). Grx2 exhibits GST activity with about 20% of the activity of purified *S. japonicum* sjGST (Table 3.2). It has relatively low catalytic efficiency (about 10% of sjGST), but its relatively high intracellular concentration compensates. Does this hybrid Grx/GST activity have physiological relevance? In the proposed reaction scheme for ArsC arsenate reductase activity, the second step is glutathionylation of the Cys12-As(V) intermediate, and step 3 is deglutathionylation of

the Cys12-As(III)-SG intermediate. We propose that Grx2 serves as a glutathione S-transferase in the glutathionylation reaction of step 2 and as a glutaredoxin in the deglutathionylation reaction of step3, a novel mechanism for participation of the hybrid Grx/GST in arsenic detoxification.

CHAPTER 4. Expression of Arsenic Resistance Genes in the Obligate Anaerobe *Bacteroides vulgatus* ATCC 8482, a Gut Microbiome Bacterium

4.1 Abstract

The response of the obligate anaerobe *Bacteroides vulgatus* ATCC 8482, a common human gut microflora, to arsenic was determined. This microbe has eight continuous genes that we demonstrate form a transcriptional unit inducible by the trivalent organoarsenical methylarsenite (MAs(III)). The first gene of this *ars* operon, *arsR*, encodes a putative ArsR As(III)-responsive transcriptional repressor. The next three genes encode proteins of unknown function. The remaining genes, *arsDABC*, have well-characterized roles in arsenic detoxification. The expression of each gene to exposure to trivalent and pentavalent inorganic and methylarsenicals was analyzed. The *arsD* gene was the most highly expressed of the *ars* operon genes and has the potential to be a novel biomarker of human oral exposure to arsenic. These results demonstrate that this anaerobic microbiome bacterium has arsenic-responsive genes that confers resistance to inorganic and organic arsenicals and may be responsible for its ability to maintain its prevalence in the gut following dietary arsenic exposure.

4. 2 Introduction

The objective of this study was to examine how the microbiome responds to arsenic exposure. Before arsenic is absorbed in the intestinal tract and methylated in liver, the

microbiome is exposed to dietary arsenic through drinking water and food. Microbiome bacterial respond to arsenic exposure by modifying their interactions with the gut (Lu et al., 2013). The human gut has at least 10^{13} microorganisms, containing 100-fold more genes than our own genome, and more than 90% are anaerobes. The majority of human gut microbiome species fall within four phyla, *Actinobacteria*, *Firmicutes*, *Proteobacteria*, and *Bacteroidetes* (Pflughoeft & Versalovic, 2011). In the rat gut microbiome arsenic affects the diversity of microbial populations (Lu et al., 2013) and their metabolism (Lu et al., 2014). Exposure to arsenite disturbs the mouse gut microbiome and substantially alters its metabolomics profile (Lu et al., 2014). The prevalence of four families of *Firmicutes* bacteria significantly decreased, whereas *Bacteroidetes* families remained relatively unchanged following arsenic exposure, suggesting that *Bacteroides* may be intrinsically resistance to arsenic. In spite of the impact of arsenic on the microbiome, nothing is known about the effect of arsenic on microbiome gene expression. Our gut microbiome is exposed to dietary arsenic before it enters our blood stream; in effect, it is potentially our first-line for sensing environmental arsenic. There are only a few studies of anaerobic arsenic-inducible gene expression in environmental microbes such as the facultative anaerobe *Shewanella putrefaciens* (Murphy & Saltikov, 2009; Stolz et al., 2006) and, to our knowledge, none in the obligate anaerobes of the gut microbiome. This lack of information limits our understanding of how the microbiome responds to the dietary arsenic.

Members of the phylum *Bacteroidetes* form the major bacterial species in the human gut microbiome (Eckburg et al., 2005). One highly abundant microbiome species is the obligate anaerobe *Bacteroides vulgatus* (Turnbaugh et al., 2009), Most if not all bacteria

and archaea have arsenic resistance (*ars*) operons responsible for a wide variety of arsenic biotransformations (Zhu *et al.*, 2014). *Bacteroides vulgatus* ATCC 8482 is a Gram-negative obligate anaerobe originally isolated in 1933 from human feces (Eggerth & Gagnon, 1933). From its genome sequence, it has a putative *ars* operon with eight genes. The first gene encodes a putative ArsR As(III)-responsive transcription repressor that might control expression of the operon. The physiological functions of the next three putative gene products are unknown. The last four correspond to the well-characterized *arsDABC* genes (Liu *et al.*, 2013). ArsD is an arsenic metallochaperone that donates As(III) to the ArsAB ATPase, an As(III) efflux pump. ArsC is an arsenate reductase that transforms As(V) into the As(III), the substrate of the efflux pump. It is interesting that an obligate anaerobe would have the gene for an enzyme that reduces pentavalent inorganic arsenic, implying that arsenate might be able to reach the lower intestine in the oxidized form.

In this study expression of each of the *ars* genes was determined by quantitative real time polymerase chain reaction (Q-RT-PCR) in response to exposure to trivalent arsenite, pentavalent arsenate and to the methylated arsenicals MAs(V) and MAs(III). The *ars* genes were induced most effectively by trivalent arsenicals. The *arsD* gene was the most highly expressed of the *ars* operon genes. Since these genes might be expressed even before arsenic enters our blood stream, we propose that expression of the *ars* genes in a gut microbe has the potential to be adapted for use as a biomarker of dietary exposure to arsenic. Knowledge of the response of a gut microbe to arsenic exposure increases understanding of how microbiome diversity is maintained.

4.3 Material and Methods.

4.31 Anaerobic gGrowth of *Bacteroides vulgatus* ATCC 8482

Bacteroides vulgatus Eggerth and Gagnon, ATCC 8482 (NCTC 11154) was purchased from the American Type Culture Collection (Manassas, VA). To revive the strains, cells were initially grown on 1.5% agar plates consisting of tryptic soy with 5% sheep blood (Fisher Scientific, Pittsburgh, PA) in an anaerobic jar with GasPak™ for 36 hr at 37 °C. Cultures were then grown in the anaerobic jar in degassed Schneider's insect medium containing 20% heat-inactivated fetal bovine serum (Sigma-Aldrich, St. Louis, MO) for 36 hr at 37 °C with shaking.

4.32 Arsenic Resistance Assays

Cultures of *B. vulgatus* were grown at 37 °C with shaking in degassed Schneider's insect medium containing 20% heat-inactivated fetal bovine serum containing the indicated concentrations of As(V), As(III), MAs(V) or MAs(III) in an anaerobic jar with GasPak™. After approximately 24 hr, when the cultures had attained an $A_{600nm} = 1.0$, the cells were collected at 4 °C by centrifugation in tightly capped tubes filled with nitrogen gas.

4.33 Isolation of Nucleic Acids.

Genomic DNA was isolated as described by Liao et al. (2011) using a QIAamp DNA Mini Kit (QIAGEN, Valencia, CA). Total RNA from *B. vulgatus* was isolated using TRIZOL reagent (Life Sciences, ThermoFisher Scientific, Grand Island, NY) according to the manufacturer's protocol. DNA was removed from total RNA preparation using TURBO DNFree Kit (Ambion, ThermoFisher Scientific, Grand Island, NY) according to the manufacturer's instructions. RNA integrity was confirmed by denaturing agarose gel electrophoresis.

4.34 Reverse Transcription (RT) and Quantitative Real-time (Q-RT) Polymerase Chain Reaction (PCR)

cDNA synthesis was carried out with 0.5 µg of total RNA using an AccuScript High Fidelity 1st strand cDNA synthesis kit (Agilent, Santa Clara, CA) according to the manufacturer's instructions. The reactions were incubated at 25 °C for 10 min and 42 °C for 30 min, followed by heat inactivation at 95 °C for 5 min. The reaction mix was then neutralized with sodium acetate and purified using a Qiagen PCR purification kit according to the manufacturer's instructions.

For RT-PCR total RNA isolated from exponentially growing *B. vulgatus* in the absence or presence of 13 µM MAs(III) was used as template in a reverse transcriptase reaction to generate cDNAs that were then amplified with the indicated primer sets (Table 4.1). The primers covering all eight genes with overlap between them were purchased from Integrated DNA technologies (IDT), Coralville, IA. cDNA of 16S rRNA was used as control to normalize gene expression. Genomic DNA of *B. vulgatus* used as positive control for all primers amplicons.

For Q-RT-PCR, 2 µl of diluted purified cDNA reaction corresponding to 6 ng of template RNA, were used in a 10 µl reaction containing forward and reverse primers for the target genes (Table 4.2) and 1X iQSYBR Green supermix (Bio-Rad Laboratories, Inc., Hercules, CA). Primer sets were designed using IDT's PrimerQuest software and only efficiency between 98% to 102% primers were used in the current study. qPCR reactions were analyzed using an Eppendorf Realplex2 PCR instrument with the following thermal cycling conditions: initial denaturation at 95 °C for 3 min followed by 40 cycles of 95 °C for 15 sec and 65 °C for 20 sec. A final melting curve analysis was performed for each reaction to confirm that the PCR generated a single amplification

product. The relative abundance of target amplicons between samples was estimated using 16S rRNA as control by the $2^{-\Delta\Delta CT}$ method. Error bars were calculated from the mean \pm SD of three independent experiments in triplicate.

Table 4.1 Oligonucleotide primers for RT-PCR. For determination of operon structure the indicated pairs of forward and reverse RT-PCR primers were used for amplification of overlapping regions of the operon. The dashes before or after the target gene indicate that the primers began or ended outside of the gene.

RT-PCR primer	Target	Primer sequences	Amplicon length (bp)
1	- <i>arsR</i>	Sense: 5'- TACTACACTGGCAGAAGGATTG -3' Antisense: 5'- AATCTGTTCCCTGCTCTACTGTG -3'	617
2	<i>arsR-rf1</i>	Sense: 5'- CACAGTAGAGCAGGAACAGATTGC -3' Antisense: 5'- CTTGAACACCCGGTTCCTAATAC -3'	422
3	<i>orf1-orf3</i>	Sense: 5'- TATTAGGAACCGGGTGTTC AAG -3' Antisense: 5'- TAATAGCGGGAACAGAACTGC -3'	723
4	<i>orf3-arsD</i>	Sense: 5'- CAGTTCTGTTCCCGCTATTAC -3' Antisense: 5'- TGTTCACAAAGACCCGTAG -3'	774
5	<i>arsD-arsA</i>	Sense: 5'- CTACGGGTCTTTGTGGAACA -3' Antisense: 5'- GGGACGACTTACCAACTAAG -3'	948
6	- <i>arsA</i> -	Sense: 5'- TTGGTAAGTCGTCCCGAAATC -3' Antisense: 5'- CTTACCGAAGCAGGAGAATCG -3'	830
7	<i>arsA-arsC</i>	Sense: 5'- CGATTCTCCTGCTTCGGTAAG -3' Antisense: 5'- ACTTCACCAGTAAACATAGGACAG -	518

		3'	
8	<i>arsC</i> - <i>acr3</i>	Sense: 5'- CTGTCCTATGTTTACTGGTGAAGT -3' Antisense: 5'- CTATCCCGAACATGGTAAACGG -3'	484
9	- <i>acr3</i> -	Sense: 5'- CCGTTTACCATGTTCGGGATAG -3' Antisense: 5'- ACTTCAACCAGCACACCTAC -3'	703
10	-16s rRNA-	Sense: 5'- GGGATGCGTTCCATTAGATAGT -3' Antisense: 5'- CGGAGTTAGCCGATCCTTATTC -3'	309

Table 4.2: Oligonucleotide primers for Q-RT-PCR. Expression of each gene was determined using the indicated Q-RT-PCR primers.

Q-RT-PCR primer	Target	Primer sequences	Amplicon length (bp)
11	<i>orf1</i>	Sense: 5'- GCTGCGATGCAACTAATC -3' Antisense: 5'- CCAAAGCCGGAAGACCTAAA -3'	78
12	<i>orf2</i>	Sense: 5'- GATTGCGATGCTGGTCTTAATG -3' Antisense: 5'- CAATGCGGTTAGGCTGTATTTTC -3'	92
13	<i>orf3</i>	Sense: 5'- AAACAGGCGGCTATCTCTTAC -3' Antisense: 5'- CCAGTCCGGCAACTATAA -3'	99
14	16s rRNA	Sense: 5'- GACTGCCCTATGGGTTGTAAA -3' Antisense: 5'- CGGAGTTAGCCGATCCTTATTC -3'	98
15	<i>arsA</i>	Sense: 5'- CTTGCAGTTCAGGCAGGTATTA -3' Antisense: 5'- GTCATTGTTTCCGCAGCTTTAC -3'	101

16	<i>acr3</i>	Sense: 5'- GCAAACGTCCGAGAGGAATAA -3' Antisense: 5'- AGCTATCCCGAACATGGTAAAC -3'	77
17	<i>arsC</i>	Sense: 5'- GGGACTATGTTATTACGGTCTGC -3' Antisense: 5'- TGGGTCGTCAAACCCAATATG -3'	98
18	<i>arsD</i>	Sense: 5'- ACGTAAGCAATAAGACGGTGAA -3' Antisense: 5'- TTGGTGGTAGGATAGGTTTGTG -3'	112

4.4 Results

4.4.1 Anaerobic Resistance to Inorganic and Organic Arsenicals in *B. vulgatus*

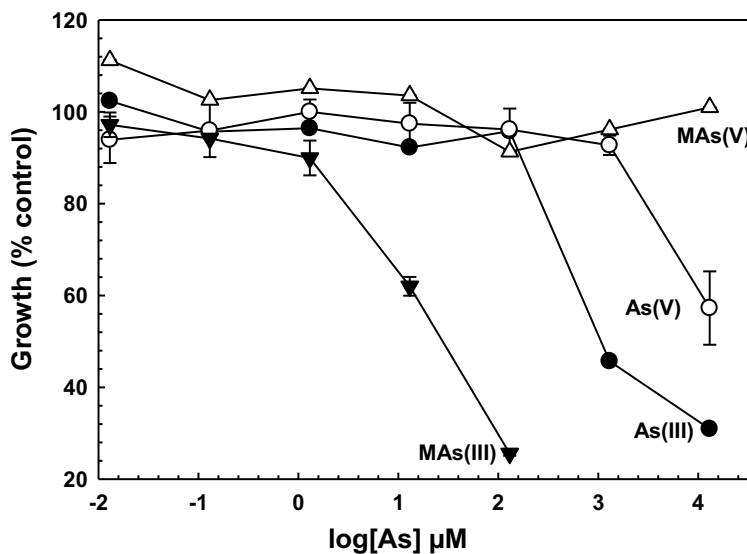


Figure. 4.1 Resistance to inorganic and organic arsenicals in anaerobic growth *B. vulgatus* strain ATCC 8482. Cultures were grown anaerobically in Schneider's insect medium containing 20% heat-inactivated fetal bovine serum containing the indicated concentrations. After 24 h at 37 C with shaking, the A600nm was measured. Data are the mean \pm SE (n=3).

was examined (Fig. 4.1). *B. vulgatus* is quite resistant to the pentavalent species arsenate and methylarsenate, with growth in medium with 10-100 mM. The cells are somewhat more sensitive to trivalent arsenite but still grow at millimolar concentrations.

ATCC8482

There have been few studies of arsenic resistance in anaerobes, and, to our knowledge, none in inhabitants of the gut microbiome. The effect of increasing

concentrations of As(III), As(V), MAs(III) and MAs(V) on anaerobic growth *B. vulgatus* in pure culture

Trivalent methylarsenite is the most toxic, with complete growth inhibition at 100 μ M MAs(III). It should be pointed out that the free concentration of arsenicals was probably less than the added concentration since the cells were grown in very rich Schneider's insect medium with 20% fetal bovine serum, which has many molecules that can bind arsenicals. The relative tolerance to arsenate and arsenite implies the presence of arsenic resistance genes in *B. vulgatus*.

4.42 The *ars* Operon of *B. vulgatus* ATCC8482

The complete sequence of the *B. vulgatus* ATCC8482 includes eight contiguous genes that might compose an *ars* operon (Fig. 4.2A). The first product encodes a putative 108-residue ArsR As(III)-responsive transcriptional repressor (accession number ABR41782.1). Three different As(III) binding sites, each of which is composed of three cysteine residues located in different regions of the protein, have been identified in ArsRs (Ordóñez *et al.*, 2008; Qin *et al.*, 2007; Shi *et al.*, 1994). BvArsR does not have any of the three known As(III)-binding motifs, so it cannot be assumed that it is in fact an As(III)-responsive repressor. It has five cysteine residues, three of which (Cys99, Cys106 and C107) near the C-terminus might form a new ArsR As(III) binding site. The next gene, *orf1*, encodes a typical 77-residue thioredoxin (accession number ABR41783.1). While its role in arsenic resistance is not known, it could serve as the source of reducing potential for both the *orf3* and *arsC* gene products, as described

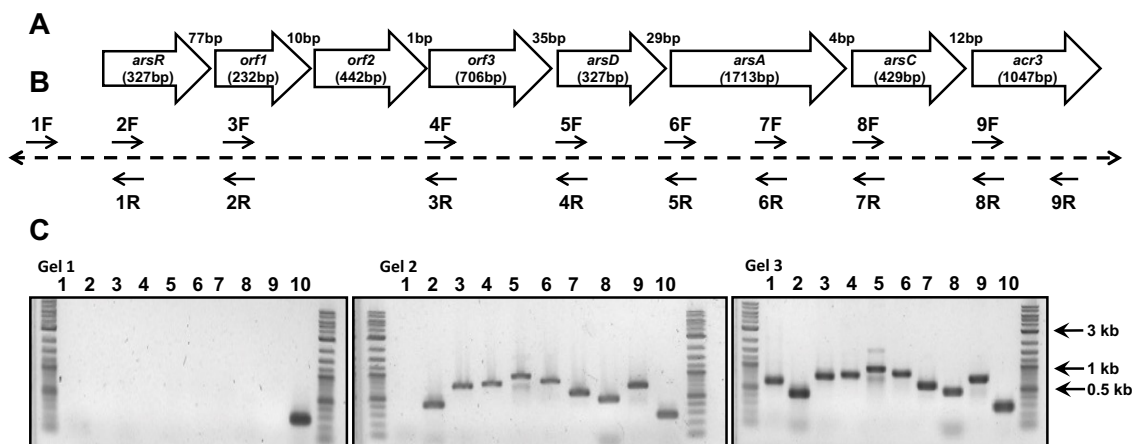


Figure. 4.2 The *ars* operon of *B. vulgatus* strain ATCC 8482. (A) The putative *ars* operon. From the complete genome *Bacteroides vulgatus* strain ATCC 8482 (Accession number CP000139) eight continuous genes encoding a putative *ars* operon were identified. (B) The *ars* transcriptional unit. Shown are the location of the primers for RT-PCR analysis of the *ars* transcriptional unit relative to the eight putative *ars* genes. (C) RT-PCR analysis. Total RNA isolated from *B. vulgatus* strain ATCC 8482 in the exponential phase ($A_{600nm}=1.2$) in the absence (gel 1) or presence (gel 2) of 1 ppm (13 mM) MAs(III) was used as the template in a reverse transcriptase reaction to generate cDNA and then amplified with the indicated primers. PCR products were resolved on a 1% agarose gel stained with ethidium bromide. Genomic DNA from *B. vulgatus* was used as a positive control for each primer set (gel 3). In each gel lane 10 is the 16S rRNA (Accession number NC_009614.1) PCR product.

below. The next gene, *orf2*, potentially encodes a 147-residue protein (accession number ABR41784.1) annotated as a thioredoxin-like protein. The *orf3* gene product (accession number ABR41785.1) is annotated as a putative cytochrome c biogenesis protein and has similarities to the disulfide bond isomerase DsbD, which uses thioredoxin to reduce disulfide bonds in periplasmic proteins (Mavridou *et al.*, 2009). The role of the Orf2 and Orf3 proteins in arsenic resistance is unknown, but it is reasonable to assume that they are in an *ars* operon because they have an arsenic-related function. The product of the next gene is a 108-residue ArsD metallochaperone (accession number ABR41786.1) (Lin *et al.*, 2006) that ferries intracellular As(III) to the next gene product, the 570-residues ArsA ATPase (accession number ABR41787.1) (Hsu & Rosen, 1989; Yang *et al.*, 2010). The next gene encodes the 142-residue ArsC arsenate reductase (accession number ABR41788.1) that couples the reducing potential of

thioredoxin to conversion of As(V) to As(III), the substrate of the product of the eighth gene, the 348-residue Acr3 efflux permease (accession number ABR41789.1) (Ghosh *et al.*, 1999; Wysocki *et al.*, 1997). Acr3 can also form a complex with ArsA to form an As(III)-translocating efflux pump (Fu *et al.*, 2010). Since the putative *arsR* gene does not have a known As(III) binding motif, and the three genes following *arsR* cannot be assigned a role in arsenic detoxification, it is not clear if this cluster of eight genes comprise an *ars* operon. To examine this question, co-transcription of the genes after induction by MAs(III) was examined by reverse transcription polymerase chain reaction (RT-PCR). A series of oligonucleotide primers covering all eight genes with overlap between them was designed for reverse transcription polymerase chain reaction (RT-PCR) (Figure 4.2B). Cells of *B.vulgatus* were grown anaerobically in the presence or absence of MAs(III) as inducer. Total RNA was isolated and used as template for reverse transcription of cDNAs in the absence (Fig. 4.2, gel 1) or presence (gel 2) of 13 μ M (1 ppm) MAs(III). In the absence of inducer no *ars* genes were expressed (gel 1). In the presence of inducer, each gene was expressed using primers that originated from the adjacent gene (gel 2), confirming co-transcription of all eight genes. The level of transcription of the housekeeping gene for 16S rRNA with or without inducer was used to normalize expression of the *ars* genes (lane 10 in each gel). To demonstrate the specificity of the primers, genomic DNA was used in place of cDNA (gel 3). These result demonstrate that the eight genes comprise a single *ars* operon.

4.43 Induction of the *B. vulgatus* ars Operon

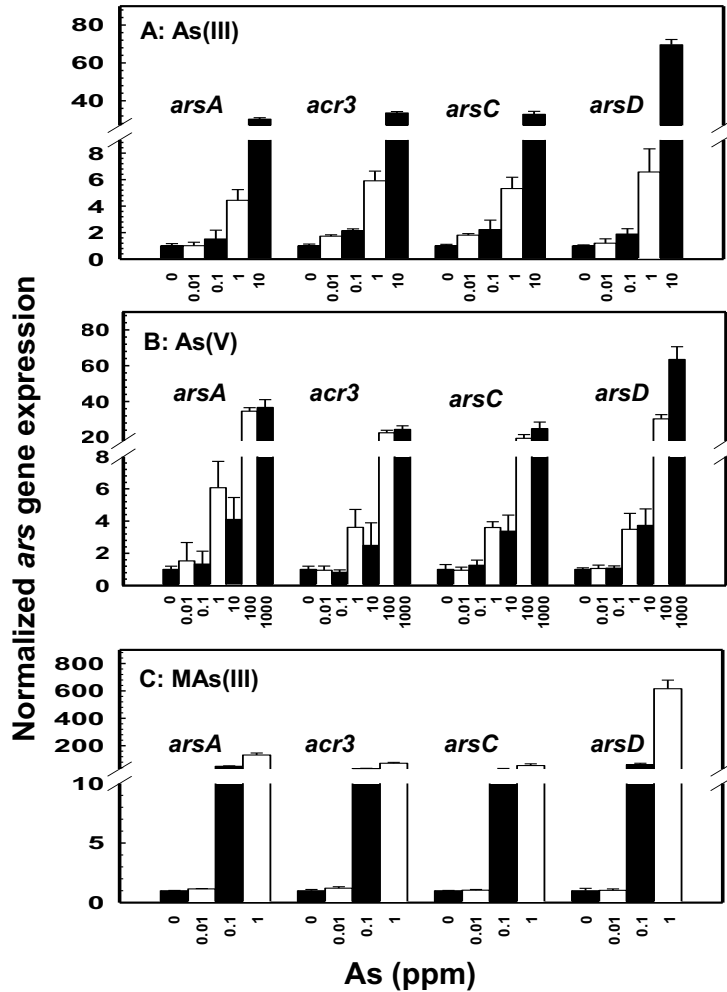


Figure 4.3. Transcriptional analysis of the *arsA*, *acr3*, *arsC* and *arsD* genes: Anaerobic cultures of *B. vulgatus* strain ATCC 8482 were induced with the indicated concentrations in ppm of As(III) (A), As(V) (B) or MAs(III) (C). Data are the mean \pm SE (n= 3).

anaerobe.

For that reason the transcription of the four *ars* genes of known function in response to addition of As(III) (Fig. 4.3A), As(V) (Fig. 4.3B) and MAs(III) (Fig. 4.3C) to pure cultures of *B. vulgatus*. Induction with 10 ppm (0.13 mM) As(III) was about 4-8 fold relative to 16S rRNA expression (to correspond with the EPA's usage, the arsenic concentrations in

In aerobic microorganisms, *ars* operons sense trivalent inorganic and organic arsenicals (Chen & Rosen, 2014). Pentavalent arsenicals do not directly induce but must first be reduced. However, to our knowledge, there have been no studies of the transcription of *ars* genes anaerobic microbiome bacteria in response to exposure to inorganic arsenic or organic arsenicals, nor has reduction of MAs(III) been demonstrated in an

Figure 4.3 are expressed as ppm). The same level of induction required 1000-fold more As(V). Differences in the rates of uptake of the arsenicals could contribute to the observed differential expression, but a more likely explanation is that the true inducers are trivalent, not pentavalent, arsenicals, and As(V) must be reduced to As(III) in order to induce. MAs(III) was by far the best inducer. It induced the *ars* genes 200-600 fold relative to 16S rRNA expression at 100-fold lower concentration than either inorganic arsenate or arsenite. The *arsA*, *acr3* and *arsC* genes were each expressed at nearly the same levels as each other, no matter which arsenical was used as inducer, but the *arsD* gene was consistently expressed at 2-4 fold higher levels than the others. The three *orfs* of unknown function are also induced by MAs(III) (Figure 4.4). The *orf1* gene, which encodes a putative thioredoxin, was transcribed nearly as well as *arsD*. The *orf3* gene was transcribed at about half the level, and *orf2* was transcribed at about 10-15 % of *arsD*. Segmental differences in mRNA stability within the polycistronic transcript of another *ars* operon was proposed to produce differential gene expression (Owolabi & Rosen, 1990) and may explain these results.

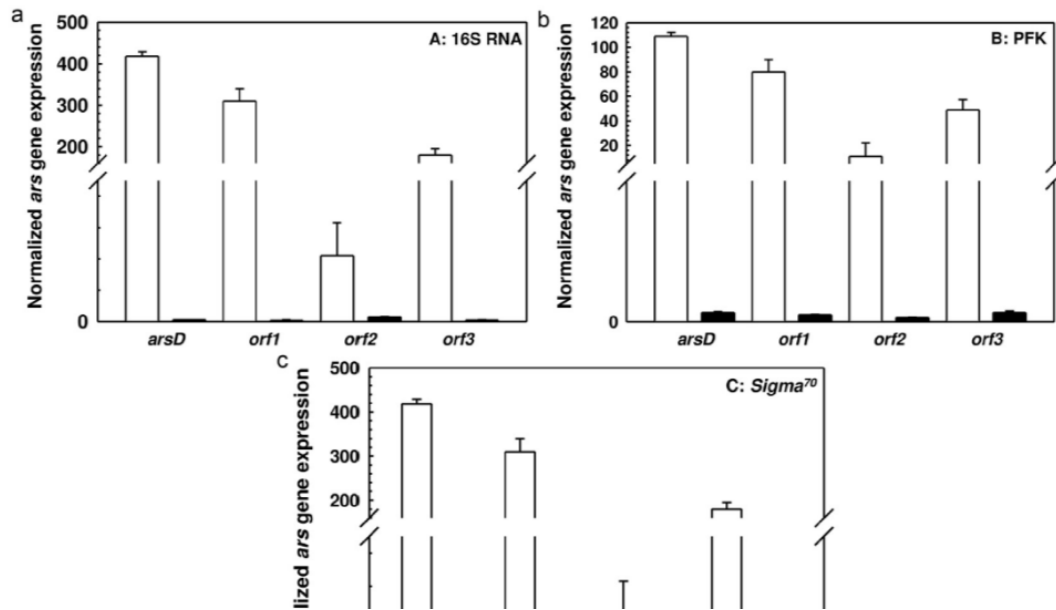


Figure 4.4. Transcriptional analysis of *orf1*, *orf2* and *orf3*: Total RNAs were isolated from cultures of *B. vulgatus* strain ATCC 8482 induced with (white bars) or without (black bars) 1 ppm (13 mM) MAs(III) and analyzed as described in the legend to Fig. 3. The *E. coli* housekeeping genes used to normalize expression were A) 16S rRNA; B) 6-phosphofructokinase (PFK) (Accession number NC_009614.1); C) σ^{70} (Accession number NC_009614.1). Data are the mean \pm SE (n = 3).

4.5 Discussion

Species within the *Bacteroides* genus, including *B. vulgatus*, colonize the intestinal tract of humans commensally. They increase nutrient utilization by their hosts by biodegradation of complex polysaccharides into more easily digested simpler sugars (Wexler, 2007). A decrease in the proportion of *Bacteroides* species in the microbiome relative to Firmicutes is associated with obesity, and weight loss is associated with an increase in the ratio of *Bacteroides* to Firmicutes (Barlow *et al.*, Ley *et al.*, 2006). In mice exposure to inorganic arsenic in drinking water produces dysbiosis, with an increase in

the relative proportion of Bacteroides relative to Firmicutes, leading to a perturbation in metabolite profiles (Lu *et al.*, 2014). The change in the gut microbiome composition is due a decrease in Firmicutes species, while Bacteroides species remain relatively constant during arsenic exposure. Here we demonstrate that *B. vulgatus* is highly resistant to inorganic arsenic, most likely the result of expression of the genes of chromosomal *ars* operon. In contrast, neither *Clostridium difficile* (Accession number NC_009089) nor *Lactobacillus gasseri* (Accession number NC_008530.1), both members of the Firmicutes phylum present in the human gut microbiome, contain an *ars* operon in its genome. We propose that the presence of *ars* operons in Bacteroides and their absence in Firmicutes leads to increased fitness of Bacteroides species relative to Firmicutes species.

A distinctive feature of *B. vulgatus ars* gene expression that is MAs(III) is by far the best inducer. Compared with As(III), MAs(III) induced the *arsD* gene to about 50-fold higher levels at three orders of magnitude lower concentration. Strikingly, *B. vulgatus* is also nearly 100-fold more sensitive to MAs(III) than to As(III) (Figure 1). This difference may be due to some degree to the higher toxicity of MAs(III) compared to As(III). However, none of the genes in the *ars* operon correspond to any of the known MAs(III) resistance genes for MAs(III): the Arsl C-As lyase (Yoshinaga & Rosen, 2014), the ArsH MAs(III) oxidase (Chen *et al.*, 2015a) or the ArsP MAs(III) efflux permease (Chen *et al.*, 2015b), so extreme MAs(III) sensitivity is not unexpected. Arsl and ArsH require oxygen, and their genes do not appear to be present in the anaerobic bacteria of the gut microbiome. Yet, humans are exposed to MAs(III) in run-offs from pesticide-treated farms, golf courses and other soils. Rice accumulates methylated arsenicals and can also be a source of dietary exposure. The

major source of MAs(III), however, may be microbiome organisms that methylate arsenic. For example, there are human gut microbiome bacteria that have an *arsM* gene encoding an As(III) S-adenosylmethionine methyltransferase, including *B. fragilis* 3_1_12 (Accession number EFR52264) and *Parabacteroides johnsonii* (Accession number WP_009860065). Thus gut microbes could methylate dietary As(III) to MAs(III), which would then induce Bacteroides to higher levels of expression, resulting in high-level resistance of the Bacteroides species in the gut microbiome. This would, in effect, be an autoamplification pathway that could augment the constancy of Bacteroides following inorganic arsenic exposure.

While these experiments utilize a single anaerobic Bacteroides species grown in pure culture, the data establish a proof of concept that the *ars* genes of this gut microbe are induced by both inorganic As(III) and organic MAs(III). Future experiments will examine expression of *ars* genes in the microbiome.

CHAPTER 5. A novel As(III) S-adenosylmethionine methyltransferase from *Aspergillus fumigatus*

5.1 Abstract

Methylation of inorganic arsenic is a central process in the organoarsenical biogeochemical cycle. Members of every kingdom have ArsM As(III) S-adenosylmethionine (SAM) methyltransferases that methylates inorganic As(III) into mono- (MAs(III)), di- (DMAs(III)) and tri- (TMAs(III)) methylarsenicals. Every characterized ArsM to date has four conserved cysteine residues. All four cysteines are required for methylation of As(III) to MAs(III), but methylation of MAs(III) to DMAs(III)

requires only the two cysteines closest to the C-terminus. Fungi produce volatile and toxic arsines, but the physiological roles of arsenic methylation and the biochemical basis is unknown. Here we demonstrate that most fungal species have ArsM orthologs with only three conserved cysteine residues. The genome of *Aspergillus fumigatus* has four *arsM* genes encoding ArsMs with only the second, third and fourth conserved cysteine residues. AfArsM1 methylates MAs(III) but not As(III). Heterologous expression of *AfarsM1* in an *Escherichia coli* conferred resistance to MAs(III) but not As(III). The existence of ArsMs with only three conserved cysteine residues suggest that the ability to methylate MAs(III) may be an evolutionary step toward enzymes capable of methylating As(III), the result of a loss of function mutation in organisms with infrequent exposure to inorganic As(III) or as a resistance mechanism for MAs(III).

5.2 Introduction

The objective of this project was to characterize a novel class of ArsM enzymes with only three of the four previously-known conserved cysteine residues. Bacterial, algal and mammalian ArsMs studied to date all have four conserved cysteine residues. Functionally, the two closest to the N-terminus form one pair, and the other two form a second pair (Marapakala *et al.*, 2015, Marapakala *et al.*, 2012, Dheeman *et al.*, 2014). For clarity, they are labeled CysA, CysB, CysC and CysD according to their distance from the N-terminus. All four cysteines are required for the first methylation step (As(III)→MAs(III)), but the second methylation step (MAs(III)→DMAs(III)) requires only the second pair, CysC and CysD. CysC and CysD form the binding site for trivalent arsenicals, explaining why the enzyme is inactive in their absence. Mechanistically CysA and CysB are thought to be involved reduction of an enzyme-bound MAs(V)

intermediate, producing a disulfide bond between CysA and CysB and an enzyme-bound MAs(III) complex (Dheeman *et al.*, 2014, Marapakala *et al.*, 2015). The disulfide bond is reduced by thioredoxin, regenerating the active form of the enzyme. Thus the two conserved cysteine pairs are considered essential for ArsM function.

Aspergillus fumigatus Af293, a common soil fungus often found in decaying organic matter such as compost heaps has four *AfarsM* genes, each with only CysB, CysC and CysD. Cells of the arsenic hypersensitive *E. coli* strain AW3110 (Δars) expressing *AfarsM1* are resistant to MAs(III) but not As(III). Purified AfArsM1 rapidly methylated MAs(III), but As(III) methylation was nearly absent. A C55S derivative methylated MAs(III), but neither C143S nor C195S mutants showed activity. These results demonstrate that the AfArsM1 of *A. fumigatus* Af293 is a functional MAs(III) SAM methyltransferase with little ability to methylate inorganic As(III). We postulate that soil fungi such as *Aspergillus* evolved three-cysteine ArsMs to cope with environmental MAs(III) produced by other soil microbes.

5.3 Material and Methods

5.31 Experimental Procedures

Chemicals. Unless otherwise indicated, chemicals were purchased from Sigma-Aldrich. MAs(V) was obtained from Chem Service, Inc., West Chester, PA, respectively. MAs(V) was reduced as described (Reay & Asher, 1977).

5.32 Strains, Media and Growth Conditions. *E. coli* Stellar™ (Clontech Laboratories, Inc., Mountain View, CA) (*F*, *endA1*, *supE44*, *thi-1*, *recA1*, *relA1*, *gyrA96*,

phoA, $\Phi 80d$ *lacZ* Δ *M15*, Δ (*lacZYA-argF*)*U169*, Δ (*mrr-hsdRMS-mcrBC*), Δ *mcrA*, λ -) was used for plasmid DNA construction and replication. *E. coli* AW3110 (Δ *ars::cam F IN(rrn-rrnE)*) (Carlin *et al.*, 1995), which is hypersensitive to As(III), was used for complementation studies. *E. coli* strains were grown aerobically at 37 °C in either Luria-Bertani (LB) medium or M9 medium (Sambrook *et al.*, 1989), as noted, supplemented with 125 μ g/ml ampicillin or 34 μ g/ml chloramphenicol, as required. Bacterial growth was monitored by measuring the optical density at 600 nm (A_{600nm}).

5.33 Synthesis of *AfarsM1* Gene and Construction of Expression

An *AfarsM1* gene corresponding to the mRNA sequence in NCBI (XM_748062.1) was chemically synthesized with 5' *Nco*I and 3' *Hind*III sites and subcloned into the *Eco*RV site of pUC57-Kan (GenScript, NJ, USA). The synthetic *AfarsM1* gene was cloned as an *Nco*I/*Hind*III double digested fragment from pUC57-Kan-*AfarsM* into expression vector pET32a(+) that produces a fusion with a 109-residue thioredoxin and six-histidine tags at the 5' end. All sequences were confirmed by commercial DNA sequencing (Sequetech, Mountain View, CA).

5.34 Phylogenetic Analysis

Multiple alignment of ArsM homolog sequence was performed using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). ArsM sequences with conserved cysteines were selected for phylogenetic analysis. Acquisition of sequences was performed by searching a list of reference organisms or from the National Center for Biotechnology (NCBI) protein database by BLASTP search (Johnson *et al.*, 2008). Phylogenetic analysis was performed to infer the evolutionary relationship among the As(III) SAM methyltransferases of various organisms. The phylogenetic tree was constructed using

the Neighbor-Joining method using MEGA 6.0.1 (Tamura *et al.*, 2013). The statistical significance of the branch pattern was estimated by conducting a 1000 bootstrap (Saitou & Nei, 1987).

5.35 Mutagenesis of the *AfarsM1* Gene

AfArsM1 mutations were generated by site-directed mutagenesis using a Quick Change mutagenesis kit (Stratagene, La Jolla, CA). The mutagenic oligonucleotides were used for both strands and the respective changes. The codons for conserved residues Cys55, Cys143 and Cys195 were individually changed to serine codons, generating three single AfArsM1 mutants. The nonconserved Cys240 and Cys241 residues were simultaneously mutated to serine codons, producing a double mutant (AfArsM1_{C240/241S}). From multiple sequence alignment of putative ArsM orthologs, it appears that there is little sequence conservation in the C-terminal domain. For that reason, an AfArsM1 derivative encoding only the N-terminal 258 residues, designated AfArsM1₂₅₈, was constructed. Each AfArsM1 derivative was confirmed by commercial DNA sequencing (Sequetech, Mountain View, CA).

5.36 Arsenic Resistance Assays

For metalloid resistance assays, competent cells of AW3110 (DE3) were transformed with constructs with or without an *AfarsM1* gene. Cells were grown overnight with shaking at 37 °C in LB medium with 100 µg ml⁻¹ ampicillin. Overnight cultures were diluted 100-fold in M9 medium containing various concentrations of As(III) or MAs(III) plus 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 37 °C with shaking for an additional 24 h. Growth was estimated from the absorbance at 600 nm.

5.37 Expression and Purification of AfArsM1

E. coli BL21(DE3) bearing AfArsM1 in vector plasmid pET32a were grown in LB medium containing 100 $\mu\text{g ml}^{-1}$ ampicillin with shaking at 37 °C. At an $A_{600\text{nm}}$ of 0.6, 0.3 mM IPTG was added as an inducer, and the culture was grown for an additional 4 h at 37 °C. The cells were harvested and suspended in 5 ml per gram of wet cells in buffer A (50 mM 4-morpholinepropanesulfonic acid, 20 mM imidazole, 0.5 M NaCl, 10 mM 2-mercaptoethanol and 20% glycerol (vol/vol), pH 7.5). The cells were broken by a single passage through a French pressure cell at 20,000 psi, and immediately treated with diisopropyl fluorophosphate (2.5 μl per gram wet cell). Membranes and unbroken cells were removed by centrifugation at 150,000 x g for 1 h, and the supernatant solution was loaded onto a Ni^{2+} -nitrilotriacetic acid column (Qiagen, Valencia, CA) at a flow rate of 0.5 ml min^{-1} . The column was washed with more than 25 column volumes of buffer A. AfArsM1 was eluted with buffer A containing 0.2 M imidazole, and the purity was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE). Protein concentrations were estimated from $A_{280\text{nm}}$ ($\epsilon = 39080 \text{ M}^{-1} \text{ cm}^{-1}$). AfArsM1-containing fractions were divided into small portions, rapidly frozen, and stored at -80 °C until use.

5.38 Assay of Arsenic Methylation Activity

Methylation of As(III) and MAs(III) was assayed both in cells of *E. coli* expressing *AfarsM1* genes and by purified ArsM as described previously (Marapakala *et al.*, 2012). For *in vivo* methylation activity assays, *E. coli* cells expressing *AfarsM1* or variants were grown to $A_{600\text{nm}} = 2$ with 0.3 mM IPTG as an inducer at 37°C with aeration in LB medium. The cells were harvested and washed once with M9 medium, then suspended in M9 medium supplemented with 0.2% D-glucose and cell density was adjusted to $A_{600} =$

3.0. As(III) or MAs(III) were individually added at 4 μ M, final concentration. Methylation activity was analyzed at the indicated time. Methylation activity of purified AfArsM1 was assayed at 37 °C in buffer consisting of 50 mM MOPS, pH 7.5, containing 0.3 M NaCl, 8 mM GSH and 1mM SAM. The reactions were terminated by adding 6% (v/v) H₂O₂ and heated at 80 °C for 5 min to oxidize all arsenic species. Denatured protein was removed by centrifugation using a 3 kDa cutoff Amicon ultrafilter. The filtrate was speciated by HPLC-ICP-MS. Where noted, H₂O₂ was not added to allow for determination of trivalent arsenicals. Without oxidation, the majority of the arsenic at early times was found to be bound to the enzyme. To determine the nature of AfArsM1-bound arsenicals, purified AfArsM1 (5 μ M) was incubated at 37 °C with 20 μ M As(III) or MAs(III) in 50 mM MOPS, containing 0.3 M NaCl and 8 mM GSH for 10 min. The samples were passed through a Bio-Gel P-6 column pre-equilibrated with reaction buffer, and portions (25 μ l) were immediately diluted with 6 M guanidine HCl to denature the protein and release bound arsenicals. Arsenic was speciated by high pressure liquid chromatography (HPLC) (PerkinElmer Series 2000) using a C18 reversed-phase column eluted with a mobile phase consisting of 3 mM malonic acid, 5 mM tetrabutylammonium hydroxide, and 5% (v/v) methanol (pH 5.6) with a flow rate of 1 ml/min, and arsenic content was determined by inductively coupled plasma mass spectrometry (ICP-MS) using an ELAN DRC-e spectrometer (PerkinElmer, Waltham, MA) (Qin *et al.*, 2006).

5.4 Results

5.41 Phylogeny of Three-cysteine ArsMs

A phylogenetic analysis of ArsM sequences in 55 fungal species was conducted by aligning the sequence of AfArsM1 with orthologous sequences identified by BLAST sequences, these correspond to CysB, CysC and CysD (Fig.5.2).

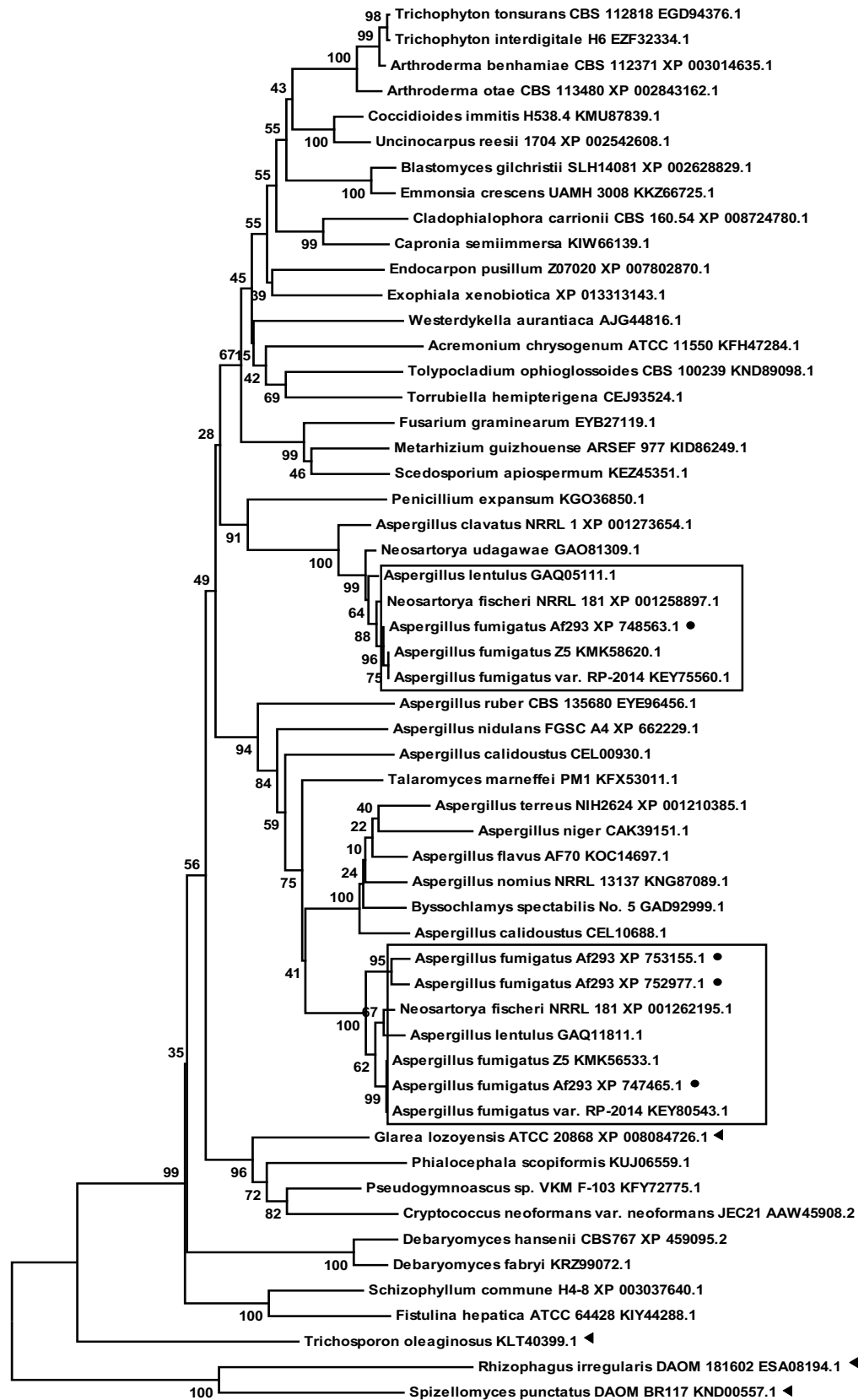


Figure. 5.1. Phylogeny of arsenite S-adenosylmethyltransferases (ArsM). A neighbor 25 joining phylogenetic tree showing the evolutionary relationships of AfArsM sequences with 26 arsenic methyltransferase sequences from other fungi. AfArsM homologs are indicated by black 27 circles. Fungi ArsM homologs with four conserved cysteines are indicated by black triangles. 28 Multiple copies of *arsM* gene in some fungi species have distant evolutionary relationships, 29 which are highlighted with closed boxes.

representative fungus is *Aspergillus fumigatus*, a common soil organism. In the eight chromosomes of *A. fumigatus* Af293 there are six clusters of *ars* genes that are similar to their bacterial counterparts (Fig. 5.3). Included in the *ars* clusters on chromosomes 1, 3, 5 and 8 are four *AfarsM* genes. AfArsM1 exhibits 93% sequence identity with AfArsM2, 79% identity with AfArsM3 and 56% identity with AfArsM4. Two closely-related orthologs with four cysteine residues are from the fungus *Glarea lozoyensis* ATCC 20868 (45%) and the bacterium *Rhodopseudomonas palustris* (43%) (Fig. 5.2). Comparing AfArsM1 with animal and plant orthologs, it shares only 26% identity with human AS3MT and 29% with algal CmArsM (*Cyanidioschyzon* sp). This low similarity indicates a more distant evolutionary relationship of fungal ArsMs with other eukaryotic As(III) SAM methyltransferases. Only four fungal species have ArsM sequences with four conserved cysteines: *Glarea lozoyensis* ATCC 20868, *Trichosporon oleaginous*, *Rhizophagus irregularis* DAOM 181602 and *Spizellomyces punctatus* DAOM BR117 (indicated by

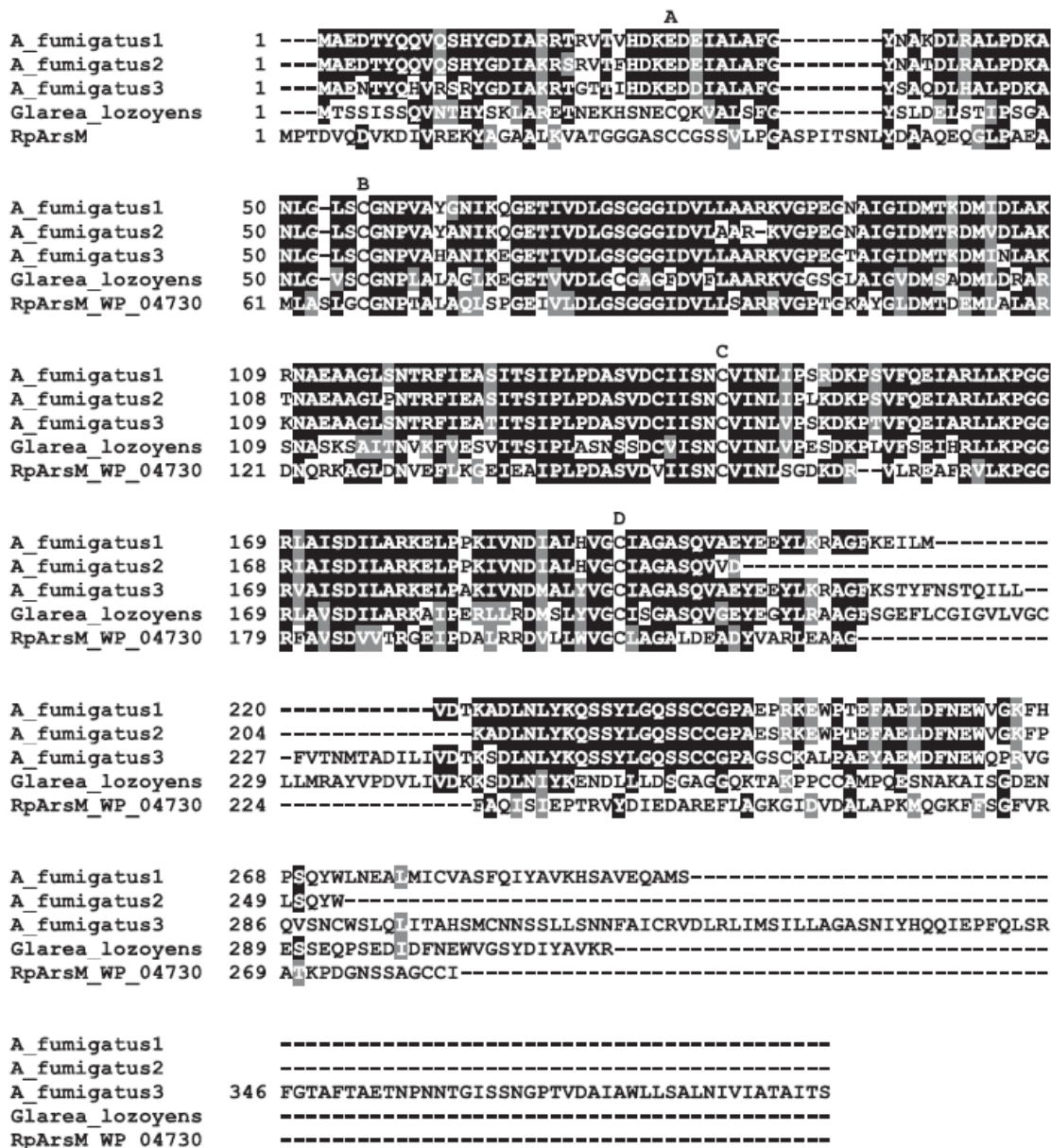


Figure 5.2. Multiple alignment of AfArsM orthologs. ArsM sequences were compared from *A. fumigatus* Af293 AfArsM1 (XP_753155.1); AfArsM2 (XP_752977.1); AfArsM3 (XP_747465.1); *Glarea lozoyensis* ATCC 20868 (XP_008084726.1) and *Rhodopseudomonas palustris* (WP_011159102.1). Identities are highlighted in black and conservative replacements in gray.

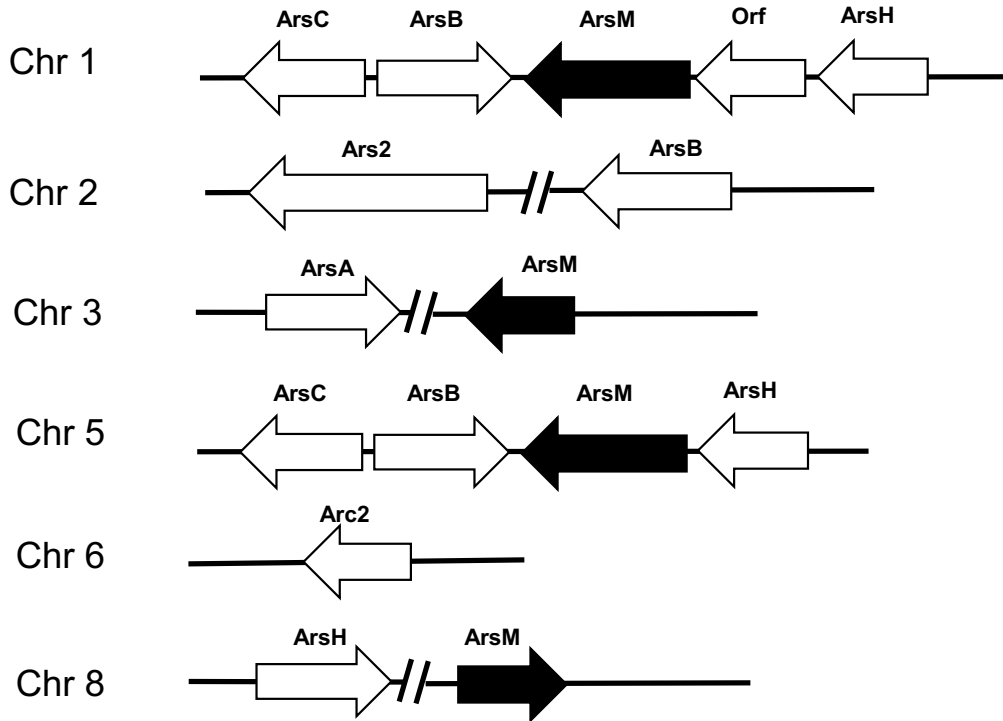


Figure 5.3. The arsenic resistance clusters in the chromosome *A. fumigatus* Af293. Four ArsM sequences are found in chromosomes 1,3,5 and 8. The sequences encoding ArsB, ArsC and ArsH are closely related to their bacterial orthologs, suggesting that the fungal genes were acquired by horizontal gene transfer.

closed triangles in Fig. 5.1). *A. fumigatus* Af293 AfArsM1, 2 and 3 cluster in one group, while AfArsM4 is in a more distant group. Similarly, two other fungal species, *A. lentulus* and *Neosartorya fischeri* NRRL 181, have multiple ArsMs that fall into the same two separate groups as the AfArsMs. This distribution in two groups in multiple organisms suggests that the ancestral genes may be acquired by separate horizontal gene transfers from bacterial sources. The ancestor of AfArsM1 may have given rise to AfArsM1, 2 and 3 by gene duplications, leading to multiple copies in *A. fumigatus* Af293. To trace the bacterial origin of AfArsM, a phylogenetic analysis of ArsM sequences of 55 representative prokaryotes and eukaryotes organisms was conducted (Fig. 1.4). On the whole, the fungal sequences were most closely related to bacterial sources from Planctomycetes and Acidobacteria, which also have three conserved cysteines. These

two bacteria sequences are quite distant from other prokaryotic ArsMs with four conserved cysteines. The fungal ArsM of *Glarea lozoyensis* ATCC 20868 (XP_008084726.1) has four conserved cysteines but clustered with other fungi. On the other hand, the fungal ArsMs from *T. oleaginous*, *R. irregularis* and *S. punctatus*, which each have four conserved cysteines, did not cluster with other fungi, suggesting that those three ArsMs were acquired from bacterial sources more recently than the initial horizontal gene transfer to fungi. This analysis suggests that the ArsM of most fungi originated from Planctomycetes and Acidobacteria.

5.42 AfArsM1 Methylates and Confers Resistance to MAs(III) but Not As(III)

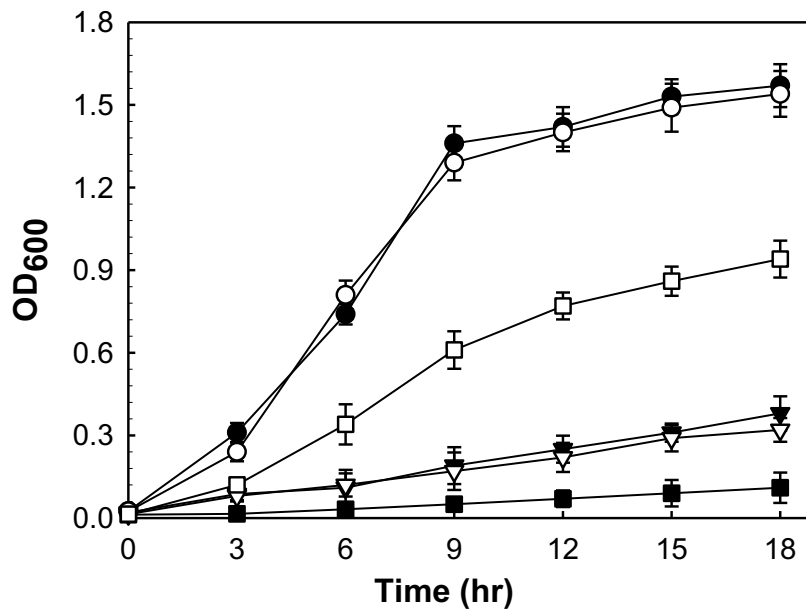


Figure. 5.4 AfArsM1 confer MAs(III) resistance, not As(III). *E. Coli* strain AW3110 bearing either pET32a-AfArsM1 (open symbols) or vector plasmid pET32a (closed symbols) were grown in M9 minimal medium with 0.3 mM IPTG for the indicated times at 37°C with shaking, as described in Experimental Procedures. Additions: none (circles), 50 mM As(III) (triangles) or 10 mM MAs(III) (squares). Growth was estimated from A600nm. Data are the mean \pm SE (n=3).

To examine the physiological role of AfArsM1 in arsenic resistance, the DNA sequence of the *AfarsM1* gene located in chromosome 5 was chemically synthesized, cloned into vector pET32a (pET32a-*AfarsM*) and expressed in the As(III)-hypersensitive *E. coli*

strain AW3110(Δars). In M9 minimal medium, *E. coli* strain AW3110 is sensitive to 50 μM As(III) and 10 μM MAs(III) (Fig. 5.4). Expression of the synthetic *AfarsM1* gene conferred resistance to MAs(III) but did not complement As(III) sensitivity. *AfarsM1* was then expressed in *E. coli* BL21(DE3), and the products of methylation analyzed. After 30 min of incubation, arsenic in the supernatant solution was speciated by HPLC ICP-MS. When cells were incubated with 4 μM MAs(III), 90% was methylated to the trimethylated species in 30 min (Fig. 5.5A). In these experiments the reactions were terminated H_2O_2 ,

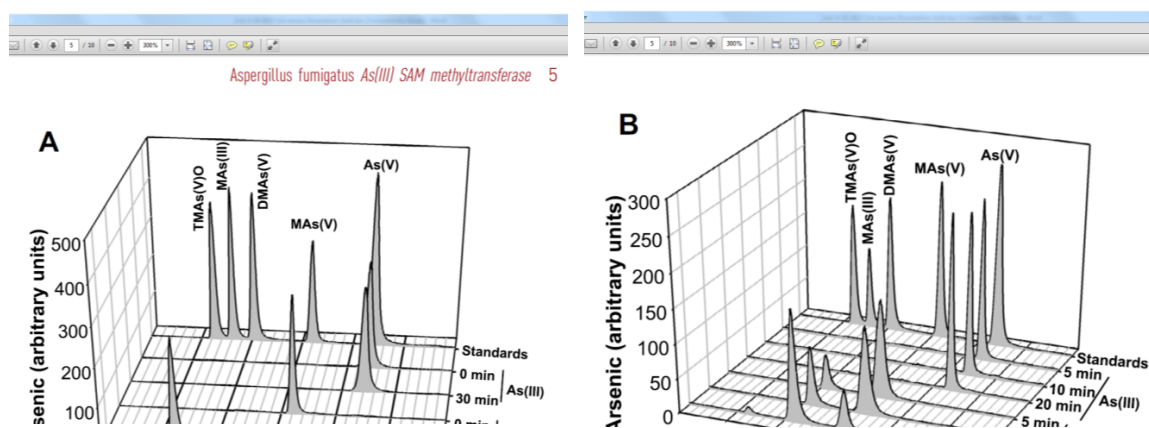


Figure 5.5. Methylation of MAs(III) or As(III) by *AfArSM1* in vivo and in vitro. Methylation reactions in vivo and in vitro were assayed as described in Experimental Procedures. All reactions were terminated by addition of 6% (v/v) H_2O_2 , final concentration. **(A)** *AfArSM1* methylates MAs(III) but not As(III) in vivo. Methylation As(III) and MAs(III) were assayed by cells of *E. coli* BL21(DE3) expressing *AfarsM1* grown in M9 minimal medium with 0.3 mM IPTG for 30 min times at 37°C, with shaking, **(B)** *AfArSM1* methylates MAs(III) but not As(III) in vitro. *AfArSM1* was purified and assayed for methylation of MAs(III) or As(III). The reaction mixture (1 ml) containing 2 IM purified *AfArSM1*, 1 mM SAM, 8 mM GSH and either 10 IM MAs(III) or As(III) was incubated at 37°C for the indicated times.

which oxidizes and solubilizes the products. In contrast, when the cells were incubated with inorganic As(III), all of the arsenic was recovered as As(V) after H_2O_2 treatment (Fig. 5.5.A). These results demonstrate that cells expressing *AfArSM1* can carry out the second and third methylation steps, transforming MAs(III) to TMAAs(III) but does not transform As(III) into MAs(III). *AfArSM1* was purified from *E. coli* and assayed for methylation activity. The enzyme methylated MAs(III) in a time-dependent manner (Fig.

4A). but As(III) methylation was nearly absent (Fig. 5.5B). These results conclusively demonstrate that the preferred substrate of AfArsM1 is MAs(III), not inorganic As(III).

5.43 Role of Conserved Cysteine Residues in MAs(III) Methylation

Most fungal As(III) SAM methyltransferases have only three conserved cysteine residues, which are Cys55 (CysB), Cys143 (CysC) and Cys195 (CysD) in AfArsM1 (Fig. 5.2). To examine their role in MAs(III) methylation, each of the three cysteines was altered individually to serine residues. When the three derivatives were expressed in *E. coli* AW3110, cells with the C55S derivative methylated MAs(III), but no methylation was observed in cells expressing either the C143S or C195S mutants, consistent with the role of CysC and CysD in binding trivalent arsenicals (Table 5.1). These results clearly demonstrate that cells expressing AfArsM1 can methylate and detoxify MAs(III) but not As(III) Each of the mutant proteins was purified, and methylation activity assayed (Table 5.1). Purified derivative C55S methylated MAs(III), but neither the C143S nor C195S enzymes exhibited methylation activity (Table 5.1). AfArsM1 has three additional nonconserved cysteine residues at positions 240, 241 and 280. To examine their contribution to methyltransferases activity, Cys240 and Cys241 were simultaneously changed to serine residues, creating a double AfArsM1C240/241S derivative. In addition, a deletion mutant containing only the first 258 residues, AfArsM1258 (which lacks Cys280), was constructed. This truncation was based on the CmArsM truncation that retains methylation activity (Marapakala et al., 2012). Each of the three purified proteins derivatives methylated MAs(III), although at a somewhat reduced levels (Table 5.1). These results clearly demonstrate that only Cys143 and Cys195, but neither conserved residue Cys55 nor the non-conserved cysteine residues, are required for MAs(III) methylation.

Table 5.1. Methylation of MAs(III) by AfArsM1 mutants in vivo and in vitro.

AfArsM1 variants	MAs(III) concentration (μM)	Products found in culture medium (μM) ^a		
		TMA _s O(V)	MA _s (V)	DMA _s (V)
WT	4 ^b	1.61±0.04 (40.3±1.0%) ^d	0.13±0.02 (3.3±0.5%)	0.58±0.03 (14.5±0.8%)
C55S		1.52±0.07 (38.0±1.8%)	0.14±0.03 (3.5±0.8%)	0.62±0.05 (15.5±1.3%)
C143S		ND ^e	1.73±0.08 (43.3±2.0%)	0.01±0.00 (0.03±0.0%)
C195S		ND	1.79±0.09 (44.7±2.3%)	0.01±0.00 (0.03±0.0%)
Control	10 ^c	ND	3.91±0.05 (97.8±1.3%)	ND
WT		0.35±0.03 (8.8±0.8%)	1.21±0.06 (30.3±1.5%)	2.19±0.13 (54.8±3.3%)
C55S		0.31±0.03 (7.9±0.8%)	1.25±0.07 (31.3±1.8%)	2.21±0.09 (55.3±2.3%)
C143S		ND	3.86±0.17 (96.5±4.3%)	ND
C195S		ND	3.87±0.15 (96.8±3.8%)	ND
258 Δ		0.11±0.02 (2.8±0.5%)	1.24±0.07 (31.0±1.5%)	2.45±0.15 (61.3±3.8%)
C240/241S		0.18±0.03 (4.5±0.8%)	1.23±0.04 (30.8±1.0%)	2.39±0.17 (59.8±4.3%)

^aMethylation activity of *E. coli* cells expressing *AfarsM1* or variants were assayed *in vivo* and *in vitro*, as described in Materials and Methods. Assays were incubated with MAs(III) at 4 μM *in vivo* or 10 μM *in vitro*, final concentration. All the samples were treated with 6% (v/v) H₂O₂, final concentration, and separated by HPLC using a C18 reverse phase column, and the amount of arsenic was estimated by ICP-MS. Data are the mean \pm SE (n = 3).

^bMAs(III) methylation *in vivo* was assayed for 20 min in cells of *E. coli* expressing wild type and the C55S, C143S and C195S mutant *AfarsM1* genes.

^cMAs(III) methylation was assayed for 40 min with purified wild type *AfarsM1* and the C55S, C143S, C195S, C240/241S double mutant and *AfarsM1* truncated at residue 258 derivatives.

^dNumbers in parentheses are the percentage of added arsenic.

^eND: non-detectable.

5.44 AfArsM1 binds but does not methylate inorganic As(III)

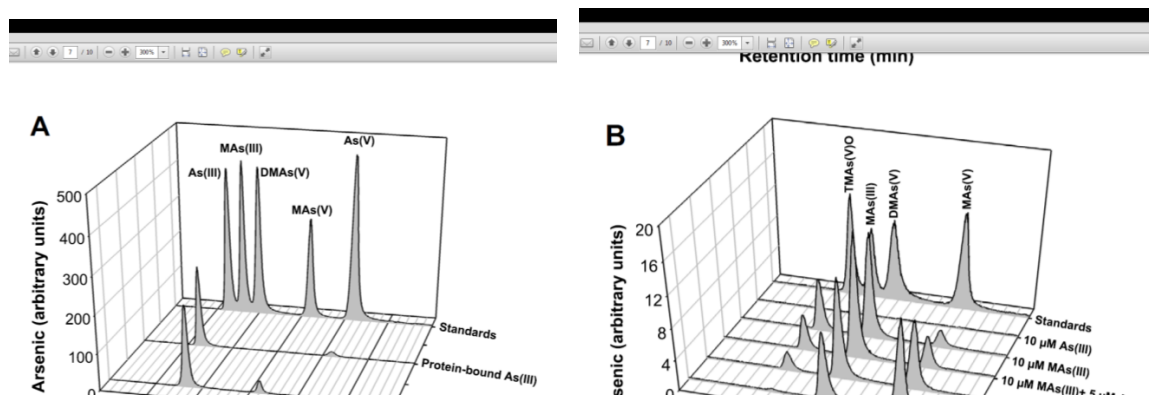


Figure 5.6. Competition of As(III) with MAs(III) for binding to AfArsM1. (A) As(III) and MAs(III) both bind to AfArsM1. (B) As(III) inhibit AfArsM1 methylation activity with MAs(III) substrate. The reaction mixture (1 ml) containing 2 IM purified AfArsM, 1 mM SAM, 8 mM GSH 10 IM MAs(III) and As(III) as indicated. Reaction was assayed at 37°C for 40 min.

Conserved residues Cys143 and Cys195 are in the CysC and CysD positions, which correspond to the arsenic binding site observed in the crystal structure of CmArsM (Ajees *et al.*, 2012). Although AfArsM1 does not methylate As(III), it is possible that it still binds inorganic As(III). To examine this possibility, 5 μ M AfArsM1 was incubated with either 20 μ M As(III) or MAs(III) for 10 min at 37°C to obtain the substrate-bound form. Unbound arsenic was removed by gel filtration using a Bio-Gel P-6 spin column, and the protein with bound As(III) or MAs(III) was denatured with guanidine HCl to release protein-bound arsenic (Dheeman *et al.*, 2014). Denatured protein was removed by filtration, and the filtrate was speciated by HPLC-ICP-MS (Fig. 5.6). The results show that either As(III) or MAs(III) bind to AfArsM1. In addition, As(III) inhibited MAs(III) methylation (Fig. 5.6B). MAs(III) methylation activity decreased as the concentration of inorganic As(III) increased. These results indicate that As(III) competes with MAs(III) for binding to Cys143-Cys195 even though it is not itself methylated. These results are

consistent with the observation that both CysA and CysB residues are required for methylation of inorganic As(III).

5.45 MAs(V) is Reduced to MAs(III) prior to Methylation

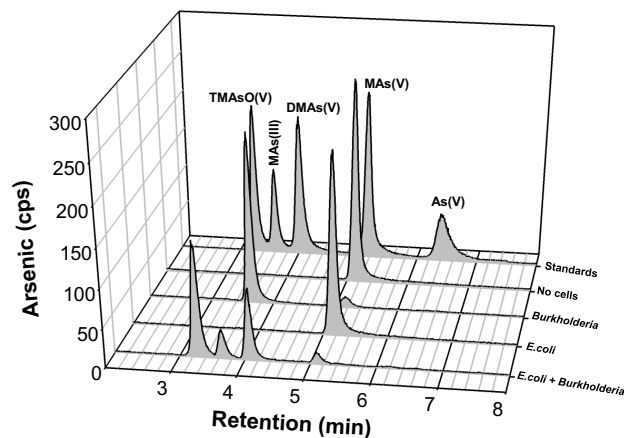


Figure 5.7. Methylation of MAs(V) by *E. coli* expressing *AfarsM1* co-cultured with *Burkholderia sp.* MR1. *E. coli* BI21(DE3) harboring the *AfarsM1* gene or *Burkholderia sp.* MR1 or both in co-culture was grown for 14 hr in M9 medium supplemented with 0.2% glucose as carbon source, 0.3 mM IPTG and 2 μ M MAs(V).

What is the physiological role of a three-cysteine ArsM? We speculate that soil microbes reduce MAs(V) to generate MAs(III) as an antibiotic that kills off their competitors (Li *et al.*, 2016). In response, the target organisms have evolved resistance mechanisms to escape MAs(III)

toxicity. While application of the commercial herbicide MSMA (monosodium methylarsenate) introduces MAs(V) into soil, how is

MAs(V) produced biologically? Soil microbes with four-cysteine ArsMs synthesize MAs(III), which is excreted and oxidized in soil and water to MAs(V). In microbial soil communities some bacteria such as the environmental isolate *Burkholderia sp.* MR1 reduce pentavalent arsenicals, and other soli organisms methylate or demethylate MAs(III) (Yoshinaga *et al.*, 2011). To test this idea, we created an artificial microbial community by co-culturing *E. coli* BI21(DE3) harboring the *AfarsM1* gene with *Burkholderia sp.* MR1. When the mixed culture was incubated with MAs(V), both DMAs(V) and TMAAs(V)O were produced (Fig 5.7). In the absence of *Burkholderia sp.* MR1, no methylation was observed. These results are consistent with our hypothesis

that the physiological role of three-cysteine ArsMs is detoxification of MAs(III) generated by microbial soil communities.

5.5 Discussion

Genes for ArsM As(III) SAM methyltransferases are widespread and found in members of every kingdom. A relatively small number of ArsM orthologs have been characterized. The majority of those have four conserved cysteine residues in the A, B, C and D positions from the N-terminus. The role of those cysteine residues has been examined in the algal CmArsM from *Cyanidioschyzon sp* (Marapakala *et al.*, 2015, Marapakala *et al.*, 2012) and human hAS3MT ortholog (Dheeman *et al.*, 2014). In both enzymes CysA and CysB are both required for methylation of As(III) but not MAs(III). From the crystal structure of CmArsM, two cysteines are located in the SAM binding domain. CysC and CysD are in the arsenic binding domain and are required for methylation of As(III) and MAs(III). In this study we identify a class of enzyme with only three of the conserved cysteine residues, CysB, CysC and CysD. For the most part these are fungal enzymes, although there are a small number of bacterial ones that cluster with them, suggesting horizontal transfer of the gene for a three-cysteine ArsM from bacteria to fungi. Some fungi have four-cysteine *arsM* genes, but those appear to be divergent from other fungal ArsMs, perhaps the result of a second horizontal gene transfer event.

In this report we show that the three-cysteine AfArsM1 does not methylate inorganic As(III) at a measurable rate even though it binds As(III). In contrast, AfArsM1 methylates MAs(III) at rates comparable with four-cysteine orthologs. Even though binding of As(III) inhibits MAs(III) methylation, *A. fumigatus* has three *arsB* genes that encoded As(III)

efflux permeases, which would reduce the intracellular concentration of As(III) to levels that would not inhibit MAs(III) methylation. From our previous studies with three-cysteine derivatives of four-cysteine CmArsM and hAS3MT, it appears that the three-cysteine enzymes methylate As(III) very slowly, perhaps a few percent of the activity of the wild type, while the rate of MAs(III) methylation is nearly wild-type (Dheeman *et al.*, 2014, Marapakala *et al.*, 2015). Without either CysA or CysB, the enzymes are unable to form a disulfide bond that reduces a putative pentavalent enzyme-bound intermediate. In the absence of the disulfide bond, the rate at which the enzyme progresses to the second methylation is extremely slow, and we predict that AfArsM1 and other three-cysteine enzymes can methylate As(III) but at only a fraction of the rate of MAs(III). In this regard, an ArsM ortholog from a soil fungus, *Westerdykella aurantiaca*, was recently shown to methylate As(III) even though it only has three conserved cysteine residues (Verma *et al.*, 2016). It is not clear if the rate of methylation of inorganic As(III) methylation is comparable with the rate of MAs(III) with this fungal enzyme.

What is the pathway of evolution of three- and four-cysteine ArsMs? Since AS3MT with only CysC and CysD is capable of methylating MAs(III) rapidly and As(III) slowly (Dheeman *et al.*, 2014), it is reasonable to consider that a two-cysteine ArsM with only CysC and CysD was the ancestor of three- and four-cysteine enzymes. A mutation generating a third cysteine (either CysA or CysB) may have been a neutral event, but a subsequent mutation introducing the fourth cysteine was a beneficial event, and that *arsM* gene would have rapidly replaced the genes for two- and three-cysteine enzymes. Are the three-cysteine ArsMs found in most fungi and some bacteria the predecessor of the four-cysteine enzymes or the result of a loss-of-function mutation? While there are no data to distinguish between these two possibilities, since fungal *arsB* and *acr3* genes

could confer substantial As(III) resistance, we infer that the primary function of fungal ArsMs is to confer MAs(III) resistance. We have shown that a number of soil microbes produce MAs(III), either by methylation of As(III) (Qin *et al.*, 2009, Qin *et al.*, 2006) or by reduction of MAs(V) (Yoshinaga *et al.*, 2011). We proposed that these organisms produce MAs(III) as a primitive antibiotic (Li *et al.*, 2016). In 1942 Selman Waksman defined “*antibiotic*” as any substance produced by one microbe that inhibits growth of other microbes. By this definition, MAs(III) is an antibiotic, and a three-cysteine ArsM that transforms it to TMAs(III), which is nontoxic (Cullen, 2005), is an antibiotic resistance. Thus we envision that the role of three-cysteine ArsMs is to protect against the antibiotic generated by MAs(III) producers.

Finally, why does *Aspergillus fumigatus* have so many clusters of *ars* genes, including four *arsM* genes? It is not unusual to find multiple *ars* clusters in a single organism (Andres & Bertin, 2016), and many fungi have multiple copies of the *arsM* gene. On the one hand, several *ars* clusters may confer additive tolerance to arsenic (Ordóñez *et al.*, 2005). On the other hand, multiple clusters may make it possible to express a cellular activity in a wider range of environmental conditions. For example, *Pseudomonas putida* KT2440 bears two virtually identical *arsRBCH* operons, but they are expressed at different temperatures (Paez-Espino *et al.*, 2014). Thus, the number and expression of arsenic-related genes, and their acquisition, may depend on the selective pressure of varying environmental stresses.

CHAPTER 6. The Biological Consequences of Polymorphisms in the Human

AS3MT Arsenic Methylation Gene

6.1 Abstract

Arsenic methylation, the primary biotransformation in the human body, is catalyzed by the enzyme As(III) S-adenosylmethionine (SAM) methyltransferases (hAS3MT). This process is thought to be protective from acute high-level arsenic exposure. However, with long term low-level exposure, hAS3MT produces intracellular methylarsenite (MAs(III)) and dimethylarsenite (DMAs(III)), which are considerably more toxic than inorganic As(III) and may contribute to arsenic-related diseases. Several single nucleotide polymorphisms (SNPs) in putative regulatory elements of the hAS3MT gene have been shown to be protective. In contrast, three previously identified exonic SNPs (R173W, M287T and T306I) may be deleterious. The goal of this study was to examine the effect of single amino acid substitutions in hAS3MT on the activity of the enzyme that might explain their contributions to adverse health effects of environmental arsenic. We identified five additional intragenic variants in hAS3MT (H51R, C61W, I136T, W203C and R251H). We purified the eight polymorphic hAS3MT proteins and characterized their enzymatic properties. Each enzyme had low methylation activity through decreased affinity for substrate, lower overall rates of catalysis and/or lower stability. We propose that amino acid substitutions in hAS3MT with decreased catalytic activity lead to detrimental responses to environmental arsenic and may increase the risk of arsenic-related diseases.

6.2 Introduction

The objective of this project was to identify polymorphisms in human AS3MT and

determine their enzymatic properties with the goal of understanding how these polymorphisms contribute to arsenic-related diseases. Human AS3MT (hAS3MT) produces MAs(III) and DMAs(III) primarily in liver (Vahter 2002). These trivalent products have been shown to be more toxic (Dopp et al. 2004; Petrick et al. 2000; Styblo et al. 2000) and potentially more carcinogenic (Styblo et al. 2002) than inorganic arsenic. MAs(III) and DMAs(III) are eventually excreted in urine, where they are abiotically oxidized in air to MAs(V) and DMAs(V) (Gong et al. 2001; Le et al. 2000). In this study the sums of trivalent and pentavalent inorganic and methylated urinary arsenic are termed iAs, MAs and DMAs, respectively. High urinary levels of MAs relative to DMAs are a susceptibility factor correlated with arsenic-related diseases (Antonelli et al. 2014; Engstrom et al. 2015; Huang et al. 2007; Vahter 2002). In contrast, a higher ratio of DMAs to MAs in urine is considered protective, perhaps because it reflects more rapid clearance of arsenic from the body (Schlawicke Engstrom et al. 2007).

The most frequent distribution of arsenic metabolites in human urine is 10-30% inorganic As, 10-20% MAs and 60-70% DMAs, but there are large individual variations (Huang et al. 2007; Vahter 2002). Some variations are associated with single nucleotide polymorphisms in the hAS3MT gene (Schlawicke Engstrom et al. 2007). Most SNPs are neutral and have little effect on health or development (Shastry 2007). However, a M287T SNP in hAS3MT has been associated with higher risk of arsenic-related diseases such as diabetes (Drobna et al. 2012), skin lesions (Pierce et al. 2014; Valenzuela et al. 2009) and cancer (Beebe-Dimmer et al. 2012; Engstrom et al. 2015). Polymorphisms in the hAS3MT gene have been associated with altered arsenic metabolite patterns among different populations, including those in Bangladesh, Argentina, Mexico, Taiwan and Central Europe (Agusa et al. 2011; Engstrom et al. 2011;

Hsieh et al. 2017; Lindberg et al. 2007; Schlebusch et al. 2013; Valenzuela et al. 2009). For example, one protective hAS3MT haplotype that produces low urinary excretion of MAs ($\approx 7.5\%$) and a higher percentage of DMAs ($\approx 78\%$) is found in indigenous populations in the Argentinean Andes exposed for generations to elevated arsenic in their water supply (0.8 mg/L) (Schlawicke Engstrom et al. 2007; Schlebusch et al. 2013; Schlebusch et al. 2015). The increased ratio of DMAs:MA in urine in this population may reflect natural selection for SNPs in the non-coding region of the human hAS3MT gene that increases expression of the gene, leading to increased arsenic tolerance. Inhabitants of Camarones in the Arica y Parinacota Region in Chile exposed to $>1\text{mg/L}$ of arsenic in their drinking water have four protective genetic variants of the hAS3MT gene (G12390C, C14215T, T14458C, and G35991A) (Apata et al. 2017). These variants are associated with more efficient arsenic metabolism and suggest human adaptation to persistent high levels of arsenic (Apata et al. 2017; Schlebusch et al. 2013; Schlebusch et al. 2015).

Until this study, only three SNPs that encode single amino acid changes in the hAS3MT protein, R173W, M287T and T306I, have been identified (Wood et al. 2006). These three exonic SNPs are found in the AS3MT coding region of African-Americans and Caucasian-Americans. They reported that the M287T SNP has a frequency of about 10% in both populations, the R173W SNP has a minor allele frequency of 0.8% in the African-American population, and the T306I SNP has a minor allele frequency of 0.8% in just one sample of Caucasian-Americans. The most common polymorphism, M287T, has been associated with lower overall methylation capacity, with lower primary (MA:iA) and secondary (DMA:MA) urinary methylation ratios (PMI and SMI,

respectively) in different populations (Agusa et al. 2011; Antonelli et al. 2014; Engstrom et al. 2011; Hernandez et al. 2008; Hsieh et al. 2017; Lindberg et al. 2007).

In this study we searched repositories of human genomic data and identified eight nonsynonymous missense variants for further analysis. The location of these single amino acid substitutions in the sequence of hAS3MT is shown in Fig. 1.2. The goal of the project was to determine the consequences of the single amino acid substitutions on the ability of these enzymes to methylate arsenic and whether structural information can be used to predict the effect of the substitutions. Knowledge of the enzymatic mechanism of AS3MT is crucial for understanding its paradoxical role in protection from arsenic exposure and its transformation of arsenic into more toxic methylated species, information that can only be acquired from studies with purified enzyme. We previously synthesized a gene for hAS3MT optimized for bacterial expression, which allowed purification of highly active AS3MT (Dheeman et al. 2014). We introduced each mutation into the synthetic hAS3MT gene and purified the resulting enzymes. We compared their enzymatic properties and stability with the most common form of hAS3MT (termed wild type in this study) using either As(III) or MAs(III) as substrate. The location of each substitution in the structure of hAS3MT was identified using a homology structural model of hAS3MT (Dheeman et al. 2014), allowing correlation of structure with enzymatic properties. For example, a I136T substitution could be placed in the N-terminal domain that contains the SAM binding domain, and this single amino acid substitution resulted in lower affinity for SAM. A C61W eliminates one of the four conserved cysteine residues (Cys32, Cys61, Cys156 and Cys206) involved in substrate binding and specificity. As predicted, the C61W enzyme methylated MAs(III) but was unable to methylate As(III). Each of the other hAS3MT polymorphic enzymes has lower methylation capacity, with

reduction in the V_{\max} , K_{cat}/K_m and affinity for As(III), resulting in lower DMA:MA ratios. Compared to wild type hAS3MT, those SNPs also show lower thermal stability. Thus, individuals with any of the eight variants would be predicted to have a longer total arsenic retention time in the body, leading to elimination of more iAs and MAs and less DMAs with a lower urinary DMAs:MA ratio. These individuals could be at greater risk for arsenic-related diseases.

6.3 Materials and Methods

6.31 Reagents.

SAM was purchased from Cayman Chemical Co., Ann Arbor, MI. A stock solution of tris(2-carboxyethyl)phosphine (TCEP) was prepared at 0.5 M and adjusted to pH 7.0. MAs(V) was reduced to trivalent MAs(III) using $\text{Na}_2\text{S}_2\text{O}_3$, $\text{Na}_2\text{S}_2\text{O}_5$, and H_2SO_4 , and adjusted to pH 6.5 with NaOH, as described (Chen et al. 2014). The identity of the reduction products were confirmed by high performance liquid chromatography (HPLC) coupled to inductively coupled mass spectroscopy (ICP-MS), as described (Dong et al. 2015b). The substrates of for methylation were the glutathione (GSH) conjugates $\text{As}(\text{GS})_3$ and $\text{MA}(\text{GS})_2$, which were prepared by incubation of 1 mM As(III) or MAs(III) with a 4-fold molar excess of GSH for 5 h at 23 °C in degassed buffers under argon, as described (Marapakala et al. 2012). All other reagents were purchased from commercial sources and were of analytical grade or better.

6.32 Strains, Media and Growth Conditions.

Escherichia coli Stellar™ (Clontech Laboratories, Mountain View, CA) was used for plasmid DNA construction and replication. For most experiments cultures of *E. coli*

bearing the indicated plasmids were grown aerobically in Luria-Bertani (LB) (Sambrook et al. 1989) at 37 °C supplemented with 100 µg/ml ampicillin or 50 µg/mL kanamycin, as required, with shaking (Dheeman et al. 2014). Bacterial growth was monitored by measuring the absorbance at 600 nm ($A_{600\text{nm}}$).

6.33 Construction of hAS3MT Variants

The eight mutations in the synthetic hAS3MT gene (Dheeman et al. 2014) were introduced by site-directed mutagenesis using a QuickChange® mutagenesis kit (Stratagene, La Jolla, CA). The oligonucleotides were used for mutagenesis. Each hAS3MT mutation was confirmed by commercial DNA sequencing (Sequetech, Mountain View, CA).

6.34 Protein Expression and Purification.

Wild type AS3MT and variant enzymes were purified by Ni-NTA chromatography, as described (Dheeman et al. 2014). Protein concentrations were estimated from $A_{280\text{nm}}$ using $\epsilon = 39080 \text{ M}^{-1} \text{ cm}^{-1}$ (Gill and von Hippel 1989). hAS3MT-containing fractions were rapidly frozen and stored at -80 °C until use. Thioredoxin (Trx) and thioredoxin reductase (TR) were prepared as described (Dheeman et al. 2014). All buffers were degassed by bubbling with argon for 30 min before use.

6.35 Assays of Arsenic Methylation

hAS3MT activity was assayed with two different procedures. The time-resolved Förster resonance energy transfer (TR-FRET) assay measures conversion of SAM to S-adenosylhomocysteine (SAH) at short times using an EPIgenous Methyltransferase Assay kit (Cisbio Bioassays, Bedford, MA) (Dong et al. 2015b). The assay was carried out using a low volume 384-well microtiter plate in a buffer consisting of 50 mM

NaH₂PO₄, 0.3 M NaCl, 1 μM purified hAS3MT, 0.5 mM GSH, 1 μM Trx, 0.3 μM TR and 0.03 mM NADPH and 10 μM of either As(GS)₃ or MAs(GS)₂, unless otherwise indicated. The plates were incubated at 37 °C for 2 min with shaking in an Eppendorf ThermoMixer C before adding SAM at 10 μM, final concentration (unless otherwise indicated), to initiate the reaction. The reactions incubated for 1, 2 and 5 min and were terminated and developed by addition of the SAH-d2 and anti-SAH-Lumi4-Tb detection reagents. The plates were incubated for 1 h, and fluorescence was measured at both 665 nm and 620 nm with excitation at 337 nm in a Synergy H4 Hybrid Multi-Mode microplate reader. The homogeneous time-resolved fluorescence (HTRF) was calculated from the ratio of emission at 665 and 620 nm. The concentration of SAH was calculated with a calibration curve constructed with known concentrations of SAH (Dong et al. 2015b). The reaction was linear over 5 min, and initial rates were calculated from the slope.

For measurement at longer times and for speciation of the products, HPLC was used, with arsenic concentrations determined by ICP-MS (Dheeman et al. 2014). The assay mixture contained 1 μM purified hAS3MT, 2.5 mM GSH, 10 μM Trx, 1.5 μM TR, 0.3 mM NADPH and 10 μM of either As(GS)₃ or MAs(GS)₂ (unless otherwise indicated) in a buffer consisting of 50 mM NaH₂PO₄ and 0.3 M NaCl, pH 8. SAM was added at a final concentration 0.5 mM to initiate the reaction at 37 °C. For initial rate determinations of SAM kinetics, the reactions were carried out for 5, 10 and 20 min at the indicated SAM concentrations. The reaction was linear over this time period, and the slope was used to estimate the initial rate. To recover all of the arsenic, the reactions were terminated by addition of H₂O₂ at 10% (v/v), final concentration, which also oxidizes all arsenicals, so the products will be termed MAs and DMAs. The assay solution was immediately passed through a 3 kDa cutoff Amicon ultrafilter (Millipore, Billerica, MA), and speciation of arsenic in the filtrate was determined by high pressure liquid chromatography (HPLC)

(PerkinElmer Series 2000) with a C18 300A reverse-phase column, with the arsenic concentration measured by inductively coupled plasma mass spectrometry (ICP-MS) using an ELAN DRC-e (Perkin Elmer, Waltham, MA). As(III), MAs(III), DMAs(V), MAs(V) and As(V) were used at 1 μ M as standards.

Methylation in cells of *E. coli* BL21(DE3) expressing hAS3MT wild type and mutants was assayed by growing cells in 2 ml of LB medium in the presence of 0.3 mM IPTG, 100 μ g/ml kanamycin and 10 μ M of either As(III) or 2 μ M MAs(III) at 37 °C for 12 h. The cell were harvested, washed and suspended in ST⁻¹ medium (Maki et al. 2009) with 2 μ M MAs(III), then grown for 3 hours at 37°C. Arsenicals were speciated by HPLC using a C18 reverse phase column, and the amount of arsenic was estimated by ICP-MS.

6.36 Assays of Thermal Stability.

Thermal stability was assayed by incubation of the enzyme at 42°C for the indicated times in a buffer consisting of 50 mM NaH₂PO₄ and 0.3 M NaCl, pH 8. The reaction was terminated rapid cooling on ice, and methylation activity determined with the TR-FRET assay for 5 min at 37 °C. $t_{1/2}$ values were calculated using SigmaPlot (Systat Software, Inc., San Jose, CA).

6.37 Sources of Single Nucleotide Polymorphisms.

hAS3MT SNPs were identified from the Ensembl genome browser, which contains 1000 Genomes, NHLBI GO Exome Sequencing Project (ESP), Exome Aggregation Consortium (ExAC), NHLBI Exome Sequencing Project, CLINSEQ SNP: CSAgilent and the HapMap Project repositories of human genome data.

6.38 Homology Model of the hAS3MT Structure with Polymorphic Residues

A homology model of hAS3MT (Dheeman et al. 2014) was built on the structure of PhAs(III)-bound CmArsM (PDB ID: 4KW7) from residues 44–371 (residue numbers based on the CmArsM sequence) using a fully automated protein structure homology modeling server SWISS-MODEL (<http://swissmodel.expasy.org/>) (Fig. 6.1A). The model quality was estimated based on the QMEAN scoring function. The final homology model incorporated 308 of those 328 residues. To place SAM in the model, *in silico* docking with SAM was carried out using the PATCHDOCK server. The docked hAS3MT model with SAM was superimposed with the As(III)-bound structure of CmArsM (PDB ID: 4FSD) (Ajees et al. 2012) to acquire the arsenic atom in the As(III) binding site of hAS3MT. PyMOL v1.3 was used to visualize the structural models (DeLano 2001). The model for each of the polymorphic variants was built similarly and superposed with each other to locate the position of the polymorphic residues (Fig. 6.1B).

6.4 Results

6.41 Missense Polymorphisms in hAS3MT and Population Frequencies

Because of the importance of arsenic metabolic reaction in humans, one objective was to determine whether the human hAS3MT gene, like many other human methyltransferase genes, includes additional functional genetic polymorphisms. From database searches we identified 891 sequence variants in the AS3MT gene. Eight nonsynonymous missense variants that result in single amino acid changes in the AS3MT protein were chosen for further analysis using the following criteria: 1) Previously known (R173W, M287T and T306I), 2) location in the As(III) (C61W and W203C) or SAM binding (H51R and I136T) domains, and 3) high PolyPhen2 scores (Adzhubei et al. 2013)(R251H, R173W and W203C). The PolyPhen2 algorithm predicts the possible

impact of an amino acid substitution on the structure and function of a human protein (Adzhubei et al. 2013). Based on the Exome Aggregation Consortium (ExAC) database, which has a large sample size, the order of the minor allele frequency (MAF) of these missense SNPs is as follows: WT > M287T (9.4%) > C61W (~0.5%) > R173W (~0.11%) > R251H (~0.1%) > T306I (~0.02%) > W203C (~0.004%) > H51R (~0.002%) (Table 6.1). Information on the frequency of I136T is not available.

Table. 6.1. Frequencies of the eight missense hAS3MT polymorphisms in the 1000 Genomes Project, ExAC, NHLBI Exome Sequencing Project or BUSHMAN population.

Original data were mined from the Ensembl genome browser database (<http://useast.ensembl.org/index.html>) accessed 5 April 2017. For the 1000 Genomes Project, all donors were over 18 and declared themselves to be healthy at the time of collection. The health conditions for other data source populations are unknown.

dbSNP rs# (MAF from ExAC)	Residue change	Allele change	Allele frequency (numbers of individuals): Top: Ancestral Bottom: Variants	Genotype frequency (numbers of individuals): Top: Homozygous dominant Middle: Heterozygous Bottom: Homozygous recessive	Population
rs201702937 (0.002%)	H51R	a152g	A: 0.9998 (5007) G: 0.000199 (1)	A A: 0.9996 (2503) A G: 0.00039 (1) G G: 0	1000 Genomes Project (African, American, East Asian, European, South Asian)
rs80317306 (0.50%)	C61W	t183g	T: 0.995 (120200) G: 0.005 (646)	NA	Exome Aggregation Consortium (ExAC) individuals (African/African American, Latino,

					East Asian, Finnish, Non-Finnish European, South Asian and others)
rs112056792 (NA)	I136T	t407c	T: 0.500 (1) C: 0.500 (1)	NA	BUSHMAN POP (Northern Kalahari of Africa)
rs35232887 (0.11%)	R173W	c517t	C: 0.999 (5003) T: 0.001 (5)	C C: 0.998 (2499) C T: 0.002 (5) T T: 0	1000 Genomes Project (African, American, East Asian, European, South Asian)
rs370022454 (0.004%)	W203C	g609t	G: 0.999734 (3763) T: 0.000265675 (1)	NA	NHLBI Exome Sequencing Project (African-American)
rs139656545 (0.10%)	R251H	g752a	G: 0.998 (4997) A: 0.002 (11)	G G: 0.996 (2493) A G: 0.004 (11) A A: 0	1000 Genomes Project (African, American, East Asian, European South Asian)
rs11191439 (9.4%)	M287T	t860c	T: 0.923 (4622) C: 0.077 (386)	T T: 0.854 (2138) C T: 0.138 (346) C C: 0.008 (20)	1000 Genomes Project (African, American, East Asian, European, South Asian)
rs34556438 (0.02%)	T306I	c917t	C: 0.99951 (8164) T: 0.000489716 (4)	NA	NHLBI Exome Sequencing Project (European-American)

M287T SNP has highest allele frequency (8% - 11%) and genotype (13% - 24%) frequency in the global population. A homozygous recessive genotype has been found only for the M287T polymorphism in several populations, with a frequency of about 1%. Genotype frequencies of C61W, I136T, W203C and T306I are unknown. Other missense SNPs (H51R, R173W and R251H) were identified only in heterozygous individuals. For the 1000 Genomes Project, all donors were over 18 and declared themselves to be healthy at the time of collection

(<http://www.internationalgenome.org/faq/can-i-get-phenotype-gender-and-family-relationship-information-samples/>).

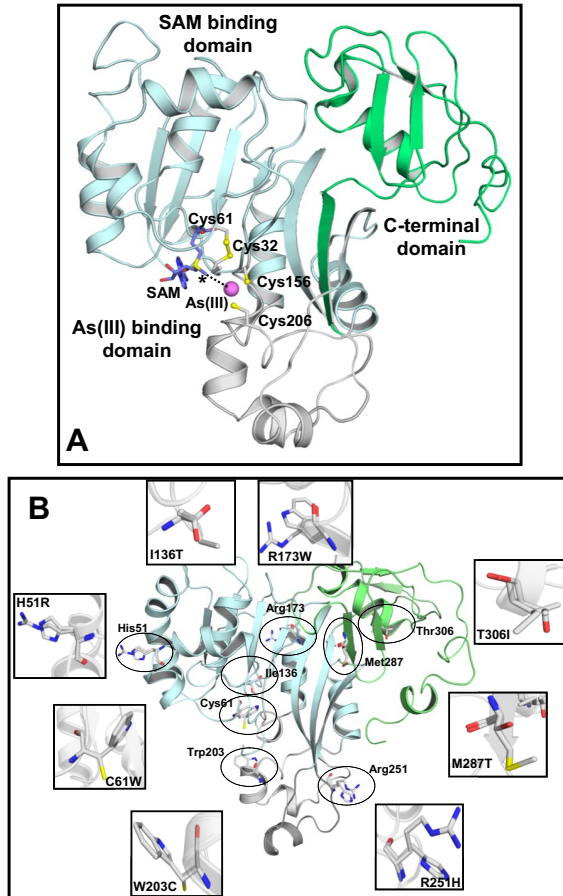


Fig. 6.1. Homology model of hAS3MT and polymorphisms. A) The structural model of the human AS3MT consists of three domain, the N-terminal SAM binding domain (light blue), the central As(III) binding domain (grey) and the C-terminal domain (green). SAM (in ball-and-stick) with its methyl group (*) poised to be donated to the arsenic atom (pink ball). B) The location of the eight residues altered by the SNPs (circled) are shown in stick form, and the predicted structure of the amino acid substitutions are shown on the boarder superimposed on the original residues.

Homology Model of hAS3MT and Polymorphisms

The structure of a protein is the key to its function, allowing visualization of substrate and allosteric binding sites. The consequences of amino acid substitutions in polymorphic variants are most easily understood by the structural changes they produce. For that reason we modeled the structure of hAS3MT and its variants on the structure of the CmArsM As(III) SAM methyltransferases from the eukaryotic alga *Cyanidioschyzon merolae* sp. 5508 (Ajees

et al. 2012). We mapped the location of the eight residues that are represented with the intragenic polymorphisms on this hAS3MT homology mode (Fig. 6.1A) and superimposed the side chains of the variants on those of the wild type residues (Fig. 6.1B). His51, Cys61 and Ile136 are in

the SAM binding domain. Arg173, Trp203 and Arg251 are in the arsenic binding domain.

Met287 and Thr306 are in the C-terminal domain. The PolyPhen2 score of each polymorphic enzyme was calculated. A high PolyPhen2 scores indicates possible deleterious effects on the structure and activity of an enzyme. The M287T enzyme has a low PolyPhen2 score, which suggests little effect on hAS3MT structure. T306I has an intermediate score. H51R, C61W, R173W, I136T, W203C have higher PolyPhen2 scores, predicting possibly harmful effects on the enzyme structure and reduced catalytic activity (Table 6.2). Arg173 is in the As(III) binding site, so we predict that the R173W substitution may affect affinity for As(III) (Ajees et al. 2012). Since Ile136T is near the SAM binding site, we predict that it may affect affinity for SAM.

Table 6.2. PolyPhen2 scores of missense variants. From the PolyPhen-2 website:

<http://genetics.bwh.harvard.edu/pph2/>

Missense variants	C61W	R173W	H51R	I136T	R251H	T306I	W203C	M287T
PolyPhen2	1	1	0.98	0.98	0.9	0.79	0.6	0
Predicted Functional Effect	Detrimental	Detrimental	Detrimental	Detrimental	Detrimental	Possibly Detrimental	Possibly Detrimental	Benign

6.43 Polymorphisms Affect hAS3MT Methylation Activity

The enzymatic activity of the polymorphic AS3MTs was determined. Each variant was expressed in *E. coli* cells in about the same amount as wild type hAS3MT, as shown by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) (Laemmli 1970). Arsenic methylation in cells of *E. coli* provides an initial screen for the effects of the mutations. Methylation in cells expressing the eight variants was compared with wild type AS3MT as a positive control and an inactive C206S mutant (Dheeman et al. 2014)

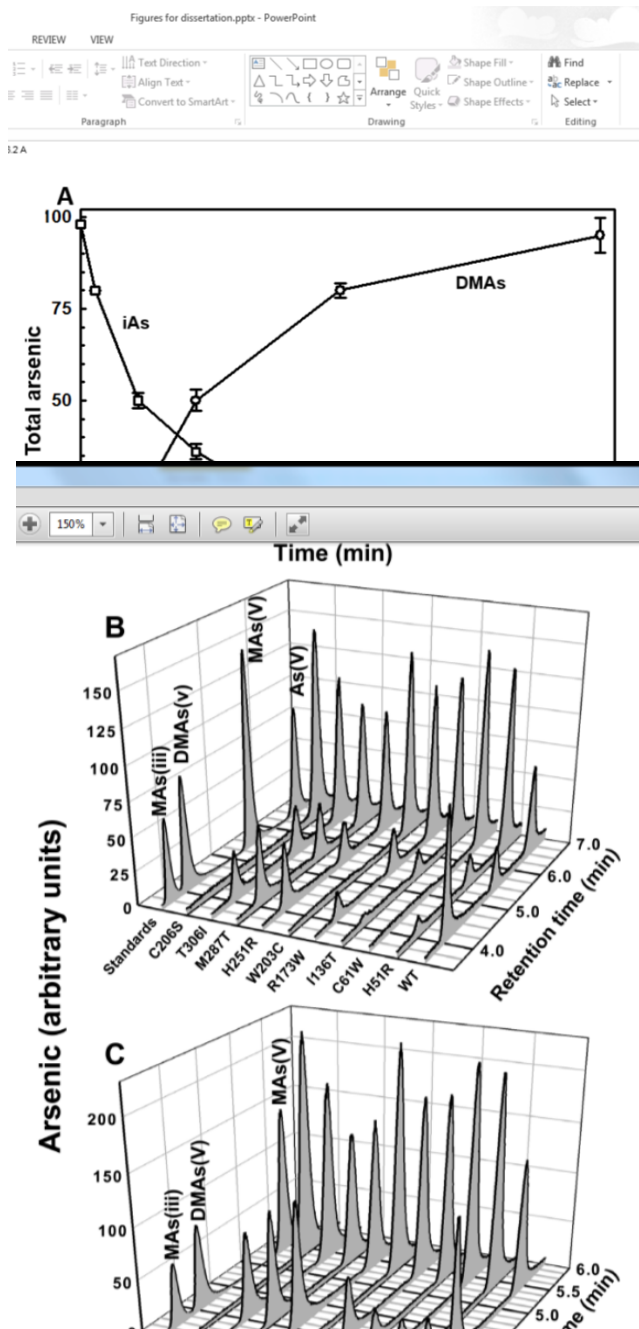


Fig. 6.2. Arsenic methylation by wild type hAS3MT and polymorphic variants. (A) Time course of As(III) methylation by wild type hAS3MT. (B) As(III) methylation by wild type hAS3MT and polymorphic variants. (C) MAS(III) methylation by wild type hAS3MT and polymorphic variants.

as a negative control. Cells expressing six of the eight SNPs (H51R, I136T, R173W, R251H, M287T and T306I) methylated As(III) to varying degrees, but only M287T activity was comparable to the wild type. Cells expressing C61W and W203C derivatives were unable to methylate As(III). Like the C206S mutant, cells expressing W203C derivative were unable to methylate MAS(III), but the cells expressing the C61W derivative retained ability to methylate MAS(III).

Purified wild type hAS3MT rapidly methylated As(III) to the methylated species and more slowly to the dimethylated species (Fig. 6.2A), consistent with our previous observations (Dheeman et al. 2014). At 30 min of reaction time, MAS accounted for approximately about 25% of total arsenic, DMAs

accounted for about 30%, with 45% remaining as iAs. The methylation activity of purified

single amino acid polymorphic enzymes were compared with wild type hAS3MT at 30 min using either As(III) (Fig. 6.2B) or MAs(III) (Fig. 6.2C) as substrate. With As(III) as substrate, the reaction measures a combination of both the first (As→MAs) and second methylation (MAs→DMAs) steps when assayed with HPLC-ICP-MS. MAs(III) is both the product of the first methylation step and is the substrate of the second methylation step, so with MAs(III) as substrate, the reaction measures only the second round of methylation (MAs→DMAs). Using As(III) as substrate, the eight polymorphic enzymes showed lower methylation activity compared to wild type hAS3MT in the following order: WT>M287T>R251H>T306I>R173W>H51R>I136T. The C61W and W203C derivatives exhibited essentially no catalytic activity and was comparable to the inactive C206S mutant.

Using MAs(III) as substrate to assay the second round of methylation, the order of methylation activity was similar to the results with As(III) as substrate with one exception. The C61W enzyme, which was inactive with As(III) as substrate, had catalytic activity with MAs(III) (Fig 6.2C). Most characterized microbial ArsMs and animal AS3MTs have four conserved cysteine residues, which are Cys32, Cys61, Cys156 and Cys206 in hAS3MT (Fig. 1.2). Cys32 and Cys61 are required for methylation of both As(III) and MAs(III), while Cys156 and Cys206 are required only for methylation of MAs(III) but not As(III) (Dheeman et al. 2014). Most fungal AS3MT have only three conserved cysteines and lack the cysteine corresponding to Cys32 in hAS3MT. The natural three-cysteine enzyme of *Aspergillus fumigatus* methylates MAs(III) but not As(III), similar to the C61W hAS3MT polymorphic enzyme (Chen et al. 2017).

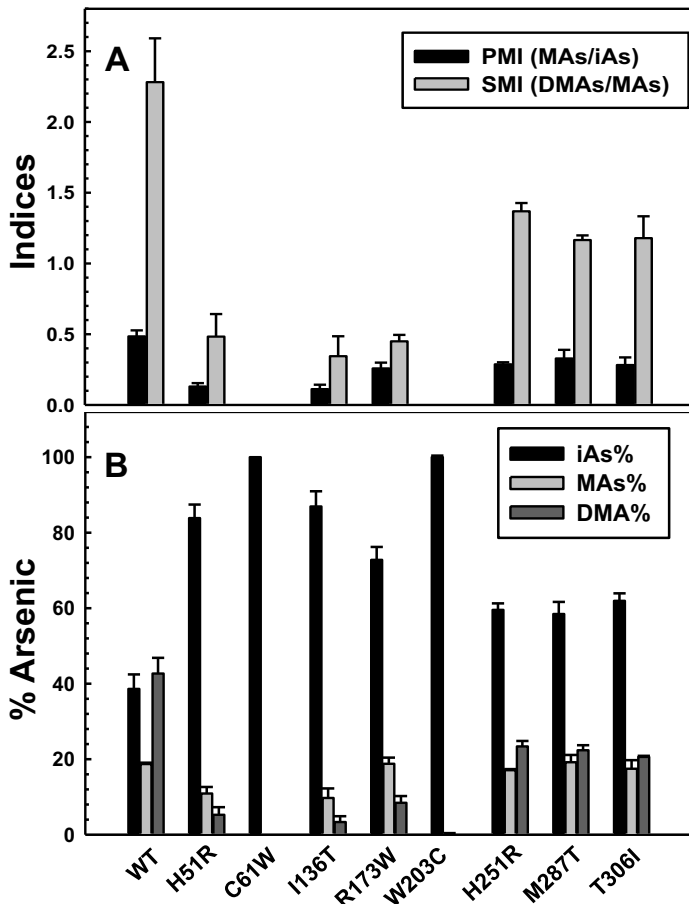


Figure 6.3. Efficiency of arsenic methylation. (A) Effect of polymorphisms on the PMI and SMI of arsenic methylation by wild type hAS3MT and variants. (B) Conversion of iAs into MAs and DMAs by each hAS3MT expressed as a percentage of total arsenic.

6.44 The Affect of Polymorphisms on the Ratio of DMAs to MAs.

The efficiency of arsenic methylation is often expressed as the ratio between the product and substrate (Huang et al. 2007; Huang et al. 2009; Schlawicke Engstrom et al. 2007). The primary methylation index (PMI) is defined as the ratio of MAs:iAs, and the secondary methylation index (SMI) is the ratio of DMAs:MAS (Huang et al. 2007). The SMI in urine has frequently been

used as an operational indicator of methylation capacity of individuals exposed to inorganic arsenic and in studies of inter-individual variability in susceptibility to adverse health effects associated with chronic exposure (Chen et al. 2003). Here we compared the *in vitro* methylation indexes of the eight polymorphic enzymes with wild type hAS3MT calculated from the data in Fig. 6.2B. The results show differences in both the PMI and SMI between the wild type and polymorphic enzymes. In every case both the SMI and PMI were lower than the wild type values (Fig. 6.3A). The

WT enzyme had the highest DMAs/MAs ratio of 2.3 ± 0.3 . The SMI of M287T, R251H and T306I (approximately 1.2) were lower than the WT enzyme but higher than the other SNPs. The SMI of the H51R, I136T, R173W enzymes were approximately 0.45. By definition the PMI and SMI of the inactive C61W and W203C enzymes are zero.

To obtain a more complete picture of the methylation cycle, the percentage of iAs, MAs and DMAs were calculated (Fig. 6.3B). The data demonstrate that SNPs with low methylation efficiency have higher percentages of iAs and MAs, and lower percentages of DMAs compared to wild type hAS3MT. After 30 min of reaction, wild type AS3MT total arsenic was 38% As, 19% MAs and 43% DMAs. The M287T, R251H and T306I enzymes showed 58-61% inorganic As, 18-20% MAs and 20-23% DMAs. H51R, I136T, R173W, C61W and W203C had the highest amount of starting material iAs (83-100%) and lowest DMAs (0-8%). These results demonstrate that the eight nonsynonymous missense variants of hAS3MT have lower arsenic methylation capacity compared with wild type hAS3MT, from which we predict may lead to individual variations in arsenic methylation and slower clearance of arsenic from the body, factors that could increase the risk of arsenic-related diseases.

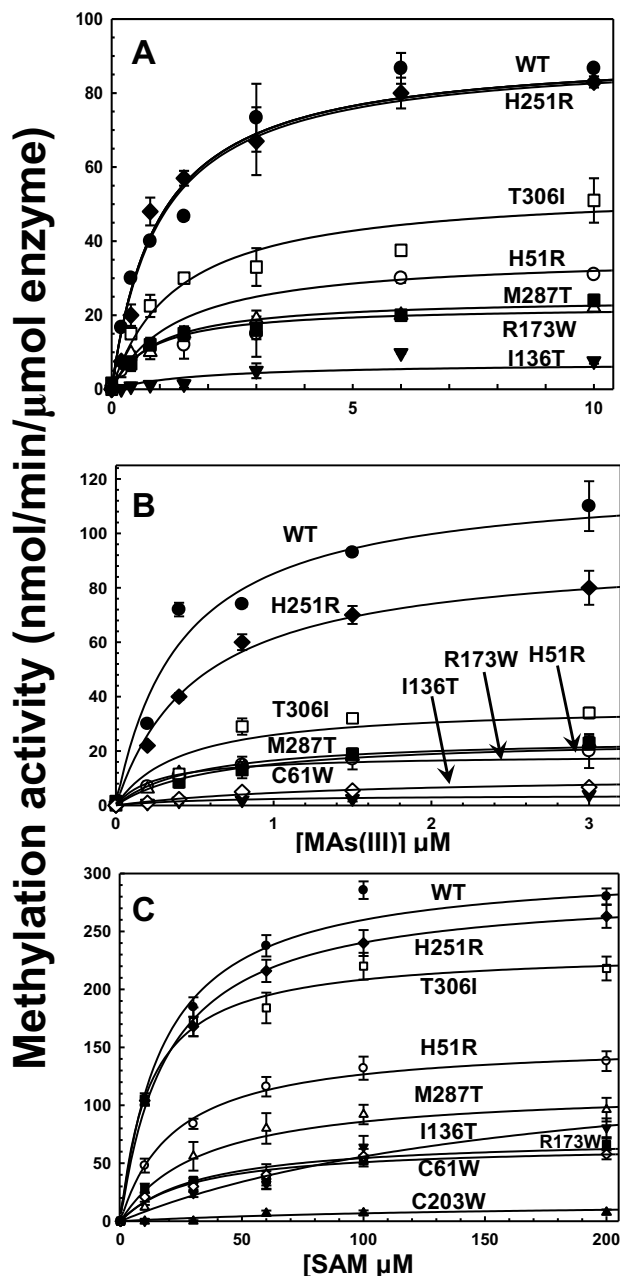


Fig. 6.4. Kinetics of wild type hAS3MT and polymorphic variants. (A) Methylation as a function of As(III) concentration. **(B)** Methylation as a function of MAs(III) concentration. Reactions were performed with. **(C)** Methylation as a function of SAM concentration.

6.45 Kinetic Analysis of hAS3MT

Polymorphic Enzymes

Kinetics govern fluxes of metabolites through intracellular pathways. Enzymes catalyze the individual steps of cellular reactions, and their rates dictate which pathways predominate. The rate of production of individual methylated species, and hence the PMI and SMI, depends on the kinetics of each step, which may vary in different polymorphic variants. How do the kinetic properties of the polymorphic enzymes compare with the wild type? The kinetic parameters for each substrate, As(III), MAs(III) and SAM were individually determined for wild type and variant

hAS3MTs. With the TR-FRET assay the first (As→MAs) (Fig. 6.4A) and second (MAs→DMAs) (Fig. 6.4B) methylation steps can be assayed

independently (Dong et al. 2015b). In both steps wild type hAS3MT exhibited the highest maximal rates (V_{max}) compared with the polymorphic enzymes. The R251H and T306I

variants showed higher V_{max} values compared to the other polymorphic enzymes. The V_{max} values of the I136T, R173W, M287T and H51R variants were significantly lower than the others in both reaction steps.

The affinity (K_m) of each enzyme for As(III) and MAs(III) was determined (Table 6.3). Wild type and polymorphic hAS3MT enzymes each had K_m values in the range of 1-2 μ M, indicating that the amino acid substitutions did not affect binding of As(III) (except for C61S and W203C, which do not methylate As(III)). The apparent K_m values for MAs(III) were lower than for As(III) for each enzyme, in the range of 0.4-0.8 μ M, suggesting that MAs(III) is a better substrate than As(III). The exceptions were W203C, which does not methylate MAs(III), and C61S which has a 3-fold reduction in affinity for MAs(III).

Table. 6.3. Kinetic parameters for arsenic methylation catalyzed by hAS3MT.

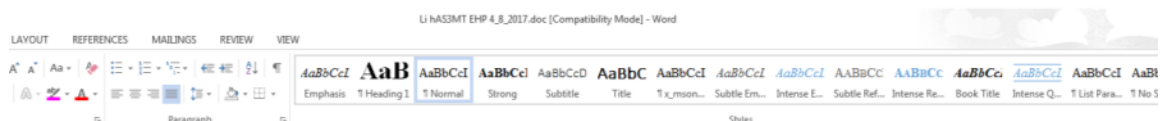


Table. 2. Kinetic parameters for arsenic methylation catalyzed by hAS3MT.

hAS3MT	K_m (μ M)			V_{max} (nmole/min/ μ mole hAS3MT)			K_{cat}/K_m ($S^{-1}M^{-1}$)		
	As(III) ^a	MAs(III) ^b	SAM ^c	As(III) ^a	MAs(III) ^b	SAM ^c	As(III) ^a	MAs(III) ^b	SAM ^c
WT	0.97 \pm 0.1	0.40 \pm 0.06	17 \pm 1	91 \pm 3	120 \pm 20	304 \pm 4	1564 \pm 50	5000 \pm 36	186 \pm 3
H51R	1.5 \pm 0.2	0.25 \pm 0.13	22 \pm 2	30 \pm 6*	18 \pm 8*	152 \pm 4*	333 \pm 10*	1200 \pm 14*	71 \pm 1*
C61W	-	1.40 \pm 0.1*	30 \pm 1*	-	10 \pm 3*	64 \pm 1*	-	119 \pm 12*	22 \pm 1*
I136T	1.0 \pm 0.25*	0.84 \pm 0.1	127 \pm 5*	8 \pm 3*	2.8 \pm 0.1*	148 \pm 8*	70 \pm 8*	20 \pm 5*	44 \pm 4*

^aActivity was measured at 0, 1, 2 and 5 min in the presence of 10 μ M SAM, 0.5 mM GSH, 1 μ M Trx, 0.3 μ M TR, 0.03 mM NADPH and As(III) at concentrations up to 10 μ M using the TR-FRET assay at 37 $^{\circ}$ C. The data are the mean \pm SE (n=3).

^bActivity was measured with the TR-FRET assay at 0, 1, 2 and 5 min in the presence of 10 μ M SAM, 0.5 mM GSH, 1 μ M Trx, 0.3 μ M TR, 0.03 mM NADPH and MAs(III) at concentrations up to 10 μ M at 37 $^{\circ}$ C.

^cActivity was measured at 0, 5, 10 and 20 min in the presence of 10 μM MAs(III), 2.5mM GSH, 10 μM Trx, 1.5 μM TR, 0.3 mM NADPH and SAM at concentrations from up to 600 μM by HPLC ICP-MS at 37 °C.

*Indicates statistically significant differences ($p < 0.02$) in V_{max} or K_{m} values between wild type hAS3MT and variants-catalyzed methylation.

The TR-FRET assay measures formation of SAH from SAM and thus cannot be used for determination of SAM kinetics (Dong et al. 2015b), so the kinetics for SAM as substrate were determined by formation of DMAs from MAs(III) as a function of SAM concentration using HPLC-ICP-MS. This measures only the second methylation step but was used because the C61S variant cannot methylate As(III). The V_{max} values for wild type hAS3MT, R251H and T306I were similar to each other and were higher than other polymorphic enzymes (Fig. 6.4C). The affinity of each polymorphic enzymes for SAM was similar to that of the wild type, in the range of 14-30 μM with one exception. The K_{m} for I136T for SAM was 137 μM , 8-fold lower affinity than wild type hAS3MT. From the homology model of hAS3MT, Ile136 is in the SAM binding domain, suggesting that a threonine substitution affects folding of the SAM binding domain.

The specificity constant or catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) of an enzyme is a useful metric for comparing the relative rates of an enzyme for multiple substrates. The $k_{\text{cat}}/K_{\text{m}}$ of the wild type and SNPs was calculated (Table 6.3). The wild type specificity constant for MAs(III) was three-fold higher than for As(III), clearly showing that the methylated species is a better substrate for hAS3MT than inorganic arsenic. In every case the catalytic efficiency of the SNPs was lower than wild type hAS3MT, unambiguously demonstrating that each single amino acid substitution resulted in a less active enzyme.

6.46 Thermal Stability of the Eight Polymorphic hAS3MTs

A possible consequence of a single amino substitution in a protein is improper folding that results in decreased stability (Adzhubei et al. 2013; Fagain 1995; Horton et al. 2001). The thermal stability of the seven active missense variants was compared with wild type hAS3MT. The enzymes were incubated at 42 °C for varying lengths of time followed by determination of methylation activity (Fig. 6.5). Wild type hAS3MT had a half-life of 20 min at 42 °C. The SNPs lost activity much faster, with half-lives ranging from 1- 5 min. R251H has relatively higher stability compared to other polymorphic enzymes, which may in part account for its higher catalytic efficiency compared to the others.

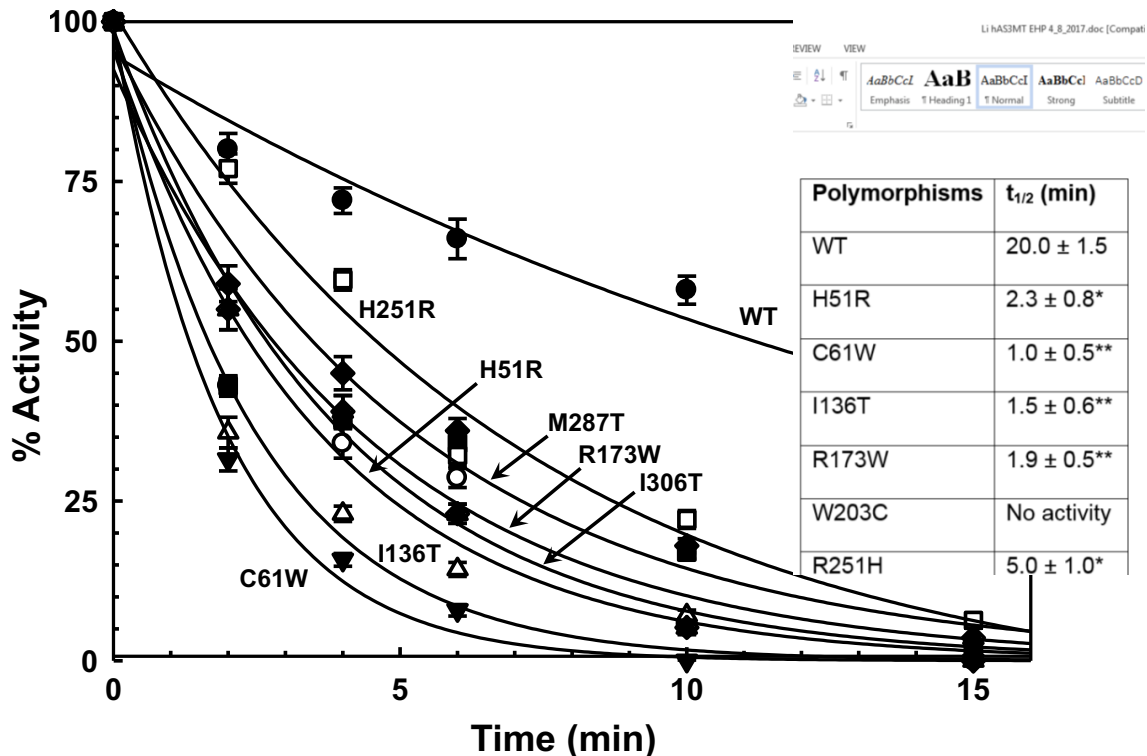


Fig. 6.5. Temperature stability of wild type hAS3MT and polymorphic variants. Wild type hAS3MT and variants were heated at 42 °C for the indicated times, following which methylation activity was assayed with the TR-FRET assay at 37 °C. The table shows temperature-dependent half-life ($t_{1/2}$) of hAS3MT and polymorphic variants.

6.5 Discussion

A growing number of single nucleotide polymorphisms in the gene for human AS3MT have been identified. The goal of this study was to characterize the products of SNPs that produce single amino acid changes in the AS3MT enzyme. To date, only the three most frequent in human populations, R173W (rs35232887), M287T (rs11191439) and T306I (rs34556438) have been studied. M287T, the most frequent polymorphism, has been associated with a lower SMI, i.e., higher urinary MAs and lower DMAs, in various populations (Drobna et al. 2012; Engstrom et al. 2011; Hernandez et al. 2008; Valenzuela et al. 2009). This SNP has been associated with higher negative outcomes. For example, in individuals with the M287T polymorphism there is an increased risk of diabetes (Drobna et al. 2012), premalignant arsenic skin lesions (Valenzuela et al. 2009), basal cell carcinoma (Engstrom et al. 2015) and perhaps bladder cancer (Beebe-Dimmer et al. 2012). On the other hand, there are no epidemiological studies of the R173W or T306I SNPs in human populations. Most individuals carrying these haplotypes are heterozygous, so each expresses a wild type AS3MT gene, which affects the phenotype. A few homozygous M287T individuals have been identified (Beebe-Dimmer et al. 2012; Lindberg et al. 2007), and these have higher MAs and lower DMAs compared with wild type hAS3MT.

There are only a few biochemical studies of these SNPs. In a ground-breaking study, Wood et al. (Wood et al. 2006) identified three nonsynonymous SNPs in hAS3MT with single amino changes R173W, M287T and T306I and examined the enzymatic activity of the three variants. In that study the human cDNA was altered by site directed mutagenesis to introduce these three changes, which were then expressed in COS-1 cells, which exhibited little endogenous AS3MT expression. The advantage of this

approach is that the enzyme is likely to have natural post-translational modifications. Cytosol containing the T306I variant had only 5% the amount of immunoreactive protein and no detectable methylation activity. The R173W variant had only 20% of the amount of immunoreactive protein and about 31% of the activity of the wild type, although it did not appear to be degraded more rapidly than the wild type protein. The K_m of the R173W variant was approximately 3 μM for As(III) and 9 μM for SAM, compared with approximately 5 μM for As(III) and 12 μM for SAM for the wild type. These values were not significantly different from each other. There was twice as much M287T immunoreactive protein in the cytosol, which did not appear to result from less rapid degradation. The M287T enzyme had 3.5-fold more activity than the wild type after correction for the amount of immunoreactive protein. The K_m of this variant was approximately 11 μM for As(III), about half the affinity of the wild type and 5 μM for SAM, about twice the affinity of the wild type. These values are significantly different from each other, but not enough to explain any physiological differences.

There was one other notable biochemical study of the M287T variant (Ding et al. 2012). In this study, wild type hAS3MT cDNA and a site-directed M287T mutant were expressed in *E. coli*, purified and assayed using an assay including Trx and TR with or without GSH. In the absence of GSH, the ratio of DMAs to MAs was low. In the presence of GSH, significantly more DMAs(III) was produced from As(III). GSH is the major intracellular thiol, so the activity in the presence of GSH is more likely to reflect the physiological activity. This is a significant finding, which led us to include Trx, TR and GSH in our assays. In the presence of GSH, the K_m for both wild type hAS3MT and M287T enzymes was approximately 1.6 μM for As(III) and 0.7-0.8 μM for MAs(III). The V_{max} values for both were relatively the same for both enzymes as well, approximately 6 pmol/ μg protein/min with As(III) and 14-16 pmol/ μg protein/min with MAs(III). From that

study, the authors concluded that there were insufficient differences between the wild type and M287T enzymes to account for differences in the SMI, and the reason for the increased susceptibility in individuals with the SNP could not be attributed to differences in catalytic activity between the variant and wild type hAS3MT.

As valuable as these contributions were, they were not conclusive. Wood et. al. (Wood et al. 2006) used crude cytosolic preparations. Enzyme kinetics are meaningful only when conducted with purified enzymes. The hAS3MT enzymes used by Ding et. al. (Ding et al. 2012) had little activity, and methylation of iAs required equimolar or excess enzyme over substrate over long time periods. These are not catalytic conditions. At most a few turnovers would occur during the reaction, not sustained catalysis. Our approach was to use highly active purified enzymes to analyze the catalytic properties of the three most frequent exonic SNPS, as well as five additional less common ones, and correlate their activity and stability with structural information. We used a synthetic hAS3MT gene as the starting material for expression and mutant construction (Dheeman et al. 2014). We showed previously that the product of the synthetic gene can be used in catalytic amounts with the Trx/TR/GSH assay. A large excess of substrate over enzyme ensures that there are multiple rounds of methylation during the assay time, which, with the TR-FRET assay, is linear up to 5 min (Dong et al. 2015b). In addition, the TR-FRET assay allows the two methylation reactions (iAs→MAs and MAs→DMAs) to be determined independently, an accomplishment not possible in previous studies. With this assay it is clear that six of the variants had similar affinity for As(III) and MAs(III) as wild type hAS3MT. The W203C enzyme lacked measurable activity. The C61W variant was not active with As(III), and Cys61 has been shown to be required for methylation of As(III) (Dheeman et al. 2014).

The kinetics of AS3MT sheds light on its role in metabolism, how its activity is controlled, and how a drug or an agonist might affect the metabolism of arsenic. Kinetic analyses have predictive physiological value. Kinetics governs the rates of metabolic pathways *in vivo*, which, in turn, allows our body to respond to arsenic exposure. The K_m for SAM of the active variants was also similar to that of the wild type except for I136T, consistent with the location of Ile136 near the SAM binding site. With each variant the V_{max} was reduced, and none of the variants approached the catalytic efficiency (k_{cat}/K_m) of the wild type enzyme.

From the kinetics, we can conclude that each single amino acid substitution produces a less active enzyme. What is the connection between the substitution and the effect on activity? The homology structural model is informative (Dheeman et al. 2014). His51 is located near the start of the N-terminal domain but not near the SAM binding site, and the H51R substitution does not affect the K_m for SAM. In contrast, Ile is near the SAM binding site, and the I136T substitution reduces the affinity for SAM 8-fold. The C61W substitution prevents formation of a disulfide bond between Cys44 and Cys61 that is required for As(III) methylation.

All of the substituted residues are on the surface of the protein except for Thr306, which is buried inside the enzyme, so a T306I substitution is likely to disrupt the structure. The M287T SNP has been proposed to lower the SMI by specifically reducing the rate of the second methylation step (Agusa et al. 2011). However, our results indicate that the first and second methylation steps are reduced by about the same amounts (78% and 75%, respectively), so there must be some other explanation. Met287 is located on the surface at the entrance to a cleft in AS3MT to which small molecule inhibitors bind (Fig. 6.6) (Dong et al. 2015a). We proposed that this cleft is an allosteric site that binds

physiological molecules that regulate methylation activity by modulating a conformational change at the cleft. We speculate that the M287T substitution hinders binding of putative modulators and/or retards the allosteric conformational change, reducing the rate of methylation of this variant. This property is consistent with observed epidemiological studies of individuals with the M287T polymorphism.

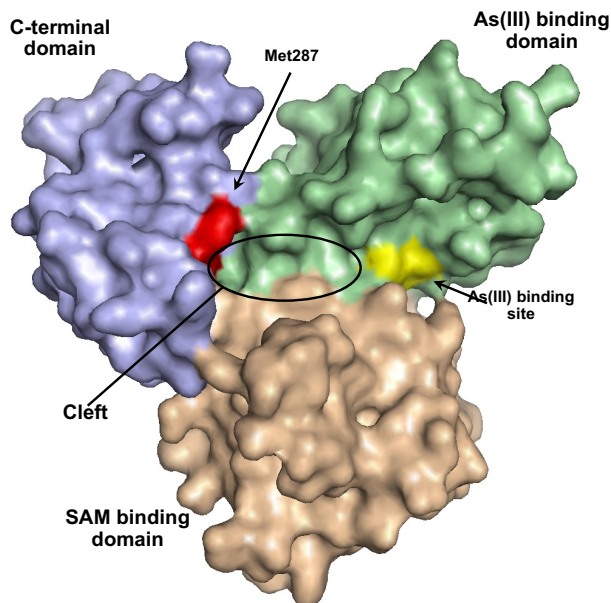


Fig. 6.6. Molecular surface model of hAS3MT. The surface model of hAS3MT was built on the structure of PhAs(III)-bound CmArsM (PDB ID: 4KW7) from residues 44–371 (residue numbers based on the CmArsM sequence) using a fully automated protein structure homology modeling server SWISS-MODEL (<http://swissmodel.expasy.org/>). PyMOL v1.3 was used to visualize the structural model (DeLano 2001). The putative allosteric site (circled) is located at the interface of the N-terminal (SAM binding) domain (wheat) and As(III) binding domain (light green). The C-terminal domain is in light blue, Met287 in red and the As(III) binding site in yellow.

Another factor that contributes to the reduced activity of most the variants is that they are less stable than the wild type. From measurements of temperature stability, the variants denature between 4- and 20-fold faster than wild type hAS3MT. Each of the eight identified SNPS is thus deleterious to one degree or another

because it destabilizes the structure of the enzyme. The few examples of protective AS3MT polymorphisms are located outside of the coding sequence in putative regulatory elements. We conclude that the identified amino acid

substitutions in AS3MT lead to decreased catalytic activity and increased harmful responses to environmental arsenic. Since arsenic levels in food and water in the United States and other developed countries are generally below the EPA and WHO recommended levels, there is little selective pressure against detrimental genotypes. In

contrast, in regions of the world with high arsenic exposure, increased expression of AS3MT is protect because it leads to faster clearance of arsenic from the body.

Future Perspectives

Microbiome bacteria have chromosomal *ars* operons that are required for detoxification of arsenic, and the microbiome responds to arsenic exposure by induction of their *ars* genes long before levels of blood or urine arsenic increase. To demonstrate the feasibility of the approach, future research could use a mouse model to measure *ars* genes expression levels, although the long-term goal is to apply this method to the human microbiome.

The effects of polymorphic hAS3MT on arsenic methylation could also be evaluated in human cells. Future research can compare the effect of their expression in cultured human HaCaT cells with wild type AS3MT. The human HaCaT cell line could be chosen for expression for several reasons. First, skin is highly susceptible to the toxic effects of arsenic, which causes degenerative, inflammatory and neoplastic changes in skin. Second, HaCaT keratinocytes do not express AS3MT and do not methylate arsenic.

References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., . . . Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**:1948-1954.
- Adzhubei I, Jordan DM, Sunyaev SR. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7:Unit7* 20.
- Agusa T, Fujihara J, Takeshita H, Iwata H. (2011). Individual variations in inorganic arsenic metabolism associated with AS3MT genetic polymorphisms. *Int J Mol Sci* **12**:2351-2382.
- Ajees AA, Marapakala K, Packianathan C, Sankaran B, Rosen BP. (2012). Structure of an As(III) S-adenosylmethionine methyltransferase: insights into the mechanism of arsenic biotransformation. *Biochemistry* **51**:5476-5485.
- Akkari, K. H., Frans, R. E. & Lavy, T. L. (1986). Factors affecting degradation of MSMA in soil. *Weed Science* **34**:781-788.
- Akter, K.F., G. Owens, D.E. Davey & R. Naidu, (2005). Arsenic speciation and toxicity in biological systems. *Rev. Environ. Contam. Toxicol.* **184**: 97-149.
- Andres, J. & P.N. Bertin, (2016) The microbial genomics of arsenic. *FEMS Microbiol. Rev.* **40**: 299-322.
- Apata M, Arriaza B, Llop E, Moraga M. (2017). Human adaptation to arsenic in Andean populations of the Atacama Desert. *Am J Phys Anthropol* in press.
- Åslund, F., Ehn, B., Miranda-Vizueté, A., Pueyo, C., & Holmgren, A. (1994). Two additional glutaredoxins exist in *Escherichia coli*: glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant. *Proc Natl Acad Sci U S A*, **91**:9813-9817.
- Åslund, F., Nordstrand, K., Berndt, K. D., Nikkola, M., Bergman, T., Ponstingl, H., . . . Holmgren, A. (1996). Glutaredoxin-3 from *Escherichia coli*. Amino acid sequence, ¹H AND ¹⁵N NMR assignments, and structural analysis. *J Biol Chem*, **271**: 6736-6745.
- Bhattacharjee, H., Sheng, J., Ajees, A. A., Mukhopadhyay, R., & Rosen, B. P. (2010). Adventitious Arsenate Reductase Activity of the Catalytic Domain of the Human Cdc25B and Cdc25C Phosphatases. *Biochemistry* **49**: 10-21.
- Barlow, G. M., Yu, A. & Mathur, R. (2015). Role of the gut microbiome in obesity and diabetes mellitus. *Nutr Clin Pract* **30**: 787-797.
- Bekker A., Holland H.D., Wang P.L., Rumble D., Stein H.J., Hannah J.L., Coetzee L.L., Beukes N.J. (2004) Dating the rise of atmospheric oxygen. *Nature* **427**: 117-20.

- Bentley, R. & T.G. Chasteen, (2002) Microbial methylation of metalloids: arsenic, antimony, and bismuth. *Microbiol. Mol. Biol. Rev.* **66**: 250-271.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., . . . Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, **112**: 535-542.
- Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermin, L. S., Schulte, G. K., . . . Pandit, J. (2000). Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem*, **275**: 24798-24806.
- Butcher, B. G., Deane, S. M., & Rawlings, D. E. (2000). The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *Applied and Environmental Microbiology*, **66**: 1826-1833.
- Cameron, A. D., Sinning, I., L'Hermite, G., Olin, B., Board, P. G., Mannervik, B., & Jones, T. A. (1995). Structural analysis of human alpha-class glutathione transferase A1-1 in the apo-form and in complexes with ethacrynic acid and its glutathione conjugate. *Structure*, **3**: 717-727.
- Challenger, F. (1947). Biological methylation. *Science Progress*, **35**: 396-416.
- Canfield DE, Rosing MT, Bjerrum C, (2006) Early anaerobic metabolisms. *Philos Trans R Soc Lond B Biol Sci* **361**:1819-34.
- Carlin, A., W. Shi, S. Dey & B.P. Rosen, (1995) The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. *J. Bacteriol.* **177**: 981-986.
- Chen, J. & Rosen, B. P. (2014). Biosensors for inorganic and organic arsenicals. *Biosensors (Basel)* **4**: 494-512.
- Chen, J., Bhattacharjee, H. & Rosen, B. P. (2015a). ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. *Mol Microbiol* **96**:1042-1052.
- Chen, J., Madegowda, M., Bhattacharjee, H. & Rosen, B. P. (2015b). ArsP: a methylarsenite efflux permease. *Mol Microbiol* **98**: 625-635
- Chen J, Sun S, Li CZ, Zhu YG, Rosen BP. (2014). Biosensor for organoarsenical herbicides and growth promoters. *Environ Sci Technol* **48**:1141-1147.
- Chen J, Li J, Jiang X, Rosen BP. (2017). Conserved cysteine residues determine substrate specificity in a novel As(III) S-adenosylmethionine methyltransferase from *Aspergillus fumigatus*. *Mol Microbiol* **104**:250-259.

- Chen, J., Qin, J., Zhu, Y. G., de Lorenzo, V., & Rosen, B. P. (2013). Engineering the soil bacterium *Pseudomonas putida* for arsenic methylation. *Applied and Environmental Microbiology*, **79**: 4493-4495.
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., . . . Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**: 12-21.
- Cullen, W.R., (2005). The toxicity of trimethylarsine: an urban myth. *J Environ Monit* **7**: 11-15.
- DeLano, W. L. (2001). *The PyMOL user's manual*. San Carlos, CA: DeLano Scientific.
- Dheeman, D.S., C. Packianathan, J.K. Pillai & B.P. Rosen, (2014). Pathway of human AS3MT arsenic methylation. *Chem. Res. Toxicol.* **27**: 1979-1989.
- Dong H, Madegowda M, Nefzi A, Houghten RA, Giulianotti MA, Rosen BP. (2015^a). Identification of small molecule inhibitors of human As(III) S-adenosylmethionine methyltransferase (AS3MT). *Chem Res Toxicol* **28**:2419-2425.
- Dong H, Xu W, Pillai JK, Packianathan C, Rosen BP. (2015^b). High-throughput screening-compatible assays of As(III) S-adenosylmethionine methyltransferase activity. *Anal Biochem* **480**:67-73.
- Drobna, Z., L.M. Del Razo, G.G. Garcia-Vargas, L.C. Sanchez-Pena, A. Barrera-Hernandez, M. Styblo & D. Loomis, (2012). Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. *J Expo Sci Environ Epidemiol* **23**: 151-155.
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.
- Eggerth, A. H. & Gagnon, B. H. (1933). The Bacteroides of human feces. *J Bacteriol* **25**: 389-413.
- Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.
- Schlawicke Engström K, Broberg K, Concha G, Nermell B, Warholm M, Vahter M. (2007). Genetic polymorphisms influencing arsenic metabolism: evidence from Argentina. *Environ Health Perspect* **115**:599–605.
- Engstrom K, Vahter M, Mlakar SJ, Concha G, Nermell B, Raqib R, et al. (2011). Polymorphisms in arsenic(+III oxidation state) methyltransferase (AS3MT) predict gene expression of AS3MT as well as arsenic metabolism. *Environ Health Perspect* **119**:182-188.

- Feng, M., Schrlau, J. E., Snyder, R., Snyder, G. H., Chen, M., Cisar, J. L. & Cai, Y. (2005). Arsenic transport and transformation associated with MSMA application on a golf course green. *J Agric Food Chem* **53**: 3556-3562.
- Fernandes, A. P., Fladvad, M., Berndt, C., Andresen, C., Lillig, C. H., Neubauer, P., Vlamis-Gardikas, A. (2005). A novel monothiol glutaredoxin (Grx4) from *Escherichia coli* can serve as a substrate for thioredoxin reductase. *J Biol Chem*, **280**: 24544-24552.
- Fernandes, A. P., & Holmgren, A. (2004). Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antioxid Redox Signal*, **6**: 63-74.
- Fisher, D. J., Yonkos, L. T. & Staver, K. W. (2015). Environmental concerns of roxarsone in broiler poultry feed and litter in Maryland, USA. *Environ Sci Technol* **49**: 1999-2012.
- Fitzpatrick DA. (2012). Horizontal gene transfer in fungi. *FEMS Microbiol Lett* **329**:1-8.
- Fu, H. L., Rosen, B. P. & Bhattacharjee, H. (2010). Biochemical characterization of a novel ArsA ATPase complex from *Alkaliphilus metalliredigens* QYMF. *FEBS Lett* **584**: 3089-3094.
- Garbarino, J. R., Bednar, A. J., Rutherford, D. W., Beyer, R. S. & Wershaw, R. L. (2003). Environmental fate of roxarsone in poultry litter. I. Degradation of roxarsone during composting. *Environ Sci Technol* **37**: 1509-1514.
- Garcia-Saez, I., Parraga, A., Phillips, M. F., Mantle, T. J., & Coll, M. (1994). Molecular structure at 1.8 Å of mouse liver class pi glutathione S-transferase complexed with S-(p-nitrobenzyl)glutathione and other inhibitors. *J Mol Biol*, **237**: 298-314.
- Ghosh, M., Shen, J. & Rosen, B. P. (1999). Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. *Proc Nat Acad Sci USA* **96**: 5001-5006.
- Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**:319-326.
- Gille, C., & Frommel, C. (2001). STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics*, **17**: 377-378.
- Habig, W. H., Pabst, M. J., & Jakoby, W. B. (1974). Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. *J Biol Chem*, **249**: 7130-7139.
- Hamdi, M., M. Yoshinaga, C. Packianathan, J. Qin, J. Hallauer, J.R. McDermott, H.C. Yang, K.J. Tsai & Z. Liu, (2012). Identification of an S-adenosylmethionine (SAM) dependent arsenic methyltransferase in *Danio rerio*. *Toxicol. Appl. Pharmacol.* **262**: 185-193.

- Hamdi, M., Sanchez, M. A., Beene, L. C., Liu, Q., Landfear, S. M., Rosen, B. P., & Liu, Z. (2009). Arsenic transport by zebrafish aquaglyceroporins. *BMC Mol Biol*, **10**: 104.
- Hernandez A, Xamena N, Sekaran C, Tokunaga H, Sampayo-Reyes A, Quinteros D, et al. (2008). High arsenic metabolic efficiency in AS3MT287Thr allele carriers. *Pharmacogenet Genomics* **18**:349-355.
- Hsu, C. M. & Rosen, B. P. (1989). Characterization of the catalytic subunit of an anion pump. *J Biol Chem* **264**: 17349-17354.
- Holm, L., Kaariainen, S., Wilton, C., & Plewczynski, D. (2006). Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics, Chapter 5, Unit 5 5*.
- Holmgren, A. (1976). Hydrogen donor system for Escherichia coli ribonucleoside-diphosphate reductase dependent upon glutathione. *Proc Natl Acad Sci U S A*, **73**: 2275-2279.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph*, **14**: 33-38, 27-38.
- Huang JL. (2013). Horizontal gene transfer in eukaryotes: the weak-link model. *Bioessays* **35**: 868-75.
- Jomova, K., Jenisova, Z., Feszterova, M., Baros, S., Liska, J., Hudecova, D., Rhodes, C. J. & Valko, M. (2011) Arsenic: toxicity, oxidative stress and human disease. *J Appl Toxicol* **31**: 95-107.
- Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N., & Gilliland, G. L. (1995) Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. *Biochemistry*, **34**: 5317-5328.
- Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis & T.L. Madden, (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5-9.
- Kozbial, P. Z., and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins, *BMC structural biology* **5**: 19.
- Kozul, C. D., Hampton, T. H., Davey, J. C. & other authors (2009) Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. *Environ Health Perspect* **117**: 1108-1115.
- Kuo, C. C., Howard, B. V., Umans, J. G. & other authors (2015) Arsenic exposure, arsenic metabolism, and incident diabetes in the strong heart study. *Diabetes Care* **38**: 620-627.
- Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.

- Le, X.C., X. Lu, M. Ma, W.R. Cullen, H.V. Aposhian & B. Zheng, (2000) Speciation of key arsenic metabolic intermediates in human urine. *Anal. Chem.* **72**: 5172-5177.
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* **444**: 1022-1023.
- Li, G., Sun, G. X., Williams, P. N., Nunes, L. & Zhu, Y. G. (2011). Inorganic arsenic in Chinese food and its cancer risk. *Environ Int* **37**: 1219-1225.
- Li, J., S.S. Pawitwar & B.P. Rosen, (2016). The organoarsenical biocycle and the primordial antibiotic methylarsenite. *Metallomics* **8**: 1047-1055.
- Lin, S., Q. Shi, F.B. Nix, M. Styblo, M.A. Beck, K.M. Herbin-Davis, L.L. Hall, J.B. Simeonsson & D.J. Thomas, (2002). A novel S-adenosyl-L-methionine:arsenic(III) methyltransferase from rat liver cytosol. *J. Biol. Chem.* **277**: 10795-10803.
- Lin, Y. F., Walmsley, A. R. & Rosen, B. P. (2006). An arsenic metallochaperone for an arsenic detoxification pump. *Proc Natl Acad Sci U S A* **103**: 15617-15622.
- Liu, Z., Rensing, C. & Rosen, B. P. (2013). Resistance pathways for metalloids and toxic metals. In *Metals in Cells*, pp. 429-442. Edited by V. Culotta & R. A. Scott. Hoboken, NJ Wiley & Sons, Inc.
- Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.
- Lomax, C., Liu, W. J., Wu, L. & other authors (2011). Methylated arsenic species in plants originate from soil microorganisms. *New Phytol* **193**: 665-672.
- Lu, K., Cable, P. H., Abo, R. P. & other authors (2013). Gut microbiome perturbations induced by bacterial infection affect arsenic biotransformation. *Chem Res Toxicol* **26**: 1893-1903.
- Lu, K., Abo, R. P., Schlieper, K. A., Graffam, M. E., Levine, S., Wishnok, J. S., Swenberg, J. A., Tannenbaum, S. R. & Fox, J. G. (2014). Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. *Environ Health Perspect* **122**: 284-291.
- Lundstrom-Ljung, J., Vlamis-Gardikas, A., Åslund, F., & Holmgren, A. (1999). Reactivity of glutaredoxins 1, 2 and 3 from Escherichia coli and protein disulfide isomerase towards glutathionyl-mixed disulfides in ribonuclease A. *FEBS Lett* **443**: 85-88.
- Maizel, D., Utturkar, S. M., Brown, S. D., Ferrero, M. A. & Rosen, B. P. (2015). Draft Genome Sequence of *Brevibacterium linens* AE038-8, an Extremely Arsenic-Resistant Bacterium. *Genome Announc* **3**.

- Maki T, Hirota W, Ueda K, Hasegawa H, Azizur Rahman M. (2009). Seasonal dynamics of biodegradation activities for dimethylarsinic acid (DMA) in Lake Kahokugata. *Chemosphere* **77**:36-42.
- Mandal BK, Suzuki KT. (2002). Arsenic round the world: a review. *Talanta* **58**:201-235.
- Mannervik, B., & Danielson, U. H. (1988). Glutathione transferases--structure and catalytic activity. *CRC Crit Rev Biochem* **23**: 283-337.
- Martin, J. L. (1995). Thioredoxin--a fold for all reasons. *Structure* **3**: 245-250.
- Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D. L., Rosen, B. P., & Edwards, B. F. (2001). Insights into the structure, solvation and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure* **9**: 1071-1081.
- Marapakala, K., C. Packianathan, A.A. Ajees, D.S. Dheeman, B. Sankaran, P. Kandavelu & B.P. Rosen, (2015). A disulfide-bond cascade mechanism for As(III) S-adenosylmethionine methyltransferase. *Acta Crystallogr. D. Biol. Crystallogr.* **71**: 505-515.
- Marapakala, K., J. Qin & B.P. Rosen, (2012). Identification of catalytic residues in the As(III) S-adenosylmethionine methyltransferase. *Biochemistry* **51**: 944-951.
- Matteson, A. R., Gannon, T. W., Jeffries, M. D., Haines, S., Lewis, D. F. & Polizzotto, M. L. (2014). Arsenic retention in foliage and soil after monosodium methyl arsenate (MSMA) application to turfgrass. *J Environ Qual* **43**: 379-388.
- Mavridou, D. A., Stevens, J. M., Goddard, A. D., Willis, A. C., Ferguson, S. J. & Redfield, C. (2009). Control of periplasmic interdomain thiol:disulfide exchange in the transmembrane oxidoreductase DsbD. *J Biol Chem* **284**: 3219-3226.
- Murphy, J. N. & Saltikov, C. W. (2009). The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. strain ANA-3. *J Bacteriol* **191**: 6722-6731.
- Mukhopadhyay, R., & Rosen, B. P. (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environ Health Perspect*, *110 Suppl 5*: 745-748.
- Nadar, S. V., Yoshinaga, M., Kandavelu, P., Sankaran, B., & Rosen, B. P. (2014). Crystallization and preliminary X-ray crystallographic studies of the ArsI C-Ar lyase from *Thermomonospora curvata*. *Acta Crystallogr F Struct Biol Commun* **70**: 761-764.
- Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C, et al. (2013). The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. *Environ Health Perspect* **121**:295-302.

- Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr.* **57**: 1367-1372.
- Nordstrand, K., Åslund, F., Holmgren, A., Otting, G., & Berndt, K. D. (1999). NMR structure of *Escherichia coli* glutaredoxin 3-glutathione mixed disulfide complex: implications for the enzymatic mechanism. *J Mol Biol* **286**: 541-552.
- Oden, K. L., Gladysheva, T. B., & Rosen, B. P. (1994). Arsenate reduction mediated by the plasmid-encoded ArsC protein is coupled to glutathione. *Mol. Microbiol.* **12**: 301-306
- Ordóñez, E., M. Letek, N. Valbuena, J.A. Gil & L.M. Mateos, (2005). Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. *Appl. Environ. Microbiol.* **71**: 6206-6215.
- Owolabi, J. B. & Rosen, B. P. (1990). Differential mRNA stability controls relative gene expression within the plasmid-encoded arsenical resistance operon. *J Bacteriol* **172**: 2367-2371.
- Paez-Espino, A.D., G. Durante-Rodriguez & V. de Lorenzo, (2014) Functional coexistence of twin arsenic resistance systems in *Pseudomonas putida* KT2440. *Environ Microbiol.* **17**:229-38.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem* **26**: 1781-1802.
- Pierce, B. G., Hourai, Y., & Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* **6**: e24657.
- Pflughoeft, K. J. & Versalovic, J. (2011). Human microbiome in health and disease. *Annu Rev Pathol* **7**: 99-122.
- Qin, J., C.R. Lehr, C. Yuan, X.C. Le, T.R. McDermott & B.P. Rosen, (2009). Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 5213-5217.
- Qin, J., B.P. Rosen, Y. Zhang, G. Wang, S. Franke & C. Rensing, (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 2075-2080.
- Qin, J., Fu, H. L., Ye, J., Bencze, K. Z., Stemmler, T. L., Rawlings, D. E. & Rosen, B. P. (2007). Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. *J Biol Chem* **282**: 34346-34355.

- Raghunathan, S., Chandross, R. J., Kretsinger, R. H., Allison, T. J., Penington, C. J., & Rule, G. S. (1994). Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity. *J Mol Biol* **238**: 815-832.
- Reay, P.F. & C.J. Asher, (1977). Preparation and purification of ⁷⁴As-labeled arsenate and arsenite for use in biological experiments. *Anal. Biochem.* **78**: 557-560.
- Rini, J., Szumlanski, C., Guerciolini, R., and Weinshilboum, R. M. (1990) Human liver nicotinamide N-methyltransferase: ion-pairing radiochemical assay, biochemical properties and individual variation, *Clinica chimica acta; international journal of clinical chemistry* **186**: 359-374.
- Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., . . . Parker, M. W. (1998). Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. *Structure* **6**: 309-322.
- Rossjohn, J., Polekhina, G., Feil, S. C., Allocati, N., Masulli, M., Di Illio, C., & Parker, M. W. (1998). A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure* **6**: 721-734.
- Rosen, B. P., Weigel, U., Karkaria, C., & Gangola, P. (1988). Molecular characterization of an anion pump. The *arsA* gene product is an arsenite(antimonate)-stimulated ATPase. *Journal of Biological Chemistry* **263**: 3067-3070.
- Ryan, D., & Collieran, E. (2002). Arsenical resistance in the IncHI2 plasmids. *Plasmid* **47**: 234-240.
- Saitou, N. & M. Nei, (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.
- Sambrook, J., E.F. Fritsch & T. Maniatis, (1989). *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory, New York.
- San Francisco, M. J., Chen, C. M., & Rosen, B. P. (1988). Identification of the membrane component of the anion pump encoded by the arsenical resistance operon of R-factor R773. *Progress in Clinical and Biological Research* **252**: 311-316.
- San Francisco, M. J., Hope, C. L., Owolabi, J. B., Tisa, L. S., & Rosen, B. P. (1990). Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon. *Nucleic Acids Res* **18**: 619-624.
- Sanders, S. L., Portoso, M., Mata, J., Bahler, J., Allshire, R. C., and Kouzarides, T. (2004) Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage, *Cell* **119**: 603-614.

- Schirrmeister BE, Gugger M, Donoghue PCJ (2015). Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* **58**:769-785.
- Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. (1995) Universal catalytic domain structure of AdoMet-dependent methyltransferases, *Journal of molecular biology* **247**: 16-20.
- Sforna M.C., Philippot P., Somogyi A., Zuilen M.A.V., Medjoubi K., Schoepp-Cothenet B., Nitschke W. and Visscher P.T. 2014. Evidence for arsenic metabolism and cycling by microorganisms 2.7 billion years ago. *Nature Geoscience* **7**: 811-815.
- Shastry BS. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. *J Hum Genet* **52**:871-880.
- Sheng, J., Ye, J., & Rosen, B. P. (2007). Crystallization and preliminary X-ray crystallographic analysis of *Escherichia coli* glutaredoxin 2 in complex with glutathione and of a cysteine-less variant without glutathione. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**: 280-282.
- Shi, J., Vlamis-Gardikas, A., Aslund, F., Holmgren, A., & Rosen, B. P. (1999). Reactivity of glutaredoxins 1, 2, and 3 from *Escherichia coli* shows that glutaredoxin 2 is the primary hydrogen donor to ArsC-catalyzed arsenate reduction. *J Biol Chem* **274**: 36039-36042.
- Shi, W., Wu, J. & Rosen, B. P. (1994). Identification of a putative metal binding site in a new family of metalloregulatory proteins. *J Biol Chem* **269**, 19826-19829.
- States, J. C., Srivastava, S., Chen, Y. & Barchowsky, A. (2009). Arsenic and cardiovascular disease. *Toxicol Sci* **107**: 312-323.
- Stolz, J. F., Basu, P., Santini, J. M. & Oremland, R. S. (2006). Arsenic and selenium in microbial metabolism. *Annu Rev Microbiol* **60**: 107-130.
- Stroher, E., & Millar, A. H. (2012). The biological roles of glutaredoxins. *Biochem J* **446**: 333-348.
- Styblo M, Drobna Z, Jaspers I, Lin S, Thomas DJ. (2002). The role of biomethylation in toxicity and carcinogenicity of arsenic: a research update. *Environ Health Perspect* 110 Suppl **5**:767-771.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725-2729.
- Tars, K., Olin, B., & Mannervik, B. (2010). Structural basis for featuring of steroid isomerase activity in alpha class glutathione transferases. *J Mol Biol* **397**: 332-340.

- Thomas DJ, Rosen BP. (2013). Arsenic methyltransferases. In: Encyclopedia of Metalloproteins, (Kretsinger RH, Uversky VN, Permyakov EA, eds). New York:Springer New York, 138-143.
- Turnbaugh, P. J., Hamady, M., Yatsunencko, T. & other authors (2009). A core gut microbiome in obese and lean twins. *Nature* **457**: 480-484.
- Van Lis, R., Nitschke, W., Duval, S., & Schoepp-Cothenet, B. (2013). Arsenics as bioenergetic substrates. *Biochimica et Biophysica Acta* **1827**: 176-188.
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**: 2184-2195.
- Vahter M. (2002). Mechanisms of arsenic biotransformation. *Toxicology*:211-217.
- Valenzuela OL, Drobna Z, Hernandez-Castellanos E, Sanchez-Pena LC, Garcia-Vargas GG, Borja-Aburto VH, et al. (2009). Association of AS3MT polymorphisms and the risk of premalignant arsenic skin lesions. *Toxicol Appl Pharmacol* **239**:200-207.
- Verma, S., P.K. Verma, A.K. Meher, S. Dwivedi, A.K. Bansiwali, V. Pande, P.K. Srivastava, P.C. Verma, R.D. Tripathi & D. Chakrabarty, (2016). A novel arsenic methyltransferase gene of *Westerdykella aurantiaca* isolated from arsenic contaminated soil: phylogenetic, physiological, and biochemical studies and its role in arsenic bioremediation. *Metallomics* **8**: 344-353.
- Vlami-Gardikas, A., Åslund, F., Spyrou, G., Bergman, T., & Holmgren, A. (1997). Cloning, overexpression, and characterization of glutaredoxin 2, an atypical glutaredoxin from *Escherichia coli*. *J Biol Chem* **272**: 11236-11243.
- Vlami-Gardikas, A., Potamitou, A., Zarivach, R., Hochman, A., & Holmgren, A. (2001). Characterization of *Escherichia coli* null mutants for glutaredoxin 2. *J Biol Chem* **10**: 10861-10868.
- Wang, G., S.P. Kennedy, S. Fasiludeen, C. Rensing & S. DasSarma, (2004) Arsenic resistance in *Halobacterium sp.* strain NRC-1 examined by using an improved gene knockout system. *J. Bacteriol.* **186**: 3187-3194.
- Wang, P.P., G.X. Sun & Y.G. Zhu, (2014). Identification and characterization of arsenite methyltransferase from an archaeon, *Methanosarcina acetivorans* C2A. *Environ Sci Technol* **48**: 12706-12713.
- Waksman, S. A. (1947). What is an antibiotic or an antibiotic substance? *Mycologia* **39**: 565-569.
- Wallden K, Rivera-Calzada A, & Waksman G. (2010). Type IV secretion systems: versatility and diversity in function, *Cell Microbiol.***12**:1203-12.

- Waters, S.B., V. Devesa, M.W. Fricke, J.T. Creed, M. Styblo & D.J. Thomas, (2004). Glutathione modulates recombinant rat arsenic (+3 oxidation state) methyltransferase-catalyzed formation of trimethylarsine oxide and trimethylarsine. *Chem. Res. Toxicol.* **17**: 1621-1629.
- Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, 593-621.
- Weinshilboum, R. M., and Raymond, F. A. (1977) Inheritance of low erythrocyte catechol-o-methyltransferase activity in man, *American journal of human genetics* **29**: 125-135.
- Weinshilboum, R. M., and Sladek, S. L. (1980) Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity, *American journal of human genetics* **32**: 651-662.
- Wilce, M. C., & Parker, M. W. (1994). Structure and function of glutathione S-transferases. *Biochim Biophys Acta* **1205**: 1-18.
- Wlodarski, T., Kutner, J., Towpik, J., Knizewski, L., Rychlewski, L., Kudlicki, A., Rowicka, M., Dziembowski, A., and Ginalski, K. (2011) Comprehensive structural and substrate specificity classification of the *Saccharomyces cerevisiae* methyltransferase, *PLoS one* **6**, e23168.
- Wysocki, R., Bobrowicz, P. & Ulaszewski, S. (1997). The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. *J Biol Chem* **272**: 30061-30066.
- Xia, B., Chung, J., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E., & Dyson, H. J. (1999). Assignment of ¹H, ¹³C, and ¹⁵N resonances of reduced *Escherichia coli* glutaredoxin 2. *J Biomol NMR* **14**: 197-198.
- Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E., & Dyson, H. J. (2001). Solution structure of *Escherichia coli* glutaredoxin-2 shows similarity to mammalian glutathione-S-transferases. *J Mol Biol* **310**: 907-918.
- Xia, T. H., Bushweller, J. H., Sodano, P., Billeter, M., Bjornberg, O., Holmgren, A., & Wuthrich, K. (1992). NMR structure of oxidized *Escherichia coli* glutaredoxin: comparison with reduced *E. coli* glutaredoxin and functionally related proteins. *Protein Sci* **1**: 310-321.
- Xue, X. M., Yan, Y., Xu, H. J., Wang, N., Zhang, X., & Ye, J. (2014). ArsH from *Synechocystis* sp. PCC 6803 reduces chromate and ferric iron. *FEMS Microbiology Letters* **356**: 105-112.
- Yang, J., Rawat, S., Stemmler, T. L. & Rosen, B. P. (2010). Arsenic binding and transfer by the ArsD As(III) metallochaperone. *Biochemistry* **49**: 3658-3666.

- Yang, H. C., Fu, H. L., Lin, Y. F., & Rosen, B. P. (2012). Pathways of arsenic uptake and efflux. *Curr Top Membr* **69**: 325-358.
- Ye J, Rensing C, Rosen BP, Zhu YG.(2012). Arsenic biomethylation by photosynthetic organisms. *Trends Plant Sci* **17**:155-162.
- Ye, J., Yang, H. C., Rosen, B. P., & Bhattacharjee, H. (2007). Crystal structure of the flavoprotein ArsH from *Sinorhizobium meliloti*. *FEBS Letters* **581**: 3996-4000.
- Yin, X.X., J. Chen, J. Qin, G.X. Sun, B.P. Rosen & Y.G. Zhu, (2011) Biotransformation and volatilization of arsenic by three photosynthetic cyanobacteria. *Plant Physiol.* **156**: 1631-1638.
- Yoshinaga, M., Y. Cai & B.P. Rosen, (2011). Demethylation of methylarsonic acid by a microbial community. *Environ Microbiol* **13**: 1205-1215.
- Yoshinaga, M. & Rosen, B. P. (2014). A C-As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. *Proc Natl Acad Sci U S A* **111**: 7701-7706.
- Zhao, F. J., Harris, E., Yan, J., Ma, J., Wu, L., Liu, W., McGrath, S. P., Zhou, J. & Zhu, Y. G. (2013). Arsenic methylation in soils and its relationship with microbial arsM abundance and diversity, and as speciation in rice. *Environ Sci Technol* **47**: 7147-7154.
- Zhu, Y.G., M. Yoshinaga, F.J. Zhao & B.P. Rosen, (2014). Earth abides arsenic biotransformations. *Annu Rev Earth and Planet Sci* **42**: 443-467.

CURRICULUM VITAE

JIAOJIAO LI

EDUCATION AND CONFENCES

2005-2010 M.D Clinical Medicine
Kuming Medical University
Kunming, Yunnan, China

2012-2017 Doctoral Candidate
Florida International University
Miami, Florida, USA

- The 50th Miami Winter Symposium Diabetes: today's research-tomorrow's therapies (January 22, 2017)
- The 5th annual Herbert and Nicole Wertheim Community Healthcare Conference (September 20, 2016)
- The 4th annual international conference on tropical and emerging disease (February 26, 2014)
- The 3rd annual international conference on tropical medicine: breakthroughs & tropical diseases. (February 21-22, 2013)

PUBLICATIONS AND PRESENTATIONS

1. Li, J. J., Packianathan, C., Rossaman, T.G., and Rosen, B.P. Nonsynonymous polymorphisms in the human AS3MT arsenic methylation gene: implications for arsenic toxicity. (*in press* Chemical Research in Toxicology, April 29, 2017). This paper has been chosen for the cover art.
2. Chen, J., Li, J. J., Jiang, X. and Rosen, B.P. Conserved cysteine residues determine substrate specificity in a novel As(III) S-adenosylmethionine methyltransferase from *Aspergillus fumigatus*. *Molecular Microbiology*. 104(2):250-259 (2017).
3. Li, J. J., Pawitwar S.S. and Rosen, B.P. The organoarsenical biocycle and the primordial antibiotic methylarsenite. *Metallomics* 1;8(10):1047-1055 (2016).

4. Li, J. J, Mandal G, and Rosen, B.P. Expression of arsenic resistance genes in the obligate anaerobe *Bacteroides vulgatus* ATCC 8482, a gut microbiome bacterium. *Anaerobe* 39:117-123 (2016).
 5. Ye, J., Nadar. V., Li, J. J. and Rosen, B.P. Structure of *Escherichia coli* Grx2 in complex with glutathione: a dual-function hybrid of glutaredoxin and glutathione S-transferase. *Acta Cryst.* 70: 1907-1913 (2014).
 6. Shashank S Pawitwar; Supurna Dhar; Sneham Tiwari; Chetraj Ojha; Jessica Lapierre; Kyle Martins; Alexandra Rodzinski; Tiyaash Parira; Iru Paudel; JiaoJiao Li; Rajib Dutta; Monica R Silva; Ajeet Kaushik, PhD; Nazira El-Hage, PhD. Developments in animal models towards Zika-virus infection pathogenesis and diagnosis. (Accepted by *Journal of Neuroimmune Pharmacology*, March 25, 2017).
 7. Charles Packianathan, Jiaojiao Li, Kavitha Marapakala, Palani kandavelu, Abdul Ajees Abdul Salam, Banumathi Sankaran and Barry P. Rosen. A novel structure of As(III) S-adenosylmethionine methyltransferase bound toxic methyl arsenite and phenylarsenite. (Manuscript in Preparation)
- 11/15/2016: 2016 HWCOP PhD program open house presentation: "Improve your logical thinking from here".
 - 03/10/2016: 2016 Journal Club presentation: paper "DNA methylation dynamics of the human preimplantation embryo. (Smith ZD, et al. *Nature*, 2014)"
 - 11/05/2016: Physiology and Immunology course assignment: present paper "Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome. (Maurice CF, et al. *Cell* 2012)"
 - 01/14/2015: 2015 Journal Club presentation: paper "Artificial sweeteners induce glucose intolerance by altering the gut microbiota. (Suez J, et al. *Nature* 2014)"
 - 03/02/2014: 2014 Journal Club presentation: present paper "The structure of native influenza virion ribonucleoproteins. (Arranz R, et al. *Science* 2012)"
 - 01/28/2013: Microbiology course assignment: Genetic evolution of HIV: Consequence to drug therapy and vaccine development.