

https://dspace.org/

Arsenic resistance genes in nocardioform bacteria

Leach, Jacobus Gerhardus Johannes

2008-07-10T08:56:56Z
http://hdl.handle.net/10539/5021

Downloaded from a DSpace repository. https://dspace.org/

Arsenic resistance genes in nocardioform bacteria

Jacobus Gerhardus Johannes Leach

submitted in fulfillment of the degree of

Master of Science

to the School of Molecular and Cellular Biology at the

University of Witwatersrand

October 2007

Supervisor: Professor Eric R. Dabbs

Declaration

I declare that this research report is utterly my own, personal, unaided work. It is being submitted for the Masters degree in Science at the University of Witwatersrand, Johannesburg and that it has not been presented for any other degree or examination at another learning institution elsewhere.

.....

J.G.J. Leach

15th of October 2007

Abstract

Arsenic contamination is becoming a global dilemma making it imperative to study natural biomechanisms involved in arsenic detoxification in humans and other living organisms to gain insight in ways how arsenic poisoning can be treated. Arsenic resistance genes are being detected in ever increasing microorganisms with diverse gene organizations. Here, in this report the presence of *ars* genes in *Rhodococcus rhodochrous*, *Arthrobacter oxydans*, and *Gordonia rubropertincta* was investigated. Genomic libraries of these As^R actinomycete species were constructed and then screened for arsenic resistance and unique carbon source utilization genes. More than one hundred positive arsenate tolerant clones were obtained from a partially *Pst*I digested library of *Gordonia rubropertincta* of which ten random clones (pKL1-10) clearly showed the arsenic resistance phenotype in an As^S Gram-positive recipient strain. Nine of these ten clones demonstrated a conserved ∼3.5 kbp DNA fragment with only clone 1 containing a 5100 bp nucleotide segment. Restriction mapping and subcloning was performed on the designated pKL335 and pKL1 recombinant plasmid constructs after which the minimum DNA required for conferring arsenic resistance were sent for sequencing. Only the pKL316 and pKL120 subclones were successfully sequenced as putative secondary structures prevented the sequencing of the rest of pKL335 and pKL1. Four and three putative cistrons were predicted from the edited contiguous sequences of pKL335(ORF1-4) and pKL120(ORF5-7), respectively. Blastn,x,p and CD alignment searches revealed that ORF1 was a putative Acr3-type ArsB homolog, ORF2,3,4 and 7 showed significant homology with Trx-type ArsCs, and ORF5 and 6 are flavin-binding monooxygenase and thioredoxin reductase analogs, respectively. From the alignment data it can be deduced that pKL335 carries an *ars*RBCCC operon structure, such as *C. efficiens* YS-314, and that pKL1 probably comprises a unique *ars*RBCCCTO gene organization. This is only the third report of a thioredoxin reductase and a flavoprotein detected within an arsenic resistant gene cluster. Multiple ArsCs, an ArsT/TrxB and an ArsO/FMO-like homolog seems to be a unique feature amongst high GC-rich Gram(+) bacteria.

Acknowledgements

I would firstly like to express my appreciation to my supervisor Prof. Eric R. Dabbs for allowing me to learn under his supervision. Moreover, I would also want to thank him for obtaining researching funding from NRF for me as well as his minimal approach in advising as it only bolsters my individualistic nature in working independently without much help, salute!!!.

Secondly, I would like to convey my gratitude to all the denizens of GH700 for their support, especially Youtaro, Roland and Jarosite (Yards) with whom I had several interesting conversations with.

Lastly, but most importantly, I want to express my appreciation and thanks to my father and brother Jannie for all their support and encouragement as well as instilling the inquisitive desire for knowledge and to question everything. Therefore, I would like to dedicate this work to them, Thanks Dad !!! :-).

Abbreviations

A	amperes
aa	amino acids
Ac	acetate
Amp	ampicillin
AP	alkaline phosphatase
As	arsenic
ATCC	American type culture collection
BHI	brain heart infusion
bp	base pairs
cfu	colony forming units
Cm	chloramphenicol
CD	conserved domain
Da	dalton
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetic acid
EtBr	ethidium bromide
GL	genomic library
Gly	glycine
hr	hour
HEPES	N-hydroxyethylpiperazine-N-2-ethanesulfonic acid
k	kilo (10^3)
Kan	kanamycin sulphate
L	liter
LA	Luria Bertani agar
LG	low gelling

m	milli (10^{-3})
M	molar
MCS	multiple cloning site
MIC	minimum inhibitory concentration
μ	micro (10^{-6})
min	minute(s)
Nal	nalidixic acid
ND	not determined
nt	nucleotides
Nys	nystatin
OD	optical density
ORF	open reading frame
PEG	polyethylene glycol
RE	restriction endonuclease
Rif	rifampicin
RNA	ribonucleic acid
Rnase	ribonuclease
rpm	revolutions per minute
sec	second(s)
SDS	sodium dodecyl sulphate
sdH_2O	sterile distilled water
Sm	streptomycin
TES	N-tris(hydroxymethyl)-methyl-2-aminoethane sulfonic acid
Tris	Tris(hydroxymethyl)-aminomethane
UV	ultraviolet
V	volt
vol	volume

Contents

Declaration i

Abstract ii

Acknowledgements iii

Abbreviations iv

List of Figures xii

List of Tables xiii

1 Introduction 1

1.1 Historical perspectives . . . 1
1.2 Properties of arsenicals . . . 1
1.3 Applications . . . 3
1.4 Sources, occurrence, and exposure . . . 3
1.5 Distribution and environmental transformation . . . 6
1.6 Toxicology . . . 7
1.6.1 Basics . . . 7
1.6.2 Mode of action . . . 7
1.7 Kinetics and metabolism . . . 7
1.8 Remediation technologies . . . 8
1.8.1 Chemical treatment methods . . . 8
1.8.2 Bioremediation . . . 9
1.9 Uptake systems . . . 10
1.10 Arsenic detoxification mechanisms . . . 10
1.10.1 Efflux-mediated systems . . . 11
1.10.2 Arsenic oxidation and respiration . . . 17

1.10.3 Sequestration 20
1.10.4 Methylation-mediated detoxification 20
1.11 Nocardioform actinomycetes 21
1.11.1 *Rhodococcus* 22
1.11.2 *Gordonia* 23
1.12 *Arthrobacter* 23

Aims and Objectives **24**

2 Materials and Methods **25**
2.1 Maintenance of bacterial specimens and their DNA 25
2.1.1 Culture Maintenance 25
2.1.2 Storing of chromosomal and plasmid DNA 26
2.2 Spot testing 26
2.3 DNA Isolation 26
2.3.1 *E. coli* plasmid DNA isolation 26
2.3.2 Nocardioform DNA isolation 27
2.4 DNA purification techniques 29
2.4.1 Phenol-chloroform extraction 29
2.4.2 Butanol extraction 29
2.5 DNA precipitation methods 29
2.5.1 Salt-Ethanol precipitation 29
2.5.2 Isopropanol-Ethanol precipitation 29
2.5.3 Ethanol precipitation 30
2.6 DNA Manipulations 30
2.6.1 DNA digestions 30
2.6.2 Phosphatase treatment 30
2.6.3 DNA ligation 30
2.7 Agarose gel electrophoresis 31
2.8 DNA extractions methods 31
2.8.1 Low gelling agarose method 31
2.8.2 Freeze-squeeze method [**68**] 32
2.9 Transformation procedures 32
2.9.1 Transformation into *E. coli* 32
2.9.2 Nocardioform transformation 33
2.10 DNA Screening 34

3 Results **36**

3.1 Analysis of *Rhodococcus rhodochrous* for arsenic resistance 36

3.1.1 Transformation of *Sfu*I library of *Rhodococcus rhodochrous* 12674 into an As-susceptible Gram(+) background 36

3.1.2 Screening of *Sfu*I library of *R. rhodochrous* 12674 for arsenic resistance . 37

3.1.3 As^R phenotypic assay of nocardioform strains and libraries 37

3.2 Construction of *Arthrobacter oxydans* C1 library 38

3.2.1 Properties of cloning vectors pDA37 and pDA71 38

3.2.2 Restriction digestion of *A. oxydans* C1 genomic DNA 39

3.2.3 Optimization of the vector to genomic DNA ratio for ligation 40

3.2.4 Transformation of the cloned genomic DNA into *E. coli* 40

3.2.5 Transformation and screening of *Bgl*II and *BamH*I *A. oxydans* C1 libraries in As^S Gram(+) background 42

3.3 Cloning of pDA37 DNA fragments into cloning vector pUC18 43

3.4 Analysis of a *Bgl*II *Gordonia rubropertincta* library for arsenic resistance and other genes. 44

3.4.1 Transformation of *Bgl*II digested *G. rubropertincta* genomic library into an arsenic-sensitive Gram(+) recipient strain 44

3.4.2 Selection for arsenic resistance in *Bgl*II *G. rubropertincta* library. 44

3.4.3 Confirmation of As^R phenotype for the putative clones 45

3.4.4 Searching the *Bgl*II *G. rubropertincta* library for biodegradation and unique carbon source metabolizing determinants 46

3.5 Construction and analysis of a partially *Pst*I digested library of *G. rubropertincta* 47

3.5.1 MIC determination of *G. rubropertincta* and negative controls for As . . . 47

3.5.2 Optimization of the partial *Pst*I digestion conditions for *G. rubropertincta* genomic DNA 48

3.5.3 Optimization of ligation conditions and transformation of pDA71 constructs. 49

3.5.4 Screening of *Pst*I *G. rubropertincta* ATCC 25593 genomic library for unique catabolic genes 50

3.5.5 Selecting for arsenic resistance in *Pst*I *G. rubropertincta* library 51

3.5.6 As^R phenotypic assay of the putative clones in SQ1 52

3.6 Restriction mapping of putative As^R cloned DNA sequences 55

3.6.1 Construction of a partial restriction map of the 3.5 kbp arsenic resistant insert (pKL335) 55

3.6.2 Construction of a partial restriction map of the 5.1 kbp arsenic resistant insert (pKL1) 58

3.6.3 Subcloning of the putative arsenic resistant DNA inserts 59
3.7 Sequence analysis of putative As^R DNA fragments 61
3.7.1 pKL316 sequence analysis 62
3.7.2 pKL120 sequence analysis 74

4 Discussion **86**
4.1 Perspective on the cloning efficiency of nocardioform DNA into vectors 86
4.2 False positive clones obtained during screening of *Bgl*II *G. rubropertincta* library 87
4.3 As^R phenotype comparison of resistant clones and subclones of *G. rubropertincta* 88
4.4 Properties of the partial *ars* operons structures for both pKL316 and pKL120 . . 90
4.5 Arsenic resistance genes 91
4.5.1 Arsenite efflux protein (ArsB) 91
4.5.2 Arsenate reductase (ArsC) 91
4.5.3 Thioredoxin reductase 95
4.5.4 Putative flavoproteins involved in arsenic resistance 96
4.6 Perspective on the failure of sequencing of pKL319 and pKL125 98

5 Conclusion **99**

A Solutions **101**

B Molecular weight markers **108**

C Restriction maps **110**

D Primers **112**

References **113**

List of Figures

1.1 Arsenic geocycle in the environment. 6
1.2 Arsenical detoxification in prokaryotes and eukaryotes. 11
1.3 Schematic representation of *ars* genes found in various microorganisms. 12
1.4 Model of ArsR-ArsD metalloregulatory circuit in *E. coli*. 14
1.5 Proposed reaction mechanisms for Grx/GSH and Trx-dependent ArsCs. 16
1.6 Models of arsenite oxidase and respiratory arsenate reductase. 18
1.7 Arsenite oxidase genes from several microorganisms. 18

3.1 Restriction digestions of *A. oxydans* genomic DNA. 39
3.2 Insert analysis of *Bgl*II *A. oxydans* C1 genomic library clones. 41
3.3 Insert analysis of *BamH*I *A. oxydans* C1 genomic library clones 41
3.4 *Pst*I digestion of pDA37. 43
3.5 Putative As^R clones from the screening of *Bgl*II digested *G. rubropertincta* ATCC 25593 genomic library in SQ1. 45
3.6 Optimization of the partial *Pst*I digestion of *G. rubropertincta* genomic DNA. . . 49
3.7 Insert analysis of partially *Pst*I digested *G. rubropertincta* genomic library clones. 50
3.8 Positive As^R clones from screening of the partially *Pst*I digested *G. rubropertincta* library in SQ1. 51
3.9 Insert analysis of arsenate clones from *Pst*I *G. rubropertincta* library. 52
3.10 Characterization of the As^R phenotype for the selected positive clones in SQ1. . 54
3.11 Single digestions of 3.5kbp arsenic resistant DNA fragment. 56
3.12 Single and double digestions of 3.5kbp As^R DNA fragment. 57
3.13 Preliminary restriction map of 3.5kbp fragment of arsenic resistant clone 3. . . . 57
3.14 Single digestions of 5.1kbp arsenic resistant DNA fragment. 58
3.15 Preliminary restriction map of 5.1kbp insert of arsenic resistant clone 1. 59
3.16 DNA sequence of pKL316. 62
3.17 ORF prediction and restriction map of pKL316 63
3.18 Blastn data of some significant alignment matches to pKL316 nucleotide sequence. 63

3.19 Nucleotide alignment of ORF1 and ORF2 with *R. erythropolis* plasmid pBD2. . 64
3.20 Nucleotide alignment of ORF3 and ORF4 with *R. erythropolis* plasmid pBD2 . . 65
3.21 Blastx data of significant alignment matches to the translated sequence of pKL316. 66
3.22 Translated protein alignment of ORF2 with *R. erythropolis* pBD2. 66
3.23 Translated protein alignment of ORF3 and ORF4 with *R. erythropolis* pBD2 . . 67
3.24 Translated protein alignment of ORF1 with *M. tuberculosis* 67
3.25 Translated protein alignment of ORF2,3 and 4 with *M. tuberculosis* 68
3.26 Putative conserved domain(s) found within ORF1. 68
3.27 Blastp data of some significant alignment matches to predicted sequence of ORF1 69
3.28 Protein alignments of ORF1 with some relevant matches. 69
3.29 Putative conserved domain(s) found within ORF2. 70
3.30 Blastp data of some significant alignment matches to predicted sequence of ORF2. 70
3.31 Protein alignment of ORF2 with *R. erythropolis* plasmid pBD2. 70
3.32 Protein alignments of ORF2 with *Arthrobacter aurescens* TC1 plasmid pTC2 and *Corynebacterium glutamicum* ATCC 13032. 71
3.33 Putative conserved domain(s) found within ORF3. 71
3.34 Blastp data of some significant alignment matches to predicted sequence of ORF3. 72
3.35 Protein alignments of ORF3 with some relevant matches. 72
3.36 Putative conserved domain(s) found within ORF4. 73
3.37 Blastp data of some significant alignment matches to predicted sequence of ORF4. 73
3.38 Protein alignment of ORF4 with *C. efficiens* YS-314. 73
3.39 Protein alignments of ORF4 with *Nocardia farcinica* 10152 plasmid pNF2 and *Rhodococcus erythropolis* plasmid pBD2. 74
3.40 DNA sequence of pKL120. 75
3.41 ORF prediction and restriction map of pKL120 76
3.42 Blastn data of some significant alignment matches to pKL120 nucleotide sequence. 76
3.43 Nucleotide alignment of ORF5 and ORF7 with *R. erythropolis* PR4 plasmid pREL1 77
3.44 Nucleotide alignment of ORF6 with *R. erythropolis* PR4 plasmid pREL1. 78
3.45 Blastx data of some significant alignment matches to the translated sequence of pKL120. 79
3.46 Translated protein alignment of ORF5 with *R. erythropolis* PR4 plasmid pREL1. 79
3.47 Translated protein alignment of ORF6 with *R. erythropolis* PR4 plasmid pREL1. 80
3.48 Putative conserved domain(s) found within ORF5. 80
3.49 Blastp data of some significant alignment matches to predicted sequence of ORF5 81
3.50 Protein alignments of ORF5 with *Rhodococcus erythropolis* PR4 plasmid pREL1 and *Streptomyces* sp. FR-008 plasmid pHZ227. 81

3.51 Protein alignment of ORF5 with *C. glutamicum* ATCC 25593. 82
3.52 Putative conserved domain(s) found within ORF6. 82
3.53 Blastp data of some significant alignment matches to predicted sequence of ORF6 83
3.54 Protein alignments of ORF6 with *Rhodococcus erythropolis* plasmid pBD2 and *Rhodococcus erythropolis* PR4 plasmid pREL1. 83
3.55 Protein alignment of ORF6 with *Streptomyces* sp. str. FR-008 plasmid pHZ227. 84
3.56 Blastp data of some significant alignment matches to predicted sequence of ORF7. 84
3.57 Protein alignments of ORF7 with some relevant matches. 85

4.1 Partial ArsB protein sequence alignment from *G. rubropertincta* and other diverse microorganisms. 91
4.2 Alignment of the predicted putative arsenate reductase protein sequences detected in the *G. rubropertincta* arsenic resistant clones. 92
4.3 Protein sequence comparison of ArsCs from *G. rubropertincta* and several selected actinomycetes species. 94
4.4 Proposed energy coupling of Gram(+) and Gram(-) bacterial arsenate reductases. 95

5.1 Arsenic gene organization prediction within pKL335 and pKL1 plasmid constructs.100

B.1 Molecular weights of λII/III DNA markers 108
B.2 Molecular weights of Generuler DNA Ladder mix marker 109
B.3 Molecular weights of Generuler 1kbp DNA Ladder Plus marker 109

C.1 Partial restriction map of *E. coli-Rhodococcus* shuttle vector pDA71 110
C.2 Partial restriction map of *E. coli-Rhodococcus* shuttle vector pDA37 111
C.3 Partial restriction map of cloning vector pUC18 111

D.1 Compilation of all the sequencing primers used 112

List of Tables

1.1 Common natural and synthetic arsenical compounds. 2
1.2 Uses and application of arsenical compounds. 4
1.3 Significant events during the historical use of arsenic as a therapeutic agent. . . . 5
1.4 Reactions catalyzed by As-metabolizing bacteria. 20

2.1 Bacterial strains used in study 25

3.1 Spot test results of nocardioform strains and libraries for arsenic resistance. . . . 37
3.2 Averaged results of screening of the *Bgl*II digested *G. rubropertincta* ATCC 25593 genomic library in SQ1 for arsenic tolerance. 45
3.3 Preliminary carbon source utilization data for *G. rubropertincta* ATCC 25593 and *R. erythropolis* ATCC 4277. 46
3.4 Arsenic resistance profiles of *G. rubropertincta* ATCC 25993 and the Gram(-) and Gram(+) recipient strains. 48
3.5 Results of screening of the partially *Pst*I digested *G. rubropertincta* ATCC 25593 genomic library in SQ1 for arsenic resistance. 52
3.6 Characterization of the arsenic sensitivity of the selected resistant clones in SQ1. 53
3.7 Characterization of the As-sensitivity of pKL1, pKL335 and their subclones in SQ1. 60
3.8 Internet program utilized for sequence analysis. 61
3.9 Summary of ORF prediction of pKL316 62
3.10 Summary of ORF prediction of pKL316 75

4.1 Summary of protein alignments of the predicted ORFs with similar gene products from selected nocardioform bacteria and related species 97

A.1 Table of stock solutions and their concentrations. 101
A.2 Table of antibiotic stock solutions and their concentrations. 101

Chapter 1

Introduction

1.1 Historical perspectives

The word "Arsenic" is borrowed from the Persian word Zarnik, meaning "yellow orpiment", which was derived from Greek arsenikon meaning "potent" [**4, 53**]. The presence of arsenic in nature was recognized since ancient times where Aristotle made first reference to sandaling (arsenic trisulphide) in the 4th century BC [**9**]. Albertus Magnus is credited to be the discoverer of metallic arsenic in 1250 AD, but it was not until 1649 that Johann Schroeder clearly reported the preparation of metallic arsenic by reducing arsenic trioxide with charcoal [**4, 6, 9, 117**].

1.2 Properties of arsenicals

Elemental Arsenic (As) is a member of Group V of the periodic table with an atomic mass of 74.91 amu. Arsenic occurs in several allotropic modifications - yellow (As_4), amorphous black (As_n), and grey metallic (As) with specific gravities of 1.97, ∼4.7, and 5.73, respectively [**5, 6**]. Grey arsenic is very brittle, crystalline, semi-metallic solid that tarnishes in air, and when heated is readily oxidized to arsenious oxide with the odour of garlic [**6**]. Common valency states of arsenic in natural environments include -3, 0, +3, and +5. Even though elemental arsenic is immiscible in water, its salts exhibit a broad range of solubilities depending on pH and ionic strength of the surrounding area. Inorganic arsenic usually exist naturally as arsenite (As^{3+}) or arsenate (As^{5+}) [**1**]. Aqueous arsenate species are ionized with three pKa's of 2.2, 7.0 and 11.5 (comparable to 2.1, 7.2, and 12.7 for phosphate).

$$H_3AsO_4 \rightarrow H_2AsO_4^- \rightarrow HAsO_4^{2-} \rightarrow AsO_4^{3-}$$

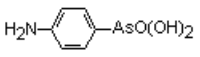
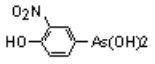
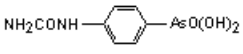
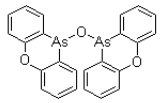
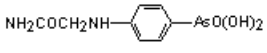
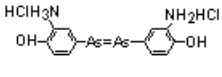
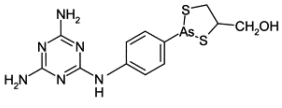
Name	Alternative name	Formula
Inorganic As		
As(III) sulfide	Arsenic trisulfide, Orpiment, Auripigment	As_2S_3
As(III) oxide	Arsenic trioxide, Arsenolite	As_2O_3
Arsenenous acid	m-Arsenious acid	$HAsO_2$
Arsenous acid	o-Arsenious acid	$As(OH)_3$
As(V) oxide	Arsenic pentoxide	As_2O_5
Arsenic acid	o-Arsenic acid	H_3AsO_4
Arsine		AsH_3
Organic As		
Methylarsines		$(CH_3)_{3-x}AsH_x$
Methylarsonic acid	MMA^{III}	$CH_3As(OH)_2$
Methylarsinic acid	MMA^{V}	$CH_3As{=}O(OH)_2$
Dimethylarsonous acid	DMA^{III}	$(CH_3)_2AsOH$
Dimethylarsinous acid	DMA^{V}, Cacodylic acid	$(CH_3)_2As{=}O(OH)$
Trimethylarsine oxide	$TMAO^{V}$	$(CH_3)_3As{=}O$
Arsenobetaine		$(CH_3)_3As^+CH_2COO^-$
Arsenocholine		$(CH_3)_3As^+CH_2CH_2OH$
Synthetic As compounds		
2-Chlorovinyldichloroarsine	Lewisite	$ClCH{=}CHAsCl_2$
Potassium arsenite		K_3AsO_3
Copper arsenite	Scheele's Green	$CuHAsO_3$
Copper acetoarsenite	Paris Green	$3CuO.As_2O_3.Cu[OOCCH_3]_2$
Copper chrome arsenate	CCA	$CrO_3.CuO.As_2O_5$
(4-Amimophenyl)-arsonic acid	p-Arsanilic acid, p-Aminobenzene-arsonic acid	
3-Nitro-4-hydroxy -phenylarsonic acid	Roxarsone	
[4-[Aminocarbonyl-amino] phenyl]-arsonic acid	Carbarsone, N-carbamoyl-arsonic acid	
10,10-Oxybisphenoxarsine	OBPA, Thenarsazine Diphenoxarsin-10-yl oxide	
[4-[(2-Amino-2-oxoethyl)amino]- phenyl]-arsonic acid	Tryparsamide	
4,4-arsenobis(2-aminophenol) dihydrochloride	Arsphenamine, Salvarsan 606	
2-[4-(4,6-Diamino-1,3,5-triazin -2-yl)amino]phenyl-1,3,2 -dithiarsolane-4-methanol	Melarsoprol	

Table 1.1: Common natural and synthetic arsenical compounds (Compiled from **[1,3,4,20,144]**)

In contrast, arsenite appears mostly unionized as $As(OH)_3$ at neutral pH with a pKa value of 9.3 **[32, 100]**.

$$\begin{array}{ccccccc} H_3AsO_3 & \rightarrow & H_2AsO_3^- & \rightarrow & HAsO_3^{2-} & \rightarrow & AsO_3^{3-} \\ & & HAsO_2^- & \rightarrow & AsO_2^- & & \end{array}$$

1.3 Applications

The infamous reputation of arsenic compounds as a poison has been recognized since antiquity **[53]**. It was historically used as a homicidal agent by discretely disposing of the old owner of the land, and was therefore referred to as "inheritance powder" **[4, 8, 53, 104]**. Most of the artificially obtained arsenic is used in agriculture and forestry for the production of pesticides, weed killers, and wood preservatives **[1–3, 32, 39]**. Smaller quantities of arsenic are utilized in glass manufacturing and applied as a metallurgical additive metal to non-ferrous alloys and feed supplement to poultry and swine **[1–3, 6, 8, 9, 39, 107, 139]**. Highly pure arsenic is used in the electronics industry for the production of semiconductor material and devices **[4, 6, 8, 9]**. Arsenic has also found applications in the preparations of dyes and pigments, poisonous gases, pyrotechnics, as catalyst and preservative in tanning **[2, 5, 6, 53]**.
Despite arsenic being synonymous to poison, it has been used as a therapeutic agent for over 2400 years. Through the 18th to the 19th century, physicians prescribed a variety of arsenide and arsenic salt preparations as antiperiodics, antipyretics, antiseptics, antispasmodics, caustics, cholagogues, hematinics, sedatives, and tonics **[96, 157]**. In 1786, Thomas Fowler prepared a 1% w/v potassium arsenite (K_3AsO_3) solution that was subsequently used to treat numerous conditions for over 150 years **[8, 96, 157]**. The organoarsenical salvarsan 606 (arsphenamine) discovered by Paul Ehrlich in 1908 was the standard therapy for syphilis for about 40 years and is still presently used for treating trypanosomid infections **[8, 96, 100, 116, 117, 157]**. Recently, renewed interest in arsenic as a medicine came about with the use of arsenic trioxide (As_2O_3) for the treatment of acute promyelocytic leukemia **[3, 96]** .

1.4 Sources, occurrence, and exposure

Arsenic is naturally found in igneous and sedimentary rocks, fossil fuels, and over 200 mineral species **[1, 78, 138, 144]**. These As-bearing minerals are commonly associated with copper, nickel, and iron and usually occur as sulfides (realgar (As_2S_2) and orpiment (As_2S_3)), sulfoarsenides (arsenopyrite (FeAsS) and enargite (Cu_3AsS_4)), arsenides (niccolite (NiAs)), arsenates (erythrite ($CO_3(AsO_4)_2.8H_2O$)), and as oxides (arsenolite (As_2O_3)) in nature **[1, 9, 32, 144]**.

Industry	Use	Arsenical and its purpose	
Forestry	Wood preservation	CCA, Na_3AsO_4	- Wood preservatives
Agriculture	Pesticides	MSMA, DSMA	- Herbicides
		$PbHAsO_4$, $CaHAsO_4$	- Insecticides
		Na_3AsO_3, OBPA	- Fungicides
		DMA^V	- Defoliant
Farming	Feed additives	Roxarsone, p-Arsanilic acid	- Promote growth and improve pigmentation in poultry
			- Intestinal palliative for swine
			- Prevent coccidiosis and control swine dysentery
		Carbarsone, Nitarsone	- Antihistomonad in turkey
Metallurgy	Pb-containing alloys	As_2O_3, As_2O_5	- Anti-friction agent in lead bearings
			- Strengthen lead acid battery grids
			- Improve sphericity of leadshot
			- Special solders
	Cu-containing alloys	As_2O_3, As_2O_5	- Bronzing
			- Improve corrosion resistance
			- Increases recrystallization temperature of Cu
Electronics	Semiconductors	Ultrapure As	- Doping agent in Ge/Si devices: transistors
			- GeAs: LED, Laser diodes and solar cells
			- InAs: IR detectors, lasers, and Hall effect applications
Glass	Glassware	As_2O_3	- Decolorizer and fining agent
			- Low melting glass
Other	Dyes and pigments	Paris green, Scheele's green	- Wallpapers, printing, and colouring agent for foodstuffs
		As_2S_2, As_2S_3	- Pigments for wall and ornamental paint
	Poisons	Lewisite, AsH_3 and derivatives	- Chemical warfare gas
		Paris green, Scheele's green, H_3AsO_3	- Rat poisons
	Pyrotechnics	As_2S_2	- Pyrotechnic (fireworks) mixtures

Table 1.2: Uses and application of arsenical compounds (Compiled from **[1, 4, 8, 9, 53, 107, 144]**).

Year	Main events
400 BC	Hippocrates used realgar(As_2S_2) and orpiment (As_2S_3) as remedies for ulcers
50 AD	Dioscorides utilized orpiment as a depilatory
14th century	Angelus Salva applied arsenic against the plaque
16th century	Jean de Gorris used arsenic as a sudorific
17th century	Lentilius and Friceius utilized arsenic as a anti-malarial agent
18th century	Numerous arsenical preparations were used therapeutically, such as Aiken's Tonic Pills, Andrew's Tonic, Arsenauro, Gross' Neuralgia Pills, Cholor-Phosphide of Arsenic, Sulphur Compound Lozenges etc. with some still in circulation at the end of the 19th century
19th century	Arsenic was a mainstay of materia medica Fowler's solution (K_3AsO_3) became renowned for its success in treating asthma, chorea, eczema, Hodgkin's disease, pemphigus, pernicious anemia, and psoriasis etc.
20th century	Ehrlich discovered an organic arsenical (salvarsan) that remedied syphilis and was used to treat trypanosomiasis Chinese utilized arsenic for hundreds of years in traditional medicine and derivatives are still applied to devitalize the pulp of diseased teeth as well as in regimens for psoriasis, rheumatic diseases, and syphilis
Present	Arsenic trioxide is being used as a chemotherapeutic agent for the treatment of acute promyelocytic leukemia

Table 1.3: Significant events during the historical use of arsenic as a therapeutic agent (Adapted from **[96]**).

Alkylated arsenic compounds such as methylarsonic (MMA^{III}), methylarsinic (MMA^{V}), and dimethylarsinous acids (DMA^{V}) as well as methyl arsines and other organoarsenicals also exist in ecological systems due to biological activity **[1, 32, 107]**. Volcanic action is the main natural source of arsenic, but weathering of mineral deposits, low-temperature volatilization, and hydrothermal activity also contribute to the availability of this element to surroundings **[1, 32, 39, 78]**.

Most of the commercially available arsenic is obtained as a by-product from the smelting of copper, lead, cobalt, and gold ore **[1, 3]**. Anthropogenic sources such as smelting slag, flue gas and fly ash from coal and fossil fuel combustion, runoff from mine tailings, and the past agricultural use of pesticides etc. contribute significantly to As contamination of air, water, and soil **[1,32,100,107,138]**. Evidently, working in several industries poses the greatest potential of arsenic exposure, while the ingestion of foodstuffs and water are the primary non-occupational means that humans are exposed to it in the environment **[1, 10, 55, 116]**.

1.5 Distribution and environmental transformation

Even though arsenic is only the twentieth most abundant element with a crustal abundance of about 1.8 ppm, its toxicity is problematic in many parts of the world [**39**]. Arsenic occurs in soil at concentrations ranging from 0.1 - 1000ppm, while that in atmospheric dust it is around 50 - 400ppm. The average arsenic levels in seawater and freshwater are about 2.6ppb and 0.4ppb, respectively .

Marine organisms usually contain arsenic residues ranging from 1 - >100ppm, chiefly as organic species such as arsenosugars (macroalgae) and arsenobetaine (marine invertebrates and fish), whereas concentrations in freshwater and terrestrial biota is typically less than 1ppm (fresh weight) [**1, 100, 144**].

A wide variety of diverse chemical and biological reactions, which include wet and dry deposition, adsorption, precipitation, natural weathering, oxidation, reduction, methylation and demethylation, dissolution, and volatilization all participate actively in the biogeocycling of this element [**1, 107, 109**]. Mobilization and speciation of arsenical compounds depends mainly on the pH, redox potential (E_h), and native biota within the soil-water interface of a particular environment [**1**].

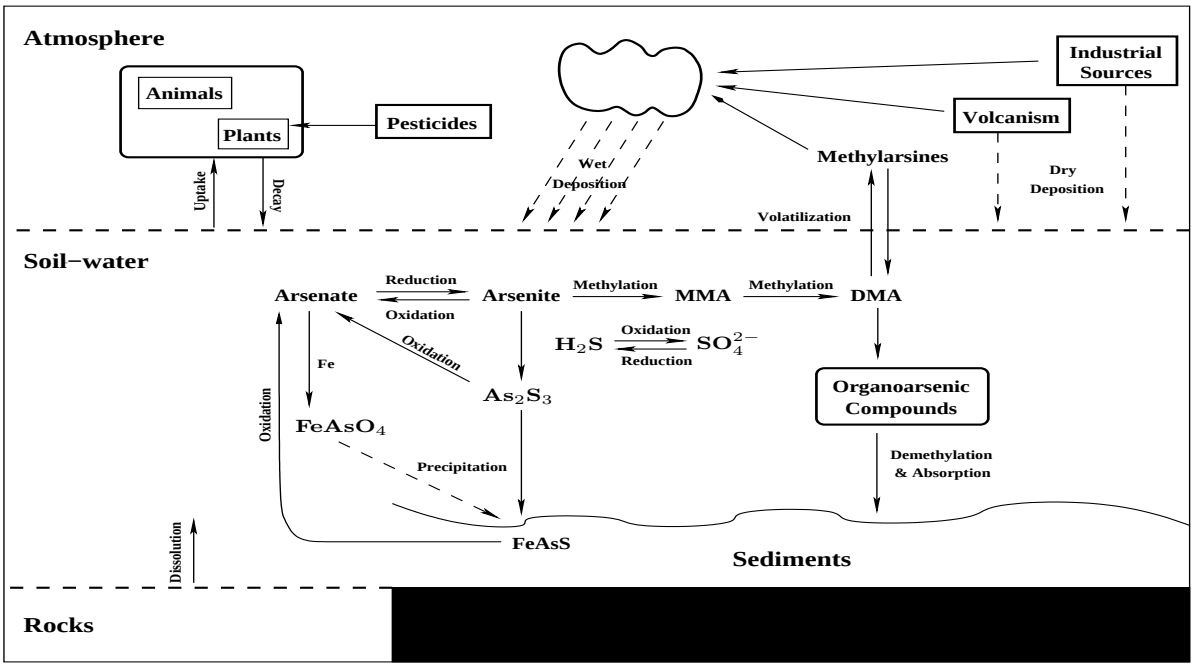


Figure 1.1: Arsenic geocycle in the environment (Adapted from [**109**]).

Although, both arsenate and arsenite are adsorbed to the surface of several iron and aluminium minerals, arsenite adsorbs less strongly making it the more mobile and bioavailable form [**1,107**]. Thus, the amount of iron, aluminium, and manganese oxides/hydroxides as well as organic content clearly affects the mobilization of different arsenical species by influencing the sorption capacity of arsenic in a particular soil or sediment type [**1, 138**]. The temperature, $[PO_4^{3-}]$, as well as the charge on the mineral surface can substantially alter the adsorption efficiency for arsenic by a certain soil [**1**].

1.6 Toxicology

1.6.1 Basics

Living organisms demonstrate a wide variety of sensitivities to different arsenic species [**1**]. Cellular toxicity of arsenical compounds depends on the form, i.e. inorganic or organic, oxidation state, as well as on the physical and chemical properties, such as solubility and stability of a particular compound [**1, 7, 32, 144**]. Generally, inorganic arsenicals are more toxic than organoarsenic compounds to biota, where the trivalent species are usually more harmful than the pentavalent species for both inorganic and organic arsenic forms [**1, 7, 32**].

1.6.2 Mode of action

Due to the structural similarity of arsenate (AsO_4^{3-}) to phosphate (PO_4^{3-}), its main toxicity stems from its ability to substitute for this essential bioelement in enzymatic reactions and cellular transport [**1, 3, 12, 66, 104, 107, 139, 145**]. Arsenolysis is the process where arsenate uncouples phosphorylation activities by forming unstable esters that spontaneously hydrolyze and thereby short circuits the ATP-generating system [**12, 66, 107, 145**].
Acting as a soft metal ion, arsenite (AsO_2^-) readily forms strong covalent bonds with thiolates of cysteine residues and imidazolium nitrogens of histidines [**66, 118, 119, 145**]. Since thiols and vicinal sulfhydryls often constitute an essential part of the active site of proteins, the binding of trivalent arsenic has the potential of inhibiting numerous enzymes [**66, 145**].

1.7 Kinetics and metabolism

Biotransformation can chiefly be classified in three modes, namely; redox reactions, methylation, and biosynthesis of organoarsenic compounds. Methylated derivatives and organic arsenical species are subject to microbial demethylation and degradation back to IAs forms [**1**].

Inorganic arsenic, especially IAs^{III}, can also be sequestered via conjugation to glutathione and related thiol compounds, chelation by small metal-binding proteins, as well as non-specific binding to proteins in various organisms [**13, 32, 145, 147**].
Even though arsenic metabolism often involves methyl transfer reactions, the principal end product in addition to the rate and extent of methylation may differ substantially among all living organisms. Biotransformation of arsenic in bacteria illustrate an immense diversity in metabolic end points. Certain microorganisms are able to metabolize arsenical compounds through diverse pathways and the importance of the individual transformation reactions are highly dependent on the environmental conditions.
It can therefore clearly be seen that the bioconversion reactions of arsenic are characterized by significant qualitative and quantitative interspecies variation. Humans are a prime example of this metabolic variation where the average relative distribution of urinary arsenic metabolites are between 10-20% IAs, 10-20% MMA^{III+V}, and 60-80% DMA^{III+V} [**152**].

1.8 Remediation technologies

The remediation procedures for the removal of metals from soil employ a number of different approaches, including extraction, immobilization, toxicity reduction, physical separation, and isolation. The site characteristics, concentration, type of pollutants to be removed, and the end use of the contaminated medium should all be taken into account when selecting the most suitable soil and sediment remediation method [**102**].

1.8.1 Chemical treatment methods

Various chemical treatment methods, such as coagulation/filtration, ion exchange, adsorption onto activated carbon or alumina, lime softening, and reverse osmosis have been applied for the partial removal of arsenic from waters [**74, 109, 175**]. Because most chemical treatment technologies are ineffective in the removal of As(III), a pre-oxidation step is often required to convert it from the trivalent state to pentavalent form. Although the addition of chemicals such as potassium permanganate, chlorine, ozone, and hydrogen peroxide effectively oxidizes trivalent arsenic, they could cause several secondary problems which will only push up the operational cost [**74, 175**]. Vitrification and electrokinetic processes have been utilized on large scale for arsenic polluted soils, while pyrometallurgical separation is another feasible method for arsenic removal but may require a pretreatment step [**102**].

1.8.2 Bioremediation

Biochemical treatments such as bioleaching and phytoremediation involves the use of microorganisms to reduce, oxidize, or eliminate contaminants, either alone or in combination with other physicochemical technologies [**175**]. With the increasing importance of heavy metals as pollutants, biological treatments offers advantages of operational simplicity, absence of chemicals, environmentally friendly and more economical [**115**]. Bioleaching involves the utilization of iron- and sulfur-oxidizing microorganisms, mainly *Acidobacilli* and *Leptospirilium* sp., in order to solubilize metal complexes under acidic conditions from soil, rock, and minerals which can subsequently be extracted and removed [**102**]. Gold recovery from refractory ore has been significantly enhanced by introducing an initial biooxidation step before cyanidation or thiourea chemical leaching [**45, 79**]. Biooxidation of As(III) not only reduces its toxicity, but it also facilitates its removal since pentavalent species adsorb more strongly to iron oxides.

Zouboulis & Katsoyianis [**175**] realized that the bacterial oxidation of dissolved iron by microbes, such as *Gallionella ferruginea* and *Leptothrix ochracea*, from natural water sources is an effective strategy for the simultaneous removal of arsenic. The biooxdation of iron primarily produces a mixture of poorly defined iron oxides that often contains considerable quantities of organic matter. Arsenic, chiefly as As^{3+}, is removed by direct adsorption or co-precipitation on the deposited biogenic iron oxides. Since the elimination of trivalent arsenic is inefficient, the nonreacted As(III) is biologically oxidized by the same microorganisms to As(V), which are consequently adsorbed onto the preformed iron oxides, promoting the efficient overall removal of arsenic.

The sorption of metals into biomass of algae or bacteria whether viable or not is another possible biological method for eradicating arsenical pollutants in water [**102**]. Phytoremediation through the use of naturally "hyperaccumulating" plants, which include the Chinese brane fern (*Pteris vittata L*) and Indian mustard (*Brassica juncea*), is a promising approach for removal of various metals from soil [**102, 149**]. Recently, the application of bacterial arsenate reductase (ArsC) and overexpression of γ- glutamylcysteine synthetase (γ-ECS) in transgenic plants for potential phytoremediation via intracellular sequestration after reduction of As^{5+} to As^{3+} were reported [**48, 81, 93**].

1.9 Uptake systems

Because AsO_4^{3-} is a structural analog of PO_4^{3-}, it is taken up by phosphate transport systems in bacteria and yeast and even in other organisms [**27, 54, 100, 118, 145**]. *E. coli* accumulate arsenate through two distinct pathways; a fast constitutive and unspecific Pit system or an inducible Pst system that comprises of a translocating ABC-type ATPase complex [**54**]. Willsky and Malamy [**159**] demonstrated that *pit* mutants that rely on the Pst system accumulate only small amounts of arsenate and exhibit an increased tolerance to arsenic. In yeast, arsenate crosses the cell membrane through the Pi transporter Pho84 in association with accessory proteins Pho87 and Pho88 [**27**]. Mutations in these proteins together with mutants of Pho86, that mediates the translocation of Pho84 from the endoplasmic reticulum to the Golgi, display the arsenate resistance phenotype [**27**]

As(III) enters the cell at physiological relevant pH via aquaglyceroporins in bacteria [**94, 126**], yeast [**166**], and mammals [**87**], as $As(OH)_3$, which resembles the inorganic equivalent of nonionized glycerol. It has been shown that As(III) uptake in *E. coli* is facilitated by the glyceroporin membrane channel GlpF and the disruption of the gene confers resistance to both arsenite and antimonite [**94, 126**]. The Fps1p gene product was identified as a glyceroporin homolog of GlpF that participates in the active accumulation of trivalent arsenic into the cells of *S. cerevisiae*. Deletions of the *fps1* gene result in reduced accretion of arsenite and consequently leads to enhanced tolerance to arsenite [**56, 165**]. Recently, two closely related mammalian homologs AQP7 and AQP9 were shown to be involved in arsenite uptake and could functionally complement for Fps1p in an *fps1*Δ yeast strain [**87**].

1.10 Arsenic detoxification mechanisms

Because living organisms are continuously exposed to heavy metals and metalloids, which are frequently toxic, they have evolved resistance mechanisms to tolerate these elements [**120**]. Various detoxification strategies are employed for arsenicals among the different types of living species, which includes active efflux, oxidation, sequestration, and possibly methylation-related reactions. Reduction reactions are often associated with certain resistance systems to mediate the rapid excretion of arsenical compounds or coupled to other metabolic processes (e.g. energy production, methylation, and chelation) for ultimately detoxification purposes. Several distinct organisms possess more than one tolerance mechanisms or sometimes multiple sets of resistance genes from one or more sources.

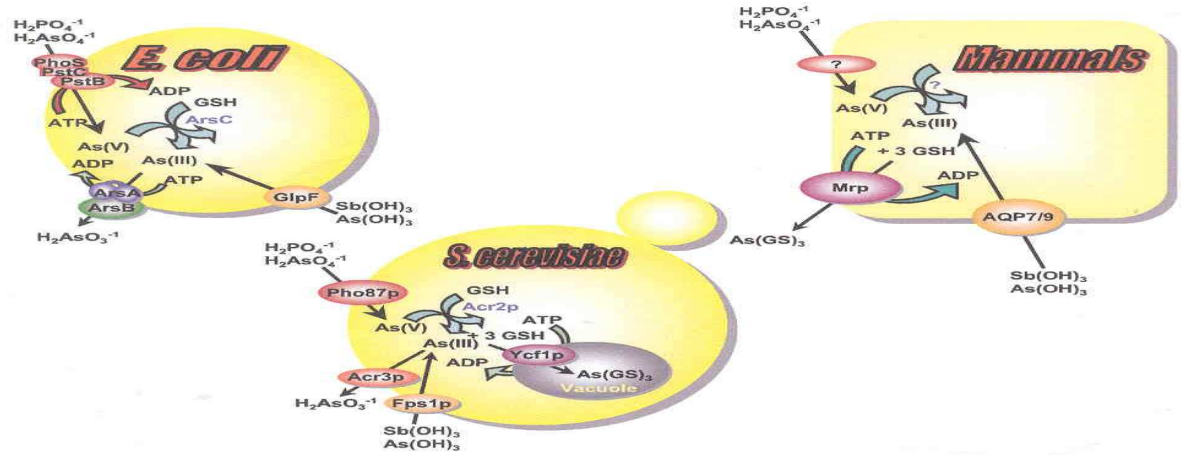


Figure 1.2: Arsenical detoxification in prokaryotes and eukaryotes (From [**118**]).

1.10.1 Efflux-mediated systems

Efflux pumps are currently the most predominant known group of resistance systems that can be either ATPases or chemiosmotic [**136,137**]. The best characterized metalloid resistance pathway is the ArsC- related extrusion system encoded by the *ars* operon found in the chromosomes and on plasmids of bacteria as well as similar resistance determinants in the genomes of some yeast, archaeal and protozoan species [**107, 117, 139**]. Fundamentally, arsenical resistance in these organisms utilizes basically the same genes and encoded biochemical mechanism even though the number of genes and functional details can differ from each other [**134, 136**].
This is illustrated by the fact that the respective Ars proteins for the various species perform the same basic role, while they only have moderate to poor sequence homology and amino acid identities. Plasmid-borne systems confer higher levels of arsenic resistance than those coded for chromosomally [**31**].

Gene organization and function of ars determinants

It is becoming increasingly clear that the *ars* operon is widely found within the chromosome and on transmissible plasmids from both Gram (+) and Gram (-) microorganisms. Indeed it seems that the *ars* operon is more abundant in newly sequenced genomes larger than 2Mb than the genes for tryptophan synthesis [**135, 137**].
The *ars* operons can be mainly grouped into three prototypes, the three-, four-, and five gene arsenic resistance determinant systems, even though unique *ars*-like operons have been reported [**123**].



Figure 1.3: Schematic representation of *ars* genes in *M. tuberculosis* chromosome (NC_000962.2, [**129**]), *B. subtilis* chromosome [**129**], *Staphylococcal* plasmids pI258 [**71**] and pSX267 [**122**], *E. coli* K12 (NC_000913, [**49**]), *E. coli* plasmids R773 [**35**] and R46 [**26**], *A. multivorum* AIU plasmid pKW301 [**143**], *Y. enterocolita* plasmid pYV [**103**], *S. marcescens* IncHI2 plasmid R476 [**123**], *A. ferrooxidans* 33020 chromosome [**29**], *Synechocystis* str. PCC6803 chromosome [**88**], *P. putida* KT2440 chromosome (NC_002947), *C. glutamicum* chromosome (NC_003450, [**106**]), *Sinorhizobium meliloti* Rm1021 chromosome [**168**], *Shewanella* sp. ANA-3 chromosome [**124**], *Halobacterium* sp. megaplasmid pNRC100 [**155**], *F. acidarmanus* chromosome [**58**], *S. cerevisiae* and *douglasii* chromosomes [**24, 90**].

The *ars* operons normally consist of at least the minimal set of genes needed for arsenical resistance known as the *ars*RBC genes arranged into a single polycistronic unit [**137, 150**]. The *ars*R,B, and C genes encodes a metalloid inducible repressor [**162**], a transmembrane arsenite efflux pump [**161**], and an arsenate reductase [**70**], respectively. Occasionally two additional genes, *ars*A and *ars*D are also found in the operons of Gram- negative bacteria. ArsA is a metalloid responsive ATPase [**121**] and ArsD a second regulator [**163**]. Recently, another gene, ArsH, have been found in several bacteria with still unknown function [**28, 88, 106, 123, 155, 168**] and ArsM, a putative As (III) methyltransferase [**155**] has been established in the archaeon, *Halobacterium*. More recently, AqpS, a putative As(III)- antiporter that belongs to the aquaglyceroporin family, have been shown to participate in a novel *ars*-like detoxification pathway in the legume symbiont *Sinorhizobium meliloti* [**168**]. A cluster of three genes involved in arsenic resistance has been identified, namely; *acr*1, *acr*2 and *acr*3 in *Saccharomyces* species that translate a regulator, reductase, and permease, respectively [**24, 90**].
A wide variety of chromosomal resistance operons have been characterized, including *E. coli* [**31**], *Pseudomonas aeruginosa* [**30**], *Bacillus subtilis* [**129**], *Corynebacterium glutamicum* [**106**], *At. ferrooxidans* [**29**], and *Sinorhizobium meliloti* [**168**] as well as the archaeal genome of *Ferroplasma acidarmanus* [**58**]. Southern blot analysis by Diorio et al. [**49**] has revealed *ars* homologs in a number of enterobacteria, such as *Shigella sonnei*, *Citrobacter freundii*, *Enterobacter coloacea*, *Salmonella arizonae*, *Erwinia carotovora*, and *Klebsiella pneumoniae*. Saltikov & Newman [**124**] have even discovered a four gene *ars* operon in the As(V)-respiring *Shewanella* sp. str. ANA-3.
Plasmid-borne arsenic resistance determinants for a variety of microorganisms have been reported, which include the staphylococcal plasmids pI258 and pSX267 [**71, 122**], *E. coli* plasmids R46 and R773 [**26, 35**], *P. aeruginosa plasmid* pUM310 [**33**], *Ac. multivorum* plasmid pKW301 [**143**], *Y. enterocolita* plasmid pYV [**103**], and *S. marcesens* IncHI2 plasmid R476 [**123**] as well as the archaeon *Halobacterium* megaplasmid pNRC100 [**155**].

Varieties of genes and operons Unlike the typical gene organization, certain operons consist of two divergent elements, for example, *ars*CR-*ars*BH in *At. ferrooxidans* [**29**] and *ars*AD-*ars*RC in the pNRC100 megaplasmid of *Halobacterium* [**155**]. Also in some cases, *ars*R and *ars*H are transcribed in the opposite direction from the other arsenic determinants [**103, 106, 123**]. The chromosomes of *C. glutamicum* and *P. putida* contain two similar sets of *ars*R-BC and *ars*RBCH operons, respectively [**106**]. Multiple versions of ArsC, that may be representatives of a single or different clade of arsenate reductases, in a single microbe have also been documented. Duplication of the *ars*AD genes have been observed in two transposon *ars*-like operons of *At. caldus* [**150**].

Moreover, individual *ars* genes apart from those within the operon have been found at distant locations in bacterial genomes **[88, 106, 155]**. The *ars*B and *ars*C genes of *M. tuberculosis* are fused into a single protein where the N- and C- terminal regions show significant homologies to the ArsB and ArsC gene products of *B. subtilis*, respectively **[100, 129]**.

Regulation of the ars operon

The *ars* operons of both Gram(+) and Gram(-) bacteria are inducible by antimonite, (Sb(III)), arsenite (As(III)), arsenate (As(V)), as well as Bi(III) **[70, 122, 160, 167]**. Both ArsR and ArsD are trans-acting, dimeric repressors that are localized in the cytoplasm **[162, 163]**. ArsRs belong to a new subfamily of helix-turn-helix (HTH) DNA-binding proteins with a conserved metal binding box, ELCVCDL, although a few atypical ArsR-like translation products have been described **[28, 32, 58, 106, 167, 168]**. When inducer is absent, transcription of the *ars* operon is repressed by a basal level of ArsR expression **[36, 162]**.The binding of As^{3+} or Sb^{3+} to the thiolates of Cys32 and Cys34 are sufficient to distort the DNA-binding domain of ArsR, which then dissociate from the operator, permitting transcription to commence **[117, 133]**.

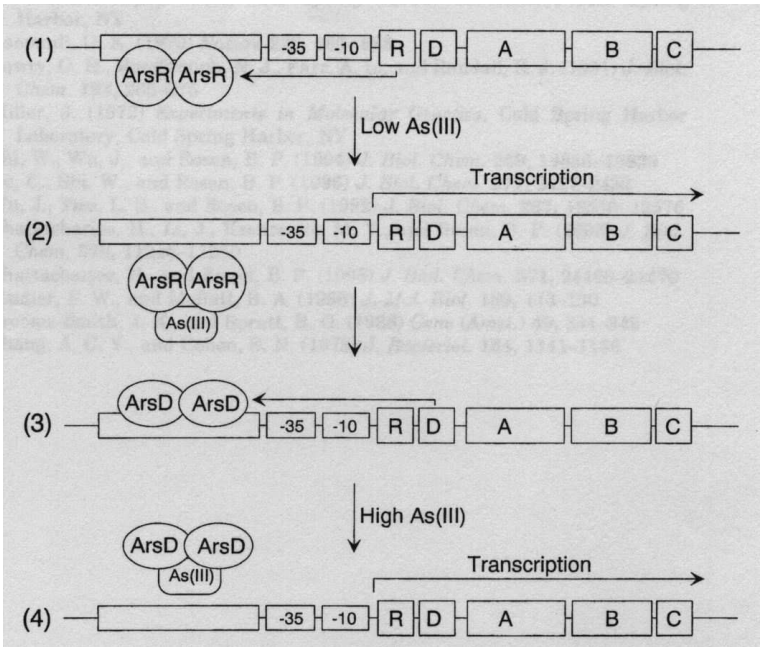


Figure 1.4: Model of ArsR-ArsD metalloregulatory circuit in *E. coli* (From **[36]**).

Because ArsD has a 2 orders of magnitude lower affinity for the same operator site, considerable levels of the protein are necessary before it can bind to the *ars* operon, such as after prolonged exposure to inducer. Therefore, so when sufficient amounts of ArsD has been expressed to saturate the operator, the operon is yet again repressed in *E. coli* [**36, 83**]. In ArsD the inducer binds to the two vicinal cysteine pairs (Cys12-Cys13 and Cys112-Cys113) in both monomers in a seemingly cooperative manner [**83, 84**].
As a result, it has been suggested that ArsR and ArsD controls the basal and maximal levels of expression of the operon, possibly to prevent the overexpression of ArsB, which can be toxic to the cell [**36, 83**]. Even though both ArsR and ArsD proteins are both roughly 13kDa homodimers that bind to the same promoter region, they share no sequence similarity and act independently of each other [**36**]. Transcriptional activation of the yeast metalloid tolerance genes, *acr*2, *acr*3 and even *ycf*1, requires the Yap1 and Yap8 (Acr1) proteins, which contains conserved cysteine residues in the N- and C- terminal CRD domains that are critical for their function [**24, 164**].

Arsenate reductase, ArsC

Once arsenate has been taken up by the cell, there is a problem with its detoxification because of its structural resemblance to phosphate which makes the active extrusion of As(V) difficult and ineffective [**97, 104, 119**]. Therefore, AsO_4^{3-} detoxification requires an initial step that will discriminate it from phosphate [**104**]. This preliminary step is performed by the small, soluble, cytoplasmic ArsC protein that reduces arsenate to arsenite, which is the transport substrate [**70**].

Currently, three unrelated clades of arsenate reductases are recognized that share a common biological function but no evolutionary relationship [**100, 135**]. These As(V)-reductases have no enzymatic activity on their own, but require other proteins to carry out the reduction process [**32, 60, 72**]. The first and the best studied group of arsenate reductases, represented by the ArsC from *E. coli* plasmid R773, necessitates reduced glutathione and glutaredoxin for activity [**60, 91, 117**]. The GSH/Grx ArsC family has a single catalytic cysteine residue, Cys12, as well as three conserved arginines (Arg60, Arg94, and Arg 107) in its close proximity that participate in the anion binding site and stabilizes the Cys-S-AsO_3 intermediate through hydrogen bond interactions [**85, 86, 91, 132**].
The second clade of ArsCs has the arsenate reductase of *S. aureus* plasmid pI258 as prototype, derive their reducing power from thioredoxin and utilizes several essential cysteines (Ser/Cys10, Cys82, and Cys89) in the catalytic process [**72, 95**]. The Trx-linked arsenate reductases are related to the low-molecular weight phosphotyrosine phosphatases (lmwPTPases) and exhibit low level phosphatase activity [**172**].

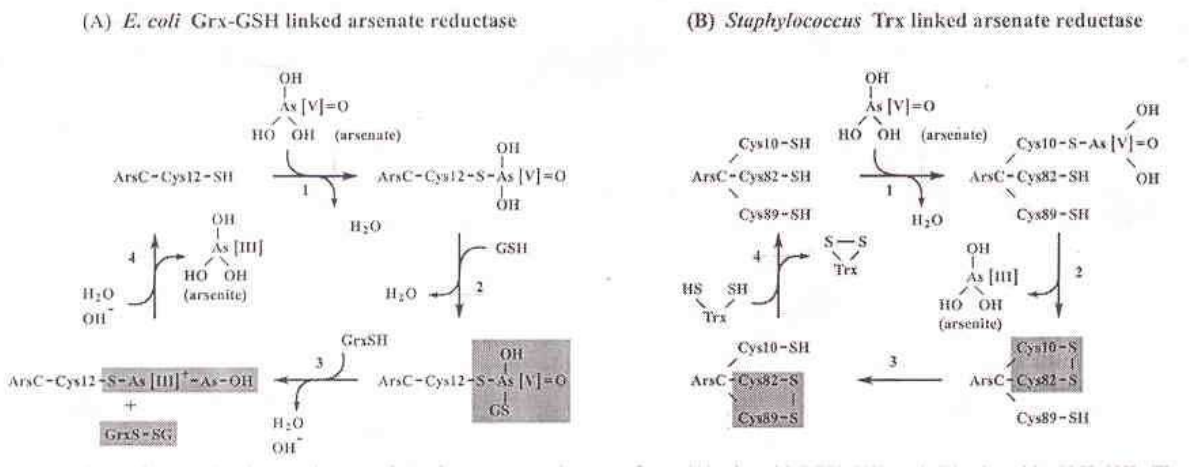


Figure 1.5: Proposed reaction mechanisms for ArsC arsenate reductases from (A) plasmid R773 ([**91**]) and (B) plasmid pI258 ([**172**]) (From [**100**]).

The dimeric yeast Acr2 protein is as yet the only member of the third family of arsenate reductases and demonstrates the same GSH and Grx requirement as the *E. coli* plasmid R773 ArsC group [**24, 99**]. Acr2 belongs to the Cdc25A superfamily of PTPases which share a consensus active site $HC(X)_5R$ motif [**98**]. Even though ArsC and Acr2 share any significant sequence homology, the yeast reductase can complement the As(V)-reductase phenotype in an R773 ArsC mutant [**99, 145**].
Interestingly, the different clades of reductase enzymes show distinct substrate specificities, where certain substances can exert dissimilar effects to the separate ArsC classes. Recently, As(V) reductase activity has been observed in liver extracts from various animals, particularly rabbit [**169**], human [**113**], and hamster [**127**]. Two putative arsenate reducing proteins have been characterized in mammalian species. A purine nucleoside phosphorylase (PNP) that was purified from rat and human liver demonstrated to catalyze the reduction of IAs^V [**114**]. MAs^V reductase from rabbit and human liver has been purified and illustrated significant sequence homology with human glutathione-S-transferase omega (hGSTω) [**169, 170**].

Arsenic efflux pumps

It might seem counter-intuitive to convert a less toxic compound to a more toxic form, but ArsC activity is closely associated with the efflux system so that arsenite never accumulates [**70**]. ArsB homologs that are found in almost every bacterial *ars* operon codes for a integral membrane channel protein that possesses 12 membrane spanning regions [**161**].

The yeast Acr3, that shows significant similarity to the ArsB homologs of *B. subtilis* and cyanobacterium *Synechocystis* sp. str. PCC 6803, is proposed to contain only ten membrane spanning segments [**165**].
By itself, ArsB functions as a secondary uniporter that is energy driven by the membrane potential of the cell. However, when ArsA is coexpressed with ArsB, the resulting $ArsA_2B$ complex is converted to an obligatorily ATP-coupled arsenite pump [**46**]. ArsB anchors the otherwise soluble cytosolic ArsA protein to the membrane that actively functions as a dimer with four ATP binding sites [**148**]. The ArsB of the *E. coli* plasmid R773 *ars* operon was shown to actively export arsenite in the absence of the ArsA ATPase, which is consistent to that established in the Gram(+) plasmid-borne operons [**46**]. The dual mode of energy coupling is unique among known bacterial uptake and efflux systems.
The metalloid stimulated translocating ATPase ArsA consist of two homologous halves A1 and A2, connected by a 25-residue linker peptide, with each half containing a glycine-rich consensus nucleotide binding sequence (NBS) consisting of G^{15}KGGVGKTSIS^{25} and G^{334}KGGVGKTTMA^{344}, respectively [**82, 121, 174**]. ArsA is allosterically activated through metalloid binding to the three thiolates of Cys113, Cys172, and Cys422 that induces a conformational change which brings the two ATP-binding domains (NBDs) of both halves closer together [**22, 121**]. The localized changes that are propagated from the allosteric metal binding domain (MBD) to the NBDs are mediated by the signal transduction regions that are recognized by a consensus 12-residue signature sequence ($D_{142/447}$TAPTG$H_{148/453}$TIRLL) [**121**]. This DTAP motif that is found on both halves of ArsA are also conserved in the ArsA1 homologs from archaea, fungi, plants, and animals, including humans [**173**].

1.10.2 Arsenic oxidation and respiration

As(III) oxidation that has been recognized since 1918 represents a potential detoxification pathway which allow microorganisms to tolerate higher levels of arsenite [**32, 100, 137, 139**]. At present more than 30 strains representing at least 9 genera have been documented and include α-, β- and γ- Proteobacteria, Deinococci (i.e. *Thermus*), and Crenarchaeota. These physiologically diverse microbes include heterotrophic as well as chemolithotrophic species that uses oxygen or even nitrate as a terminal electron acceptor (TEA) [**107, 139**].
Anderson et al. [**12**] characterized two contiguous genes, *aso*A and *aso*B, encoding the α- and β- subunits of a putative arsenite oxidase in the pseudomonad *Alcaligenes faecalis*. The large 85-kDa- catalytic subunit contains molybdenum bound to two pterin cofactors and a [3Fe-4S] cluster while the ∼14kDa accessory subunit has a Rieske type [2Fe-2S] cluster [**12, 51**].

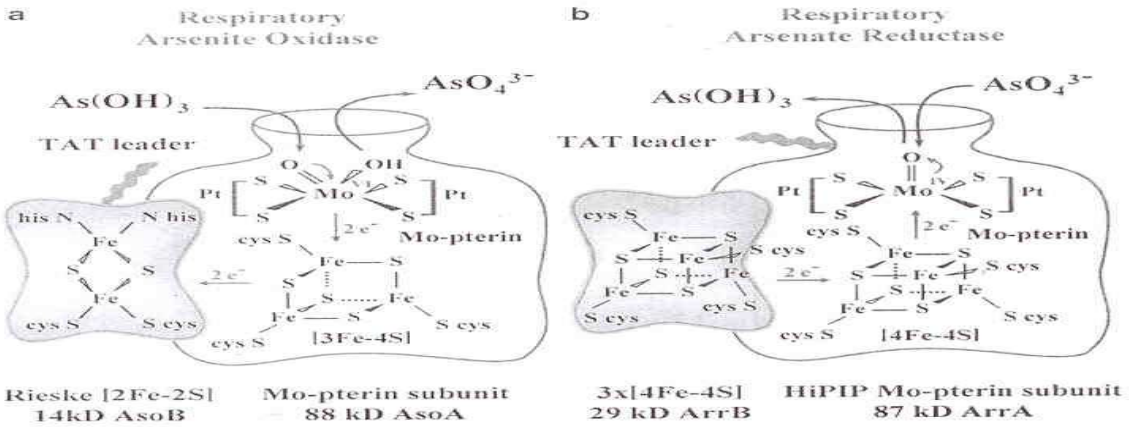


Figure 1.6: Models of arsenite oxidase and respiratory arsenate reductase (From [**137**]).

Biochemical characterization and sequence analysis indicate that arsenite oxidase belong to the DMSO family of molybdopterin Fe-S oxidoreductases [**51**] Two adjacent genes, denoted *aox*A and *aox*B, encoding an arsenite oxidase was identified from the β-Proteobacterium *Cenibacterium arsenoxidans* that was isolated from industrial wastewater [**101**].

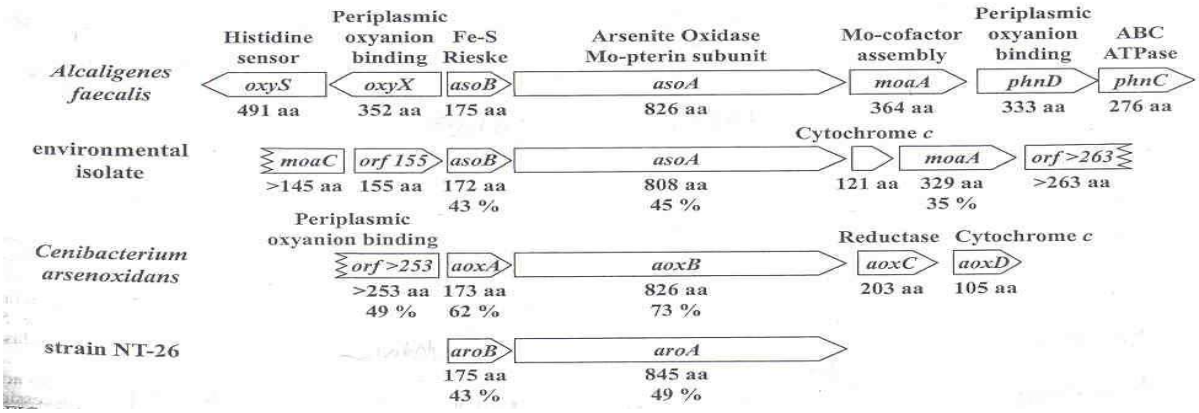


Figure 1.7: Arsenite oxidase genes in *A. faecalis* (NCBI accesion no. AY297781; [**12**]), a Sargasso sea meta-genome environmental isolate (NCBI AACY01082423; [**154**]), *C. arsenoxidans* (NCBI AF509588; [**101**]) and chemolithotrophic α -proteobacterial strain NT-26 (NCBI AY345225, [**128**]) and upstream and downstream genes (From [**135**]).

Yet another set of polypeptides, referred to as AroA and AroB, was obtained from a chemolithotrophic α-Proteobacterium strain NT-26 that was thought to produce an $\alpha_2\beta_2$ heterotetramer arsenite oxidase [**128**]. The arsenite oxidase of the heterotrophic β-Proteobacterium *Hydrogenophaga* sp. str. NT-14 appears to form an $\alpha_3\beta_3$ hexameric enzyme [**153**]. Putative relating genes for arsenite oxidase have been identified in the genomes of two hyperthermophilic archaea *Aeropyrum pernix* and *Sulfolobus tokadaii* and the phototroph *Chloroflexus auranticus* [**128**].

Up to date, at least 16 prokaryotic species are capable of using arsenate as a terminal electron acceptor and include representatives from Crenarchaeota, thermophilic Eubacteria, γ-, δ- and ϵ-Proteobacteria and low and high GC content Gram(+) bacteria. These microbes that are collectively known as dissimilatory arsenate reducing prokaryotes (DARPs) have been isolated from a wide diversity of environments and are able to utilize various electron donors [**107, 139**]. In a sense, arsenate respiration can be considered as a resistance system since it contributes to the reduced accumulation of AsO_4^{3-} and contribute to energy production [**137**]. Krafft and Macy [**77**] was the first to characterize a respiratory arsenate reductase from the strict anaerobe *Chrysiogenes arsenatis* str. BAL-1T isolated from gold mine wastewater.

These authors showed that the respiratory arsenate reductase, like arsenite oxidase, is a heterodimeric, periplasmic protein consisting of a large 87-kDa molybdopterin subunit (ArrA) and a smaller 29-kDa subunit with both translation products containing at least one Fe-S cluster [**77**]. Recently, *arr*AB-like determinants for respiratory arsenate reductases have been described for the Gram-negative γ-Proteobacterium *Shewanella* sp. str. ANA-3 [**124**] and the Gram(+) *Bacillus selenitrireducens* strain MLS10 [**11**]. Furthermore, genomic sequencing of the As(V)-respiring bacterium *Desulfitobacterium hafniense* revealed two hypothetical proteins that demonstrated significant similarity to the ArrA and ArrB gene products of *Shewanella* sp. str. ANA-3 [**105**].

N-terminal sequence data indicate that respiratory arsenate reductases are also presumably related to the DMSO reductase family of mononuclear molybdenum enzymes [**124**]. Interestingly, Tat leader sequences have been found at the N-termini of the small subunit of arsenite oxidase and large subunit of respiratory arsenate reductase [**124, 135**]. The twin arginine translocation (TAT) motifs are implicated in the export of these proteins in their already preassembled state to the periplasm [**21**]. More recently, a *Thermus* species was shown to be capable of rapid As(III) oxidation under aerobic conditions and grew anaerobically via As(V) respiration coupled to lactate oxidation [**57**].

Metabolism	Reaction	Genes	Representative Organism
Oxygen-dependent AsO_2^- oxidation	$2HAsO_2 + 2H_2O + O_2 \rightarrow 2HAsO_4^{2-} + 4H^+$	AroAB, AsoAB and AoxAB	*A. faecalis*
Nitrate-dependent AsO_2^- oxidation	$HAsO_2 + H_2O + NO_3^- \rightarrow AsO_4^{2-} + 2H^+$	None known	Eubacterial strain MLHE-1
Lactate-dependent AsO_4^{3-} reduction	$CH_3CH(OH)COO^- + 2HAsO_4^{2-} + 3H^+ \rightarrow 2HAsO_2 + 2H_2O + HCO_3^-$	ArrAB	*B. selenitireducens*
Acetate-dependent AsO_4^{3-} reduction	$CH_3COO^- + 4HAsO_4^{2-} + 7H^+ \rightarrow 4HAsO_2 + 4H_2O + 2HCO_3^-$	ArrAB	*C. arsenatis*

Table 1.4: Reactions catalyzed by As-metabolizing bacteria. (Adapted from **[39]**)

1.10.3 Sequestration

The production of metal binding proteins may play an important role in tolerance acquisition in certain organisms. Metal binding proteins can be classified into two groups, the metallothioneins (MTs) found frequently in various eukaryotic species and the phytochelatins (PCs) present in plants and some fungi. Both the metallothionein and phytochelatin encoding genes are stimulated by arsenite **[38, 76]**.

Metallothioneins are small thiolate-rich proteins that sequester a wide variety of metals in mammalian cells. The best characterized prokaryotic MTs are those from the cyanobacterial species of the genus *Synechococcus*. The *smt*A gene codes a metallothionein of 56 amino acid long peptide containing 9 cysteine residues. Metallothionein (SmtA) expression is negatively controlled by the repressor SmtB which belongs to the same helix-turn-helix (HTH) family as ArsR and CadC **[151]**.

1.10.4 Methylation-mediated detoxification

Even though biomethylation of arsenic is widespread among bacteria, fungi, algae, plants and animals, including humans, it is not universal **[20, 109]**. Organoarsenicals are usually more stable, less toxic and extensively metabolized, and excreted more rapidly than inorganic arsenic compounds, especially in mammals **[1]**. Normally, bacteria are less sensitive to methylated arsenic derivatives than inorganic arsenical species **[144]**. Since the toxic effects of arsenic compounds are correspondingly related to the rate of clearance, biomethylation was regarded as a detoxification mechanism in animals **[116]**. Lately, experimental studies have demonstrated that reactive trivalent organoarsenic metabolic intermediates are more harmful, which suggest that methylating reactions may not solely act as an arsenical tolerance system **[141]**.

Biomethylation involves the sequential and alternating $2e^-$ reduction of IAs^V to IAs^{III} followed by oxidative methylation to pentavalent organic arsenic which can then be further reduced and metabolized **[66, 147]**. S-adenosyl methionine (SAM or AdoMet) has been identified as the typical methyl donor in eukaryotic and prokaryotic systems, though methylcobalamin (CH_3-B_{12}) might also serve as a methyl donor in microorganisms **[20]**.

These methyl cofactor dependent reactions are enzymatically catalyzed by As(III)-methyltransferases in the presence of GSH [**109**]. Trivalent arsenic methyltransferases have been purified and partially characterized from liver extracts of rabbit, hamster, and rhesus monkey [**158, 171**]. Arsenic methylating activity have also been detected in primary human and rat hepatocytes as well as in different tissues of mice [**63, 142**].
Thomas and colleagues [**146**] isolated a 42-kDa protein, which showed close similarity to the homologous translation products of *cyt19* in mouse and human, that was capable of catalyzing the entire conversion of arsenite to its methylated metabolites. Very recently, while examining numerous microbial genomes Qin and colleagues [**111**] identified 125 bacterial and 16 archaeal homologs of Cyt19. Twelve homologous genes, that belong to the UbiE/Coq5 S-adenosyl-L-methionine-dependent C-methyltransferases group, were designated *ars*M and their gene products ArsM as they were found adjacent to the archetypal As-inducible transcriptional regulatory gene *ars*R, suggesting a putative role for their respective proteins in arsenic detoxification. These Cyt19 As(III) S-adenosylmethyltransferases (As3MT) homologs originate from a broad diversity of microorganisms, including aerobic and anaerobic archaeal and bacterial mesophiles, thermophiles and halophiles.
The purified 283-residue ArsM of *Rhodopseudomonas palustris* CGA009 was shown to catalyze the SAM-dependent conversion of trivalent arsenic to di- and trimethylated species with TMA(III) gas being the final product. Interestingly, TMA(III) serves as an intermediate in the formation of other organoarsenicals, such as O-phosphotidyltrimethylarsonium lactic acid which is found in marine algae and animals. Arsenobetaine that are found in marine animals such as the Western rock lobster is one of the degradation products of this arsenolipid [**111**].
In the anaerobic, methanogenic microbes, IAs methylation is linked to methane biosynthesis and could be a widely occurring detoxification pathway. Arsenate can be partially detoxified through the generation of significant amounts of stable organoarsenical compounds . The biosynthesis of arsenolipids has been proposed as a possible arsenic tolerance system in marine algae and other organisms. In fact, formation of lipid soluble organoarsenicals could be an adaptive mechanism for marine phytoplankton to compensate for limited nitrate availability [**144**].

1.11 Nocardioform actinomycetes

Actinomycetes are chemoheterotrophic Gram(+) microorganisms that resemble fungi in their filamentous growth to produce mycelial colonies and is reflected in the name that translates from Greek meaning “ray mushroom” (fungus) [**65**]. The term “nocardioform” refers to the formation of rudimentary or extensive branching hyphae that may persist as stable mycelium or fragment into rod shaped and coccoid elements.

Nocardioform actinomycetes include the following genera: *Nocardia, Rhodococcus, Gordonia, Tsukamurella, Pseudonocardia, Nocardiodes, Promicromonospora, Saccharopolyspora, Micropolyspora, Intrasporangium* and *Oerskovia*. The nocardioform group of bacteria are generally aerobic, catalase positive, non-motile actinomycetes that show great morphological diversity.

The cells of nocardioforms are typically 0.5-1.2 μm in diameter with all containing a high (63-79%) G + C content. The various genera of the nocardioform group are differentiated primarily on the basis of cell wall, whole-cell sugar patterns, and lipid composition. Most nocardioforms have a type IV cell wall that contains the taxonomically important components: meso-diaminopimelic acid (meso-DAP), arabinose, and galactose. Certain genera contains mycolic acids with 34-90 carbons atoms and up to six double bonds like those found in the cell walls of *Mycobacterium* and *Corynebacterium* species **[65, 80]**. Nocardioforms often form brightly coloured colonies **[17]**.

Microorganisms of this actinomycetes group are widely distributed in nature and are frequently found in soil and aquatic environments **[65, 80]**. While most nocardioforms are nonpathogenic, free-living saprophytes, some species are associated with infections in humans, animals, and plants **[92, 110]**. Human and equine infections primarily occur in the lungs resulting in pneumonia and abscesses, but can disseminate to cause lesions in other organs and bacteraemia **[92]**.

Nocardioform bacteria are noteworthy for their ability of degrading a wide range of toxic compounds such as aromatic hydrocarbons **[61]**, chlorinated phenols **[62]**, PCBs **[25]**, lignin-related compounds **[50]**, acrylamide **[14]**, pesticides **[18,34,108]**, sulphonated azo dyes **[64]**, and natural rubber **[16]**. Certain species from this mycolate-containing subgroup of nocardioform actinomycetes are also capable of desulphurizing coal and petroleum **[44, 69]** as well as the synthesis of biosurfactants and other anabolic substances **[89]**. It is evidently clear that nocardioform bacteria have tremendous potential in the industrial and environmental biotechnology fields.

1.11.1 *Rhodococcus*

The genus name *Rhodococcus*, meaning red coccus, was first introduced by Zopf in 1891 and was revised and redefined in 1977 in order to accommodate the "rhodochrous" complex observed in several actinomycetes strains **[19, 80]**. Currently, the *Rhodococcus* genus comprises of 12 established species such as *R. rhodochrous*, *R. erythropolis, R. equi, R. fascians* and *R. opacus*. Rhodococci are taxonomically distinguished by mycolic acids consisting of 34-54 carbon atoms with up to 3 double bonds and the dihydrogenated menaquinone with 8 isoprene units as its predominant isoprenolog **[65, 80]**.

R. equi has been recognized as an equine pathogen for many years and result in commonly respiratory infections in a variety of animals. *R. fascians* causes leaf gall in numerous plant specimens and leads to fasciation in sweet peas [**92**].

1.11.2 *Gordonia*

In 1971 Tsukamura proposed a new genera for corynebacteria called *Gordona* that was chosen to pay tribute to the American bacteriologist Ruth E. Gordon. This actinomycetes genus was reintroduced in 1988 by Strackenbrandt et al. for four formerly classified *Rhodococcus* species and was subsequently renamed to *Gordonia* in 1997. Presently, the *Gordonia* genus consist of 19 valid species, which include *G. rubropertincta*, *G. sputi*, *G. terrae*, and *G. bronchialis* [**15**]. *Gordonia* species are characterized by mycolic acids of 48-66 carbon atoms with 1-4 double bonds and MK-9(H_2) as the dominant menaquinone [**15,65**]. *G. bronchialis* has been found to be associated with the sputa of patients with cavitary pulmonary tuberculosis and brochiectasis [**15**].

1.12 *Arthrobacter*

Arthrobacter, meaning jointed rod, presently consist of 15 recognized species such as *A. globiformis, A. citreus, A. nicotianae* an *A. oxydans*. These microorganisms are irregular, non-sporing, aerobic, catalase positive, Gram(+) bacteria that exhibits an oxidative metabolism. The genus *Arthrobacter* are characterized by a marked rod-coccus growth cycle. In the early stages of the growth cycle the cells are irregular rods, often V-shaped, with clubbed ends and no filaments. However, as growth proceeds these rods fragment into small cocci of 0.6-1.0μm in diameter. These chemoorganotrophic microorganisms has L-lysine as their diamino acid in the cell wall peptidoglycan and MK-9(H_2) as the major menaquinone isoprenolog. *Arthrobacter* bacteria contains a G + C content of about 59 - 70% in their genomic DNA and occurs widely in the environment, particularly in soil [**65, 73**].

Aims and Objectives

Arsenic contamination of groundwater due to leaching of geological formations and past agricultural use of pesticides poses considerable health risks worldwide. The abundance of arsenic in nature has accordingly directed the evolution of enzymes for its own detoxification. This has led to much interest in investigating and unraveling these resistance mechanisms and the genes conferring increased arsenic tolerance.

This project aims to continue the search for arsenic resistance genes in nocardioform and related species. Firstly, this would involve identifying putative arsenic resistant nocardioforms. Existing genomic libraries of possible As^R nocardioform strains, most notably from *Rhodococcus* and *Gordonia*, will be screened for arsenic resistance genes. It would also be necessary to construct genomic libraries of other prospective resistant organisms in *E. coli* and subsequently transforming it into an arsenic-susceptible Gram-positive strain. Positive clones from screening on arsenate and arsenite-containing media will be subjected to endonuclease digestion in order to construct a restriction map of the inserts. Subcloning of the smaller fragments will be analyzed for the As^R phenotype and will assist in determining the minimum amount of DNA required to confer arsenic tolerance. Finally, the DNA fragments will be sequenced to characterize and compare the putative resistance genes with previously isolated genes available in genomic databases.

Furthermore, since nocardioforms are noteworthy for their ability to degrade and utilize numerous compounds, it was decided to screen these above mentioned libraries also for biodegradation genes of several recalcitrant or unusual carbon sources.

Chapter 2

Materials and Methods

2.1 Maintenance of bacterial specimens and their DNA

2.1.1 Culture Maintenance

E. coli and *Rhodococcus* bacterial strains were maintained on Luria-Bertani agar (LA) plates containing appropriate amounts of required antibiotic(s) at 4°C for short periods of time. *Gordonia* species were sustained in Brain-Heart-Infusion (BHI) media. Liquid LBSG cultures of *Rhodococcus* strains were also stored in the cold room. Bacterial strains were preserved for long term in 33% glycerol at -70°C.

E. coli strains were grown on LA plates or in LB broth at 37°C overnight, except for λ lysogens that were incubated at 30°C in order to prevent the induction of the lytic cycle. Rhodococcal strains were grown on LA plates at 30°C. Single colonies were inoculated into 5ml of LBSG medium and incubated with agitation for 2-3 days. *Gordonia* species were grown on BHI agar or broth at 30°C while shaking for about 48 hrs.

Strain	**Characteristics**	**Origin**
E. coli MM294-4	*endA1*, *hsdR17*, *gyrA*	E. Dabbs
E. coli λMM294-4	lysogenic for λcI857	Quan 1993
R. erythropolis SQ1	Highly transformable derivative of strain ATCC 4277, Rif^R	Quan & Dabbs 1993
R. rhodochrous Ri8	iri^- mutant of strain ATCC 12674, Rif^R	Andersen 1996
G. rubropertincta ATCC 25593	Rif^R	E. Dabbs
A. oxydans C1		E. Dabbs

Table 2.1: Bacterial strains used in study

2.1.2 Storing of chromosomal and plasmid DNA

Plasmids can be preserved within suitable bacterial strains in 33% glycerol at -70°C. Plasmids and genomic DNA were stored in either 4mM Tris.HCl or sterile distilled water (sdH_2O) at -20°C. Chromosomal DNA and plasmids can be kept in cesium chloride (CsCl) or suspended in TE buffer for long term storage.

2.2 Spot testing

Two drops (100-200μl) of sterile distilled water were transferred to each well. A small portion of cells of the different bacteria or clones were taken with a toothpick and transferred and resuspended in the separate wells. The replicator was sterilized over a bunsen burner and placed into the wells which was then transferred onto the appropriate test plates for a few seconds. The bacterial samples were also replicated onto a control plate. The liquid was allowed to diffuse into the solid agar medium and incubated at the desired temperature until the cultures have grown up sufficiently.

2.3 DNA Isolation

2.3.1 *E. coli* plasmid DNA isolation

Plasmid DNA mini preparation

A single colony of a bacterial culture was used to inoculate 1ml LB broth containing 100μg/ml ampicillin and grown at 37°C, except for lysogenic strains, with aeration for overnight. The suspended cells were transferred to sterile Eppendorf tubes and pelleted by microfuging for 30 sec. The supernatant was decanted and the pellet was resuspended in 80μl of solution I by aspirating the mixture. To this 160μl of solution II was added and mixed through gentle inversion and then allowed to stand for 15 min at room temperature. Following this, 120μl of solution III was added to the tubes and shaken vigorously for 15-20 sec and placed on ice for 5 min. The Eppendorf tubes were then microfuged at 4°C, after which the supernatant was carefully transferred to fresh tubes. After the solution was warmed up in the waterbath briefly, 220μl of isopropanol was added to the mixture and inverted gently for a few times. Samples were left on the bench for five minutes at room temperature which was followed by microfuging the precipitated DNA for 5 min at ambient conditions. The supernatant was carefully decanted and blotted on a paper towel. The DNA pellet was rinsed by adding 150μl of ethanol and microfuging for 1 min.

Again, the liquid was slowly poured off and the pellet was dried in the Speedvac for 20 min. The DNA was suspended in 100μl of sdH_2O to which 1μl (10mg/ml) of ribonuclease (Rnase) was added and mixed by gentle tapping for several minutes.

Large-scale plasmid DNA isolation

Bacterial cells containing plasmid were inoculated into 100ml of LB broth with the appropriate selective agent to maintain the plasmid. The suspension of cells were incubated at 37°C, (or $<$ 33°C for lysogens) overnight on a shaker. The culture was transferred to a JA-10 and the cells were collected by centrifuging for 10 min at 6000 rpm. Supernatant was discarded and pellet was resuspended in 5ml of solution 1. Afterwards, the suspended cells were transferred to a JA-20 tube followed by adding 10ml of solution 2. The contents were gently inverted and allowed to stand for at least 15 minutes at ambient temperature, during which lysis should have occurred, resulting in increased viscosity and clearing of the mixture. This was followed by the addition of 7.5ml of solution 3 and tube shaken vigorously for 20 sec and then placed in an ice/water slurry for about 5 min. The white precipitate was removed by centrifugation for 10 min at 15000 rpm in a prechilled Beckman JA-20 rotor, after which the supernatant was carefully decanted to a fresh tube. After the liquid mixture was warmed up to room temperature, 12ml of isopropanol was added and mixed by inversion and left on the bench for 5 min. The precipitated DNA was spun down by centrifuging the JA-20 tube for 10 min at 15000 rpm. The liquid phase was discarded and the pellet was washed with 1ml of 96% ethanol by brief centrifugation, and then dried under vacuum for 20 min. The pellet was resuspended in 4ml of TE buffer through gentle agitation in a 30°C shaking waterbath for two or more hours. Following this, 4.1g of cesium chloride was added and dissolved by gentle inversion. Ethidium bromide (400μl) was added and the refractive index adjusted to between 1.387 - 1389 using a refractometer. The DNA preparation mixture was loaded into a Quickseal tube and then centrifuged for at least 16 hrs at 45k rpm. The tube was viewed under UV and the plasmid band extracted with a syringe and transferred to a sterile Eppendorf tube. The EtBr was removed by several butanol extractions and precipitated out of CsCl with the ethanol-precipitation method.

2.3.2 Nocardioform DNA isolation

Bulk DNA extraction

Cells of *Rhodococcus* or *Gordonia* strains were used to inoculate 300 ml of LBS medium supplemented with the appropriate concentration of glycine and grown in a shaking 30°C waterbath to stationary phase (2-5 days).

The bacterial cells were harvested by centrifuging in a Beckman JA-10 rotor at 6000 rpm for 10 min. The pellet was resuspended in 5ml of TE buffer containing 5mg/ml of freshly added lysozyme and then transferred to a sterile JA-20 tube, which was then incubated in a 37°C waterbath for approximately 2 hrs. A 1/10th volume (500μl) of TE + 10% SDS with a speck of Proteinase K was added and incubated at 48°C for >1 hr. The cleared viscous solution was placed into a 50Ti tube using a 2ml upside down pipette and spun down at 40000 rpm for 45-60 min in a Beckman L5-50 ultracentrifuge. Following this, the supernatant was transferred to a clean JA-20 tube to which 4.4g of CsCl was added and slowly inverted until completely dissolved. The solution was centrifuged at 15000 rpm for 15 min and the liquid was decanted under the scum to a new tube. This was followed by adding 400μl of ethidium bromide after which the refractive index was adjusted to between 1.391 -1.392. Using a Pasteur pipette, the liquid DNA mixture was loaded into a Quickseal tube and ultracentrifuged for a minimum of 16 hrs at 45000 rpm. The tube was illuminated under UV light and the DNA band was removed via needle attached to a hypodermic syringe and transferred to a sterile microfuge tube. The EtBr was eliminated by butanol extraction and precipitated using the ethanol precipitation method.

Mini plasmid DNA preparation

A small portion of bacterial culture from an agar plate was suspended in about 1ml of glycine supplemented LBS broth and grown at 30°C for 2-3 days on a rotary shaker. A 200-300μl aliquot of the cell suspension was transferred to a fresh Eppendorf tube and microfuged for 1 min. The supernatant was poured off and the pellet resuspended in 200μl of TE buffer containing 5mg/ml of freshly added lysozyme. The mixture was then incubated in a 37°C water bath for 1 hr with inversion after every ten minutes. This was followed by the addition of 40μl of 10% SDS -TE buffer and gently mixed by inversion and allowed to stand at room temperature for 10 min. After 40μl of 5M KAc pH6.0 was added and shaken vigorously for around 15 sec, the tube was placed on ice for 5 min. The contents were centrifuged for five minutes at 4°C, after which the supernatant was transferred to a sterile microfuge tube. Protein and other compounds was removed with phenol-chloroform extraction and the DNA precipitated using the salt ethanol method.

2.4 DNA purification techniques

2.4.1 Phenol-chloroform extraction

The volume of the DNA solution was adjusted to 300μl with TE buffer. Then 100μl (1/3 vol) of the bottom layer of TE- saturated phenol was added to the Eppendorf tube and inverted for about a minute. The mixture was microfuged for 5 min to separate the aqueous phase from the organic layer that now contains the denatured proteins. The upper layer was extracted and transferred to a fresh microfuge tube to which 133μl (1/3 of total volume V_T) of chloroform was added. The contents were inverted for a few times and then centrifuged for 30 sec at room temperature. The top phase was removed and placed into a new Eppendorf tube and the DNA was precipitated using the salt-ethanol method.

2.4.2 Butanol extraction

The ethidium bromide after CsCl centrifugation was removed by adding 100 - 150μl of butanol to the DNA solution. The Eppendorf tube was inverted several times and then the upper layer was extracted and transferred to a new sterile Eppendorf. The addition of 100μl aliquots of butanol and removing the top layer was repeated until a clear lower phase was achieved.

2.5 DNA precipitation methods

2.5.1 Salt-Ethanol precipitation

A 1/10th volume of 1M NaCl and two volumes of 96% ethanol were added to the DNA solution. The mixture was microfuged for 20 min at 4°C, after which the supernatant was carefully decanted and the Eppendorf tube briefly blotted on a paper towel. The precipitated DNA was dried under vacuum for 20 min which was then resuspended in an appropriate volume of sterile distilled water or 4mM Tris.HCl.

2.5.2 Isopropanol-Ethanol precipitation

One third volume of isopropanol was added to the DNA solution and mixed by gently inverting the microfuge tube. The mixture was left to stand at ambient temperature for 5 min and then microfuged for a further five minutes. The supernatant was slowly discarded and the DNA pellet was rinsed with 96% ethanol by brief centrifugation. Liquid was removed and the tube was blotted on a paper towel which was then dried in the Speedvac for 20 min. Following this, the pellet was resuspended in a suitable amount of sdH_2O or 4mM Tris.HCl.

2.5.3 Ethanol precipitation

An aliquot of the purified DNA in cesium chloride was transferred to fresh microfuge tube. Two volumes of sterile distilled water was added to the DNA-CsCl mixture and then inverted gently for a few times. More than double this combined volume of 96% ethanol was then added and the contents centrifuged at 4°C for 20 min. The supernatant was carefully poured off and the microfuge tube was blotted and then dried under vacuum for another 20 min.

2.6 DNA Manipulations

2.6.1 DNA digestions

For diagnostic purposes, a small amount (13.5μl) of DNA was aliquoted to an Eppendorf tube to which a 1/9th volume of 10× restriction buffer was added. The tube was briefly tapped and then 0.5μl of the particular restriction endonuclease was added. The mixture was again tapped and briefly microfuged for a few seconds after which the tube was incubated at the optimum temperature, as specified by manufacturer, for overnight. Bulk digestions was prepared in the same manner, but after incubation the restriction enzyme was deactivated and DNA purified by phenol/chloroform extraction. Double digestions were performed in the most appropriate buffer in which both enzymes showed suitable activity. Otherwise, if no compatible buffer could be found, the digestion then had to be carried out sequentially.

2.6.2 Phosphatase treatment

Linearized plasmid DNA was treated with calf intestinal alkaline phosphatase (CIP) to remove 5' phosphate groups to prevent the recirculation of the vector. To the digested plasmid DNA, a ninth volume of 10× dephosphorylating buffer was added or just left in any of the restriction buffers. One unit of phosphatase enzyme was added and gently mixed by tapping, after which the solution was incubated at 37°C for overnight. After incubation, the phosphatase was heat inactivated at 65°C for 10 min followed by phenol-chloroform extraction.

2.6.3 DNA ligation

Various ratios of vector and insert DNA were mixed together in different Eppendorf tubes and the volume adjusted to 18μl with sterile distilled water where necessary. A 1/9th volume of ligation buffer was added and tube tapped briefly to mix the solution.

Thereafter, 1μl of T4 DNA ligase (Roche/Fermentas) was added and mixed by gentle tapping and then microfuged for a couple of seconds. The ligation mixture was then usually incubated at 14°C for 16 - 22 hrs.

2.7 Agarose gel electrophoresis

A 20 - 30μl aliquot of a molten, autoclaved agarose stock solution was transferred to a conical flask to which about 2μl of ethidium bromide was then added. The contents were poured into a gel casting tray and allowed to solidify at 4°C for 20 - 30 min. The comb was removed from the gel and the tray placed in the electrophoresis unit. Then 5 - 15μl of sample was usually mixed with 2μl of tracking dye and loaded into the wells of the gel. A DNA molecular weight marker was also loaded into a separate well. The agarose gel was submerged with around 200ml of 0.5× TBE running buffer containing 1μg/ml EtBr and then electrophoresed at 10 - 12Vcm^{-1} for about 1 to 1.5 hrs at room temperature. Agarose gels were viewed using a UV Gel Image Master System and photographed with a polaroid camera. The sizes of the unknown DNA fragments were determined by extrapolating from a standard curve of molecular weight versus distance migrated plotted on a semilog paper. The concentration of the DNA was approximated by comparing the fluorescence intensity of the DNA molecular weight marker bands of known concentration with those of the DNA samples.

2.8 DNA extractions methods

2.8.1 Low gelling agarose method

Twenty milliliters of 0.6 - 0.8% low melting agarose in 0.5× TBE buffer was prepared and autoclaved at 121°C for 20 min. Two milliliters of ethidium bromide was added, agarose solution poured onto a tray with a large well comb and then left to polymerize at 4°C for 30 minutes. Aliquots of the DNA samples were mixed with bromophenol blue tracking dye and then loaded into the wells of the gel. The agarose gel was ran at 100V and 4°C for around 2 hrs in a prechilled electrophoresis tank. The gel was illuminated under long wavelength UV light and the particular DNA band were excised with a sterile scalpel and placed into a microfuge tube. The pieces of agarose were melted at 65°C for 30 min and the final volume adjusted to 300μl with TE buffer. Three phenol and one chloroform extractions were performed at room temperature to purify the DNA fragment followed by salt-ethanol precipitation.

2.8.2 Freeze-squeeze method [68]

A 20μl aliquot of 0.4 - 0.8% agarose containing 1μg/ml of EtBr was prepared, poured onto a tray with a large well comb, and allowed to solidify at 4°C for 20 min. Then 35μl aliquots of the DNA samples was mixed with 5μl of tracking dye and loaded into the wells. Electrophoresis tank was filled with a suitable volume of running buffer and then ran at 10 V/cm and 20mA for 1.5 hrs. The DNA fragments of interest were cut out of the gel and agarose slices placed into fresh Eppendorf tubes. The gel pieces were crushed with a sterile spatula or pipette tip and frozen at -80°C for several hours. Following this, the mashed agarose was thawed with a gentle flicking until no ice was visible. After this, the contents were microfuged at 4°C for six minutes. The supernatant was aspirated and transferred to a new sterile tube, where after the agarose was again crushed and frozen at -80°C for another couple of hours. The gel slices was then thawed, microfuged (20000g, 6 min, 4°C) and the supernatant removed to another Eppendorf. The liquid fractions from the separate extractions were pooled together and centrifuged for a further 6 min. The supernatant was yet again transferred to an Eppendorf tube to which a 1/10th volume of 3M sodium acetate (NaAc) and 2 volumes of 96% ethanol was added. The solution was mixed by inversion and placed at -20°C for overnight. The mixture was then microfuged for 15 min at 4°C, after which the supernatant was decanted and tube inverted on a paper towel. This was followed by the addition of 500μl of cold 70% ethanol and the tube vigorously vortexed for a few seconds. Thereafter, the contents were microfuged at 4°C for 5 min and then the supernatant discarded. The Eppendorf tube was blotted on a paper towel and the DNA dried under vacuum for 20 min, after which the pellet was resuspended in 20 - 50μl of sdH_2O.

2.9 Transformation procedures

2.9.1 Transformation into *E. coli*

$CaCl_2$-mediated transformation procedure

E. coli cells were grown in 5ml of LB broth at 37°C overnight on a rotary wheel. A hundredth volume of preculture was used to inoculate 20ml of prewarmed LB containing 0.5% glucose and incubated for 2 hrs at 37°C on a shaker. The culture was transferred to a JA-20 tube and cooled down for 5 min on a ice/water slurry. The cells were then pelleted by centrifugation for five minutes at 10000 rpm in a prechilled rotor. Supernatant was discarded and cell pellet resuspended in a 10ml of $CaCl_2$ transformation buffer and allowed on ice for a minimum of 15 min. The cell suspension was yet again centrifuged at 10000 rpm for 5 min. The liquid was decanted off and pellet resuspended in a 1/15th of transformation buffer, after which the JA-20 tube was left on ice for 2 - 24 hrs.

After the incubation time has elapsed, 100μl of competent cells were transferred to chilled Eppendorf tubes. Then, 2-10μl of foreign DNA was added and mixed by bubbling air though the solution. The tubes were allowed to stand on ice for 10 min or more for diffusion to occur and then heat shocked at 42°C for 90 sec. Following this, 500μl of prewarmed LB was added to the microfuge tubes and then incubated for 1 hr at 37°C with lids open to allow phenotypic expression of antibiotic resistance markers. The cells were suspended by inversion and then spread on LA plates containing 100μg/ml ampicillin which were subsequently incubated for overnight at 37°C.

High efficiency transformation method [67]

A 5ml preculture of *E. coli* MM 294-4 was prepared in LB and grown at 37°C for overnight. One milliliter of preculture was used as inoculum for every 50ml of SOB medium required and then incubated at 30°C for 5 - 7 hrs. Transfer 50ml of culture to a JA-20 and centrifuge for 10 min at 4000 rpm in a prechilled rotor (4°C). The supernatant was decanted and the cell pellet resuspended in 16ml of ice cold TB buffer. After 10 min of incubation, the cells were again collected by centrifugation at 4000 rpm for a another ten minutes. The liquid was discarded and the pellet resuspended in 15ml of chilled TB buffer and 1.12ml of DMSO (7%) which was mixed by gentle swirling. The cells were chilled on ice for a further 10 min or more, after which 500μl aliquots of competent cells were transferred to several microfuge tubes. These tubes were frozen in liquid N_2 for a minute or so and then placed afterwards at -70°C for storage. When needed, the cells were thawed on ice and then 100μl samples were transferred to fresh Eppendorf tubes. To this, 10μl of foreign DNA was added and mixed by bubbling air through the solution and then allowed on ice for 30 min, where after the cell suspension was heat shocked for 30 sec at 42°C. After the addition of 800μl of SOC medium, the mixture was incubated at 37°C while shaking for an hour. The suspension of cells were inverted and spread on LA Amp 100 plates which were then incubated at 37°C for overnight.

2.9.2 Nocardioform transformation

***Rhodococcus* PEG-mediated transformation method**

Precultures of the recipient *R. erythropolis* strains SQ1 and Ri8R were prepared in 5ml of 2% and 3% glycine supplemented LBSG media containing 50μg/ml of rifampicin, respectively. The cells suspensions were grown at 30°C until the cultures have reached stationary phase (2 - 3 days). A 0.5 - 1ml aliquot of preculture was transferred to a sterile Eppendorf tube and microfuged for 30 sec. The cell pellet was washed with 1ml of B buffer and then yet again centrifuged for another 30 sec.

The B buffer was decanted and cells were gently resuspended in 1ml of basal medium containing a speck of lysozyme (5μg/ml). The microfuge tube was then incubated in a 37°C waterbath for 1.5 - 2 hrs with inversion every 10 min. During the incubation time, a suitable volume of P buffer was prepared and an appropriate amount of polyethylene glycol (PEG) weighted out. The PEG was UV sterilized for 10min and then dissolved by vigorous vortexing in the correct volume of P buffer to obtain a 50% w/v P-PEG solution. After incubation, the resultant protoplast suspension was microfuged for 15 sec. The buffer was discarded and pellet rinsed in 1ml of P buffer and then centrifuged for 30 sec. The pellet was again resuspended in the required volume of protoplast medium, where after 100 - 200μl was transferred to separate Eppendorfs. Small amounts(5 - 20μl) of the DNA were added to the microfuge tubes and the contents mixed by gently bubbling air through the solution. The tubes were then left on the bench for about 15 min for diffusion to occur. Following this, an equal volume of P-PEG was added to each tube and then bubbled until homogeneity was achieved. The PEG-protoplast suspension mixture was spotted and spread on chilled regeneration plates. The agar plates were incubated at 30°C for 10 - 15 hrs for phenotypic expression of the Cm^R selective marker gene. Afterwards, the plates were then underlayed with 220 - 250μl of a 4μg/ml stock solution of chloramphenicol ($V_f = 40\mu$g/ml). The plates were then incubated at 30°C until colonies were visible, which was usually after 4 - 7 days.

2.10 DNA Screening

Ten milliliters subcultures of the *EcoR*I disrupted pDA71* vector and the mixed genomic library in SQ1 was prepared in 2%-glycine supplemented LBSG media containing 50μg/ml rifampicin and then grown for 2-3 days at 30°C. A small portion (200-500μl) of these cultures were briefly spun down in a microfuge and then resuspended in a similar volume of distilled water to remove media components so that no growth can be attributed to any residual quantities of these contaminants during the screening process. Serial dilutions (10^{-1} - 10^{-6}) were prepared for these two cultures, of which 100μl aliquots were used to spread on the agar plates containing chloramphenicol (20-40μg/ml). The original pooled library could alternatively also be used for making up the dilutions. A chloramphenicol stock solution was made in methanol as SQ1 is able to utilize ethanol as sole carbon source.
For the screening for arsenic resistance the typical volume of the 10^{-6} dilution for the library and control was plated on LA Cm40 plates to ascertain the cell viability (i.e. cfu/ml) of the cultures and to compare the growth difference with and without arsenic. Usually, the 10^{-2} dilution of the two cultures was used to spread on half of LA Cm40 plates containing either 10 mM AsO_2^- or 100mM AsO_4^{3-}, but other lower dilutions can also be used as well if necessary.

With the screening for unique carbon utilization genes, the same 10^{-6} dilutions of the library and control in SQ1 was plated on minimal media (MM) containing chloramphenicol (20μg/ml) and nystatin (50μg/ml) as well as 0.5% glucose, while the 10^{-2} dilutions were spread on MM Cm20 Nys50 plates containing the desired amounts of the various selected chemicals. All the screening plates were incubated at 30°C for a maximum of ten days.

Chapter 3

Results

A brief chronological account of the failed and unsuccessful screening of first a *Sfu*I digested library of *Rhodococcus rhodochrous* ATCC 12674, second of incomplete constructed *Bgl*I and partial *BamH*I digested *Arthrobacter oxydans* C1 libraries, and third of a *Bgl*II *Gordonia rubropertincta* ATCC 25593 genomic library will be described initially in this result section. Also, the unsuccessful cloning of As^R fragments from pDA37 into pUC18 will be briefly discussed as well. Then the construction and screening of a partially *Pst*I digested library of *G. rubropertincta* 25593 will be elaborated which ultimately led to the isolation of definitive arsenic resistant clones.

3.1 Analysis of *Rhodococcus rhodochrous* for arsenic resistance

3.1.1 Transformation of *Sfu*I library of *Rhodococcus rhodochrous* 12674 into an As-susceptible Gram(+) background

A nocardioform genomic library that was made by N. Ahmed (unpublished) previously was first transformed into the As^S *R. rhodochrous* strain Ri8. Besides that the transformation was not successful, it was decided not to use this bacterial strain in the future due to the possibility of homologous recombination as this was a mutant strain of 12674. The DNA of this *Sfu*I digested genomic library was subsequently transformed into the As-sensitive *Rhodococcus erythropolis* strain SQ1 using the polyethylene glycol mediated protocol. After 7 days of incubation at 30°C approximately 5000 clones were obtained

3.1.2 Screening of *Sfu*I library of *R. rhodochrous* 12674 for arsenic resistance

A smear of all the clones were spread onto half of LA Cm40 plate containing 100mM AsO_4^{3-} and/or 15mM AsO_2^-. A *Pst*I library of the As-susceptible *Nocardia brasiliensis* in SQ1, which served as a negative control, was spread onto the other halve of the agar plates. This was also performed on a LA Cm40 plate only as a control as well as to compare the growth of the two libraries with and without supplemented arsenic. After the plates were grown for 3 days at 30°C no satisfactory conclusion could be made regarding arsenic tolerance. Therefore, it was necessary to spot test these libraries onto a LA agar plates supplemented with arsenic in order to reach a definitive confirmation of arsenic resistance.

3.1.3 As^R phenotypic assay of nocardioform strains and libraries

These two above mentioned genomic libraries together with several additional nocardioform strains, such as *R. australis* strains 448 and 554, *R. rhodochrous* Ri8, *R. erythropolis* SQ1, *Arthrobacter oxydans* C1, and *Gordonia rubropertincta* 25593, were replica plated onto a LA only agar plate and one supplemented with 50mM AsO_4^{3-}. The results after 72 hrs of incubation at 30°C are summarized in the table below.

Organism	**LA Arsenate (mM)**	
	0	**50**
*Sfu*I *R. rhodochrous* ATCC 12674 GL in SQ1	+++	-
*Pst*I *N. brasiliensis* GL in SQ1	+++	-
R. erythropolis SQ1	+++	-
R. rhodochrous Ri8	+++	-
R. australis 448	+++	+++
R. australis 554	+++	-
A. oxydans C1	+++	+++
G. rubropertincta ATCC 25593	+++	+++

where - = no growth and +++ = excellent growth

Table 3.1: Spot test results of nocardioform strains and libraries for arsenic resistance.

From the spot testing it was concluded that *R. rhodochrous* ATCC 12674 does not confer high level resistance to arsenic. Due to this finding led to the construction of a genomic library of the As^R *Arthrobacter oxydans* C1 to ultimately screen for resistance genes in this organism.

3.2 Construction of *Arthrobacter oxydans* C1 library

The construction of genomic libraries was performed in order to facilitate the isolation and characterization of arsenic resistance genes. Due to the slow growth rate as well as the lack of suitable cloning systems has hindered the isolation and investigation of the genes of nocardioform actinomycetes species in the past. Work done by Dabbs and colleagues led to the development of a series of *E. coli-Rhodococcus* shuttle vectors for cloning purposes [**42, 112**]. As a result of these vectors it was possible to construct genomic libraries initially in *E. coli* where DNA manipulation and analyzing recombinant constructs are much easier than in *Rhodococcus*. Also, transformation in *E. coli* is far more efficient and less time consuming than in Rhodococcal strains. Before detailing the methodology of constructing the library, a brief description of the cloning shuttle vector pDA71 that will be used for its construction and how it was developed would be in order.

3.2.1 Properties of cloning vectors pDA37 and pDA71

Initially a series of plasmids were created by combining the replication genes of a nocardiophage with determinants from an unstable genetic element of *R. rhodochrous* ATCC 12674. The generalized transducing phage Q4 was used to transfer the As^R phenotype from strain ATCC 12674 to a cured strain without this extrachromosomal element [**40, 41**].
pDA21 that was produced in this manner was utilized in the consequent construction of nocardioform cloning vectors pDA37 and pDA71 by joining the rhodococcal replicon with the *E. coli* vector pEcoR251. This vector carries a positive selection function which is encoded by the restriction endonuclease *EcoR*I gene. Therefore, cloning foreign DNA into this suicide gene allows for the direct selection of recombinants since the expression of the non disrupted enzyme in the unaltered pDA71 vector will ultimately result in cell death. The suicide gene was fused in frame with the regulatory elements of λ phage so that the expression is controlled by the temperature-sensitive lambda repressor [**37**]. Consequently such vectors are maintained in λ lysogens where the expression of this endonuclease is switched off by the repressor. The ampicillin resistance marker encoded by the *bla* gene was used for selection of transformants in a non lysogenic *E. coli* strain. pDA37, that was obtained by the ligation of an As^R nocardioform plasmid with pEcoR251, had several unfortunate drawbacks, such as low copy number, narrow host range, and at 14.6 kbp was large.
These shortcomings were dealt with by deleting the identified regions responsible for copy number and host range and replacing the arsenic resistance determinant(s) with a non enzymatic Cm^R marker from *R. rhodochrous* ATCC 12674 to acquire the 8.8 kbp pDA71 shuttle vector [**112**].

The chloramphenicol resistance determinant is utilized when selecting for clones in *Rhodococcus*. The one disadvantage still about this vector is the limited number of unique restriction sites in the *EcoR*I gene for cloning purposes.

3.2.2 Restriction digestion of *A. oxydans* C1 genomic DNA

In order for a particular restriction enzyme to be suitable for library construction it needs to meet certain criteria: 1) must produce DNA fragments greater than 2kbp so that the likelihood of obtaining intact genes are relatively high; 2) must cut sufficiently frequent enough; and 3) must contain a unique site in the cloning region or produce cohesive sticky ends to one. The *EcoR*I endonuclease gene has only unique restriction sites for *Hind*III, *Bgl*II, *Pst*I, and *Sfu*I. Overnight digestions of 13.5μl aliquots of *A. oxydans* genomic DNA were performed with these four enzymes and their isocaudomers (i.e enzymes that recognizes slightly different palindromic sequences but still generate compatible ends) at their respective appropriate temperatures after which these samples were run on a 0.8% agarose gel for 1.5 hrs at 10V/cm.

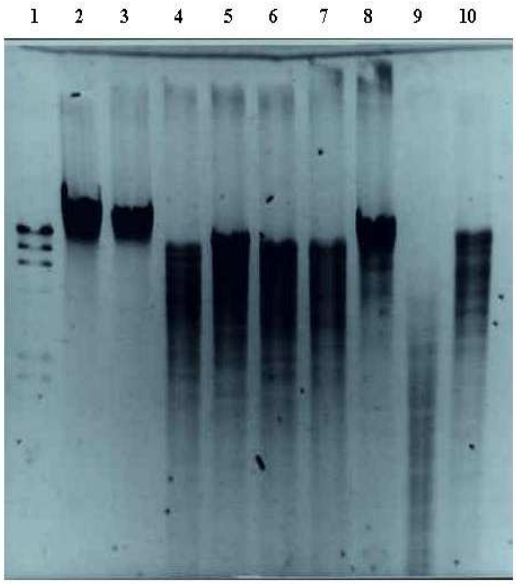


Figure 3.1: Restriction digestions of *A. oxydans* genomic DNA; Lanes: 1) λII molecular weight marker, 2) uncut DNA, 3) *Hind*III, 4) *Pst*I, 5) *Bgl*II, 6) *BamH*I, 7) *Bcl*I, 8) *Nsi*I, 9) *Acc*I, and 10) *Cla*I.

It were evident that the endonucleases *Hind*III and *Nsi*I were cleaving the genomic DNA of *A. oxydans* C1 too rarely, whereas *Acc*I were cutting it too frequently and consequently were not appropriate for library construction.

The *Cla*I restriction enzyme digested the chromosomal DNA irregularly for it to be used in the genomic library. Even though the four remaining endonucleases all were restricting the DNA of *A. oxydans* C1 sufficiently frequent enough, it was decided to construct a *Bgl*II (A↓GATCT) and a partial *BamH*I (G↓GATCC) digested library of this organism.

3.2.3 Optimization of the vector to genomic DNA ratio for ligation

Vector calibration were performed where different concentrations of pDA71 (i.e 1,3,5, and 10μl) were transformed into the non lysogenic *E. coli* strain MM294-4. The optimum amount of pDA71 were evaluated on the quantity of *endR1* mutants growing on a LA Amp200 agar plate. A concentration of pDA71 that would produce 1-3 mutants were adjudged to be a suitable amount to conduct ligations with. After several attempts of transformation it was decided to use 3μl of pDA71 in the subsequent ligations. Vector to genomic DNA ratios such as 1:1, 1:2, 1:3, 1:4 and 1:5 were tested out and it was found that the 3μl plasmid to 6μl insert DNA combination (i.e. 1:2) resulted in the greatest number of transformants.

3.2.4 Transformation of the cloned genomic DNA into *E. coli*

After the ligation conditions had been optimized, the pDA71 constructs were transformed into *E. coli* MM294-4 using the calcium chloride mediated transformation procedure. Plasmid DNA of (∼10-20) random colonies from both libraries were extracted and digested with the appropriate restriction endonuclease(s) to release the cloned fragments in order to analyze the insert frequency and average insert size. The recombinant plasmid DNA of the clones from the *Bgl*II digested *A. oxydans* C1 library were digested again with the same restriction enzyme and then ran on a 0.8% agarose gel (Fig3.2A).
As the *Bgl*II/*BamH*I hybrid site (A↓GATCC) can't be restricted by these two endonucleases anymore, it was necessary to digest the *BamH*I/*Bgl*II recombinant clones' plasmid DNA with the *Hind*III and *Pst*I restriction enzymes that flanks the *Bgl*II site to release the inserted fragments. After an overnight double digestion with these two restriction endonucleases, the DNA samples were electrophoresed on a 0.8% gel. (Fig3.3A).
Using a calibration curve the molecular weights of the insert fragments were determined and with the ligation efficiency the average insert size (a) was calculated. In case of the *BamH*I clones, 420bp had to be subtracted from the insert sizes to account for the *Hind*III/*Pst*I flanking region before calculation could be performed. Small fragments were taken as negligible and assumed as zero in the final computational result. The average insert size values for the *BamH*I and *Bgl*II digested genomic libraries were determined to be 1.8 kbp and 2.3 kbp, respectively. The genome size of *Arthrobacter oxydans* C1 were approximated to be 6.0 Mbp.

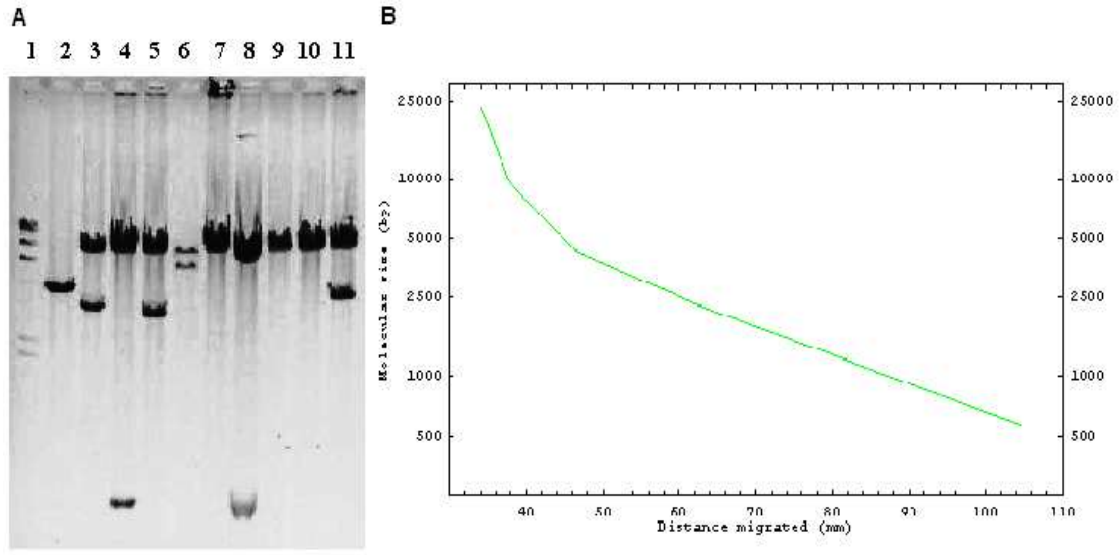


Figure 3.2: A) Insert analysis of random *Bgl*II *A. oxydans* C1 genomic library clones; Lanes: 1) λII molecular weight marker, 2) linearized pDA71 mutant, 3-11) *Bgl*II digested recombinant clones B) Corresponding molecular size vs distance migrated calibration plot of λII molecular weight markers.

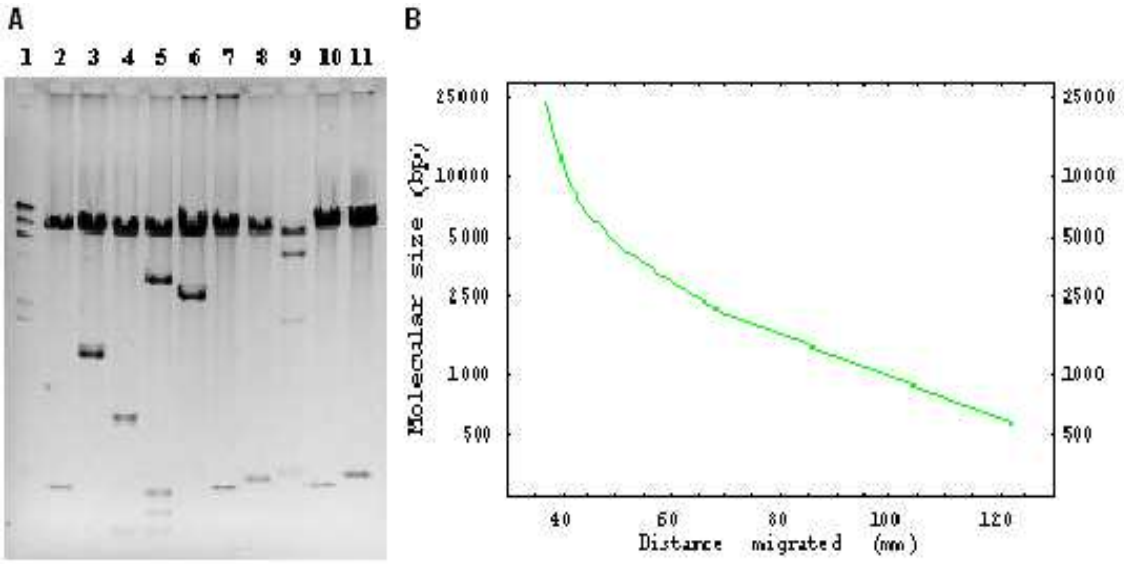


Figure 3.3: A) Insert analysis of random *BamH*I *A. oxydans* C1 genomic library clones; Lanes: 1) λII molecular weight marker, 2-11) *Hind*III/*Pst*I digested recombinant clones B) Corresponding molecular size vs distance migrated calibration plot of λII molecular weight markers.

The number of clones that would be required to give a 95% probability of finding any part of the organism's genome was calculated using the following equation:

$$N = \frac{\ln(1-P)}{\ln(1-\frac{a}{b})} \tag{3.1}$$

where N = No. of clones
P = Probability
a = Average insert size
b = Genome size

Putting in the values of the variables it was found that a minimum of 10300 and 7800 clones would be required for the *BamH*I and *Bgl*II *A. oxydans* C1 libraries at this probability setting. But due to poor ligation efficiency as well as the absence of any more live cells of *A. oxydans* to get more DNA from, only about 4000 and 3500 clones were obtained for the *BamH*I and *Bgl*II libraries, respectively. Therefore, the probabilities for these libraries were merely 68.9% and 74.1%, respectively. Regardless of this incomplete status, the clones of both these libraries were separately pooled together and the plasmid DNA extracted from 100ml subcultures which were grown overnight at 37°C.

3.2.5 Transformation and screening of *Bgl*II and *BamH*I *A. oxydans* C1 libraries in As^S Gram(+) background

The isolated plasmid DNA of both libraries was transformed in the As-sensitive *R. erythropolis* strain SQ1 as screening needed to take place in a organism which is not resistant to arsenic and which the possible As^R determinants will be expressed and function in. The probability that the arsenic resistant gene products from a high G+C gram positive bacteria would properly be expressed and function in *E. coli* is very doubtful. About 10000 clones were obtained for both libraries and their respective clones were pooled together subsequently.
The mixed pooled cultures of both *A. oxydans* C1 libraries were screened for the As^R phenotype, but after approximately 7 days of incubation no positive clones was visible on the arsenic-containing LA Cm40 plates. This could be due to the libraries not been fully representative of the entire organism's genome.

3.3 Cloning of pDA37 DNA fragments into cloning vector pUC18

It was also attempted to subclone fragments of the As^R *E. coli-Rhodococcus* shuttle vector pDA37 into pUC18 to ultimately sequence the genes conferring arsenic resistance. Firstly, preliminary overnight digestions at 37°C with *Pvu*I, *BamH*I and *Pst*I restriction endonucleases were performed on this vector and then samples ran on a 1.0 agarose gel at 10V/cm for 1.5 hrs. The *Pst*I enzyme produced three reasonably sized fragments that was adjudged to be suitable for cloning.

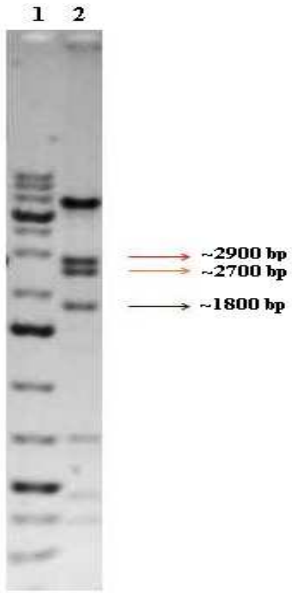


Figure 3.4: *Pst*I digestion of pDA37; Lanes: 1) Generuler DNA Ladder mix marker, 2) *Pst*I digested pDA37

Those three *Pst*I fragments of interests were cut out of the gel and purified using the freeze-squeeze procedure. Vector calibration were done to determine the optimum amount of pUC18 to used for ligations. After pUC18 was treated with phosphatase to prevent vector ligating to itself, ligations were set with this vector and the *Pst*I DNA fragments which also included several negative controls. Unfortunately, all attempts in cloning these *Pst*I inserts into pUC18 failed possibly due to their being instable as it came from an unstable element from *R. rhodochrous* ATCC 12674 or even the overexpression of a plausible ArsB protein that is lethal to bacterial cells [**163**]. Moreover, even when transformants were obtained and the vector only control demonstrated few colonies, plasmid screening showed either no or only small inserts still.

3.4 Analysis of a *Bgl*II *Gordonia rubropertincta* library for arsenic resistance and other genes

3.4.1 Transformation of *Bgl*II digested *G. rubropertincta* genomic library into an arsenic-sensitive Gram(+) recipient strain

A *Bgl*II digested genomic library of *Gordonia rubropertincta* ATCC 25593, that was constructed by a previous postgraduate S. Quan (unpublished), was transformed in the rhodococcal strain SQ1 using the PEG-mediated transformation protocol. At first the transformation were not efficient and successful and it was necessary to reconstruct this library again in *E. coli* MM294-4 to obtain more recombinant plasmid DNA of the library. The residual genomic library plasmid DNA was transformed into *E. coli* and the ∼20000 clones obtained were pooled together and mixed thoroughly, after which the recombinant constructs' DNA were isolated from a 100ml overnight subculture.

The protoplast transformation method was optimized by changing several procedural variables independently such as the lysozyme amount and incubation period, amount of cells used per transformation reaction, as well as phenotypic expression duration time. Different quantities of lysozyme were added qualitatively to the cells and incubated at 37°C for periods of 30, 60, 90, 120, and 240 min in order to establish the optimum incubation time which was found to be 1.5 hrs. The amount of competent protoplasts per transformation reaction were varied between 100-200μl but this had only minor effect on the efficiency of the transformation process. The time allowed for phenotypic expression was tested and 8, 12, and 18 hr periods were tried out and it was found that 12-14 hrs of incubation to be ideal. The genomic libraries' DNA were again transformed at these optimized conditions and the ∼20000 clones that were attained were combined and then mixed on a shaker for one hour at 30°C.

3.4.2 Selection for arsenic resistance in *Bgl*II *G. rubropertincta* library.

Even though several hundred putative arsenite resistant were obtained from the *Bgl*II *G. rubropertincta* genomic library initially, the screening was repeated to see whether the result were reproducible, which was in fact found to be the case. The averaged results from the screening of the *Bgl*II digested *G. rubropertincta* ATCC 25593 genomic library for the As^R phenotype are summarized in the following table.

LA Cm20 [Arsenic]	10^{-1} Dilution		10^{-2} Dilution		10^{-6} Dilution	
	pDA71*	GRub	pDA71*	GRub	pDA71*	GRub
0mM As	ND	ND	ND	ND	26	73
10mM AsO_2^-	3	~500	0	47	ND	ND
100mM AsO_4^{3-}	15	43	1	5	ND	ND

Table 3.2: Averaged results of screening of the *Bgl*II digested *G. rubropertincta* ATCC 25593 genomic library in SQ1 for arsenic tolerance, where GRub - *Bgl*II *Gordonia rubropertincta* library, pDA71* - SQ1 with *EcoR*I disrupted pDA71 vector and ND - not determined.

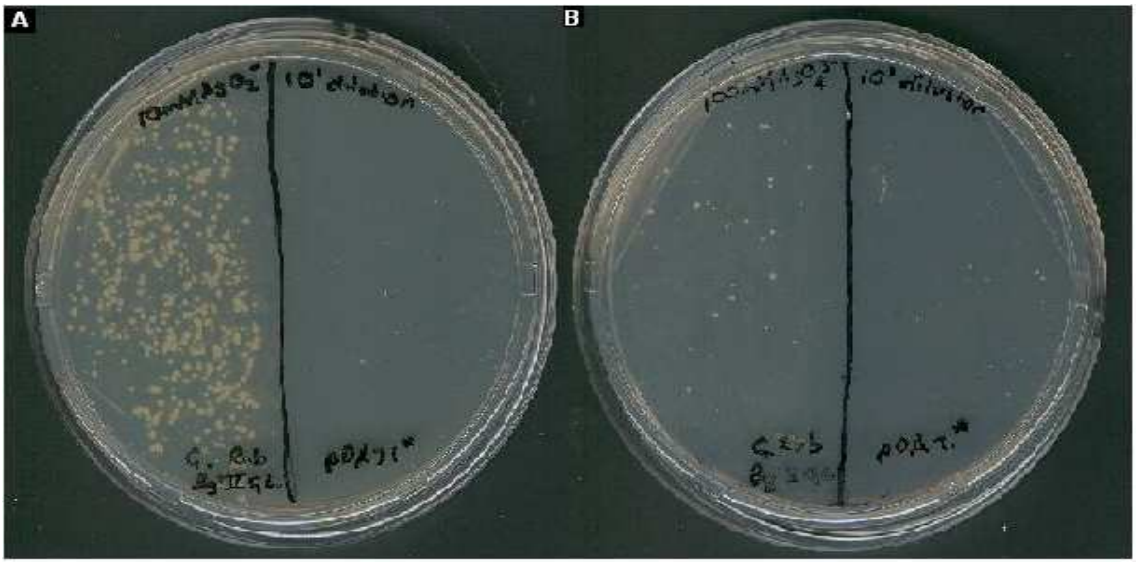


Figure 3.5: A) Positive arsenite resistant clones and B) arsenate resistant clones from screening of *Bgl*II digested *G. rubropertincta* ATCC 25593 genomic library in SQ1 with the library on the left and control on the right side of the plates.

3.4.3 Confirmation of As^R phenotype for the putative clones

After the recombinant plasmid DNA of 10 random As^R clones were isolated using the *Rhodococcus* mini plasmid preparation procedure, a small amount of each clones' plasmid DNA (5μl) was transformed into *E. coli* MM294-4. The recombinant DNA was again isolated from overnight minicultures of the putative arsenic resistant clones in *E. coli* and then digested for 16 hrs with the endonuclease *Bgl*II. The samples were ran on a 0.8% agarose gel in order to analyze the presence of inserts and their corresponding molecular sizes. Even though most of the putative clones demonstrated no inserts, it was thought that a plasmid rearrangement or mutations might account for this finding.

Therefore, the original pDA71 and the positive clones were then restricted by the enzymes *EcoR*I and *BamH*I overnight at 37°C to investigate whether there were any differences in their digestion patterns. Indeed, in some of the clones that showed initially no insert actually revealed an altered restriction pattern from that of control pDA71 vector. Regardless, several of these putative clones' DNA were transformed into 25μl of competent protoplast cells which were subsequently spotted on distinct areas of regeneration plates and then grown at 30°C for 4 - 7 days. Finally, the As^R phenotype for the individual putative clones in SQ1 were again tested but none of these recombinant constructs conferred arsenic resistance that was initially observed.

3.4.4 Searching the *Bgl*II *G. rubropertincta* library for biodegradation and unique carbon source metabolizing determinants

At the same time this *Bgl*II digested library of *Gordonia rubropertincta* ATCC 25593 was also screened for several unique catabolic genes. From a previous thorough carbon source utilization assay for various nocardioform strains done by S. Quan (unpublished), several compounds were identified for further analysis. The preliminary carbon source utilization experimental data on these selected substances for *G. rubropertincta* and the wild-type strain of SQ1, *Rhodococcus erythropolis* ATCC 4277, are listed in the following table.

Carbon Source	**Concentration (%)**	**ATCC 25593**	**ATCC 4277**
Sucrose	0.050	+	±
Mannose	0.100	+	-
Isopropanol	0.100	+	-
Sodium Benzoate	0.050	+	-
Syringic Acid	0.050	+	-
Cinnamyl alcohol	0.002	+	-
m-Cresol	0.050	+	-

where + = good growth, ± = variable growth, and - = no growth

Table 3.3: Preliminary carbon source utilization data for *G. rubropertincta* ATCC 25593 and *R. erythropolis* ATCC 4277.

Although, putative clones were obtained for all the selected compounds, except m-cresol, from the screening of this library on the minimal media plates, emphasis was placed on the positive clones for the utilization and/or degradation of the ligninolytic compounds syringic acid and cinnamyl alcohol as well as for sodium benzoate.
To insure that the growth were in fact recombinant *Rhodococcus* colonies, a 2ml sloppy agar layer containing 10μl of the combined clones for the various chemicals were poured onto LA plates containing 10mM $CaCl_2$.

After the plates had dried at 42°C for 10min, 5μl of three *Rhodococcus* specific phages, isolated by Y. Shibayama (2006), were spotted on separate areas of these agar plates and incubated at 30°C for 2-3 days to ascertain whether they plaque on these putative rhodococcal cells. As all these phages formed clearing zones within the lawn of bacterial growth, it was conclusively affirmed that those colonies were indeed recombinant SQ1 clones.
Consequently, the plasmid DNA of the pooled putative clones for these compounds were extracted and transformed into *E. coli*. The recombinant DNA was again extracted and then digested with the restriction enzyme *Bgl*II in order to investigate the presence of inserts and the corresponding size of the cloned fragments. For some of the positive clones for syringic acid and cinnamyl alcohol conserved DNA fragments were observed, while other constructs either were not linearized or showed no insert after *Bgl*II restriction.
This led to the original pDA71 vector as well as the putative recombinant constructs for these substances to be digested by *BamH*I and *EcoR*I as with the arsenic positive clones previously. Once more, altered digestion patterns were found from that of pDA71 for some of the clones that showed initially no inserts, which yet again infers the plausibility of plasmid rearrangements and mutations within the *Bgl*II site. Putative individual clones for these various compounds were separately transformed into 25μl of competent SQ1 protoplasts followed by incubation for 7-10 days at 30°C. As with the positive arsenic resistant clones, these putative recombinant constructs did not exhibit the catabolic ability to degrade these ligninolytic and other compounds it showed earlier.

3.5 Construction and analysis of a partially *Pst*I digested library of *Gordonia rubropertincta*

3.5.1 MIC determination of *G. rubropertincta* and negative controls for As

Before the genomic library was constructed, a detailed arsenic tolerance phenotypic assay on *G. rubropertincta* ATCC 25593 as well as for the Gram(-) and Gram(+) recipient strains *E. coli* MM294-4 and *R. erythropolis* SQ1 was performed. These bacterial cultures were spot tested on LA Nal20 Nys50 agar plates supplemented with various concentrations of AsO_4^{3-} or AsO_2^-. The results of this replica plating after 7 days at 30°C are recapitulated in the following table.
Therefore, the minimum inhibitory concentration for arsenite for the *E. coli*, *R. erythropolis*, and *G. rubropertincta* were found to be 2mM, 3mM and 7.5mM, respectively. Also, MICs of around 2.5 mM, 3mM, and 250mM arsenate were established for these bacterial strains. The MIC values were taken where the particular microorganisms started growing only slightly.

[Arsenic]	**Degree of growth**		
Arsenite (mM)	***E. coli* MM294-4**	***R. erythropolis* SQ1**	***G. rubropertincta***
0	+++	+++	+++
1	++	++	++
2	+	++	++
3	±	+	++
4	-	-	++
5	-	-	++
7.5	-	-	+
10	-	-	-
Arsenate (mM)			
1	+++	+++	+++
2	++	++	+++
3	+	+	+++
4	±	-	+++
5	±	-	+++
10	±	-	+++
25	-	-	+++
50	-	-	++
100	-	-	++
250	-	-	+
500	-	-	±

where +++ = excellent, ++ = moderate, + = slight, ± = scarce, and - = no growth

Table 3.4: Spot test results of *G. rubropertincta* ATCC 25593, *E. coli* MM294-4 and *R. erythropolis* SQ1 on solid agar media containing various concentrations of arsenic.

3.5.2 Optimization of the partial digestion conditions for *G. rubropertincta* genomic DNA

As the restriction enzyme *Pst*I (C↓TGCAG) cut the genomic DNA of *Gordonia rubropertincta* too frequently, it was necessary to determine the ideal conditions for a partial digestion that would result into consistent and reasonably sized DNA fragmentation pattern. Serial dilutions of this endonuclease, that initially included 2×, 4×, 16×, and 64×, were prepared in O buffer of which a 1/9 vol of each dilution was added to 9μl of genomic DNA and then incubated for either 30 or 60 minutes. The samples were run on a 0.4% agarose gel at 10Vcm^{-1} for 1.5 hrs. This gave a rough indication in which range to work with, and subsequently *Pst*I digestions with 4×, 16×, 32×, 64×, and 128× enzyme dilutions were done on the genomic DNA for overnight. The *Pst*I restricted samples were again electrophoresed on a 0.8% gel for 90 min at 80V and 20mA. From these gels it was determined that a dilution in between 16× and 32× to be suitable for library construction.

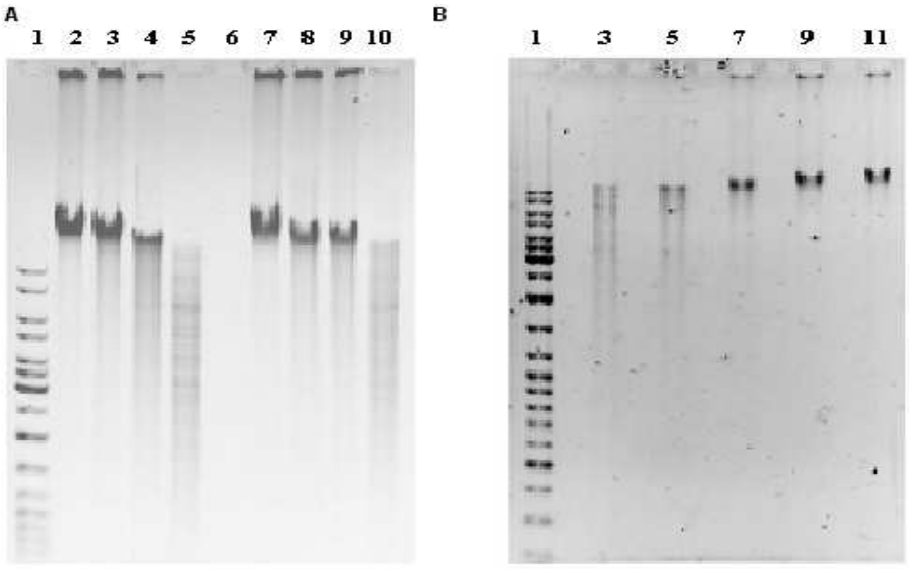


Figure 3.6: Optimization of the partial *Pst*I digestion of *G. rubropertincta* genomic DNA; A) Partial *Pst*I digestions of *G. rubropertincta* genomic DNA with time varied; Lanes: 1) Generuler DNA Ladder mix marker, 2-5) Digestion of genomic DNA with 64×, 16×, 4×, and 2× *Pst*I endonuclease dilutions for 30 min, and 7-10) for 60min, B) Partial *Pst*I digestions of *G. rubropertincta* genomic DNA with time constant; Lanes: 1) Generuler DNA Ladder mix marker, Digestions of genomic DNA with 3) 4×, 5) 16×, 7) 32×, 9) 64× and 11) 128× *Pst*I enzyme dilutions overnight.

3.5.3 Optimizing the ligation conditions and transformation of the pDA71 constructs

As before, the concentration of pDA71 for ligations was calibrated using 1-10μl, and two microliters was found to be optimum. After testing for various vector to genomic DNA ratios, such as 1:1, 1:3 and 1:5, it was established that the 2μl plasmid to 10μl insert DNA to be the best permutation. Transformations of ligations at this ideal ratio were performed and then recombinant plasmid DNA of 10 random clones were isolated from 1ml minicultures. The recombinant construct DNA of these clones were restricted with *Pst*I overnight and ran on a 0.8% agarose gel.

The molecular weights of the inserted genomic DNA fragments were determined from a calibration curve, after which the average insert size (a) was calculated to 2.9kbp. The number of clones necessary to achieve a probability of 95% of finding any part of the microbe's genome in the library was computed using equation 3.1. After substituting the value of a and assuming the genome size to be 6300 kilobase pairs, it was found that 6200 clones would be required.

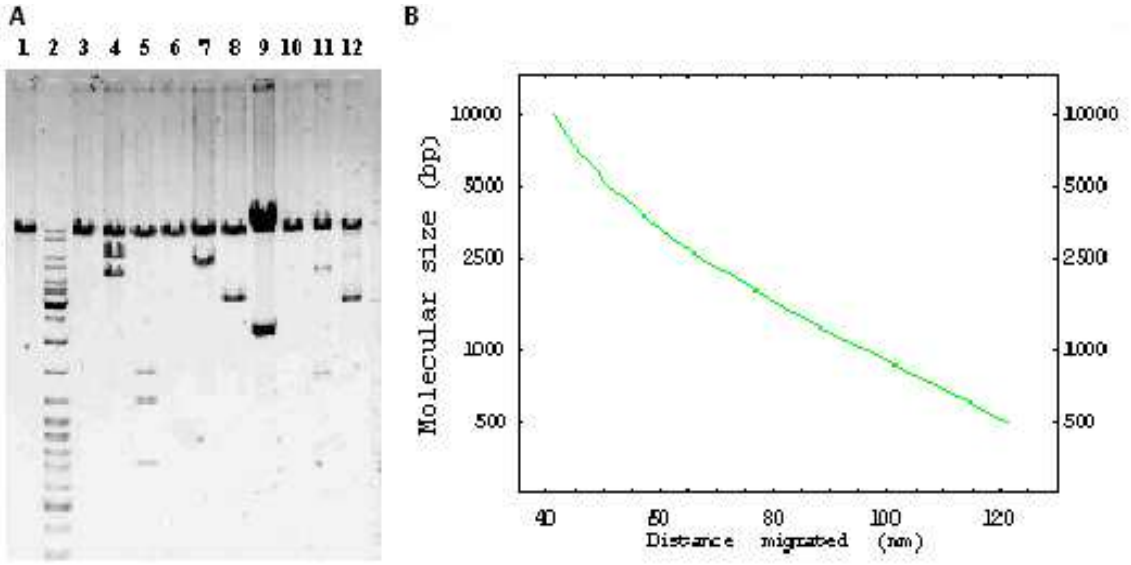


Figure 3.7: A) Insert analysis of partially *Pst*I digested *G. rubropertincta* 25593 genomic library clones; Lanes: 1) linearized pDA71 mutant, 2) Generuler DNA Ladder mix marker, 3-12) *Pst*I digested recombinant clones B) Corresponding molecular size vs distance migrated calibration plot of Generuler DNA ladder mix markers

As the transformation efficiency was very high, an approximate 18000 recombinant clones were obtained that were subsequently pooled together and agitated for an hour on a shaker at 37°C. Therefore, the probability of any DNA fragment being present in this partially *Pst*I digested library of *G. rubropertincta* ATCC 25593 was evaluated to be about 0.999. The pDA71 constructs' DNA was extracted from a 100ml overnight subculture culture and then transformed into the *R. erythropolis* SQ1 recipient strain again. The approximately 39000 clones that were attained were typically combined and agitated for 1hr at 30°C.

3.5.4 Screening of *Pst*I *G. rubropertincta* ATCC 25593 genomic library for unique catabolic genes

Yet again, this partially *Pst*I digested library of *Gordonia rubropertincta* was checked for the ability to degrade and/or utilize those previously mentioned carbon sources, but after 3 days of incubation at 30°C growth was obtained on both halves of the MM screening plates, except for m-cresol where no growth was observed for either culture. After repeating the screening and achieving the same result, no further attention was placed on this task.

3.5.5 Selecting for arsenic resistance in *Pst*I *G. rubropertincta* library

In order to fast track the isolation of As^R determinants in the *Pst*I *G. rubropertincta* genomic library, the marker rescue technique was applied to screen directly for plasmid-borne arsenical resistance in *E. coli*. The usual amount of a 100th dilution of the library and pDA71* in *E. coli* were spread on LA Amp200 supplemented with either 100mM AsO_4^{3-} or 10 mM AsO_2^- and allowed to incubate for 3-5 days at 37°C. Although, some prospective colonies were obtained, growth was also observed on the negative control side, so no clear conclusion could be reached on the validity of these putative clones and were therefore abandoned.

As a result, 100μl aliquots of various dilutions, such as 1/1 (undiluted), 1/10, 1/50, and 1/100, of both pDA71* and library in SQ1 were spread on halve of LA Cm40 Nal25 Nys50 plates supplemented with either 100mM AsO_4^{3-} or 10 mM AsO_2^-, whereas the 10^{-6} dilution were spread on the same solid agar media without arsenic. These screening plates were all incubated at 30°C for 3-10 days depending on whether the media contained arsenic as it caused a noticeable lag growth period. Even though numerous positive clones for arsenate were obtained for the undiluted subculture the screening were repeated with the original pooled library culture as cfu/ml of below 10^7 were only achieved. In this instance only the 10^{-2} dilution were used to spread on the arsenic-containing plates. The results from the arsenic resistance screening are summarized in the following table.

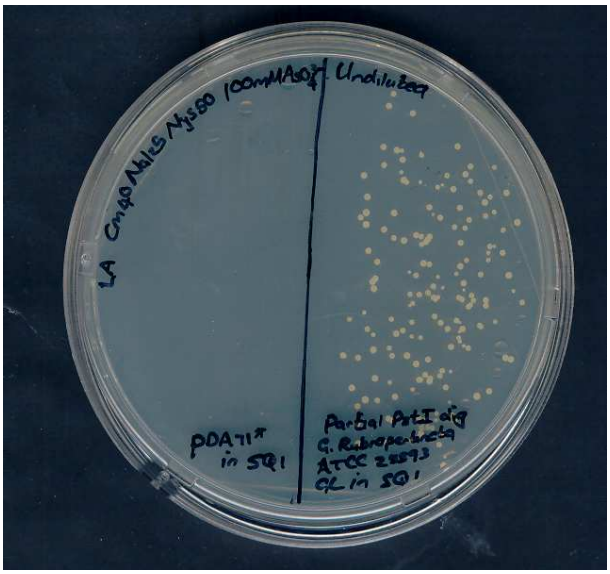


Figure 3.8: Positive As^R clones from screening of the partially *Pst*I digested *G. rubropertincta* ATCC 25593 genomic library in SQ1 with control on the left and library on the right.

LA Cm20 Nal25 Nys50 [Arsenic]	**10^{-2} Dilution**		**10^{-6} Dilution**	
	pDA71*	**GRub**	**pDA71***	**GRub**
0mM As	ND	ND	355	373
10mM AsO_2^-	0	3	ND	ND
100mM AsO_4^{3-}	41	206	ND	ND

Table 3.5: Results of screening of the partially *Pst*I digested *G. rubropertincta* ATCC 25593 library in SQ1 for arsenic resistance, where GRub - *Pst*I *Gordonia rubropertincta* library, pDA71* - SQ1 with *EcoR*I disrupted pDA71 vector and ND - not determined.

3.5.6 As^R phenotypic assay of the putative clones in SQ1

The recombinant plasmid DNA of ten random arsenate resistant clones were isolated from 1ml overnight minicultures and transformed back into *E. coli*. The extracted plasmid DNA of the individual constructs, designated pKL1-10, were then digested with *Pst*I overnight again and ran on a 0.8 % gel at 10V/cm to analyze the sizes of the inserts.

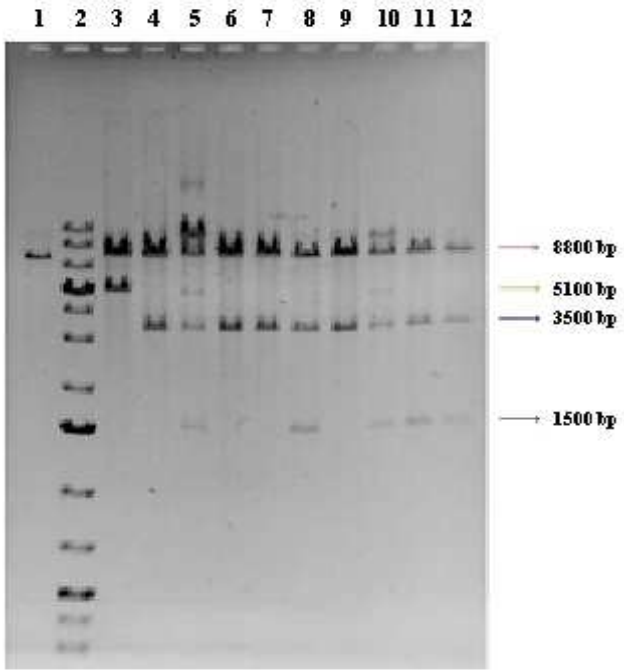


Figure 3.9: Insert analysis of putative arsenate resistant clones from *Pst*I *G. rubropertincta* library; Lanes: 1) linearized pDA71 mutant, 2) Generuler DNA Ladder mix marker, 3-12) *Pst*I digested As^R clones

The gel revealed that 90% of the recombinant constructs had a 3.5 kbp fragment in common, and 5 of the 10 clones had the same 1.5 kbp DNA piece as well. Moreover, clones 2, 4, 5 and 7 had an identical *Pst*I restriction pattern (I), as did clones 6, 9, and 10 (II) and clones 3 and 8 (III) which infers that they are multiples of the same construct. pKL3 and pKL8 showed an additional DNA band of about 15-20kbp which could, but unlikely, be an artifact of an incomplete digestion of the clones. Transformant 1 was the only one that demonstrated a totally different DNA fragmentation pattern, but it was postulated though that it could be the result of a rearrangement of the 3500 and 1500 bp DNA pieces found in the other clones.
Duplicate clones representative of the three digestion arrangements were selected together with clone 1 and retransformed into SQ1 by spotting them on distinct areas of a regeneration plate and then grown for 5-7 days at 30°C. These recombinant transformants were replica plated on LA Cm40 plates containing different amounts of arsenate or arsenite. The results of the spot testing of these putative arsenic resistant clones after 10 days of incubation at 30°C are encapsulated in the table below.

[Arsenic]	**Degree of growth**							
Arsenite(mM)	**pDA71***	**1**	**2**	**3**	**5**	**6**	**8**	**9**
0	+++	+++	+++	+++	+++	+++	+++	+++
1	++	++	++	++	++	++	++	++
2	++	++	++	++	++	++	++	++
3	+	++	++	++	++	++	++	++
4	-	++	++	++	++	++	++	++
5	-	+	++	++	++	+	++	+
7.5	-	+	+	+	+	-	+	+
10	-	-	-	-	-	-	-	-
Arsenate(mM)								
1	++	++	++	++	++	++	++	++
5	±	++	++	++	++	++	++	++
10	-	++	++	++	++	++	++	++
20	-	++	++	++	++	++	++	++
30	-	++	++	++	++	++	++	++
40	-	++	+	+	+	+	+	+
50	-	++	+	+	+	+	+	+
75	-	++	+	+	+	+	+	+
100	-	++	+	+	+	±	+	+

where +++ = excellent, ++ = moderate, + = slight, ± = scarce, and - = no growth

Table 3.6: Spot test results of pDA71* and selected As^R clones (see Fig3.9; Lanes 3-5,7,8,10,11) in SQ1 on solid agar media containing different amounts of arsenic.

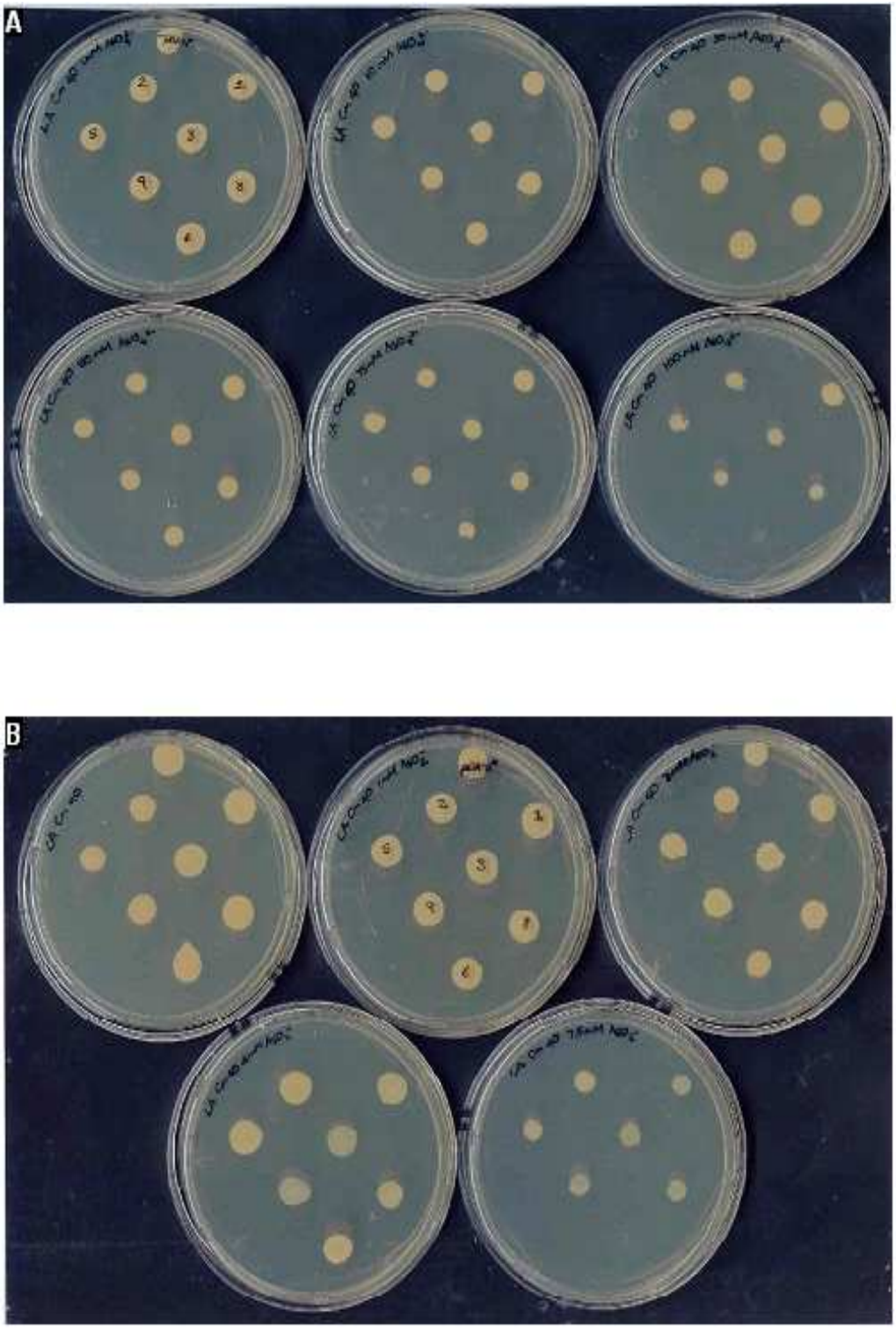


Figure 3.10: Spot test results of pDA71* and selected As^R clones 1-3,5,6,8, and 9 in SQ1 on A) LA Cm40 plates with increasing amounts of arsenate(mM); from left to right row1: 1, 10, 30, row2: 50, 75, 100 and B) LA Cm40 plates with increasing amount of arsenite(mM); from left to right row1: 0, 1, 3, row2: 4, 7.5

The recombinant plasmid DNA of these arsenic resistant clones were separately transformed back into *E. coli* MM294-4. Their arsenic tolerance capacity were investigated in order to ascertain whether their determinants gets expressed and function in a Gram(-) environment. This was done by spot testing pDA71* and these putative clones in MM294-4 on LA Amp200 agar plates containing various concentrations of AsO_4^{3-} or AsO_2^- and incubated at 37°C for 3-5 days. The recombinant clones conferred no increased arsenic resistance ability and exhibited the same growth profile as the control *E. coli* shown in table 3.4. This was as expected as the redox chemistry in gram-positive and gram-negative microorganisms are distinct and so their *ars* gene products would unlikely be functional in each other. Moreover, several studies have also demonstrated that arsenate reductases from other Gram(+) bacteria do not function either in *E. coli* due to mechanistic distinctions and the correct redox molecules (thioredoxin and thioredoxin reductase) not being available.

3.6 Restriction mapping of putative As^R cloned DNA sequences

Large-scale plasmid preparations were carried out for transformants 1, as it was unique, and clone 3, since it contained all three semiconserved fragments. As illustrated above, the 3.5kbp insert of clones 2 and 5 were sufficient in conferring the As^R phenotype and thus emphasis was place on this DNA fragment of clone 3. Bulk *Pst*I digestions of the recombinant plasmid DNA of transformants 1 and 3 were performed overnight at the recommended temperature and then run on 0.6% agarose gels for 90 min at $10 Vcm^{-1}$. The approximately 3500 bp fragment from 3 and 5100 bp insert from 1 was excised from the agarose gels and purified using freeze-squeeze method or the use of the Qiagen gel purification kit. The ultimate aim was to obtain a restriction map, which would enable subcloning to be done, so that the minimum DNA responsible for conferring the arsenic resistant phenotype can be determined.

3.6.1 Construction of a partial restriction map of the 3.5 kbp arsenic resistant insert (pKL335)

The purified 3500bp insert from clone 3, pKL335, was subjected to overnight digestions with the unique cloning site restriction endonucleases and their corresponding isocaudomers followed by electrophoresis on a 0.8% gel for 1 and a halve hours at 100V and 20mA. Evidently, it can be seen that neither *Bgl*II nor *Hind*III restricts this DNA fragment, whereas the fourbase cutter *Sau3A*I cuts it too frequently. As *Cla*I is partially inhibited by *dam* methylation, its digestion pattern could not be 100% accurate. Also, subcloning with *Cla*I in any case would be complicated as the resulting subclone(s) would need to be released from the vector with restriction enzymes flanking the *Sfu*I site.

Therefore, this would require that a unique restriction endonuclease needed to be found that was cutting the vector only just outside the *EcoR*I cloning region as the restriction site for *Sfu*I is located on the extreme right in this gene.

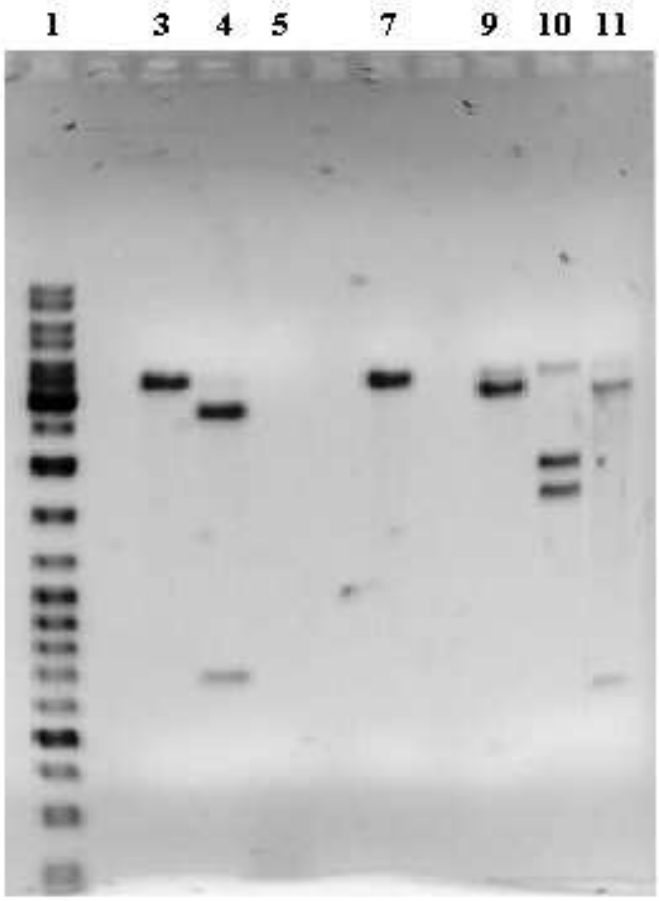


Figure 3.11: Single digestions of 3.5kbp arsenic resistant fragment; Lanes: 1) Generuler DNA Ladder mix marker 3) *Bgl*II, 4) *BamH*I, 5) *Sau3A*I, 7) *Hind*III, 9) *Nsi*I, 10) *Sfu*I, and 11) *Cla*I.

As a result, further overnight single and double digestions were carried out with only the single cutters *BamH*I, *Nsi*I, and *Sfu*I, which were then electrophoresed on a 0.6% gel for 75 min.

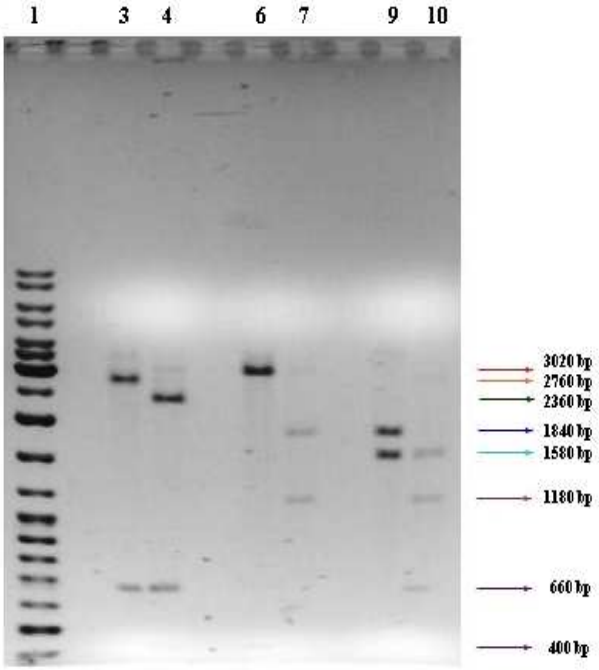


Figure 3.12: Single and double digestions of 3.5kbp As^R fragment; Lanes: 1) Generuler DNA Ladder mix marker, 3) *BamH*I, 4) *BamH*I + *Nsi*I, 6) *Nsi*I, 7) *Nsi*I + *Sfu*I, 9) *Sfu*I, 10) *Sfu*I + *BamH*I.

Consequently, the molecular sizes of the resulting DNA fragments were determined from a calibration curve and the following partial restriction map was deduced for this $\sim$ 3500 bp DNA insert as illustrated below.

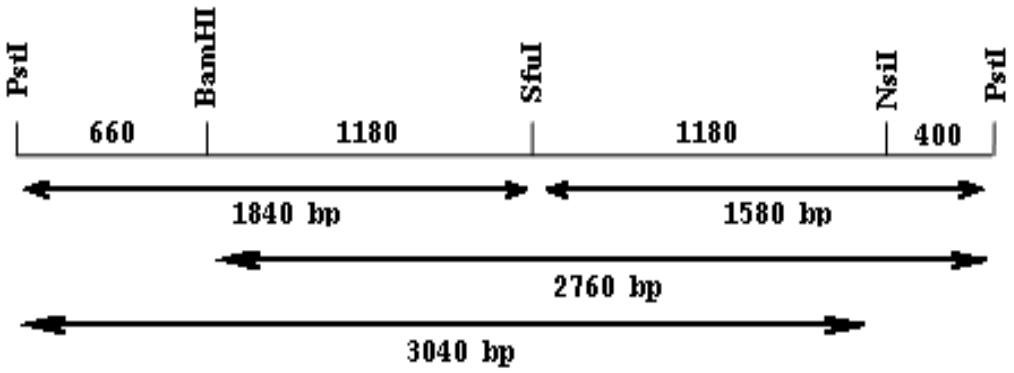


Figure 3.13: Preliminary restriction map of 3.5kbp fragment of arsenic resistant clone 3.

3.6.2 Construction of a partial restriction map of the 5.1 kbp arsenic resistant insert (pKL1)

Again, the 5.1 kbp underwent the same overnight digestions as did pKL335 and then run on a 0.8% gel at identical conditions.

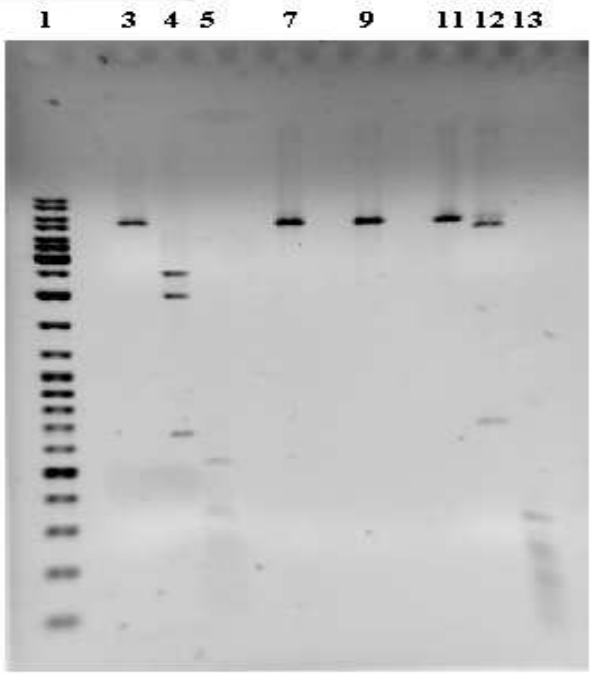


Figure 3.14: Single digestions of 5.1kbp arsenic resistant fragment; Lanes: 1) Generuler DNA Ladder mix marker 3) *Bgl*II, 4) *BamH*I, 5) *Sau3A*I, 7) *Hind*III, 9) *Nsi*I, 11) *Sfu*I, 12) *Cla*I, and 13) *Taq*I.

Surprisingly, only *BamH*I and *Cla*I digested this DNA fragment less than 3 times, whereas the fourbase cutters cleaved too frequently and the remaining endonucleases didn't digest it at all. Due to the possible complications involved with the subsequent subcloning with the *Cla*I endonuclease and the questionable restriction pattern as a result of its partial *dam* methylation sensitivity, this enzyme was not included in the restriction maps of both inserts. The size of the internal fragment was determined by digesting pDA71 and the original recombinant construct pKL1 with *BamH*I to compare and examine which of the three fragments produced during the single digestion of the insert only are released in this manner. It turned out from a calibration curve that the *BamH*I fragment is approximately 2500 bp in length and that the flanking regions are 2.0 and 0.6kbp, respectively.



Figure 3.15: Preliminary restriction map of 5.1kbp insert of arsenic resistant clone 1.

3.6.3 Subcloning of the putative arsenic resistant DNA inserts

From the partial restriction map of the pKL335 insert, it was decided to attempt to subclone the two *Pst*I-*Sfu*I fragments as well as the *BamH*I-*Nsi*I fragment into the pDA71 vector in order to ascertain whether the deleting any segment influences the As^R phenotype of the original fragment. After the ligations were transformed into *E. coli*, plasmid screens were done to confirm if the correct DNA fragments were cloned. The analysis of the transformants for the *Sfu*I subcloning showed that multiple clones had either of the two DNA fragments. However, due to the formation of *Bgl*II/*BamH*I and *Pst*I/*Nsi*I hybrid sites during the ligation of the *BamH*I-*Nsi*I fragment, the cloned inserts were released by a double digestion of *Hind*III and *Sfu*I. Surprisingly, the screening of these recombinant colonies showed either no inserts or smaller fragments than expected on numerous occasions. Therefore, only the *Sfu*I subclones called pKL316 and pKL319 were transformed into SQ1 again.

As there were only one suitable enzyme able to digest the 5.1 kbp pKL1 insert, subcloning was performed on the ∼2000 bp *BamH*I-*Pst*I fragment and the internal *BamH*I segment. After ligation, the cloned inserts were excised from the recombinant constructs by a *Hind*III/*Pst*I double digestion. The plasmid screens for these transformants yielded several clones bearing the *BamH*I-*Pst*I fragment, designated pKL120, and a single transformant with the 2.5kbp *BamH*I internal piece, pKL125. Duplicates of the pKL120 and the sole pKL125 plasmid construct were also transformed into the recipient *R. erythropolis* strain SQ1.

The *Sfu*I subclones pKL316 and pKL319 of the 3.5 kbp fragment, as well as the *BamH*I constructs pKL120 and pKL125 of the 5100 bp fragment were spot tested together with their respective original clones pKL335 (pKL2) and pKL1 on LA Cm40 plates supplemented with various concentrations of AsO_4^{3-} or AsO_2^-. The results from this replica plating after 10 days of incubation at 30°C are summarized in the following table.

[Arsenic] Arsenite(mM)	Degree of growth						
	pDA71*	**pKL1**	**pKL120**	**pKL125**	**pKL335**	**pKL319**	**pKL316**
0	+++	+++	+++	+++	+++	+++	+++
1	++	++	++	++	++	++	++
2	++	++	++	++	++	++	++
3	+	++	++	++	++	+	++
4	-	++	+	+	++	-	+
5	-	+	-	±	++	-	-
7.5	-	+	-	-	+	-	-
10	-	-	-	-	-	-	-
Arsenate(mM)							
1	++	++	++	++	++	++	++
5	±	++	±	±	++	±	++
10	-	++	-	-	++	-	++
20	-	++	-	-	++	-	++
30	-	++	-	-	++	-	++
40	-	++	-	-	++	-	++
50	-	++	-	-	++	-	+
75	-	++	-	-	++	-	+
100	-	++	-	-	+	-	±

where +++ = excellent, ++ = moderate, + = slight, ± = scarce, and - = no growth

Table 3.7: Spot test results of pDA71* as well as pKL1, pKL335 and their subclones in SQ1 on solid agar media containing different amounts of arsenic.

3.7 Sequence analysis of putative As^R DNA fragments

The inserts of pKL316 and pKL319 were ligated into the *Pst*I and *Xmi*I (*Sfu*I isocaudomer) sites of the cloning vector, pUC18. Even though the transformation efficiency was poor, releasing the inserts from the few recombinant constructs by a *Pst*I/*BamH*I double digestion yielded a single positive for both *Sfu*I subclones. Also, ligations of the DNA fragment of the recombinant constructs pKL120 and pKL125 into this vector were also undertaken. The cloning of the ∼2000bp fragment of pKL1 into the *BamH*I and *Pst*I sites of this vector was efficient and successful, as confirmed by plasmid screens of random recombinant colonies. Seemingly, the internal segment of clone 1 was ligated into the unique *BamH*I restriction site of pUC18, as numerous clones were obtained. However, examining random transformants from the subcloning of the 2.5 kbp DNA fragment showed unfortunately either no insert or smaller inserts than expected.

Subsequently, the inserts from pKL316, pKL319, and pKL120 cloned into pUC18 were sent to the Inqaba biotech company for sequencing via primer walking. However, problems were encountered with the ∼1900 bp *Sfu*I subclone and could as a result not be sequenced. The sequences received of pKL316 and pKL120 were edited using the Sequencer 4.2 program in order to resolve any mismatches and consequently obtain a contiguous consensus DNA sequence. The various internet programs that were consulted for the analysis of the consensus sequences of these two cloned inserts are tabulated below.

Program	**Website address**
NebCutter	http://tools.neb.com/nebcutter/
Genemark	http://opal.biology.gatech.edu/Genemark/
Frameplot	http:// www.nih.go.jp/∼jun/research/frameplot/
BLAST	http://ncbi.nlm.nih.gov/BLAST/
FASTA	http://ebi.ac.uk/fasta33/

Table 3.8: Internet program utilized for sequence analysis.

The contiguous consensus sequence of these constructs after editing were investigated for open reading frames (ORFs) of possible genes and whether parts of the DNA aligns to any previously obtained nucleotide (Blastn) and translated protein sequences (Blastx) from gene databases. The protein sequences of the predicted ORFs was also analyzed for alignment to putative matches (Blastp) and whether they contain any conserved domains. The results of the ORF prediction and sequence alignment analysis are reported below.

3.7.1 pKL316 sequence analysis

DNA sequence

The resulting contiguous consensus sequence of pKL316, which had a G+C content of 64.8%, after alignment of the sequenced DNA sections and editing is as follows



Figure 3.16: DNA sequence of pKL316; where the *Xmi*I/*Sfu*I hybrid site is shown in red and the *Pst*I site in green

ORF prediction

The open reading frame analysis of pKL316 using Genemark and Frameplot is summarized in the table below

ORF no. #	Frame	Left end	Right end	Gene length (bp)	Amino acids (aa)
1	+1	<1	198	198+	65+
2	+1	259	876	618	205
3	+3	918	1341	423	141
4	+2	1340	1574>	234+	78+

Table 3.9: Summary of ORF prediction of pKL316

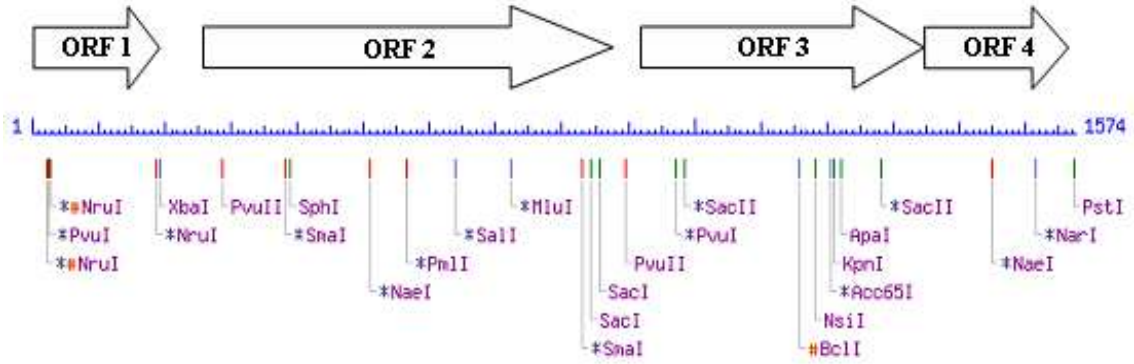


Figure 3.17: ORF prediction and restriction map of pKL316 with ORF1 being homologous to Acr3-type ArsB proteins and ORF2,3, and 4 to Trx-type ArsCs.

Nucleotide sequence analysis results of pKL316

Various parts of pKL316 sequence align significantly to multiple nucleotide matches from other nocardioform and related bacteria. Alignment data on some of the most relevant matches to this *G. rubropertincta* DNA fragment are summarized below.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
AY223810.1	Rhodococcus erythropolis linear plasmid pBD2, complete sequence	318	1103	81%	4e-83	82%
AP008931.1	Rhodococcus erythropolis PR4 plasmid pREL1, complete sequence	311	933	63%	5e-81	82%
CP000431.1	Rhodococcus sp. RHA1, complete genome	293	650	55%	1e-75	92%
AP006620.1	Nocardia farcinica IFM 10152 plasmid pNF2 DNA, complete sequence	282	863	54%	3e-72	83%
CP000476.1	Arthrobacter aurescens TC1 plasmid TC2, complete sequence	241	533	42%	8e-60	74%
DQ231520.1	Streptomyces sp. FR-008 ArsR2 (arsR2), ArsO (arsO), ArsB (arsB), ArsR1 (arsR1), ArsC (arsC), and ArsT (arsT) genes, complete cds; and unknown genes	210	436	52%	1e-50	79%
AP006618.1	Nocardia farcinica IFM 10152 DNA, complete genome	210	793	51%	1e-50	90%
CP000454.1	Arthrobacter sp. FB24, complete genome	208	466	49%	5e-50	73%
BA000035.2	Corynebacterium efficiens YS-314 DNA, complete genome	179	423	57%	2e-41	80%

Figure 3.18: Blastn data of some significant alignment matches to pKL316 nucleotide sequence.

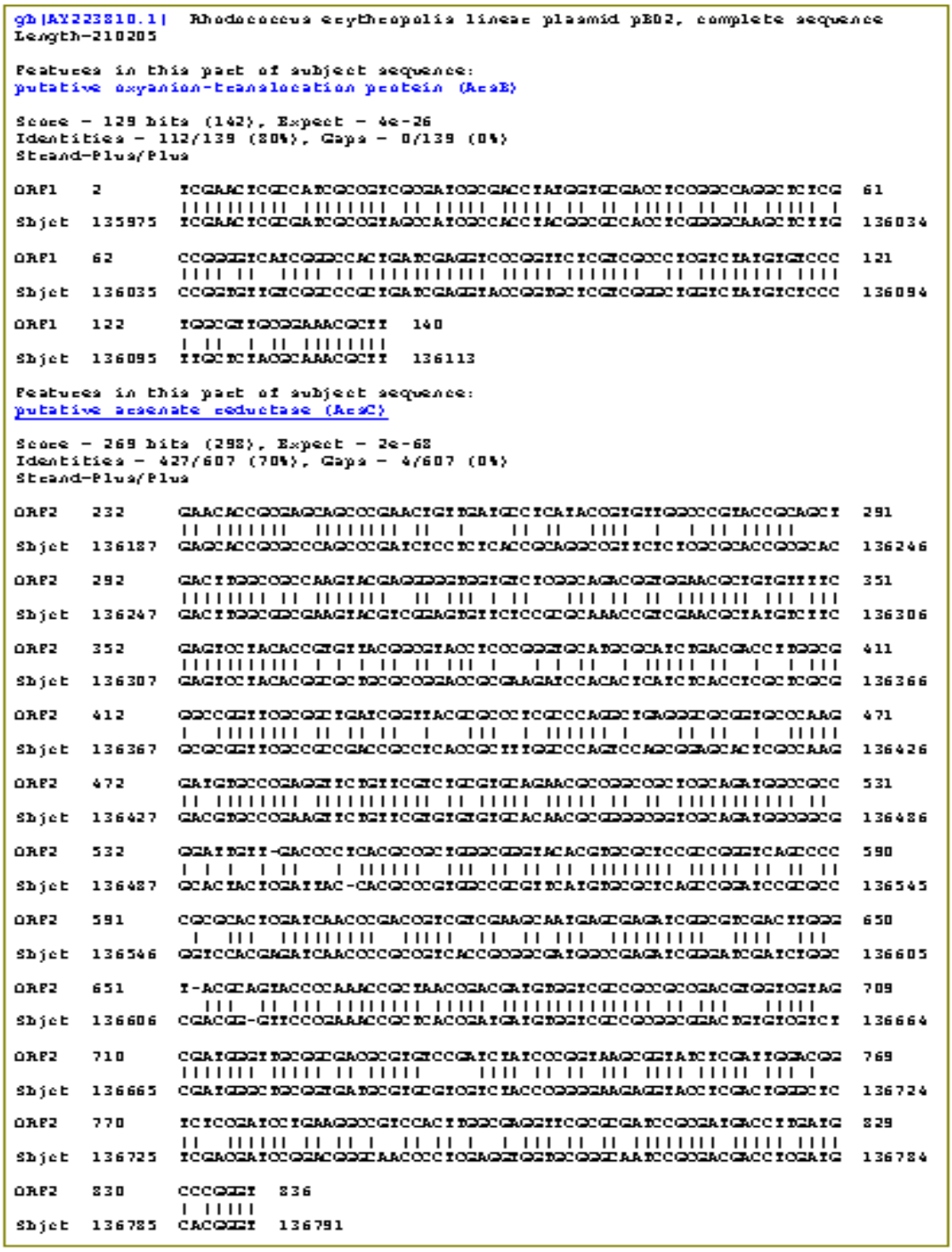


Figure 3.19: Nucleotide alignment of ORF1 and ORF2 with *R. erythropolis* plasmid pBD2 ArsB and ArsC (BD1, refer to Fig4.3), respectively

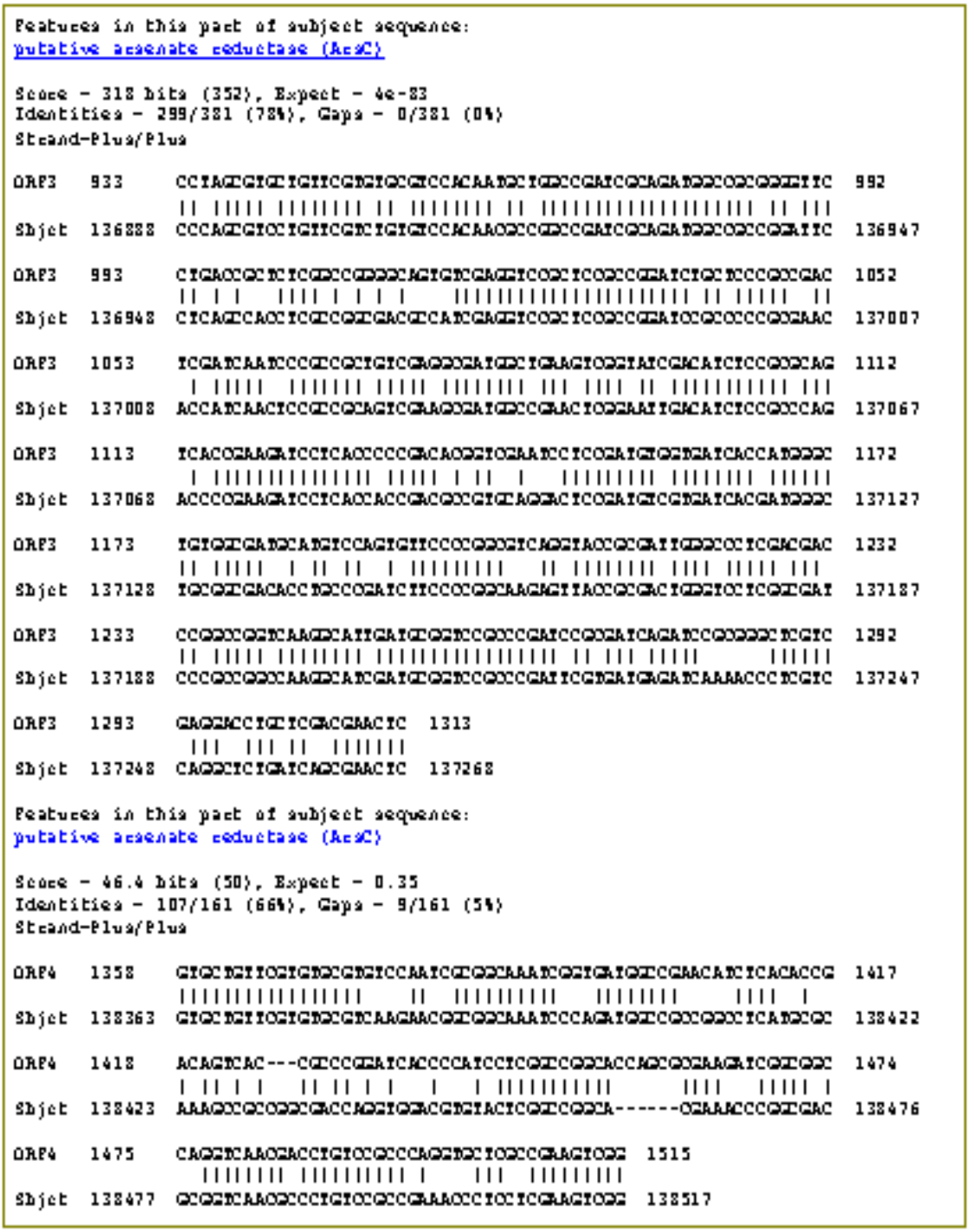


Figure 3.20: Nucleotide alignment of ORF3 and ORF4 with *R. erythropolis* plasmid pBD2 ArsCs (BD2 and BD3, refer to Fig4.3)

Translated protein sequence analysis results of pKL316

The blastx results revealed that sections of the translated protein sequence of pKL316 showed substantial similarity to multiple protein matches from numerous microorganisms. Some of these significant translated protein matches are illustrated below.

Accession No.	Alignment	Score	E value
ref|NP_898751.1|	putative arsenate reductase (ArsC) [Rhodococ...	292	4e-77
ref|YP_950225.1|	arsenate reductase [Arthrobacter aurescens T...	222	6e-56
ref|YP_703330.1|	probable arsenate reductase (glutaredoxin) [...	216	3e-54
ref|YP_122077.1|	putative arsenate reductase [Nocardia farcin...	204	1e-50
ref|NP_630906.1|	arsenate reductase [Streptomyces coelicolor ...	204	1e-50
ref|NP_737488.1|	putative arsenate reductase [Corynebacterium...	202	2e-50
ref|NP_217159.1|	PROBABLE ARSENIC-TRANSPORT INTEGRAL MEMBRANE...	174	1e-41

Figure 3.21: Blastx data of significant alignment matches to the translated sequence of pKL316.

```
ref|NP_898751.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
gb|AAP74021.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
Length=217

Score = 292 bits (747), Expect = 4e-77
Identities = 157/210 (74%), Positives = 173/210 (82%), Gaps = 0/210 (0%)
Frame = +1

ORF2    232  EHREQPELLNPHIVLARIAADLAAKYEGVVSRQIVERCVPESYIVLRRISRVHAHLIILA  411
             EHR QP+LL P  VL+RIA DLAAKY GV S QIVER VPESYI LRRI+++H HLI+LA
Sbjct  3    EHRAQPDLLSPQAVLSRIAHDLAAKYVGVPSAQIVERGVPESYIALRRIAKIHIHLISLA  62

ORF2    412  GRPAADRLRALAQAEGAVPKDVPEVLPVCVQWAGRSQGAAGLLIPHAAGRGVHVRSAGSAP  591
              RPAADRL ALAQ+ GA+ KDVPEVLPVCV WAGRSQGAA LL  HA GRGVHVRSAGSAP
Sbjct  63   ARPAADRLIALAQSSGRLAKDVPEVLPVCVHWAGRSQGAAALLDYHARGRGVHVRSAGSAP  122

ORF2    592  AHSIWPIVVEAAGEIGVDLGIQYPKPLIddvvaaadvvvaGGIGDACPIYPGKRYLDWIV  771
              H IWP V  AAG+EIG+DL    +PKPLIDDVVAAAD VV+GGIGDAC +YPGKRYLDW +
Sbjct  123  VHEIWPAVIAAAGEIGIDLADGPPKPLIDDVVAAADCVVSGGIGDACVVYPGKRYLDWAL  182

ORF2    772  SDPEGRPLGEVRAIRDDLDARVRELLAELI  861
              DP+G+PL  VRAIRDDLDARV  LL ELI
Sbjct  183  DDPDGQPLEVVRAIRDDLDARVIALLIELI  212
```

Figure 3.22: Translated protein alignment of ORF2 with *R. erythropolis* plasmid pBD2 ArsC (BD1, refer to Fig4.3).

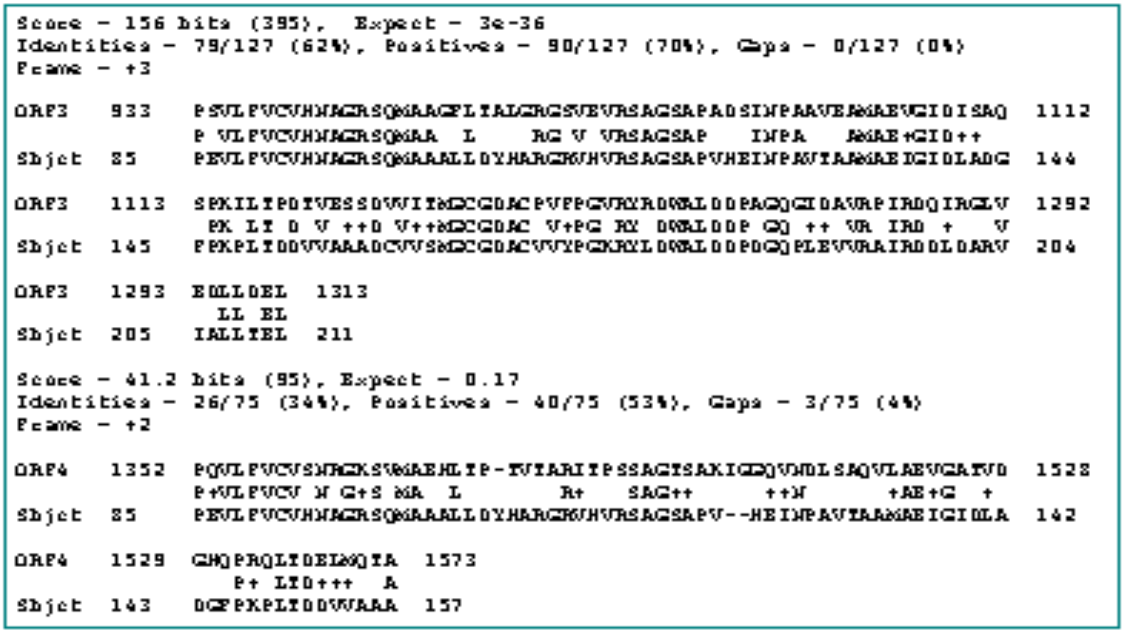


Figure 3.23: Translated protein alignment of ORF3 and ORF4 with *R. erythropolis* plasmid pBD2 ArsC (BD1, refer to Fig4.3).

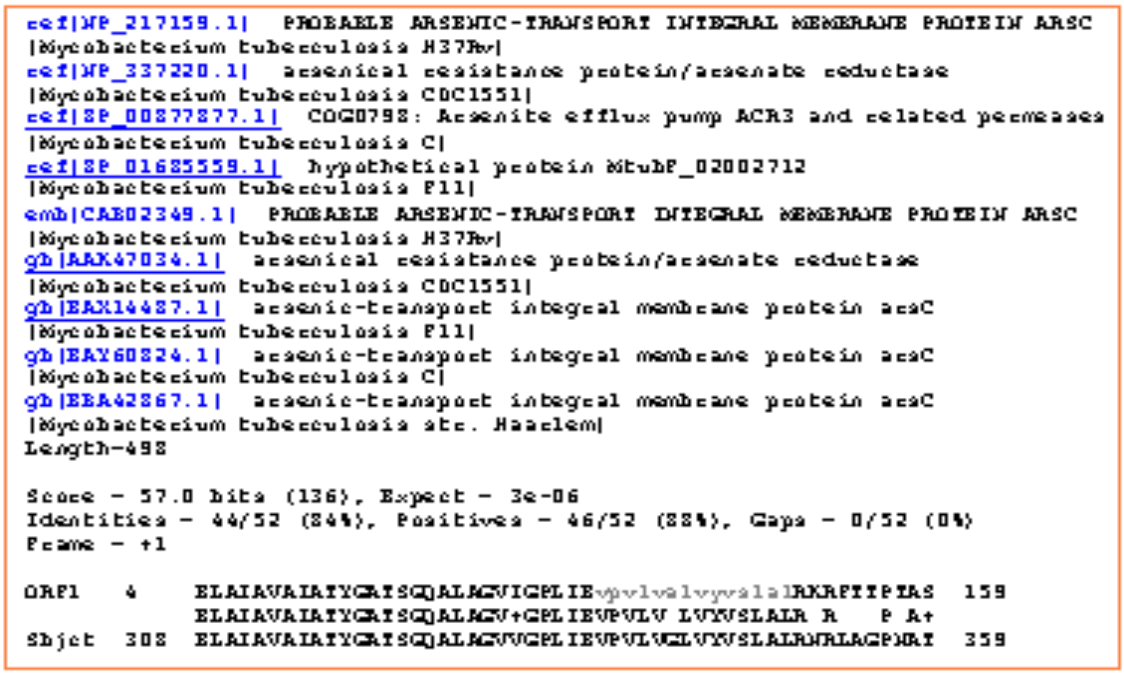


Figure 3.24: Translated protein alignment of ORF1 with *M. tuberculosis* ArsBC fusion protein.

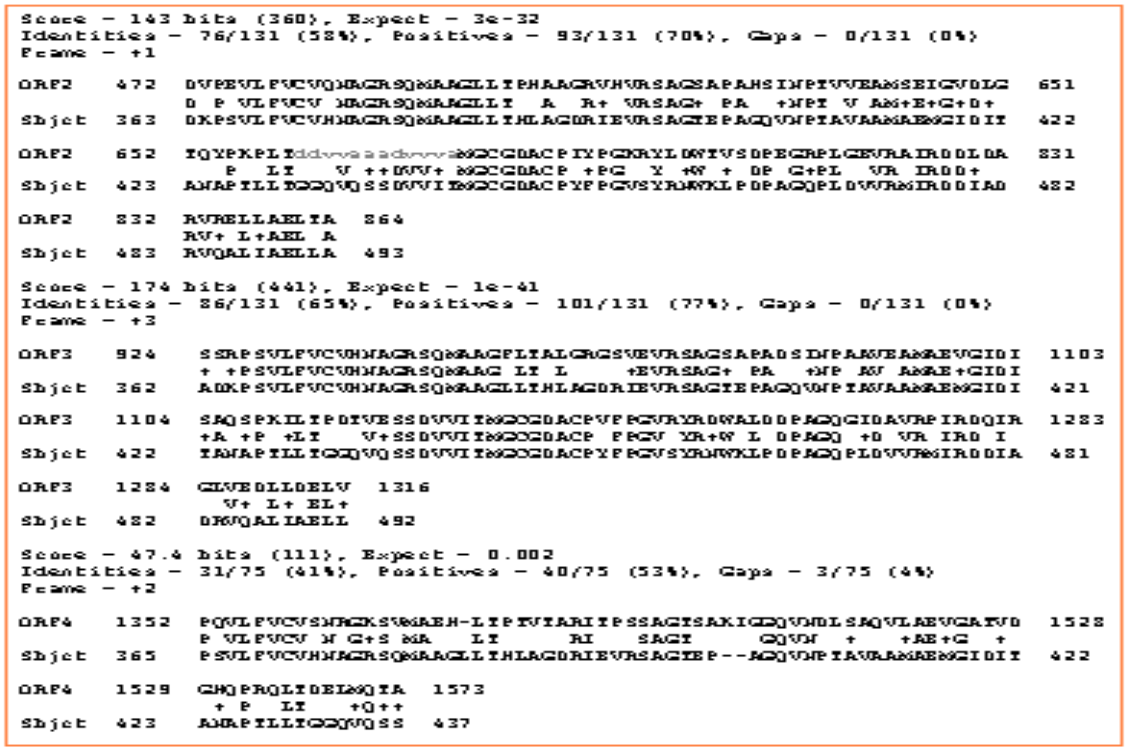


Figure 3.25: Translated protein alignment of ORF2,3 and 4 with *M. tuberculosis* ArsBC fusion protein

Protein sequence analysis of predicted ORFs in pKL316

ORF 1 (ArsB)

Conserved domains search

The CD analysis showed that the partial 65 amino acid sequence of ORF1 aligned to a portion of the ACR3 arsenite efflux protein family structural domain with an e value of $1e^{-6}$.



Figure 3.26: Putative conserved domain(s) found within ORF1.

Protein sequence alignment

The predicted amino acid sequence of ORF1 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E value
ref|YP_001069735.1|	arsenical-resistance protein [Mycobacteri...	82.4	7e-15
ref|YP_345594.1|	putative arsenite export protein [Rhodococcu...	80.5	3e-14
ref|YP_954029.1|	arsenical-resistance protein [Mycobacterium ...	79.7	4e-14
ref|YP_122074.1|	putative arsenic resistance transporter [Noc...	78.6	9e-14
ref|NP_898750.1|	putative oxyanion-translocation protein (Ars...	78.6	1e-13
ref|YP_703329.1|	probable arsenic resistance transporter [Rho...	73.6	3e-12
ref|YP_118661.1|	putative arsenite transporter [Nocardia farc...	72.0	9e-12
ref|NP_630908.1|	arsenic resistance membrane transport protei...	72.0	1e-11
gb|ABB70174.1|	ArsB [Streptomyces sp. FR-008]	68.6	1e-10

Figure 3.27: Blastp data of some significant alignment matches to predicted sequence of ORF1

```
ref|YP_001069735.1| arsenical-resistance protein [Mycobacterium sp. JLS]
gb|ABN97244.1| arsenical-resistance protein [Mycobacterium sp. JLS]
Length=369

Score = 82.4 bits (202),  Expect = 7e-15, Method: Composition-based stats.
Identities = 48/62 (77%), Positives = 51/62 (82%), Gaps = 0/62 (0%)

Query  2    ELAIAVAIATYGATSGQALAGVIGPLIEVPVLVALVYVSLALRKRFTTPTASSLDVSKEP  61
            ELAIAVAIATYGATSGQALAGV+GPLIEVPVLVALVYVSLALRKRF   T+  L  +
Sbjct  306  ELAIAVAIATYGATSGQALAGVVGPLIEVPVLVALVYVSLALRKRFGHHTSQHLPANSST  365

Query  62   AR  63
             R
Sbjct  366  TR  367

ref|NP_898750.1| putative oxyanion-translocation protein (ArsB) [Rhodococcus
erythropolis]
gb|AAP74020.1| putative oxyanion-translocation protein (ArsB) [Rhodococcus
erythropolis]
Length=365

Score = 78.6 bits (192),  Expect = 1e-13, Method: Composition-based stats.
Identities = 45/52 (86%), Positives = 46/52 (88%), Gaps = 0/52 (0%)

Query  2    ELAIAVAIATYGATSGQALAGVIGPLIEVPVLVALVYVSLALRKRFTTPTAS  53
            ELAIAVAIATYGATSGQALAGV+GPLIEVPVLV LVYVSLALRKRF     S
Sbjct  308  ELAIAVAIATYGATSGQALAGVVGPLIEVPVLVGLVYVSLALRKRFAVKEVS  359

gb|ABB70174.1|  ArsB [Streptomyces sp. FR-008]
Length=370

Score = 68.6 bits (166),  Expect = 1e-10, Method: Composition-based stats.
Identities = 36/46 (78%), Positives = 43/46 (93%), Gaps = 0/46 (0%)

Query  2    ELAIAVAIATYGATSGQALAGVIGPLIEVPVLVALVYVSLALRKRF  47
            ELAIAVAIAT+G TSGQAL+GV+GPLIEVPVL+ LVYV+LA R++F
Sbjct  316  ELAIAVAIATFGVTSGQALSGVVGPLIEVPVLIGLVYVALAWRRKF  361
```

Figure 3.28: Protein alignments of ORF1 with *Mycobacterium* sp. JLS, *R. erythropolis* plasmid pBD2 and *Streptomyces* sp. FR-008 plasmid pHZ227.

ORF 2 (ArsC1)

Conserved domains search

The CD analysis showed that part of the 205 amino acid sequence of ORF2 aligned to the low molecular weight phosphatase (LMWPc) family structural domain with an e value of $2e^{-16}$.

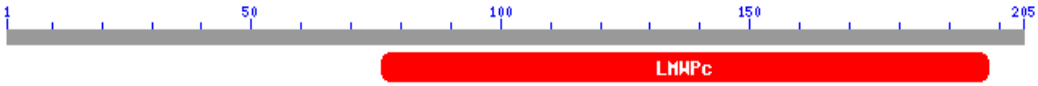


Figure 3.29: Putative conserved domain(s) found within ORF2.

Protein sequence alignment

The predicted amino acid sequence of ORF2 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E value
ref|NP_898751.1|	putative arsenate reductase (ArsC) [Rhodococ...	277	2e-73
ref|YP_950225.1|	arsenate reductase [Arthrobacter aurescens T...	233	6e-60
ref|NP_600728.1|	protein-tyrosine-phosphatase [Corynebacteriu...	217	4e-55
ref|NP_737488.1|	putative arsenate reductase [Corynebacterium...	213	7e-54
ref|YP_954028.1|	protein tyrosine phosphatase [Mycobacterium ...	182	1e-44
ref|YP_122076.1|	putative arsenate reductase [Nocardia farcin...	179	6e-44
ref|YP_345591.1|	putative arsenate reductase [Rhodococcus ery...	176	5e-43

Figure 3.30: Blastp data of some significant alignment matches to predicted sequence of ORF2.

```
ref|NP_898751.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
gb|AAP74021.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
Length=217

Score =  277 bits (709),  Expect = 2e-73, Method: Composition-based stats.
Identities = 150/200 (75%), Positives = 165/200 (82%), Gaps = 0/200 (0%)

Query  2    PHTVLARTAADLAAKYEGVVSRQTVERCVFESYTVLRRTSRVHAHLTTLAGRFAADRLRA  61
            P  VL+RTA DLAAKY GV S QTVER VFESYT LRRT+++H HLT+LA RFAADRL A
Sbjct  13   PQAVLSRTAHDLAAKYVGVFSAQTVERYVFESYTALRRTAKIHTHLTSLAARFAADRLTA  72

Query  62   LAQAEGAVPKDVPEVLFVCVQNAGRSQMAAGLLTPHAAGRVHVRSAGSAPAHSINPTVVE  121
            LAQ+ GA+ KDVPEVLFVCV NAGRSQMAA LL  HA GRVHVRSAGSAP H INP V
Sbjct  73   LAQSSGALAKDVPEVLFVCVHNAGRSQMAAALLDYHARGRVHVRSAGSAPVHEINPAVTA  132

Query  122  AMSEIGVDLGTQYPKPLTDDVVAAADVVVAMGCGDACPIYPGKRYLDWTVSDPEGRPLGE  181
            AM+EIG+DL   +PKPLTDDVVAAAD VV+MGCGDAC +YPGKRYLDW + DP+G+PL
Sbjct  133  AMAEIGIDLADGFPKPLTDDVVAAADCVVSMGCGDACVVYPGKRYLDWALDDPDGQPLEV  192

Query  182  VRAIRDDLDARVRELLAELT  201
            VRAIRDDLDARV  LL ELT
Sbjct  193  VRAIRDDLDARVIALLTELT  212
```

Figure 3.31: Protein alignment of ORF2 with *R. erythropolis* plasmid pBD2.



Figure 3.32: Protein alignments of ORF2 with *A. aurescens* TC1 plasmid pTC2 ArsC and *C. glutamicum* ATCC 13032 ArsC.

ORF 3 (ArsC2)

Conserved domains search

The CD analysis showed that most of the 141 amino acid sequence of ORF3 aligned to the low molecular weight phosphatase (LMWPc) family structural domain with an e value of $8e^{-23}$.

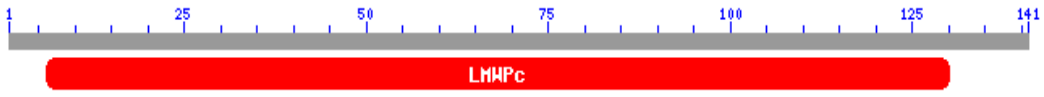


Figure 3.33: Putative conserved domain(s) found within ORF3.

Protein sequence alignment

The predicted amino acid sequence of ORF3 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E value
ref|YP_950222.1|	Arsenate reductase [Arthrobacter aurescens T...	196	4e-49
ref|YP_703330.1|	probable arsenate reductase (glutaredoxin) [...	195	7e-49
ref|NP_630906.1|	arsenate reductase [Streptomyces coelicolor ...	191	1e-47
ref|YP_122077.1|	putative arsenate reductase [Nocardia farcin...	190	2e-47
ref|NP_898752.1|	putative arsenate reductase (ArsC) [Rhodococ...	190	2e-47
ref|YP_345595.1|	putative arsenate reductase [Rhodococcus ery...	187	1e-46
gb|ABB70176.1|	ArsC [Streptomyces sp. FR-008]	181	1e-44
ref|NP_737487.1|	putative arsenate reductase [Corynebacterium...	173	3e-42

Figure 3.34: Blastp data of some significant alignment matches to predicted sequence of ORF3.

```
ref|NP_898752.1| putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
gb|AAP74022.1| putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
Length=138

Score =  190 bits (482),  Expect = 2e-47, Method: Composition-based stats.
Identities = 99/130 (76%), Positives = 114/130 (87%), Gaps = 0/130 (0%)

Query  3    SSRPSVLFVCVHNAGRSQMAAGFLTALGRGSVEVRSAGSAPADSINPAAVEAMAEVGIDI  62
            +S+PSVLFVCVHNAGRSQMAAGFL+ L    ++EVRSAGSAPA++IN AAVEAMAE+GIDI
Sbjct  2    TSKPSVLFVCVHNAGRSQMAAGFLSHLAGDAIEVRSAGSAPANTINSAAVEAMAELGIDI  61

Query  63   SAQSPKILTPDTVESSDVVITMGCGDACPVFPGVRYRDWALDDPAGQGIDAVRPIRDQIR  122
            SAQ+PKILT D V+ SDVVITMGCGD CP+FPG  YRDW L DPAGQGIDAVRPIRD+I+
Sbjct  62   SAQTPKILTTDAVQDSDVVITMGCGDTCPIFPGKSYRDWVLGDPAGQGIDAVRPIRDEIK  121

Query  123  GLVEDLLDEL  132
             LV+ L+ EL
Sbjct  122  TLVQALISEL  131

gb|ABB70176.1|  ArsC [Streptomyces sp. FR-008]
Length=140

Score =  181 bits (459),  Expect = 1e-44, Method: Composition-based stats.
Identities = 94/132 (71%), Positives = 108/132 (81%), Gaps = 0/132 (0%)

Query  3    SSRPSVLFVCVHNAGRSQMAAGFLTALGRGSVEVRSAGSAPADSINPAAVEAMAEVGIDI  62
            + +PSVLFVCVHNAGRSQMAA +LT L     +EVRSAGS P  ++NPAAVEAMAEVGIDI
Sbjct  2    ADKPSVLFVCVHNAGRSQMAAAWLTHLAGDRLEVRSAGSNPGTAVNPAAVEAMAEVGIDI  61

Query  63   SAQSPKILTPDTVESSDVVITMGCGDACPVFPGVRYRDWALDDPAGQGIDAVRPIRDQIR  122
            SA++PKILT D V+ SDV ITMGCGD CPVFPG RY DW L DPAGQG+ AVRPIRD+I+
Sbjct  62   SAETPKILTVDAVKESDVCITMGCGDTCPVFPGKRYLDWKLKDPAGQGVAAVRPIRDEIK  121

Query  123  GLVEDLLDELVP  134
             LVE L+ E+ P
Sbjct  122  TLVEGLIAEIAP  133

ref|NP_737487.1| putative arsenate reductase [Corynebacterium efficiens YS-314]
dbj|BAC17687.1| putative arsenate reductase [Corynebacterium efficiens YS-314]
Length=138

 Score =  173 bits (438),  Expect = 3e-42, Method: Composition-based stats.
 Identities = 88/134 (65%), Positives = 107/134 (79%), Gaps = 0/134 (0%)

Query  2    SSSRPSVLFVCVHNAGRSQMAAGFLTALGRGSVEVRSAGSAPADSINPAAVEAMAEVGID  61
            S ++PSVLFVCVHNAGRSQMAA +LT L   G + VRSAGS P D +NP AV AMAE GID
Sbjct  5    SDTQPSVLFVCVHNAGRSQMAAAYLTELSGGRITVRSAGSEPRDQLNPVAVAAMAEDGID  64

Query  62   ISAQSPKILTPDTVESSDVVITMGCGDACPVFPGVRYRDWALDDPAGQGIDAVRPIRDQI  121
            +++ SPKILT + V++SDVVITMGCGD CP+FPG RY DW L+DPAG+GID VR IRD I
Sbjct  65   MTSASPKILTTEAVQASDVVITMGCGDTCPIFPGKRYEDWELEDPAGKGIDDVRRIRDDI  124

Query  122  RGLVEDLLDELVPA  135
            +  +  L+DEL+P+
Sbjct  125  KHRITVLVDELLPS  138
```

Figure 3.35: Protein alignments of ORF3 with *R. erythropolis* plasmid pBD2, *Streptomyces* sp. FR-008 plasmid pHZ227 and *C. efficiens* YS-314

ORF 4 (ArsC3)

Conserved domains search

The CD analysis showed that the partial 78 amino acid sequence of ORF4 aligned to a portion of the low molecular weight phosphatase (LMWPc) family structural domain with an e value of $2e^{-11}$

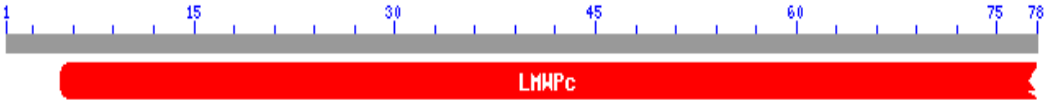


Figure 3.36: Putative conserved domain(s) found within ORF4.

Protein sequence alignment

The predicted amino acid sequence of ORF4 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E value
ref|YP_954027.1|	protein tyrosine phosphatase [Mycobacterium ...	72.8	5e-12
ref|NP_737486.1|	putative arsenate reductase [Corynebacterium...	48.9	8e-05
ref|YP_122075.1|	putative arsenate reductase [Nocardia farcin...	48.1	1e-04
ref|YP_829717.1|	protein tyrosine phosphatase [Arthrobacter s...	47.8	2e-04
ref|NP_898754.1|	putative arsenate reductase (ArsC) [Rhodococ...	45.8	6e-04

Figure 3.37: Blastp data of some significant alignment matches to predicted sequence of ORF4.

```
ref|NP_737486.1| putative arsenate reductase [Corynebacterium efficiens YS-314]
dbj|BAC17686.1| putative arsenate reductase [Corynebacterium efficiens YS-314]
Length=139

Score = 48.9 bits (115),  Expect = 8e-05, Method: Composition-based stats.
Identities = 33/77 (42%), Positives = 46/77 (59%), Gaps = 3/77 (3%)

Query  1   VSTTPQVLFVCVSNRGKSVMAEHLTPTVTA-RITPSSAGTSAKIGGQVNDLSAQVLAEVG  59
           +STTP VLFVCV N GKS MA  L       ++T  SAGT  K G ++N  S + +AEVG
Sbjct  1   MSTTPSVLFVCVGNGGKSQMAAALAEKHAGDQLTIHSAGT--KPGSKLNAQSVEAIAEVG  58

Query  60  ATVDGHQPRQLTDELMQ  76
           A +    P+ +  +L++
Sbjct  59  ADMSKGHPKGIDPQLLR  75
```

Figure 3.38: Protein alignment of ORF4 with *C. efficiens* YS-314.

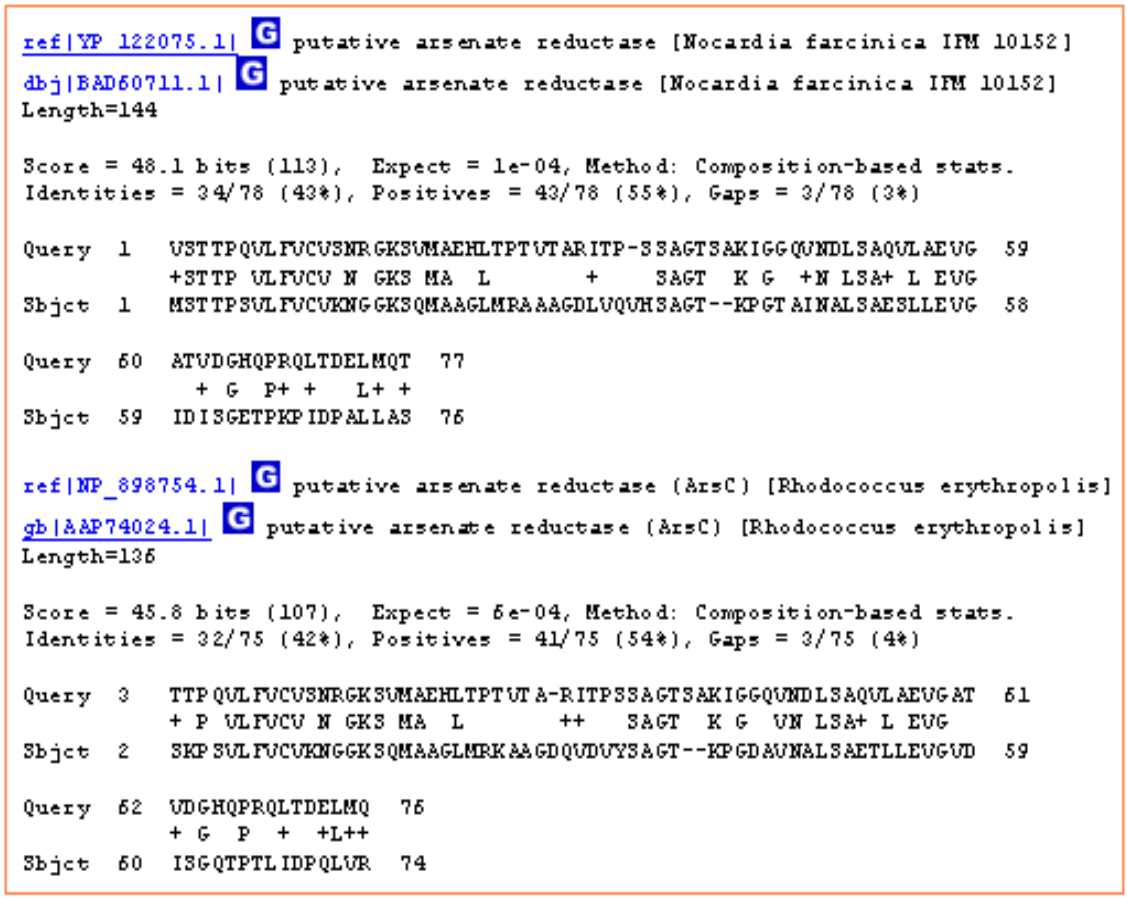


Figure 3.39: Protein alignments of ORF4 with *Nocardia farcinica* 10152 plasmid pNF2 and *Rhodococcus erythropolis* plasmid pBD2.

3.7.2 pKL120 sequence analysis

DNA sequence

The resulting contiguous consensus sequence of pKL120, which had G+C content of 70.8%, after alignment of the sequenced DNA sections and editing is as follows.

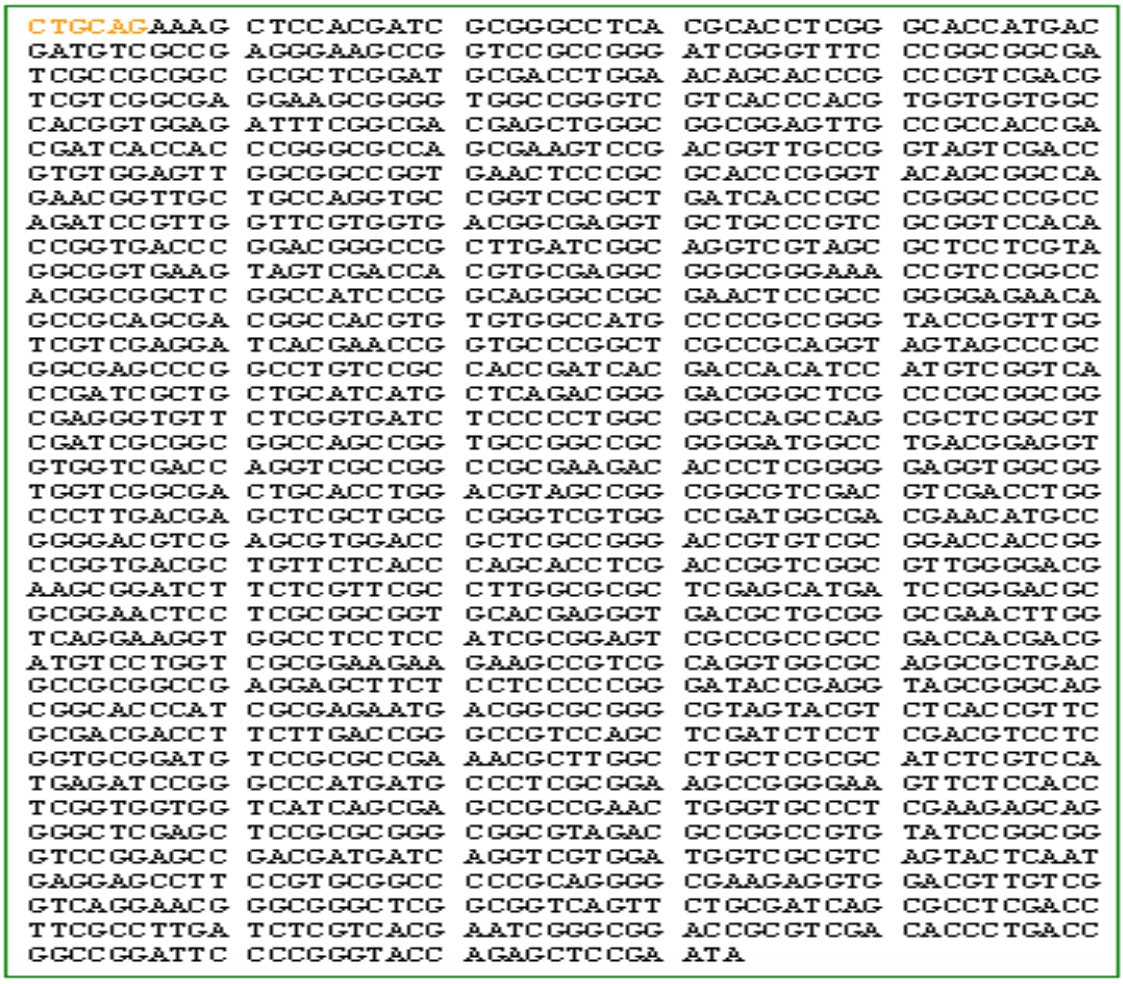


Figure 3.40: DNA sequence of pKL316; where the *Pst*I site is shown in orange

ORF prediction

The open reading frame analysis of pKL120 using Genemark and Frameplot is summarized in the table below.

ORF no. #	Frame	Left end	Right end	Gene length (bp)	Amino acids(aa)
5	-3	<3	566	564+	188+
6	-2	772	1759	978	325
7	-1	1802	1933	132	43+

Table 3.10: Summary of ORF prediction of pKL316

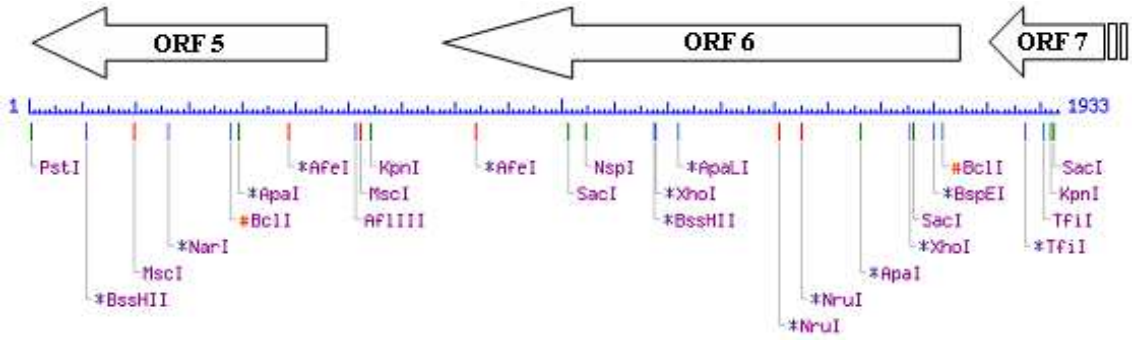


Figure 3.41: ORF prediction and restriction map of pKL120 with ORF5 being homologous to ArsO/FMO-like proteins and ORF6 and ORF7 being ArsT/TrxB and Trx-type ArsC homologs, respectively

Nucleotide sequence analysis results of pKL120

Various parts of pKL120 sequence align significantly to multiple nucleotide matches from other nocardioform and related bacteria. Alignment data on some of the most relevant matches to this *G. rubropertincta* DNA fragment are summarized below.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
CP000431.1	Rhodococcus sp. RHA1, complete genome	1018	1399	72%	0.0	88%
AY223810.1	Rhodococcus erythropolis linear plasmid pBD2, complete sequence	1018	1138	54%	0.0	86%
AP008931.1	Rhodococcus erythropolis PR4 plasmid pREL1, complete sequence	1018	1453	91%	0.0	85%
AP006618.1	Nocardia farcinica IFM 10152 DNA, complete genome	884	1282	65%	0.0	93%
CP000511.1	Mycobacterium vanbaalenii PYR-1, complete genome	681	814	49%	0.0	100%
BA000035.2	Corynebacterium efficiens YS-314 DNA, complete genome	583	837	72%	7e-163	74%
CP000717.1	Mycobacterium tuberculosis F11, complete genome	535	535	46%	3e-148	73%
BA000036.3	Corynebacterium glutamicum ATCC 13032 DNA, complete genome	453	453	45%	9e-124	71%
AL939118.1	Streptomyces coelicolor A3(2) complete genome; segment 15/29	446	446	46%	1e-121	71%
CP000476.1	Arthrobacter aurescens TC1 plasmid TC2, complete sequence	338	401	50%	5e-89	78%
CP000454.1	Arthrobacter sp. FB24, complete genome	331	827	56%	8e-87	76%
DQ231520.1	Streptomyces sp. FR-008 ArsR2 (arsR2), ArsO (arsO), ArsB (arsB), ArsR1 (arsR1), ArsC (arsC), and ArsT (arsT) genes, complete cds; and unknown genes	291	843	79%	7e-75	83%

Figure 3.42: Blastn data of some significant alignment matches to pKL120 nucleotide sequence.

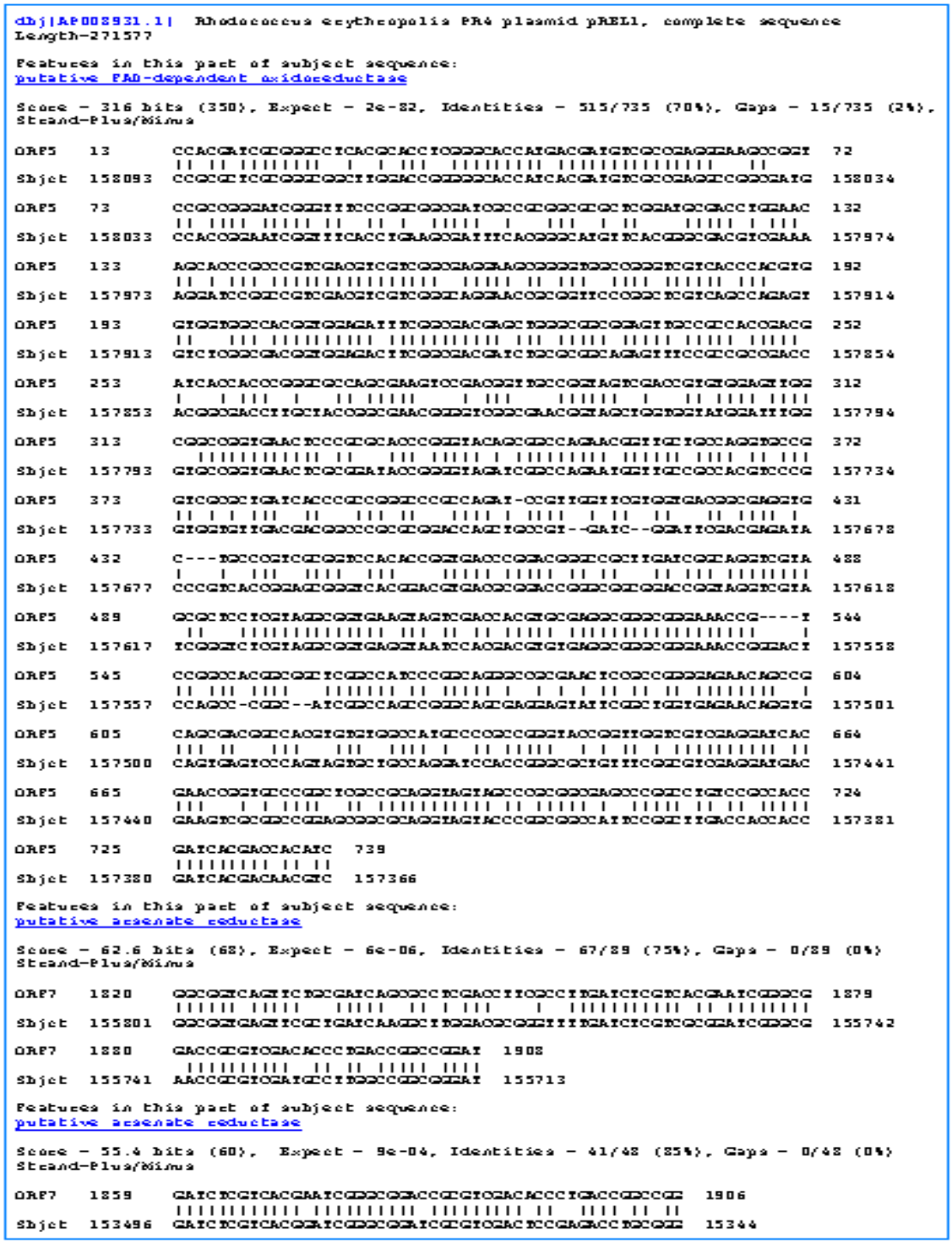


Figure 3.43: Nucleotide alignment of ORF5 and ORF7 with *R. erythropolis* PR4 plasmid pREL1

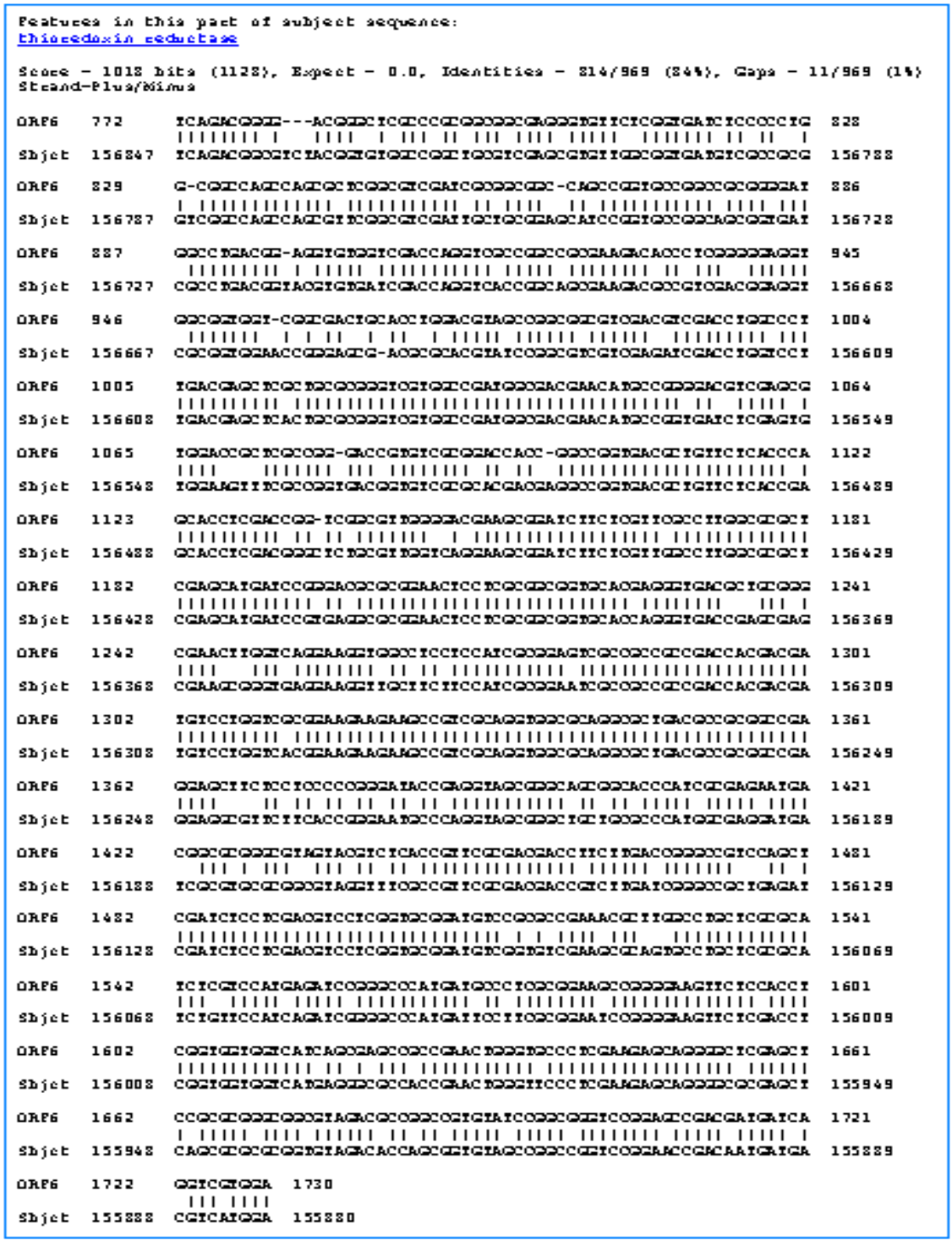


Figure 3.44: Nucleotide alignment of ORF6 with *R. erythropolis* PR4 plasmid pREL1.

Translated protein sequence analysis results of pKL120

The blastx results revealed that sections of the translated protein sequence of pKL120 showed substantial similarity to multiple protein matches from numerous microorganisms. Some of these significant translated protein matches are illustrated below.

Accession No.	Alignment	Score	E value
ref|YP_703611.1|	thioredoxin-disulfide reductase [Rhodococcus...	450	1e-124
ref|NP_898753.1|	putative thioredoxin reductase (TrxB) [Rhodo...	440	1e-121
ref|YP_345596.1|	thioredoxin reductase [Rhodococcus erythropo...	436	2e-120
ref|YP_121878.1|	putative thioredoxin reductase [Nocardia far...	413	2e-113
ref|YP_956825.1|	thioredoxin reductase [Mycobacterium vanbaal...	402	4e-110
emb|CAA65070.1|	thioredoxin [Mycobacterium tuberculosis]	378	3e-103
ref|NP_218430.1|	PROBABLE THIOREDOXIN REDUCTASE TRXB2 (TRXR) ...	374	3e-102
ref|NP_739543.1|	putative thioredoxin reductase [Corynebacter...	371	8e-101
ref|NP_602282.1|	thioredoxin reductase [Corynebacterium gluta...	369	3e-100
ref|YP_950223.1|	thioredoxin-disulfide reductase [Arthrobacte...	314	1e-83
ref|YP_829715.1|	thioredoxin reductase [Arthrobacter sp. FB24...	310	2e-82
ref|NP_628076.1|	thioredoxin reductase (NADPH) [Streptomyces ...	310	3e-82
ref|NP_630905.1|	thioredoxin reductase [Streptomyces coelicol...	307	1e-81
gb|ABB70177.1|	ArsT [Streptomyces sp. FR-008]	301	1e-79
ref|YP_345598.1|	putative FAD-dependent oxidoreductase [Rhodo...	242	2e-77

Figure 3.45: Blastx data of some significant alignment matches to the translated sequence of pKL120.

```
ref|YP_345598.1|  putative FAD-dependent oxidoreductase [Rhodococcus erythropolis PR4]
dbj|BAE46106.1|  putative FAD-dependent oxidoreductase [Rhodococcus erythropolis PR4]
Length=360

Score = 242 bits (618), Expect(2) = 2e-77,
Identities = 120/186 (64%), Positives = 142/186 (76%), Gaps = 0/186 (0%)
Frame = -3

ORF5    569  GMAEPPWPDGPPPASMVVDYPIAYEERYDLPIKRPVKVIGVDRDGQHLAVTIWQRIWRAR  390
             G   P W  GPPPASMVVDY IAYE RYDLP++RPVKVI V R G    V ++   W AR
Sbjct   68   GWPMPGWSPGPPPASMVVDYLIAYEIRYDLPVRRPVKVISVDRSGDGYLVESDHGSWSAR  127

ORF5    389  RVISAIGIWQQPPWPLYPGARPFIGRQLHIVDYRQPSDFAGARVVIVGGGNSAAQLVAEI  210
              V++  IGIW+QPPWP+YPG RPFIG Q+HI   YR P+ FAG++V +VGGGNSAAQ+VAE+
Sbjct   128  AVVWIIGIWAQPPWPIYPGIREFIGIQIHIISYRSPIPFAGSKVAVVGGGNSAAQIVAEV  187

ORF5    209  StvatttwvtIRPPRPLADDVDGRVLFQVASERAAAIAAGKPDPQGPASLGDIVGVPEVR  30
             SIVA I W+I+R PRPL DDVDGR+LF VA E A  IA+G+ D GG A LGDIVGVP V+
Sbjct   188  SIVAEILWLISREPRPLFDDVDGRILFDVAREHAREIASGEIDSGGIAGLGDIVGVPPVQ  247

ORF5    29   EARDRG  12
              AR+RG
Sbjct   248  AARERG  253

Score = 73.9 bits (180), Expect(2) = 2e-77
Identities = 46/70 (65%), Positives = 55/70 (78%), Gaps = 0/70 (0%)
Frame = -1

ORF5?   769  MGQQRSVIDMDVVVIgggqqagiaagyyLRRAGHRPVILDDQPVPQGRWPHIWPSLRLPSP  590
             M  Q +   +DVVVIGDQQAG+AAGYYLRR+G  PVILD +  PQG+W H W SL LPSP
Sbjct   1    MHSQGIPITVDVVVIGDQQAGMAAGYYLRRSGRDPVILDAETAPQGSWQHYWDSLHLPSP  60

ORF5?   589  AEPAALPGWP  560
             AE++LPGWP
Sbjct   61   AEYSSLPGWP  70
```

Figure 3.46: Translated protein alignment of ORF5 with *R. erythropolis* PR4 plasmid pREL1.

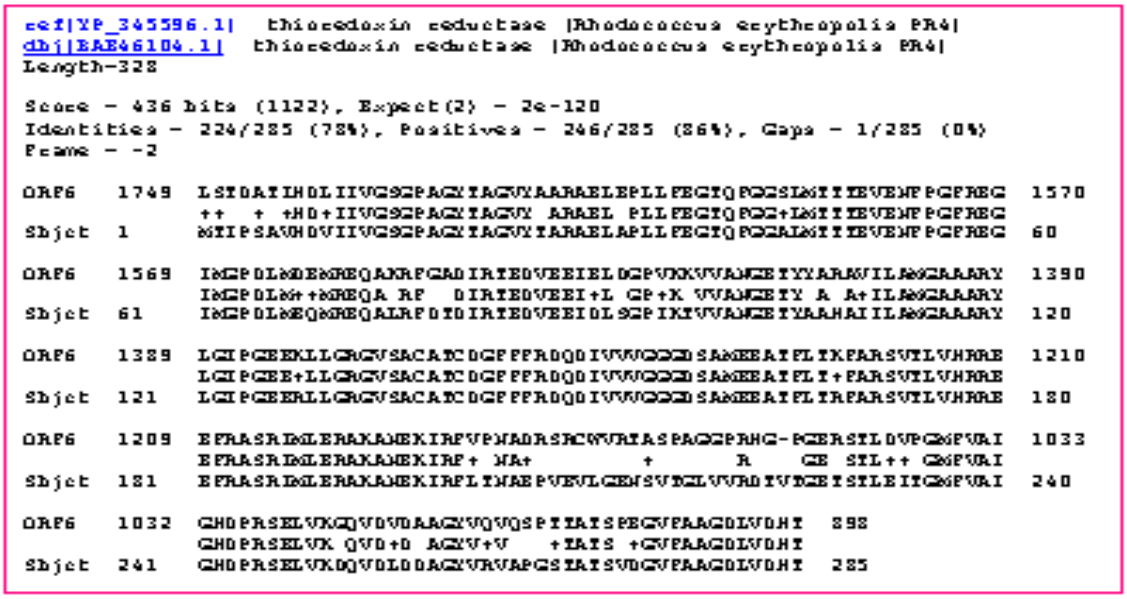


Figure 3.47: Translated protein alignment of ORF6 with *R. erythropolis* PR4 plasmid pREL1.

ORF 5 (ArsO)

Conserved domains search

The CD analysis showed that the partial 188 amino acid sequence of ORF5 aligned to portions of the TrkA, predicted flavoprotein involved in K+ transport (e value of $5e^{-19}$), FMO-like, flavin-binding monooxygenase-like (e value of $6e^{-13}$), TrxB, thioredoxin reductase (e value of $1e^{-6}$), Pyrredox2, pyridine nucleotide-disulphide oxidoreductase (e value of $3e^{-6}$), and the IucD, lysine/ornithine N-monooxygenase (e value of $2e^{-5}$) structural family domains.

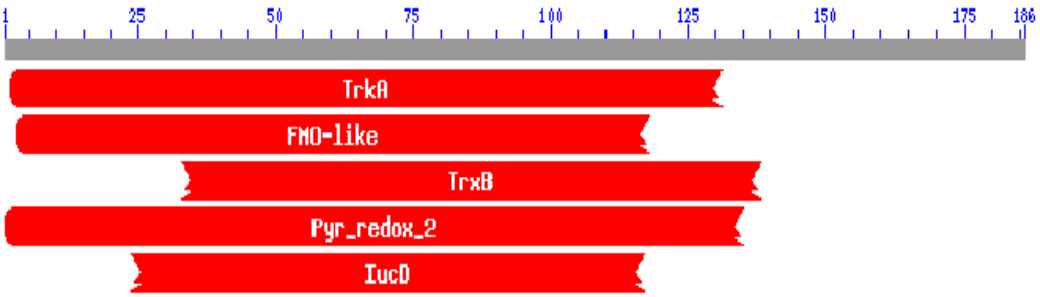


Figure 3.48: Putative conserved domain(s) found within ORF5.

Protein sequence alignment

The predicted amino acid sequence of ORF5 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E value
ref|YP_345598.1|	putative FAD-dependent oxidoreductase [Rhodo...	240	2e-62
ref|NP_630909.1|	monooxygenase [Streptomyces coelicolor A3(2)...	205	1e-51
gb|ABB70173.1|	Ars0 [Streptomyces sp. FR-008]	197	2e-49
emb|CAK50864.1|	putative monooxygenase [Streptomyces ambofaci...	197	2e-49
ref|NP_744078.1|	monooxygenase, putative [Pseudomonas putida ...	187	3e-46
ref|NP_737483.1|	putative oxidoreductase [Corynebacterium eff...	185	8e-46
gb|AAX37822.1|	pFQ25.6 [Streptomyces sp. FZ]	183	4e-45
ref|NP_627889.1|	oxidoreductase [Streptomyces coelicolor A3(2...	158	1e-37
ref|YP_704971.1|	possible oxidoreductase [Rhodococcus sp. RHA...	118	2e-25
ref|YP_831860.1|	oxidoreductase [Arthrobacter sp. FB24] >gb|A...	109	9e-23

Figure 3.49: Blastp data of some significant alignment matches to predicted sequence of ORF5

```
ref|YP_345598.1|  putative FAD-dependent oxidoreductase [Rhodococcus erythropolis PR4]
dbj|BAE46106.1|  putative FAD-dependent oxidoreductase [Rhodococcus erythropolis PR4]
Length=360

Score =  240 bits (613), Expect = 2e-62, Method: Composition-based stats.
Identities = 120/183 (65%), Positives = 142/183 (77%), Gaps = 0/183 (0%)

Query  1    MPPWPDGFPPASHVVDYFTAYEERYDLPIKRPVRVTGVDRDGQHLAVTTNQRIWRARRVI  60
            MP W  GFPPASHVVDY TAYE RYDLP++RPVRVT V R G    V ++   W AR V+
Sbjct  71   MPGWSPGFPPASHVVDYLTAYETRYDLPVRRPVRVTSVTRSGDGYLVESDHGSWSARAVV  130

Query  61   SATGTWQQPFWPLYPGAREFTGRQLHTVDYRQPSDFAGARVVIVGGGNSAAQLVAEISTV  120
            + TGTW+QPFWP+YPG REFTG Q+HT  YR P+ FAG++V +VGGGNSAAQ+VAE+STV
Sbjct  131  NTTGTWRQPFWPIYPGIREFTGTQIHTTSYRSPTPFAGSKVAVVGGGNSAAQIVAEVSTV  190

Query  121  ATTTWVTTRPPRFLADDVDGRVLFQVASERAAAIAAGKPDPGGPASLGDIVMVPEVREAR  180
            A T W+T+R PRFL DDVDGR+LF VA E A  IA+G+ D GG A LGDIVMVP V+ AR
Sbjct  191  AETLWLTSREPRFLPDDVDGRILFDVAREHAREIASGETDSGGIAGLGDIVMVPPVQAAR  250

Query  181  DRG  183
            +RG
Sbjct  251  ERG  253

gb|ABB70173.1|  Ars0 [Streptomyces sp. FR-008]
Length=356

Score =  197 bits (502), Expect = 2e-49, Method: Composition-based stats.
Identities = 105/178 (58%), Positives = 125/178 (70%), Gaps = 2/178 (1%)

Query  8    FPPASHVVDYFTAYEERYDLPIKRPVRVTGVDRDGQHLAVTTNQRIWRARRVISATGTWQ  67
            +P A HVVDY T YE+RYDLP++   RV  V RD   L V  +   WRAR VISATG+W
Sbjct  75   YPDARHVVDYLTDYEKRYDLPVQHGTRVDAVRRDAGRLLVEADTGTWRARAVISATGSWS  134

Query  68   QPFWPLYPGAREFTGRQLHTVDYRQPSDFAGARVVIVGGGNSAAQLVAEISTVATT--TW  125
            +PF P  PG   FTGRQLHTV+YR P+DF G RV++VGGGNS AQ+ A+++       TW
Sbjct  135  RPFLPAVPGREVFTGRQLHTVNYRSPADFTGQRVIVVGGGNSGAQVAADLALDGRVELTW  194

Query  126  VTTRPPRFLADDVDGRVLFQVASERAAAIAAGKPDPGGPASLGDIVMVPEVREARDRG  183
            VT RPPRFLADD+DGR LF VA+ R  A+  G+ D GG ASLGDIV VP VR+ARD G
Sbjct  195  VTQRPPRFLADDIDGRALFDVATARRRALEEGRSDTGGVASLGDIVAVPPVRQARDAG  252
```

Figure 3.50: Protein alignments of ORF5 with *Rhodococcus erythropolis* PR4 plasmid pREL1 and *Streptomyces* sp. FR-008 plasmid pHZ227.

```
ref|NP_600725.2|  predicted flavoprotein involved in K+ transport
[Corynebacterium glutamicum ATCC 13032]
ref|YP_225793.1|  FAD-dependent pyridine nucleotide-disulphide oxidoreductase
[Corynebacterium glutamicum ATCC 13032]
dbj|BAB98902.1|  Predicted flavoprotein involved in K+ transport
[Corynebacterium glutamicum ATCC 13032]
emb|CAF21517.1|  FAD-dependent pyridine nucleotide-disulphide oxidoreductase
[Corynebacterium glutamicum ATCC 13032]
Length=342

Score =  172 bits (436), Expect = 8e-42, Method: Composition-based stats.
Identities = 92/187 (49%), Positives = 117/187 (62%), Gaps = 6/187 (3%)

Query  1    MPPWP----DGFPPASHVVDYFTAYEERYDLPIKRPVRVTGVDRDGQHLAVTTNQRIWRA  56
            +P WP    DGFP +SHVVDY T YE RY++P+ R V V  VD       + +  R W A
Sbjct  73   LPGWPMPKFDGFPTSSHVVDYLTEYERRYEIPVSRSVSVDKVDFVNGAFHLQSGTRSWIA  132

Query  57   RRVISATGTWQQPFWPLYPGAREFTGRQLHTVDYRQPSDFAGARVVIVGGGNSAAQLVAE  116
            + V++ATGTW  PF P +PG  +F+G   H+ +Y     F G R+ IVGG NS AQ+ AE
Sbjct  133  KNVVAATGTWSAPFVPTFPG--KFSGTHWHSSNYPGVHPFVGQRIAIVGGANSGAQIAAE  190

Query  117  ISTVATTTWVTTRPPRFLADDVDGRVLFQVASERAAAIAAGKPDPGGPASLGDIVMVPEV  176
            +S V+  TW T   PR++ DDVDGR LF+    RA AI  G+ DPG  + LGDIVM+PEV
Sbjct  191  LSEVSEVTWFTRDKPRWMPDDVDGRELFRRNRLRALAIQRGERDPGPDSELGDIVMLPEV  250

Query  177  REARDRG  183
            R ARD G
Sbjct  251  RRARDSG  257
```

Figure 3.51: Protein alignment of ORF5 with *C. glutamicum* ATCC 25593.

ORF 6 (ArsT)

Conserved domains search

The CD analysis showed that most of the 325 amino acid sequence of ORF6 aligned to the thioredoxin reductase (TrxB) family structural domain with an e value of $1e^{-76}$, and its central protein region showed homology to another pyridine nucleotide-disulphide oxidoreductase (e value of $2e^{-6}$) conserved structural domain as well.

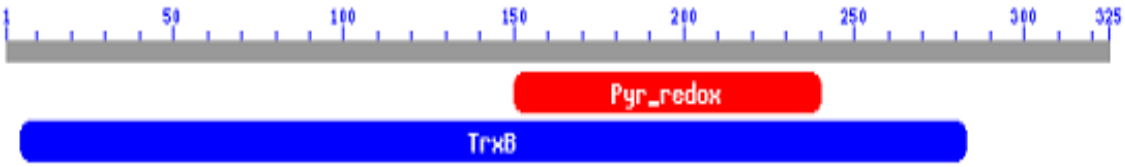


Figure 3.52: Putative conserved domain(s) found within ORF6.

Protein sequence alignment

The predicted amino acid sequence of ORF6 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.

Accession No.	Alignment	Score	E-value
ref|YP_703611.1|	thioredoxin-disulfide reductase [Rhodococcus...	442	1e-122
ref|NP_898753.1|	putative thioredoxin reductase (TrxB) [Rhodo...	430	5e-119
ref|YP_345596.1|	thioredoxin reductase [Rhodococcus erythropo...	427	6e-118
ref|YP_121878.1|	putative thioredoxin reductase [Nocardia far...	412	1e-113
ref|YP_956825.1|	thioredoxin reductase [Mycobacterium vanbaal...	381	3e-104
ref|NP_739543.1|	putative thioredoxin reductase [Corynebacter...	363	9e-99
ref|NP_602282.1|	thioredoxin reductase [Corynebacterium gluta...	360	7e-98
emb|CAA65070.1|	thioredoxin [Mycobacterium tuberculosis]	358	4e-97
ref|NP_218430.1|	PROBABLE THIOREDOXIN REDUCTASE TRXB2 (TRXR) ...	353	8e-96
ref|YP_950223.1|	thioredoxin-disulfide reductase [Arthrobacte...	302	2e-80
ref|NP_628076.1|	thioredoxin reductase (NADPH) [Streptomyces ...	298	2e-79
ref|YP_833636.1|	thioredoxin reductase [Arthrobacter sp. FB24...	296	1e-78
ref|YP_949864.1|	thioredoxin-disulfide reductase [Arthrobacte...	293	7e-78
gb|ABB70177.1|	ArsT [Streptomyces sp. FR-008]	286	1e-75

Figure 3.53: Blastp data of some significant alignment matches to predicted sequence of ORF6

```
ref|NP_898753.1|  putative thioredoxin reductase (TrxB) [Rhodococcus erythropolis]
gb|AAP74023.1|  putative thioredoxin reductase (TrxB) [Rhodococcus erythropolis]
Length=328

Score = 430 bits (1106),  Expect = 5e-119, Method: Composition-based stats.
Identities = 243/327 (74%), Positives = 268/327 (81%), Gaps = 6/327 (1%)

Query  1    LSTDATIHDLIIVGSGPAGYTAGVYAARAELEPLLFEGTQFGGSLMTTTEVENFPGFREG  60
            ++  + +HD+IIVGSGPAGYTAGVY ARAEL PLLFEGTQFGG+LMTTTEVENFPGFREG
Sbjct  1    MTIPSAVHDVIIVGSGPAGYTAGVYTARAELAPLLFEGTQFGGALMTTTEVENFPGFREG  60

Query  61   IMGPDLMDEMREQAKRFGADIRTEDVEEIELDGPVKKVVANGETYYARAVILAMGAAARY  120
            IMGPDLM++MREQA RF  DIRTEDVEEI+L GP+K VVANGETY A A+ILAMGAAARY
Sbjct  61   IMGPDLMEQMREQALRFDTDIRTEDVEEIDLSGPIKTVVANGETYAARAIILAMGAAARY  120

Query  121  LGIPGEEKLLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLTKFARSVTLVHRRE  180
            LGIPGEE+LLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLT+FARSVTLVHRRE
Sbjct  121  LGIPGEERLLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLTRFARSVTLVHRRE  180

Query  181  EFRASRIMLERAKANEKIRFVPNADRSRCGVRTASPAGGPRHG-PGERSTLDVPGMFVAI  239
            EFRASRIMLERAKANEKIRF+ NA+          +       R     GE STL++ GMFVAI
Sbjct  181  EFRASRIMLERAKANEKIRFLTNAEPVEVLGENSVTGLVVRDTVTGETSTLEITGMFVAI  240

Query  240  GHDPRSELVKGQVDVDAAGYVQVQSPTTATSPEGVFAAGDLVDHTSVRPSPRPAPAGRR-  298
            GHDPRSELVKGQVD+D AGYV+V   +TATS +GVFAAGDLVDHT  R +   A  G
Sbjct  241  GHDPRSELVKDQWDLDDAGYVRVAPGSTATSVDGVFAAGDLVDHT-YRQAITAAGTGCSA  299

Query  299  --DRRRALAGRQGEITENTLAAAGEPV  323
              D  R LA R G+IT NTL AAG  V
Sbjct 300 AIDAERWLADR-GDITANTLDAAGHTV 325

ref|YP_345596.1|  thioredoxin reductase [Rhodococcus erythropolis PR4]
dbj|BAE46104.1|  thioredoxin reductase [Rhodococcus erythropolis PR4]
Length=328

Score =  458 bits (1178),  Expect = 2e-127, Method: Compositional matrix adjust.
Identities = 242/327 (74%), Positives = 266/327 (81%), Gaps = 6/327 (1%)

Query  1    LSTDATIHDLIIVGSGPAGYTAGVYAARAELEPLLFEGTQFGGSLMTTTEVENFPGFREG  60
            ++  + +HD+IIVGSGPAGYTAGVY ARAEL PLLFEGTQFGG+LMTTTEVENFPGFREG
Sbjct  1    MTIPSAVHDVIIVGSGPAGYTAGVYTARAELAPLLFEGTQFGGALMTTTEVENFPGFREG  60

Query  61   IMGPDLMDEMREQAKRFGADIRTEDVEEIELDGPVKKVVANGETYYARAVILAMGAAARY  120
            IMGPDLM++MREQA RF  DIRTEDVEEI+L GP+K VVANGETY A A+ILAMGAAARY
Sbjct  61   IMGPDLMEQMREQALRFDTDIRTEDVEEIDLSGPIKTVVANGETYAARAIILAMGAAARY  120

Query  121  LGIPGEEKLLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLTKFARSVTLVHRRE  180
            LGIPGEE+LLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLT+FARSVTLVHRRE
Sbjct  121  LGIPGEERLLGRGVSACATCDGFFFRDQDIVVVGGGDSAMEEATFLTRFARSVTLVHRRE  180

Query  181  EFRASRIMLERAKANEKIRFVPNADRSRCGVRTASPAGGPRHG-PGERSTLDVPGMFVAI  239
            EFRASRIMLERAKANEKIRF+ NA+          +       R     GE STL++ GMFVAI
Sbjct  181  EFRASRIMLERAKANEKIRFLTNAEPVEVLGENSVTGLVVRDTVTGETSTLEITGMFVAI  240

Query  240  GHDPRSELVKGQVDVDAAGYVQVQSPTTATSPEGVFAAGDLVDHTSVRPSPRPAPAGRR-  298
            GHDPRSELVK  QVD+D AGYV+V   +TATS +GVFAAGDLVDHT  R +   A  G
Sbjct  241  GHDPRSELVKDQWDLDDAGYVRVAPGSTATSVDGVFAAGDLVDHT-YRQAITAAGTGCSA  299

Query  299  --DRRRALAGRQGEITENTLAAAGEPV  323
              D  R LA R G+IT NTL AAG  V
Sbjct 300  AIDAERWLADR-GDITANTLDAAGHTV  325
```

Figure 3.54: Protein alignments of ORF6 with *R. erythropolis* plasmid pBD2 and *R. erythropolis* PR4 plasmid pREL1.

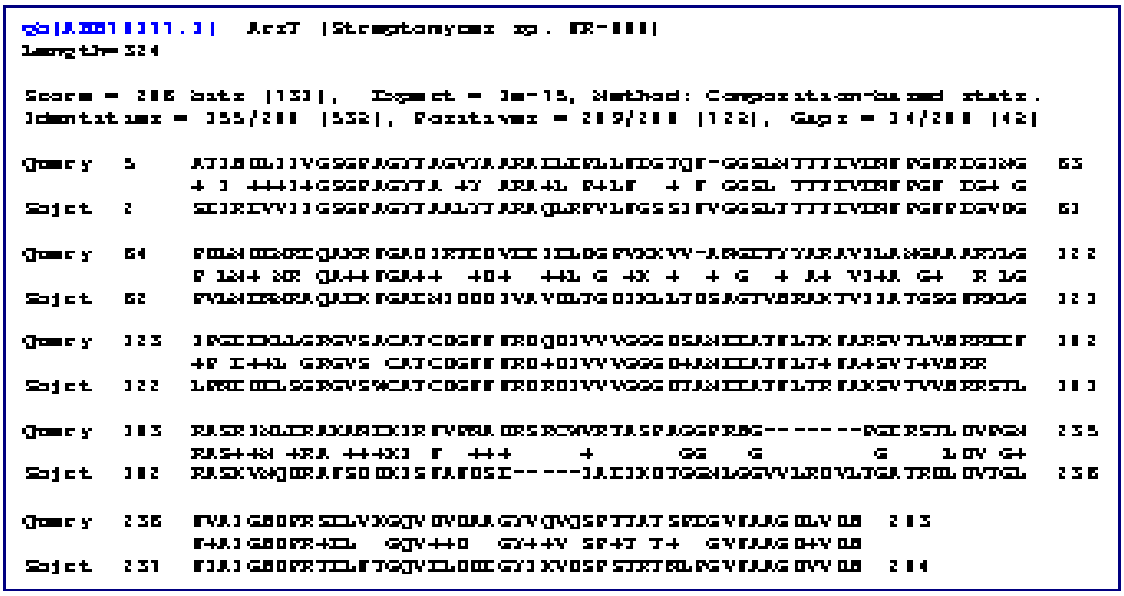


Figure 3.55: Protein alignment of ORF6 with *Streptomyces* sp. str. FR-008 plasmid pHZ227.

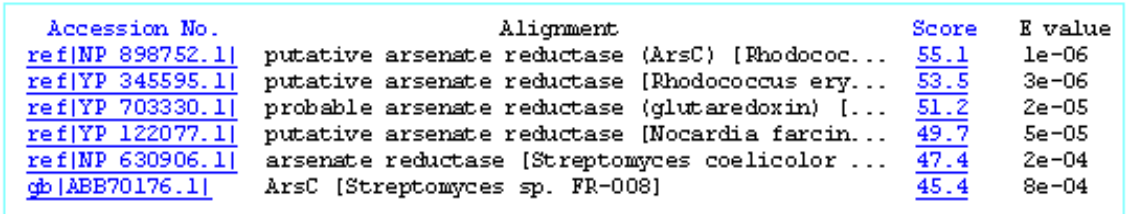
ORF 7 (ArsC4)

Conserved domains search

The CD analysis showed that the partial 43 amino acid sequence of ORF7 showed homology with no putative conserved domain probably because it was to small to constitute a full protein domain.

Protein sequence alignment

The predicted amino acid sequence of ORF7 was used to search protein databases for similar matches and some of the most significant alignments are listed below in the following table.



Accession No.	Alignment	Score	E value
ref|NP_898752.1|	putative arsenate reductase (ArsC) [Rhodococ...	55.1	1e-06
ref|YP_345595.1|	putative arsenate reductase [Rhodococcus ery...	53.5	3e-06
ref|YP_703330.1|	probable arsenate reductase (glutaredoxin) [...	51.2	2e-05
ref|YP_122077.1|	putative arsenate reductase [Nocardia farcin...	49.7	5e-05
ref|NP_630906.1|	arsenate reductase [Streptomyces coelicolor ...	47.4	2e-04
gb|ABB70176.1|	ArsC [Streptomyces sp. FR-008]	45.4	8e-04

Figure 3.56: Blastp data of some significant alignment matches to predicted sequence of ORF7.

```
ref|NP_898752.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
gb|AAP74022.1|  putative arsenate reductase (ArsC) [Rhodococcus erythropolis]
Length=138

Score = 55.1 bits (131),  Expect = 1e-06, Method: Composition-based stats.
Identities = 26/34 (76%), Positives = 30/34 (88%), Gaps = 0/34 (0%)

Query  5    WYPGNPAGQGVDAVRPIRDEIKAKVEALIAELTA  38
            W  G+PAGQG+DAVRPIRDEIK  V+ALI+ELTA
Sbjct  100  WVLGDPAGQGIDAVRPIRDEIKTLVQALISELTA  133

ref|YP_122077.1|  putative arsenate reductase [Nocardia farcinica IFM 10152]
dbj|BAD60713.1|  putative arsenate reductase [Nocardia farcinica IFM 10152]
Length=136

Score = 49.7 bits (117),  Expect = 5e-05, Method: Composition-based stats.
Identities = 26/39 (66%), Positives = 33/39 (84%), Gaps = 2/39 (5%)

Query  5    WYPGNPAGQGVDAVRPIRDEIKAKVEALIAELTAEPARS  43
            W   +PAG+GV++VRPIRDEIKA+VEAL+AEL   PAR+
Sbjct  100  WVLEDPAGKGVESVRPIRDEIKARVEALVAELL--PARA  136

gb|ABB70176.1|  ArsC [Streptomyces sp. FR-008]
Length=140

Score = 45.4 bits (106),  Expect = 8e-04, Method: Composition-based stats.
Identities = 24/35 (68%), Positives = 26/35 (74%), Gaps = 0/35 (0%)

Query  5    WYPGNPAGQGVDAVRPIRDEIKAKVEALIAELTAE  39
            W   +PAGQGV AVRPIRDEIK  VE LIAE+  E
Sbjct  100  WKLKDPAGQGVAAVRPIRDEIKTLVEGLIAEIAPE  134
```

Figure 3.57: Protein alignments of ORF7 with *R. erythropolis* plasmid pBD2, *N. farcinica* IFM 10152 plasmid pNF2 and *Streptomyces* sp. FR-008 plasmid pHZ227.

Chapter 4

Discussion

4.1 Perspective on the cloning efficiency of nocardioform DNA into vectors

Even the biotech company Inqaba was unable to clone pKL319 and residual ∼3.0 kbp of the original plasmid construct pKL1 using proven effective cloning kits. As no or only small inserts were obtained from the plasmid screens after the transformations attempts on the *Pst*I fragments from pDA37 and pKL125 into pUC18 indicates that these DNA segments might be unstable in this vector and consequently disintegrate.

The presence of secondary structures in pKL125 and pKL319 could be another contributing factor towards the problems encountered with their subcloning.

As the sequence analysis of pKL316 contained the terminal region of an ArsB protein, its possible overexpression could very plausibly explain partly if not totally the low cloning efficiency of pKL319 as this protein is toxic at high concentrations [**163**]. This would consequently clarify together with other contributing factors the inability to subclone the *BamH*I-*Nsi*I fragment of pKL335. The potential presence of this transmembrane channel protein in the *Pst*I fragments of pDA37 and pKL125 could also justify the problems encountered with their cloning as well. The higher success rate of cloning these subclones into pDA71 than pUC18 could be as a result of the former being a lower copy number plasmid and consequently the prospective toxic effect of ArsB expression being less prominent.

4.2 False positive clones obtained during screening of *Bgl*II *G. rubropertincta* library

Firstly, since this *Bgl*II digested library of *Gordonia rubropertincta* ATCC 25593 constructed by S. Quan was not for degree purposes, no clear record of the details of it was accessible. As a result, no information on the average insert size, no. of clones obtained, and most importantly the probability value of the final library was available. Even though ∼20000 clones were attained during the reconstruction of this library in *E. coli*, it can not be said for certain whether the residual recombinant plasmid DNA was truly representative of the entire original library.
Although, the false positive clones were ultimately showed not to confer either arsenic resistance or specific carbon source utilization abilities, they must have had some selective advantage that allowed them to grow on the screening plates initially. Neither mutations nor any residual growth components in the dilution cultures can fully account for the numerous putative clones achieved during the screening process. Also, the putative clones were not a contaminant as the *Rhodococcus*-specific phages were able to plaque (form clearing zones) on a lawn of these respective colonies.
In the case of these anomalous As^R clones, mutations might have occurred in *R. erythropolis* SQ1 that either led to the upregulation of some efflux protein or a decreased expression of a certain unspecific PO_4^{3-} transporter system. In both cases it would have resulted in a reduced accumulation of arsenic and therefore enabled those mutants to grow better than the wild-type SQ1 cells. Studies on arsenic resistance in *E. coli* and the yeast *Saccharomyces cerevisiae* have shown that mutations in the Pit, GlpF and Fps1p transport systems led to a higher arsenic tolerance phenotype. The fact that the ∼10-20 random clones that were chosen didn't exhibit arsenic resistance does not necessarily mean that all of them were false positive, considering that true As^R clones were obtained during the screening of the *Pst*I library of the same organism. No definitive generalization can be made on the arsenic resistant phenotype of all as the number of samples analyzed represent less than 5% of the total number of clones.
A fascinating irregularity is that most of the positive clones obtained in the *Bgl*II library were putative arsenite resistant clones whereas in the *Pst*I library the majority were arsenate tolerant clones. Ideally, you would expect the same number of resistant clones to be attained for both analyzed arsenic substances, since *ars* operons usually confer resistance to both. This suggests that most of the putative arsenite resistant clones obtained from the *Bgl*II library are most likely false positives and realistically only about 10% could genuinely be arsenic resistant, in that way the number of clones would be of a similar magnitude for both arsenical species.

The absence or disruption of intact arsenic resistance genes due to the presence of *Bgl*II and *Pst*I restriction sites in the middle of them does not explain or correlate with the experimental data found in the detailed analysis of the As^R clones of the *Pst*I library. Moreover, no *Bgl*II site was observed in either of the 3.5 or 5.1kbp arsenic resistant DNA fragments during restriction mapping. Even though no significant number of arsenite resistant clones were obtained in the *Pst*I library screening, the analyzed arsenate clones did demonstrate increased AsO_2^- tolerance levels when a detailed As^R phenotypic assay were performed on them. As the screening amount of 10mM arsenite is actually above the established MIC for *G. rubropertincta* (∼7.5mM) it not so astonishing of the lack of reputed AsO_2^--resistant clones. In hindsight, screening for arsenite resistance in this bacterium should have been conducted at a concentration of about 5mM.
Some of the erroneous putative clones attained from the screening for unique carbon source utilization genes could possibly code for agarases and similar proteins. However, the fact that growth was noticed on both sides during the screening for these catabolic genes in the *Pst*I library affirms the likely suspicion that these colonies are pure artefacts. Curiously, it would be of note to verify whether indeed *R. erythropolis* SQ1 and *G. rubropertincta* ATCC 25593 at all have genes coding for agarase-like proteins. From these inconclusive and speculative screening results for these catabolic genes it looks very doubtful that both organisms contain genes capable of utilizing ligninolytic compounds or benzoate.

4.3 As^R phenotype comparison of resistant clones and subclones of *G. rubropertincta*

From the tables 3.4 and 3.6 it can see that all the analyzed As^R clones exhibit essentially the same arsenite resistance profile than that of *G. rubropertincta*, indicating than all the necessary determinants (ArsB) conferring AsO_2^- resistance exist intact in these clones. The presence of the terminal region of this arsenite transmembrane channel protein shown by the sequence analysis of pKL316 confirms with any doubt that all the clones that contained the 3.5 kbp DNA fragment harbours a complete *ars*B gene. The partial sequence alignment to a ArsB protein in this ∼1.6kbp DNA fragment also correlates with the loss of As(III) tolerance observed in both pKL335 *Sfu*I subclones due to this gene being disrupted. The pKL120 construct in essence displayed the same arsenite resistance phenotype as the control, indicating the absence of any ArsB-like geneproduct in this cloned insert. Alignment analysis of this 2.0 kbp *Pst*I-*BamH*I DNA segment supported this statement as neither the predicted thioredoxin reductase nor a putative oxidoreductase/flavin-dependent monooxygenase would plausibly result or play a role in conferring arsenite resistance.

Since an intermediary As(III) tolerance profile was noted for the internal *BamH*I subclone of pKL1 hints to the possibility that it might contain a major part of a gene encoding for some sort of arsenite efflux protein. Fully sequencing the whole 5.1 kbp of clone 1 will definitively elucidate the exact structure and what translation products are coded for in the rest of this DNA fragment.

The recombinant plasmid construct pKL1 is the only clone which demonstrates a comparable arsenate resistance phenotype to that of *Gordonia rubropertincta* ATCC 25593. Therefore, this DNA segment most likely includes all the required components (thioredoxin reductase (TR) and arsenate reductase (ArsC)) that contribute to the host organism's tolerance to arsenate. Even though the other investigated clones exhibited enhanced resistance levels it were not strictly to the same extent as that of the original bacteria's As(V) tolerance profile. This suggests that the concomitant expression of TR with ArsC(s) found in clone 1 give rise to greater resistance, and that this is most probably the reason for the difference in arsenate tolerance capacity between pKL1 and the other clones. The pKL316 *Sfu*I subclone illustrates almost the same resistance phenotype than the original clone indicating that this DNA segment has basically all the required constituents that resulted in conferring tolerance to AsO_4^{3-} in pKL335. Examining the ∼1.6kbp DNA fragment's sequence illustrated the existence of multiple *ars*C genes that are crucial for conferring arsenate resistance, thereby substantiating the experimental data of the As^R assay for this subclone (table 3.7). However, the disruption of the ArsB gene product is perhaps the reason for the discrepancy in the arsenate tolerance levels between pKL335 and pKL316 as it appears that this protein contributes to resistance to As(V) as well in some way.

Since ∼1.9kbp *Pst*I-*Sfu*I fragment presumably then comprises of the rest of the oxyanion efflux protein ArsB and the trans-acting repressor ArsR which would predictably not lead to any substantial effect towards enhanced AsO_4^{3-} resistance. Again, the predicted gene products of pKL120 would, as the arsenic tolerance profiling of this subclone showed, generate no significant contribution to the resistance phenotype of the original recombinant construct. As it can only be speculated on the precise gene organization in the internal 2.5kbp *BamH*I fragment and the remainder of pKL1, no accurate assumptions on the lack of arsenate resistance in pKL125 can really be made. The partial sequence of a putative arsenate reductase detected in the pKL120 cloned insert indicates to the presence of at least one *ars*C gene but we can only simply guess that clone 1 possibly hold at least one intact ArsB protein in order to explain the observed As^R phenotype of the original plasmid construct pKL1.

4.4 Properties of the partial *ars* operons structures for both pKL316 and pKL120

The G+C contents of pKL316 and pKL120 that were 64.8% and 70.8% respectively are in accordance with expected value of 63-69% for *Gordonia* genomic DNA. From the partial sequences of the arsenic resistant inserts in pKL335 and pKL1 it can undoubtedly be said that *Gordonia rubropertincta* ATCC 25593 contains one or two *ars* operons. Until the complete sequences for both the 3.5kbp and 5.1kbp *Pst*I inserts are known, there is no definite way at the moment of telling whether they are different DNA fragments of the same operon or originate from two distinct ones. Although, both possess a similar ∼700bp *Pst*I-*BamH*I fragment, the fact that pKL1 is not restricted by *Sfu*I casts uncertainty on the premise that both belong to the same arsenic resistance gene cluster. However, it does not exclude the slight possibility that the pKL1 cloned insert is due to the result of a plasmid rearrangement of the ∼3.5kbp and ∼1.5kbp nucleotide segments.

Even though the complete sequence of pKL335 is not available, it is highly probable that pKL319 possess the large residual section of the ArsB-like protein and the trans-acting repressor ArsR and consequently resemble the *ars* operon structure of *C. efficiens* YS-314. It seems to be a phenomenon amongst nocardioforms and their close relatives to contain multiple arsenate reductases in a single operon as this have not been noticed in any other Gram(-) or low G+C Gram-positive bacteria. The exact role and significance of each of these ArsCs are up to further scientific work as these ORFs (2,3,4, and 7) are not the result of duplication of an existing gene since their amino acid sequences are far from identical.

The genetic arrangement of ORF5 (ArsO homolog) and ORF6 (ArsT homolog) in pKL120 next to a putative ArsC gene product infers that these proteins are likely to be coupled with an existing arsenic resistance operon. The presence of thioredoxin reductase and flavin-binding monooxygenase analogs has only recently been detected in the *Streptomyces* sp. str. FR-008 linear plasmid pHZ227 [**156**]. A putative TrxB homolog has also been found in an arsenic resistance cluster of *Rhodococcus erythropolis* linear pBD2 megaplasmid [**140**]. The absence of *ars*T homologs in other characterized operons could be that this required TR activity are drafted from distantly located, chromosomal TrxB-like proteins. The exact function of ORF5 (ArsO-like) in an arsenic efflux detoxification systems is unknown but can maybe display a similar role as ArsH which shows weak homology to oxidoreductases. More experimental studies needs to be done in order to clarify the significance of FMO-like proteins in an As^R gene cluster framework

4.5 Arsenic resistance genes

4.5.1 Arsenite efflux protein (ArsB)

From the alignment and CD results for the partial 65aa sequence clearly demonstrates a strong homology to Acr3 family of arsenite efflux ArsB proteins. Moreover, this indicates that most ArsB proteins from the mycolate-containing actinomycetes belong to the ACR3 group of arsenite transport channels. Furthermore, as this protein sequence compare highly to the last 60+ amino acids of other ArsB proteins of nocardioforms and relating species suggest clearly that the true ORF length of this putative As-inducing oxyanion translocating protein would be in the range of 365aa.

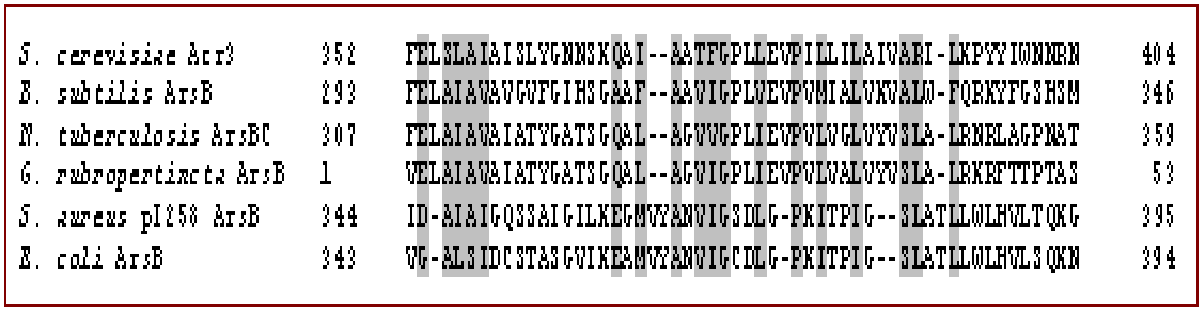


Figure 4.1: Partial ArsB protein sequence alignment from *G. rubropertincta* and other diverse microorganisms; where the conserved amino acids are highlighted in gray.

As little is known of the structures of ArsB proteins besides that from *E. coli* plasmid R773 we can not really deliberate on the conservation of any amino acids that are essential for function.

4.5.2 Arsenate reductase (ArsC)

Even though most of the arsenate reductases that have been detected and/or discovered from a diverse group of organisms contains only between 130-140aa, several ArsC homologs in other nocardioforms' *ars* operons comprises of 210-225aa that shows strong homology to the 205 amino acid sequence of ORF2. The alignment of this predicted protein sequence shows similarity ranging from 35-75% identity to various ArsC homologs and lmwPTPases paralogs from mainly actinomycetes species. CD analysis illustrates that the expected and predicted low molecular weight protein tyrosine phosphatase conserved domain basically only extend to the last 130aa of this ArsC homolog. This suggests that these ArsC-like proteins, such as ORF2, are essentially the typical arsenate reductase with a residual N-terminal region containing a well conserved ETIERFL sequence motif.

Possibly, this might be a remnant from the evolutionary transition of an ancestral phosphatase drafted for arsenate reductase duty. As the Trx-dependent ArsC proteins demonstrate partial phosphatase activity, this additional portion found in ORF2 might be a necessary part for conferring a complete phosphatase phenotype. The nucleotide based search showed that the alignment of ORF2 with one of the ArsC homologs of *R. erythropolis* plasmid pBD2 and *A. aurescens* TC1 plasmid pTC2 started from base 232 which is outside the predicted protein region that supposedly begins at position 259. This could imply that the actual size of this arsenate reductase might be slightly larger (∼215aa) than Genemark predicted. It is highly unlikely that the translated sequence just upstream of the predicted ORF2 aligning fairly well with corresponding sections of other ArsC homologs to be a mere coincidence. One thing that might have contributed to this ambiguity is a possible mismatch in the Shine-Dalgarno sequence preceding this ORF during sequencing, consequently leading to an error in the starting location prediction.

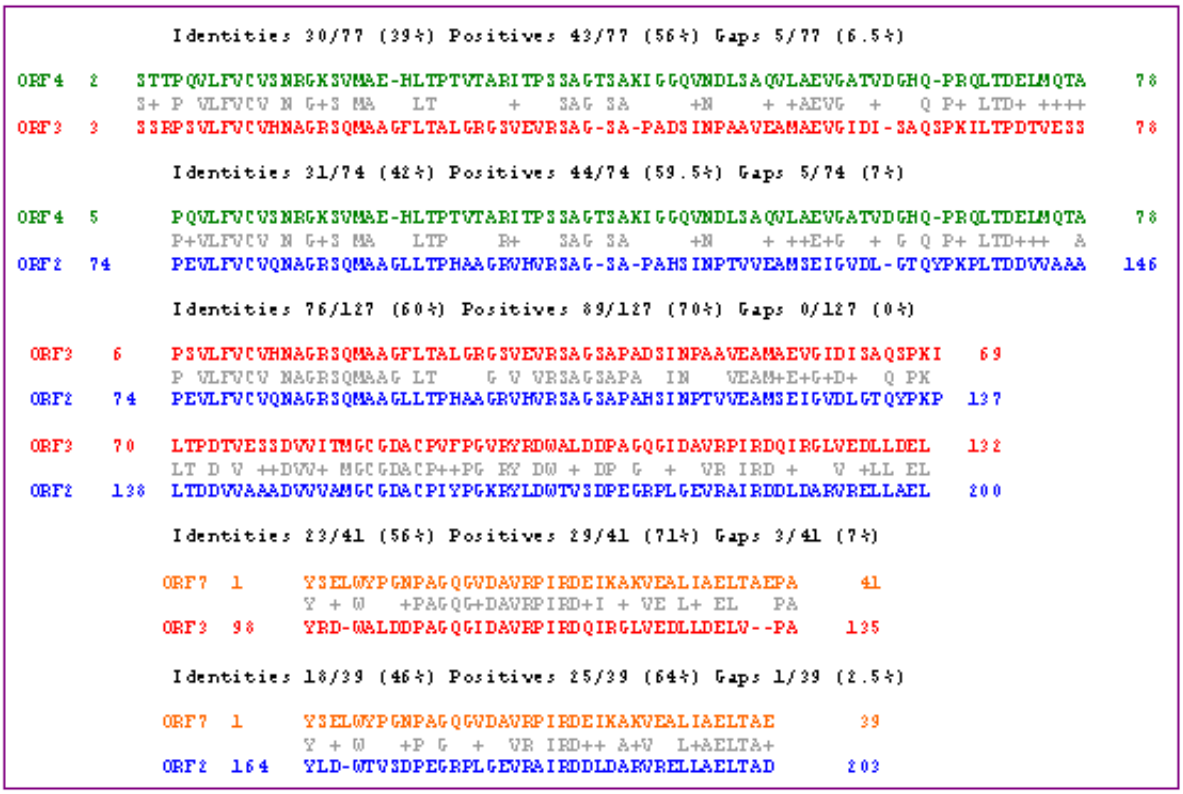


Figure 4.2: Alignment of the predicted putative arsenate reductase protein sequences detected in the *G. rubropertincta* arsenic resistant clones.

The predicted 141aa of ORF3 again align substantially with miscellaneous putative ArsCs from various organisms and is of comparable size to most of the arsenate reductases that have so far been detected. Notably, this protein were moderately analogous to the ArsCs from *S. aureus* plasmid pI258 (40% amino acid identity) and *Halobacterium* sp. NRC-1 plasmid pNRC1 (45% amino acid identity) but not significantly homologous to those of *S. cerevisiae* and *E. coli*. As expected ORF3 contain the lmwPTPase domain and basically stretches the whole protein.
The predicted 78 and 43aa protein sequences also aligned partially but considerably to different parts of numerous putative Trx-type ArsC homologs. Yet again, ORF4 consisted of the N-terminal region of the LMWPc domain, whereas ORF7 were probably too small to sufficiently demonstrate any putative conserved domain. Although, we are unable to affirm whether these partial protein sequences are sections of the same arsenate reductase, it is unlikely that they originate from a single protein.
It can clearly be seen from figure 4.2 that all of the putative arsenate reductases detected from the *G. rubropertincta* ATCC 25593 resistant clones are different and distinct. Also, ORF2,3 and 4 contain the characteristic P-loop CXXNXXRS (CX_5R) active site sequence motif of the Trx-ArsC family and their lmwPTPases paralogs. Furthermore, 2 and 3 illustrate all three essential catalytic cysteines (Cys10, 82, and 89) as well as the equivalent aspartic acid D105 found in the ArsC of *S. aureus* plasmid pI258.
From inspecting this alignment comparison of the various ArsCs from selected actinomycetes species that it seems that there occurs three different ArsC types in most of them, one of 210-225aa, one of 135-140aa with the three essential cysteines, and one of 130-150aa with only one catalytic cysteine present. It can be postulated that ORF4 might belong to the last group of arsenate reductases, whereas ORF7 appears to be another specimen of the second class but only a complete protein sequence of both of them will definitively confirm this notion. Interestingly, all the classified group 3 ArsC homologs have an arginine to the structural similar lysine amino acid change at the corresponding position R16 of *S. aureus* pI258 ArsC. As ORF4 contain that same $^{17}R \rightarrow K$ amino acid change that the group 3 Trx-type arsenate reductase has, adds value to the argument that it belongs to this group of ArsCs. Also, as all group 3 ArsCs, possibly including ORF4, only hold one critical cysteine residue in their entire protein sequence implies a different reaction mechanism resembling perhaps that of the Grx/GSH arsenate reductases. Strikingly, it is mainly the C terminus sequence (last $\sim$50aa) of group 2 and 3 that are rather divergent as the rest are fairly analogous between the two ArsC classes. These arbitrary Trx-type ArsC groups most likely originated from the same common phosphatase ancestor, especially the first two classes. These multiple ArsC within a single operon are probably a great contributing factor in why nocardioforms and related species demonstrate highly elevated arsenic resistance levels.

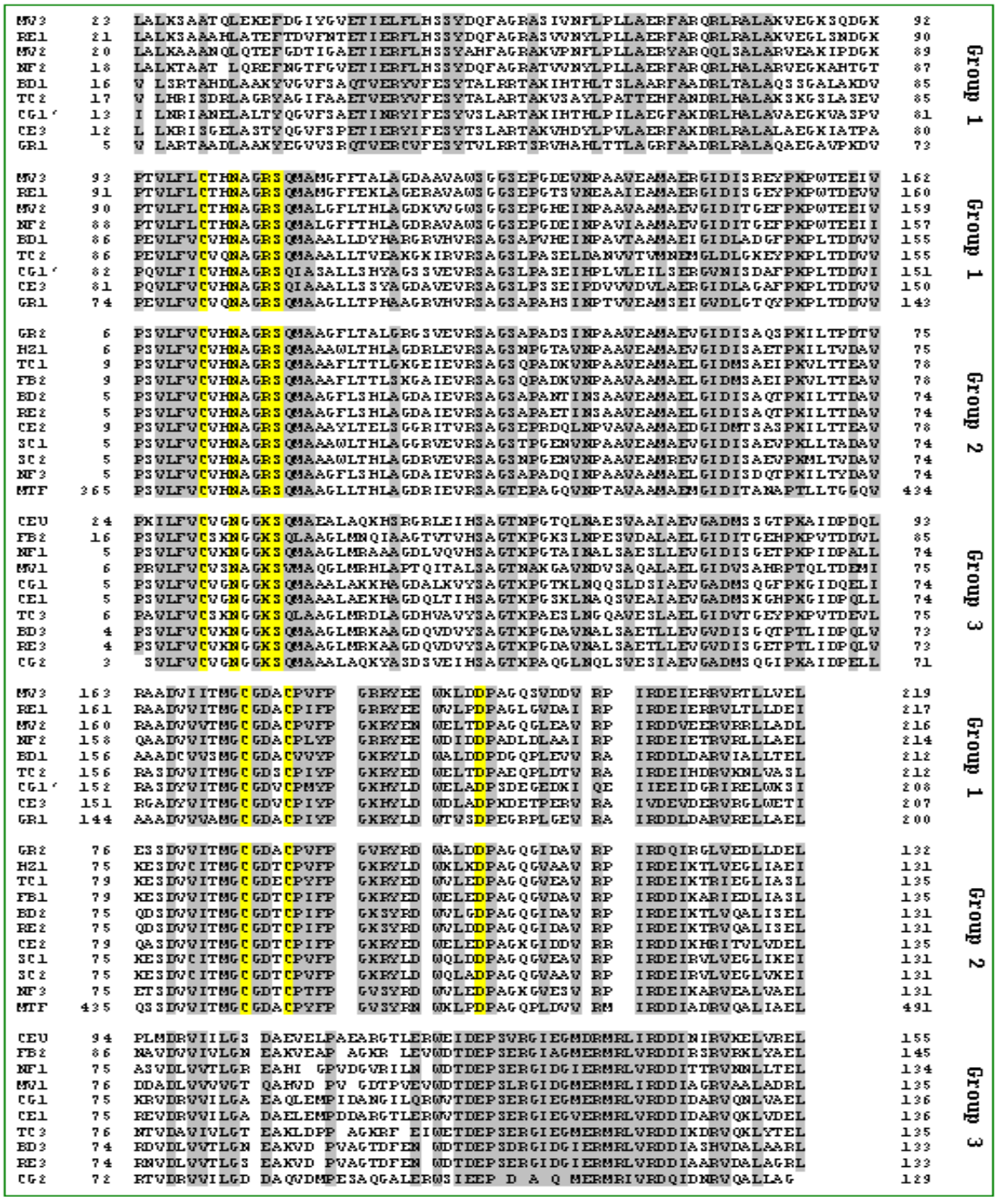


Figure 4.3: Protein sequence comparison of ArsCs from *G. rubropertincta* (GR), *R. erythropolis* plasmid pBD2 (BD)[NC_005073], *R. erythropolis* PR4 plasmid pREL1 (RE)[NC_007491], *N. farcinica* 10152 plasmid pNF2 (NF)[NC_006363], *C. efficiens* YS-314 (CE)[NC_004369], *C. glutamicum* ATCC 13032 (CG)[NC_003450], *A. aurescens* TC1 plasmid pTC2 (TC)[NC_008713], *Arthrobacter* sp. FB24 (FB)[NC_008541], *Streptomyces* sp. FR-008 plasmid pHZ227 (HZ)[DQ231520], *S. coelicolor* A3(2) (SC)[NC_003888], *M. vanbaalenii* PYR-1 (MV)[NC_008726] and *M. tuberculosis* fusion ArsBC protein (MTF)[NC_000962]; where conserved amino acids are highlighted in gray and catalytic important residues highlighted in yellow.

4.5.3 Thioredoxin reductase

ORF6 shows considerable similarity to the thioredoxin reductases of various actinomycetes, most notably with those found in the *ars* operons of *R. erythropolis* plasmid pBD2 (77% amino acid identity) and *Streptomyces* sp. FR-008 plasmid pHZ22777 (52% amino acid identity). The predicted TrxB domain extend basically the entire sequence and another pyridine nucleotide-disulphide oxidoreductase domain is detected in the central part (150-250aa) of the protein as encountered in the characterized thioredoxin reductase of *M. tuberculosis*. The length of ORF6 corresponds with the usual protein size (∼325-330 amino acids) of the typical TrxB homologs of nocardioforms and related species. The presence of a thioredoxin reductase in an arsenic resistant gene cluster and therefore the simultaneous expression of ArsC and TrxB proteins is not that surprising as the function of these arsenate reductases are dependent on the activity of TR. The coexpression of both ArsC and TrxB proteins are probably one of the reasons why nocardioforms demonstrate very high arsenic tolerances levels compared to most other microorganisms. Since the thioredoxin reductase activity is normally acquired from distant chromosomal TrxB homologs, it is therefore not typically found in many other *ars* operons.

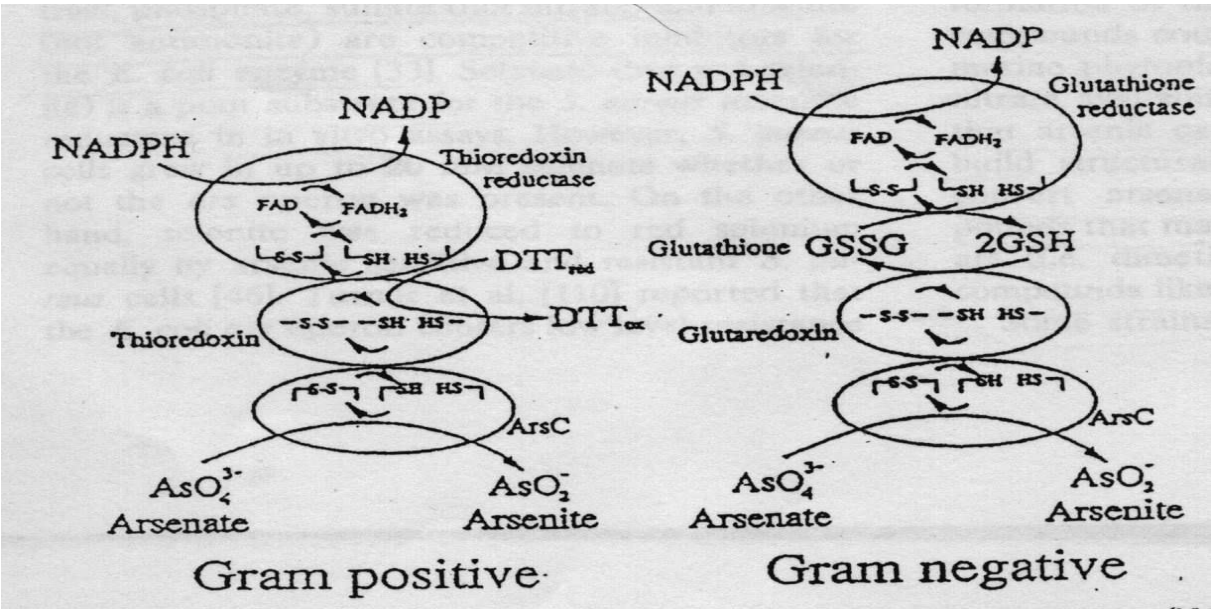


Figure 4.4: Proposed energy coupling of Gram(+) and Gram(-) bacterial arsenate reductases. (From **[32]**)

4.5.4 Putative flavoproteins involved in arsenic resistance

The CD search for ORF5 shows that its N-terminal region demonstrates strong homology particularly to the TrkA conserved domain that include flavoproteins involved in potassium transport. Moreover, this section of ORF5 also illustrates significant similarity to flavin-binding monooxygenases containing the FMO-like putative domain as well as to a different pyridine nucleotide- disulphide oxidoreductase domain 2 to which ORF6 revealed significant homology too. Not surprisingly, various proteins belonging to these putative conserved domain families showed substantial likeness to the presumably partial sequence of ORF5. Blastn analysis of pKL120 shows curiously that putative FAD-dependent oxidoreductases and monooxygenases from *R. erythropolis* plasmid pREL1, *Streptomyces* sp. FR-008 plasmid pHZ227 (ArsO), and *S. coelicolor* A3(2) aligns from base number 739. This definitely reflects more realistically that the true start position of ORF5 is likely to be in the region of 740-760. A possible mismatch during sequencing and/or error made during editing, probably between nucleotides 540-570 as this is where the break in blastx alignment occurs and the supposed start position of ORF5 predicted by Genemark, might be a possible explanation to this discrepancy. From the sizes of the proteins which ORF5 shows significant homology it can be assumed that it has an ORF length of about 340-370aa. Interestingly, ORF5 shows extensive amino acid similarity to the putative monooxygenase from *P. putida* KT2440 (53%) to which ArsO from *Streptomyces* sp. FR-008 plasmid pHZ227 also shows 48% identity with.

Another FMO-like protein from *Aeromonas salmonicida* A449 with which ORF5 shows 35% sequence identity with has also being proposed to be involved in arsenic resistance. The possible function and whether ORF5 are of utmost necessity to conferring enhanced arsenic tolerance is currently unknown and up for debate. It has been suggested that ArsO-like might serve a similar role as the ArsH genes discovered in several Gram(-) bacteria which shows slight homology to pyridine nucleotide-dependent oxidoreductases. Alternatively, ORF5 and ArsO could participate in the energy coupling cascade in returning ArsC to its reduced state. Thirdly, as it shows strong similarity to flavoproteins involved in potassium transport, it might act in delivering the crucial K^+ ion to its binding site in Trx-type arsenate reductases in which it reduces the pKa of the catalytic cysteine, stabilizes and increases the specific activity of these ArsC enzymes. However, further experimental work needs to be done in order to establish the exact role of these FAD-dependent proteins within an As^R framework.

Organism	Protein identity (%)						
	ORF1	**ORF2**	**ORF3**	**ORF4**	**ORF5**	**ORF6**	**ORF7**
R. erythropolis plasmid pBD2	86(308-359)	75(13-212)[1] 57(5-136)[2] 37(4-134)[3]	76(2-131)[2] 59(80-211)[1] 40(2-133)[3]	42(2-74)[3]		77(1-271)	76(100-133)[2] 45(180-216)[1]
*R. erythropolis*PR4 plasmid pREL1	95(307-353)	47(23-217)[1] 58(3-135)[2] 37(4-134)[3]	75(2-131)[2] 57(87-223)[1] 40(2-133)[3]	42(2-74)[3] 32(91-163)[1]	65(71-253) 34(70-198)	76(1-271)	73(100-133)[2] 55(184-217)[1]
N. farcinica IFM 10152 plasmid pNF2	85(306-361)	46(20-214)[2] 57(5-135)[3] 41(5-135)[1]	73(2-134)[3] 57(88-218)[2] 43(2-140)[1]	43(1-76)[1] 35(87-160)[2] 34(1-77)[3]			66(100-136)[3] 50(181-216)[2] 50100-131[1]
C. efficiens YS-314	53(305-361)	57(12-203)[3] 56(9-135)[2] 38(5-137)[1] 35(18-155)[u]	65(5-212)[2] 50(76-207)[3]	42(1-75)[1] 42(22-88)[u] 40(7-81)[2] 35(78-153)[3]	51(70-254)	63(6-273)	50(102-135)[2]
C. glutamicum ATCC 13032	80(307-348) 73(327-368) 80(292-333)	53(10-204)[1′] 37(5-138)[1] 37(4-128)[2]	50(77-204)[1′]	30(79-154)[1′] 41(4-73)[1] 38(1-75)[2]	49(73-253) 39(73-256)	63(12-277)	
Arthrobacter sp. FB24	84(312-356)	57(9-136)[1] 45(16-146)[2]	70(5-138)[1] 42(11-145)[2]	42(15-85)[2]	47(84-210)	56(1-273) 54(6-277)	71(102-133)[1]
A. aurescens TC1 plasmid TC2		59(17-215)[2] 55(9-135)[1] 41(6-135)[3]	70(5-138)[1] 55(81-212)[2] 40(1-135)[3]	38(3-77)[1] 33(6-81)[2]		58(1-273)	67(102-135)[3] 51(179-217)[2]
Streptomyces sp. FR-008 plasmid pHZ227	78(316-361)	54(-135)	71(2-133)	38(1-70)	58(75-252)	52(2-271)	68(100-134)
S. coelicolor A3(2)	82(311-357)	55(3-134)[1] 55(3-131)[2]	73(2-136)[1] 71(4-136)[2]	36(1-77)[1] 39(5-70)[2]	59(66-249) 57(66-214) 47(80-226)	55(2-270) 53(1-274)	63(100-135)[1] 58(100-135)[2]
M. tuberculosis F11	84(308-359)	58(363-493)	65(362-492)	41(365-437)		62(14-304)	58(460-493)
M. vanbaalenii PYR-1	86(306-358)	48(22-216)[2] 45(25-219)[3]	64(92-223)[3] 62(89-216)[2]	60(4-78)[1] 33(90-162)[2]		67(1-271)	58(186-219)[3] 52(183-216)[2]

Table 4.1: Summary of protein alignments of the predicted ORFs with similar gene products from selected nocardioform bacteria and related species

4.6 Perspective on the failure of sequencing of pKL319 and pKL125

Unfortunately, the rest of the two As^R plasmid constructs pKL1 and pKL335 could not be sequenced even though several experimental strategies were attempted and ultimately failed. The use of sequencing kits that are specially optimized for high GC-rich DNA, which sometimes can cause problems relating to the quality of sequencing, did not make any difference. Also, not even the addition of chemical agents such as BSA and DMSO that normally reduces or prevents the formation of secondary structures improved the sequencing. The cloning of the ∼1.9kbp *Pst*I-*Sfu*I fragment of pKL335 and the ∼2.5 *BamH*I fragment of pKL1 into another plasmid using cloning kits to see whether this would change the outcome and quality of sequencing came to no avail. Amplifying these two arsenic resistant cloned inserts from the original vector pDA71 and sequencing the fragments like a PCR product had no affect.

An interesting feature of several arsenic resistant gene clusters are that they contain inverted repeat (IR) sequences at the start and end of their operons, possibly to indicate the initiation and termination sites of these polycistronic units. Furthermore, these IR sequences upstream of the -35 and -10 promoter regions of *ars*R acts as a putative operator site where this repressor protein can possibly bind. Moreover, additional dyad sequences have also been encountered in numerous cases, notably in *C. glutamicum* ATCC 13032, within the promoter regions of *ars*B genes which overlap the putative binding site of ArsR. Therefore, it has been suggested that these *ars*B mRNA hairpin loops acts as potential regulatory sites in order to prevent overexpression of this arsenite efflux pump.

Consequently, this IR phenomenon within *ars* operons can invariably account for these putative secondary structures that have been proposed to be the main cause in the poor quality and unsuccessful sequencing of pKL319 and the rest of pKL1. Considering that these direct repeats have been detected in various As^R operons it is strange that not similar problems have been encountered by other researchers. The mostly likely reason to this is probably because they used different, more advanced and expensive sequencing techniques and equipment. The conventional sequencing via primer walking that is done by Inqaba Biotech company is very inaccurate especially the longer the DNA segment becomes. Several sequencing reactions of the same section of insert needs to be performed in order to reduce the inherent error of this typical method. Due to financial constrains, pKL1 and pKL335 could not be sequenced by the high throughput technique based on GS technology that most likely would have overcome the secondary structure problem and any GC-rich effect.

Chapter 5

Conclusion

In future, it might be prudent to utilize a cloning vector with a lower copy number when a certain gene product is suspected to be toxic to the cell at high concentrations. Brief sonication of ligations just before transformation together with the addition of BSA to the ligation mixture may also be advisable in order to minimize 2° structure formation during cloning.
The false positives attained during the screening of the *Bgl*II *G. rubropertincta* library mostly likely contained a mutation, possibly as result of some residual ethidium bromide, a mutagen, which afforded these particular recombinant clones some growth advantage on the initial screening plates. The fact that several of the analyzed clones weren't linearized with *Bgl*II but showed different *BamH*I and *EcoR*I digestion patterns than the control support the notion that some sort of DNA alteration must have occurred.
From the As^R assay of the putative arsenate resistant clones from the partial *Pst*I library it is clearly evident that only pKL1 basically confer the full arsenic resistant phenotype of the original bacterium *G. rubropertincta*. The ORF prediction of pKL316 and pKL120 totally corroborated the As^R profiles obtained for these subclones. Clearly, from the partial sequences of pKL316 and pKL120 it can be perceived that *Gordonia rubropertincta* ATCC 25593 definitely contains arsenic resistant determinants encompassed within one or two operon structures. With an MIC of >250mM and ∼7.5mM for arsenate and arsenite, respectively, makes *G. rubropertincta* ATCC 25593 one of the most arsenic resistant microorganism characterized so far. It is becoming ever more obvious that arsenic resistance genes are ubiquitous amongst bacterial species and that they evolved early in life. More and more unique *ars* operon structures are being discovered with the actinomycetes' operons especially exhibiting novel features. Many of the As^R gene cluster of nocardioforms and related species contain multiple, distinct arsenate reductases of different sizes and properties.

The CX_5R sequence motif however is highly conserved with all the Trx-dependent ArsC homologs and lmwPTPases paralogs. Putative arsenate reductase or low molecular protein tyrosine phosphatases of 210-225aa are found exclusively in actinomycetes species. Seemingly, there appears to exist three arbitrary sets of Trx-type ArsCs, where the arsenate reductase from *S. aureus* plasmid pI258 belongs to group2. The presence of a thioredoxin reductase (ArsT/TrxB) and a FAD-binding monooxygenase homolog (ArsO/FMO-like) appears to be another unusual phenomenon specific to the *ars* operons of some high GC-rich Gram-positive bacteria.

Even though the alignment analysis provided a definite indication to the designation of the proposed cistrons, there was however anomalies with the starting position of some ORFs between the nucleotide sequence based search versus the predicted protein sequence analysis. Ambiguous mismatches and background signals from sequencing plausibly led to a few errors in the edited contiguous sequence to creep in, resulting in these discrepancies in the predicted start position of particular cistrons. Just a single in frame mutation that results in a premature termination signal or adjustment of initiation codon could potentially lead to an erroneous ORF prediction and ultimately giving either a truncated or elongated protein. Subsequently, this incorrect ORF sequence will then be analyzed during the protein based search. In contrast, even with this early terminal signal or altered initiation codon, the nucleotide sequence based search would still give an alignment which would include supposedly non-translated areas. If areas outside the predicted protein sequence align to the same type of protein as the ORF itself it could hint that there are maybe some mistake somewhere. As a result, I will now provide my own ORF prediction of the two characterized arsenic resistant plasmid constructs based on all the information gathered in this report.

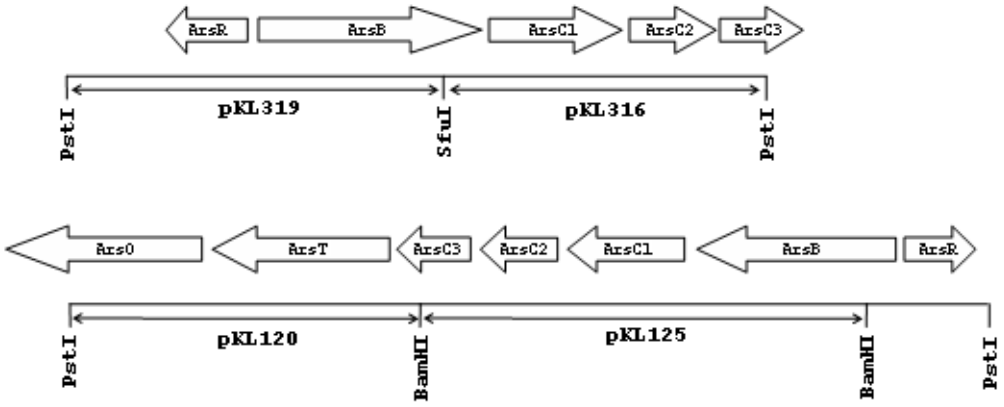


Figure 5.1: Arsenic gene organization prediction within pKL335 and pKL1 plasmid constructs.

Appendix A

Solutions

Miscellaneous stock solutions

Solution	Typical volume (ml)	Amount (mg/ml sdH_2O)
1M NaCl	10	58
0.25M KCl	10	18.6
1M $MgCl_2$	100	202
1M $CaCl_2$	100	219
0.5% KH_2PO_4	50	5
1M $C_6H_{12}O_6$	20	180
10M NaOH	50	400
3M CH_3COONa	50	246
1% EtBr	10	10
2M $Na_2AsO_4.7H_2O$	20	625
1M $NaAsO_2$	20	130
5% C_6H_5COONa	10	50
5% $C_9H_{10}O_5$	10	50

Table A.1: Table of stock solutions and their concentrations.

Antibiotic stock solutions

Antibiotic	Concentration (mg/ml)	Solvent(s)
Ampicillin	200	7:3 Ethanol/sdH_2O
Chloramphenicol	4	Ethanol/Methanol
Kanamycin	100	sdH_2O
Nalidixic acid	10	7:3 Ethanol/sdH_2O
Nystatin	10	Methanol
Rifampicin	10	Ethanol
Streptomycin	100	sdH_2O

Table A.2: Table of antibiotic stock solutions and their concentrations.

General growth media

Luria Bertani broth (LB)
1.0% Tryptone
0.5% Yeast extract
0.5% NaCl

LB agar
1.0% Tryptone
0.5% Yeast extract
0.5% NaCl
1.5% Technical agar

LBSG
1.0% Tryptone
0.5% Yeast extract
0.5% NaCl
10.3% Sucrose
1-3% Glycine

BHI
3.7% BHI

BHI agar
3.7% BHI
1.5% Technical agar

BHIG
3.7% BHI
1-3% Glycine

Minimal Media

10X Stock III solution
7.0% K_2HPO_4
2.7% KH_2PO_4
0.1% $MgCl_2.6H_2O$

Solution A
3.0% Technical agar

Solution B
20% v/v 10X Stock III
0.2% NH_4Cl
Combine solution A and B after autoclaving

***E. coli* plasmid extraction solutions**

Solution I
10mM EDTA pH 8.0
25mM Tris.HCl pH 8.0
50mM Glucose

Solution II
0.2M NaOH
1.0% SDS

Solution III
29.4% w/v KAc
11.5% v/v glacial acetic acid
pH 5.0

Nocardioform plasmid extraction solutions

TE buffer
10mM EDTA pH 8.0
10mM Tris.HCl pH 8.0

TE-SDS
TE
10% SDS

5M KAc
49.1% KAc
pH to 6.0 with glacial acetic acid

DNA purification

TE-saturated phenol
10ml TE
10g Phenol

Agarose gel electrophoresis buffers and solutions

5X TBE buffer
5.4% Tris base
2.75% Boric acid
2% v/v 0.5M EDTA pH 8.0

Running Buffer
10% v/v 5X TBE
0.1% v/v 10mg/ml EtBr

Agarose Gels
0.4 - 1.2% Agarose
10% 5X TBE
0.1% v/v 10mg/ml EtBr

Tracking dye
30% v/v Glycerol
0.025% Bromophenol blue

Two-phase tracking dye
50% v/v Glycerol
0.25mM EDTA pH 8.0
0.25% Bromophenol blue
0.25% Xylene cyanol

λII/III molecular weight marker
5μl λII/III molecular weight marker DNA
10μl 10X Restriction buffer
85μl sdH_2O
15μl Tracking dye

Transformations

***E. coli* $CaCl_2$-mediated transformation**

Transformation buffer
100mM $CaCl_2.6H_2O$
10mM Tris.HCl pH 8.0

High Efficiency *E. coli* transformation

TB Buffer
10mM HEPES
15mM $CaCl_2.6H_2O$
250mM KCl
adjust pH to 6.7 with 10M KOH and autoclave
55mM $MnCl_2$

SOB
2.0% Tryptone
0.5% Yeast extract
10mM NaCl
2.5mM KCl
10mM $MgCl_2.6H_2O$
10mM $MgSO_4$

SOC
SOB
20mM Glucose

Nocardioform PEG-mediated transformation

0.25M TES
5.73% TES
pH to 7.2 with 10M NaOH

Basal Buffer
10.3% Sucrose
0.2% $MgCl_2.6H_2O$
0.025% K_2SO_4
10% v/v 0.25M TES pH 7.2

Protoplast buffer
B buffer
1.0% v/v 5mg/ml KH_2PO_4
2.5% v/v 1M $CaCl_2$

P-PEG
P buffer
50% w/v PEG

Regeneration medium agar
1.0% Tryptone
0.5% Yeast extract
0.3% NaCl
10.3% Sucrose
0.33% Glucose
0.33% $MgCl_2.6H_2O$
Dissolve solutes in 0.86 vol. of dH_2O in the microwave for 1min
Add 1.83% Technical agar and autoclave
Allow medium to cool to about 60°C and then add the following
3.3% v/v 0.25M TES pH 7.2
2% v/v 1M $CaCl_2$
1% v/v 5mg/ml KH_2PO_4
50μg/ml Rifampicin
50μg/ml Nystatin

Appendix B

Molecular weight markers

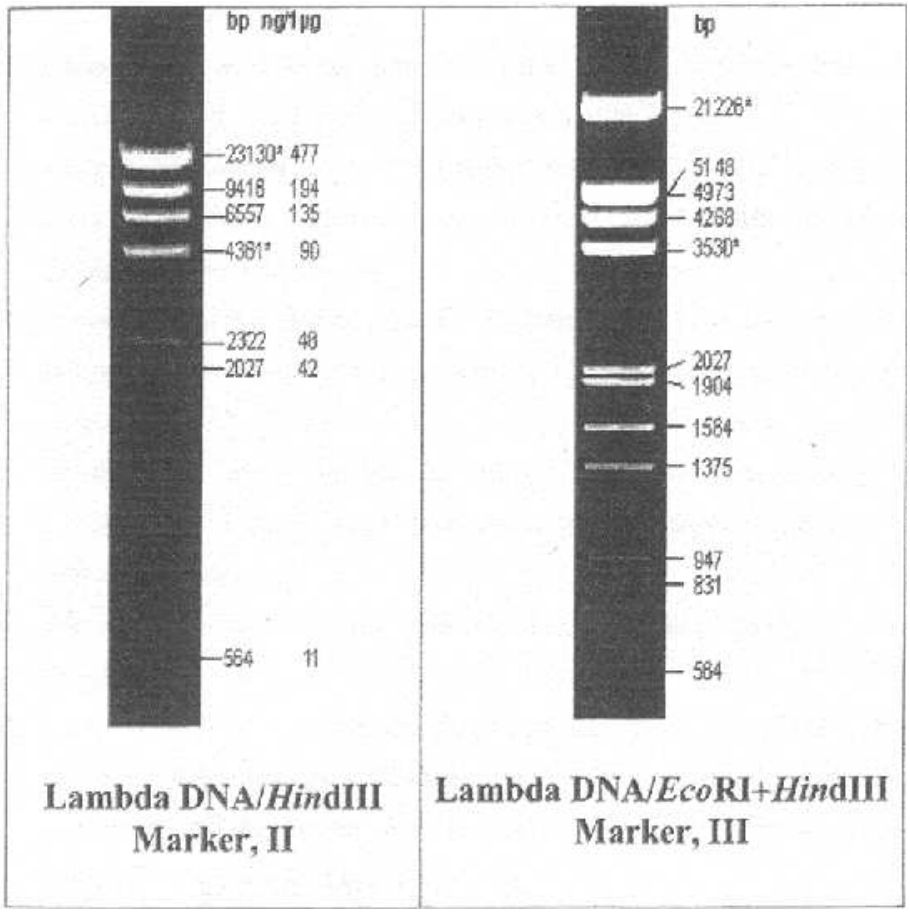


Figure B.1: Molecular weights of λII/III DNA markers

.

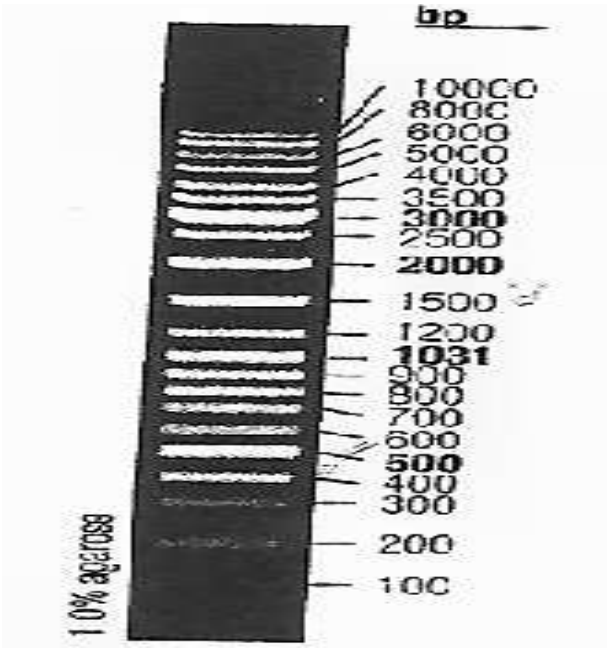


Figure B.2: Molecular weights of Generuler DNA Ladder mix marker

.

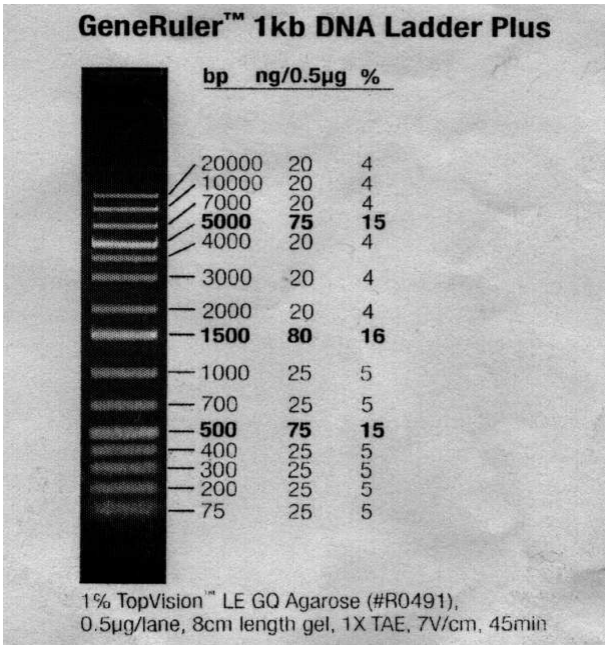


Figure B.3: Molecular weights of Generuler 1kbp DNA Ladder Plus marker

Appendix C

Restriction maps

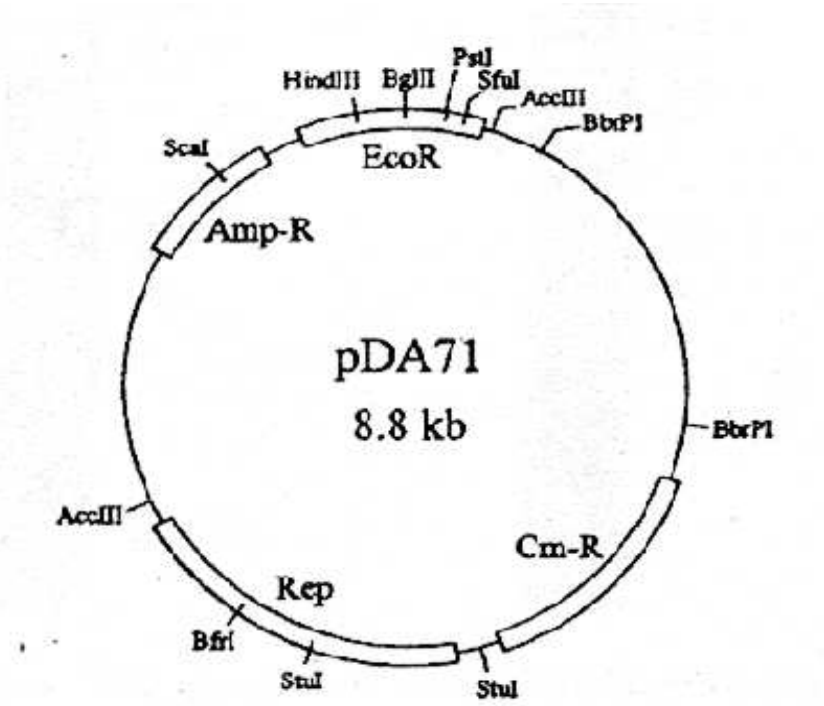


Figure C.1: Partial restriction map of *E. coli-Rhodococcus* shuttle vector pDA71

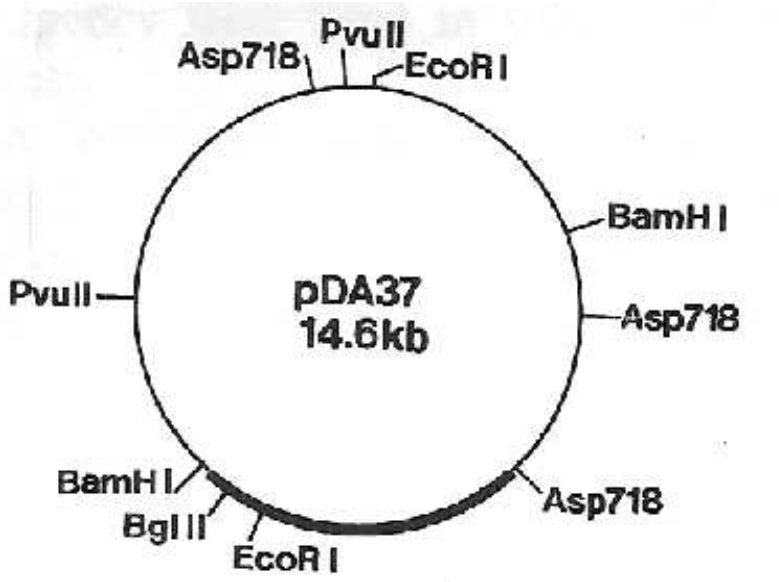


Figure C.2: Partial restriction map of *E. coli-Rhodococcus* shuttle vector pDA37

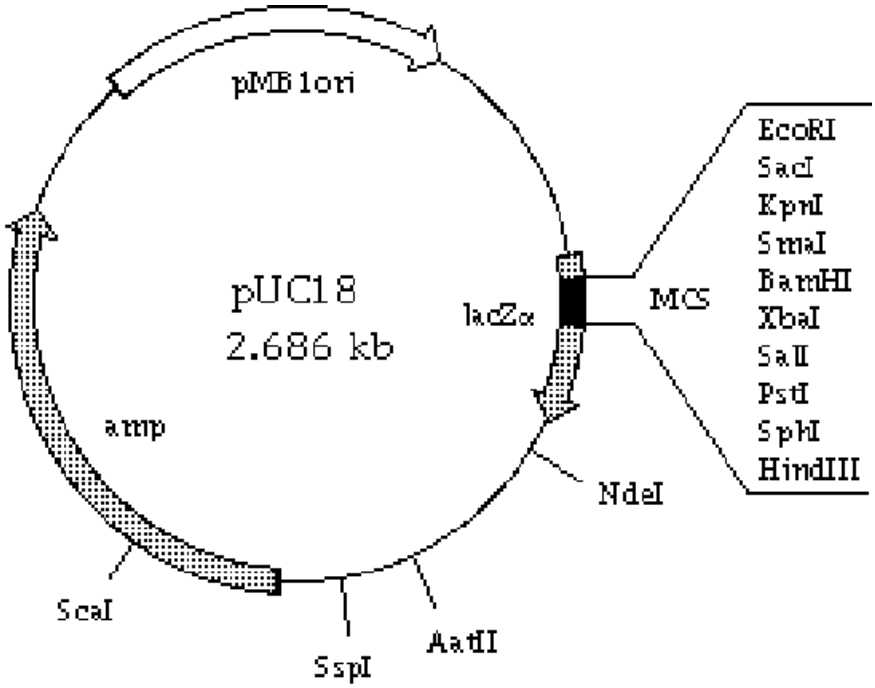


Figure C.3: Partial restriction map of cloning vector pUC18

Appendix D

Primers

M13/pUC	F	5′ GTAAAACGACGGCCAGT	3′
M13/pUC	R	5′ CAGGAAACAGCTATGAC	3′
316int.	F	5′ CAGCACGCTAGGTCGAGA	3′
316int.	R	5′ CACTCGATCAACCCGACCGT	3′
120int.	F	5′ AGGTCGTAGCGCTCCTCGTA	3′
pDAPstI	F1	5′ GGTTGTTAATCTTGAGTATAATTC	3′
pDAPstI	F2	5′ CCTTACGTCCTTTTCTTAGAGGG	3′
pDAPstI	R1	5′ GATTTACAAATTTGTTAATACATAG	3′
pDAPstI	R2	5′ CAACACTCTGAGCGAAGTCGTTGG	3′
3.5int1	F	5′ GCTCGCGGTGTTCACCGCGA	3′
3.5int2	F	5′ CACGGTCGAATCCTCCGATG	3′
3.5int3	F	5′ CCAGACCTCGATCGAGGTCTG	3′
3.5int4	F	5′ GATGACCATCAGCACGTCCGCAC	3′
5.1int1	F	5′ GGCAGGTCGTAGCGCTCCTCGTAGG	3′

Figure D.1: Compilation of all the sequencing primers used

References

[1] http://www.inchem.org

[2] http://www.dchtrust.org

[3] http://www.ats.cdc.gov/toxprofiles/tp2-c5.pdf

[4] http://en.wikipedia.org/wiki/Arsenic

[5] http://www.du.edu/∼jcalvert/phys/arsenic.html

[6] http://www.scescape.net/∼woods/elements/arsenic.html

[7] http://risk.lsd.ornl.gov/tox/profiles/arsenic.shtml

[8] http://scifun.chem.wisc.edu/chemweek/arsenic/arsenic.html

[9] http://nautilus.fic.pt/st2.5/scenes-e/elem/e03310.html

[10] Abernathy CO, Liu Y-P, Longfellow D, Aposhian HV, Beck B, Fowler B, Goyer R, Menzer R, Rossman TG, Thompson C, and Waalkes M. Arsenic: Health effects, mechanisms of actions, and research issues. Environ. Health Perspect. 1999, **107**:593-597.

[11] Afkar E, Lisiak J, Saltikov C, Basu P, Oremland RS, and Stoltz JF. The respiratory arsenate reductase from *Bacillus selenitireductens* str. MLS-10. FEMS Microbiol. Lett. 2003, **226**:107-112.

[12] Anderson GL, Williams J, and Hille R. The purification and characterization of arsenite oxidase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase. J. Biol. Chem. 1992, **267**:23674-23682.

[13] Aposhian HV. 1989. Biochemical toxicology of arsenic, pp265-269. In: Reviews in biochemical toxicology, Hodgson E, Bend JR, Philpot RM (Eds), Elsevier, NY, USA.

[14] Arai T, Kuroda S, and Watanabe I. Biodegradation of acrylamide monomer by a *Rhodococcus* strain. In "Actinomycetes 261 Bakt suppl. II" (KP Schaal and G Pulverer, Eds), pp. 297-308. Gustav Fischer, Stuttgart.

[15] Arenskötter M, Bröker D, and Steinbüchel A. Biology of the metabolically diverse genus *Gordonia*. Appl. Environ. Microbiol. 2004, **70**:3195-3204.

[16] Arenskötter M, Baumeister R, Berekaa MM, Pötter G, Kropenstedt RM, Linas A, and Steinbüchel A. Taxonomic characterization of two rubber degrading bacteria belonging to the species *Gordonia polyisoprenivorans* and analysis of hypervariable regions of 16S rDNA sequences. FEMS Microbiol Lett. 2001, **205**:277-282.

[17] Atlas RM. Principles of microbiology, 2ed, 1997, Wm. C. Brown publishers, London, pp.1035-1041.

[18] Bekhi RM, Topp, EE, and Blackwell BA. Ring hydroxylation of N-methylcarbamate insecticides by *Rhodococcus* TE.1. J. Agricult. Food Chem. 1994, **42**:1375-1378.

[19] Bell KS, Philp JC, Aw DWJ, and Christofi N. The genus *Rhodococcus*. J. Appl. Microbiol. 1998, **85**:195-210.

[20] Bentley R and Chasteen TG. Microbial methylation of metalloids: arsenic, antimony and bismuth. Microbiol. Mol. Biol. Rev. Jun. 2002, **66**(2):250-271.

[21] Berks B, Sargent CF, and Palmer T. The Tat protein export pathway. Mol. Microbiol. 2000, **35**:260-274.

[22] Bhattacharjee H, Li J, Ksenzenko MY, and Rosen BP. Role of cysteinyl residues in metalloactivation of the oxyanion-translocating ArsA ATPase. J. Biol. Chem. 1995, **270**:11245-11250.

[23] Bhattacharjee H, Ho YS, and Rosen BP. Genome organization and chromosomal localization of the *asna*-1, a mouse homolog of the bacterial arsenic translocating ATPase gene. Gene 2001, **272**:291-299.

[24] Bobrowicz P, Wysocki R, Owsianik G, Gouffeau A, and Ulaszewski S. Isolation of three contiguous genes, *acr*1, *acr*2 and *acr*3, involved in resistance to arsenic compounds in the yeast *Saccharomyces cerevisiae*. Yeast 1997, **13**:819-828.

[25] Boyle AW, Silvin CJ, Hassett JP, Nakas JP and Tanenbaum SW. Bacterial PCB degradation. Biodegradation 1992, **3**:285-298.

[26] Bruhn DF, Li J, Silver S, Roberto F, and Rosen BP. Arsenical resistance operon of IncN plasmid 46. FEMS Microbiol. Lett. 1996, **139**:149-153.

[27] Bun-ya M, Shikata K, Nakade S, Yompakdee C, Harashima S and Oshima Y. Two new genes, *pho*86 and *pho*87, involved in inorganic phosphate uptake in *Saccharomyces cerevisiae*. Curr. Genet. 1996, **29**:344-351.

[28] Butcher BG and Rawlings DE. The divergent chromosomal *ars* operon of *Acidothiobacillus ferrooxidans* is regulated by an atypical ArsR protein. Microbiology 2002, **148**:3983-3992.

[29] Butcher BG, Deane SM, and Rawlings DE. The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. Appl. Environ. Microbiol. 2000, **66**:1826-1833.

[30] Cai J, Salmon K, DuBow MS, A chromosomal *ars* operon homologue of *Pseudomonas aeruginosa* confers increased resistance to arsenic and antimony in *Escherichia coli*. Microbiology 1998, **144**:2705-2714.

[31] Carlin A, Shi W, Dey S, and Rosen BP. The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. J. Bacteriol. 1995, **177**:981-986.

[32] Cervantes C, Ji G, Ramirez JL, and Silver S. Resistance to arsenic compounds in microorganisms. FEMS Microbiol. Rev. 1994, **15**:355-367.

[33] Cervantes C and Chavez J. Plasmid-determined resistance to arsenic and antimony in *Pseudomonas aeruginosa*. Antonie van Leeuwenhoek 1992, **62**:333-337

[34] Chacko CI, Lockwood JL, and Zabik M. Chlorinated hydrocarbon pesticides degradation by microbes. Science 1966, **154**:1005-1008.

[35] Chen CM, Misra TK, Silver S, Rosen BP. The nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenic resistance operon. J. Biol. Chem. 1986, **261**:15030-15038.

[36] Chen Y and Rosen BP. Metalloregulatory properties of the ArsD repressor. J. Biol. Chem. 1997, **272**:14257-14262.

[37] Cheng SC and Modrich P. Positive selection cloning vehicle for overproduction of hybrid proteins. J. Bacteriol. 1983, **154**:105-108.

[38] Cobbett CS. Phytochelatins and their roles in heavy metal detoxification. Plant. Physiol. 2000, **123**:825-832.

[39] Croal LR, Gralnick JA, Malasarn D, Newman DK. The genetics of geochemistry. Annu. Rev. Genet. 2004, **38**:175-202.

[40] Dabbs ER, Gowan B, and Andersen SJ. Nocardioform arsenic resistance plasmids and construction of *Rhodococcus* cloning vectors. Plasmid 1990, **23**,242-247.

[41] Dabbs ER and Sole GJ. Plasmid-borne resistance to arsenate, arsenite, cadmium, and chloramphenicol in a *Rhodococcus* species. Mol. Gen. Genet. 1988, **211**:148-154.

[42] Dabbs ER, Gowan B, Quan S, and Andersen SJ. Development of improved *Rhodococcus* plasmid vectors and their use in cloning genes of potential commercial and medical importance. Biotechnologia 1995b, **7-8**,129-135.

[43] de Groot P, Deane SM, and Rawlings DG. A transposon-located arsenic resistance mechanism from a strain of *Acidothiobacillus caldus* from commercial arsenopyrite biooxidation tanks. Hydrometallurgy 2003, **71**:115-123.

[44] Denome SA, Oldfield C, Nash LJ, and Young KD. Characterization of the desulfurization genes from *Rhodococcus* sp. strain IGTS8. J. Bacteriol. 1994, **176**:6707-6716.

[45] Deng T and Liao M. Gold recovery enhancement from a refractory flotation ore by sequential bioleaching thiourea leach. Hydrometallurgy 2002,**63**:249-255.

[46] Dey S and Rosen BP. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. J. Bacteriol. 1995, **177**:385-389.

[47] Dey S, Dou D, Tisa LS, and Rosen BP. Interaction of the catalytic and the membrane subunits of an oxyanion-translocating ATPase. Arch. Biochem. Biophys. 1994, **311**:418-424.

[48] Dhankher CP, Li Y, Rosen BP, Shi J, Salt D, Sinecoff JF, Sashti NA, and Meagher RB. Engineering tolerance and hyperaccumulating of arsenic plants by combining arsenate reductase and γ-glutamylcysteine synthetase expression. Nature 2002, **20**:1140-1145.

[49] Diorio C, Cai J, Marmor J, Shinder R, and DuBow MS. An *Escherichia coli* chromosomal *ars* operon homolg is functional in arsenic detoxification and is conserved in gram-negative bacteria. J. Bacteriol. 1995, **177**:2050-2056.

[50] Eggeling L and Sahm H. Degradation of coniferyl alcohol and other lignin-related aromatic compounds by *Nocardia* sp. DSM 1069. Arch. Microbiol. 1980, **126**:141-148.

[51] Ellis PJ, Conrads T, Hille R, and Kuhn P. Crystal structure of the 100kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64Å and 2.03Å. Structure 2001, **9**:125-132.

[52] Ferreira NP, Robson PM, Bull JR, and van der Walt WH. The microbial production of a 3aα-H-4α(3-propionic acid)-5α-hydroxy-7$\alpha\beta$-methylhexahydroindan-1-one-σ-lactone from cholesterol. Biotechnol. Lett. 1984, **6**:517-522.

[53] Frankenberger WT Jr. (Ed). 2001. Environmental chemistry of arsenic. Marcel Dekker, NY.

[54] Gatti D, Mitra B and Rosen BP. *Escherichia coli* soft metal ion-translocating ATPase. J. Biol. Chem. 2000. **275**:34009-34012.

[55] Gebel T. Confounding variables in the environmental toxicology of arsenic. Toxicology 2000, **144**:155-162.

[56] Ghosh M, Shen J and Rosen BP. Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA. 1999, **96**:5001-6.

[57] Gihring TM and Banfield JF. Arsenite oxidation and arsenate respiration by a new *Thermus* isolate. FEMS Microbiol. Lett. 2001, **204**:335-340.

[58] Gihring TM, Bond PL, Peters SC, and Banfield JF. Arsenic resistance in the archaeon "*Ferroplasma acidarmanus*": a new insight into the structure and evolution of the *ars* genes. Extremophiles 2003, **7**:123-130.

[59] Gilbert SC, Morton J, Buchanan S, Oldfield C, and McRoberts A. Isolation of a "unique" benzothiophene-desulphurization bacterium, Gordona sp. 213E, and characterization of the desulphurization pathway. Microbiology 1998, **144**:2545-2553.

[60] Gladysheva TB, Oden KL, and Rosen BP. Properties of the arsenate reductase of plasmid R773. Biochemistry 1994, **33**:7288-7293.

[61] Haider K, Jugnow G, Kohnen R and Lim SV. Degradation of chlorinated benzenes, phenols, and cyclohexane derivatives by soil bacteria under aerobic conditions. In "Decomposition of toxic and non-toxic organic compounds in soil", pp.207-226. Ann Arbor, MI.

[62] Häggblom MM, Janke D, and Salkinoja-Salonen MS. Transformation of chlorinated phenolic compounds in the genus *Rhodococcus*. Microbial Ecol. 1989, **18**:147-159.

[63] Healy SM, Casrez EA, Ayalo-Fierro F, and Aposhian HV. Enzymatic methylation of arsenic compounds V. arsenite methyltransferase activity in tissues of mice. Appl. Pharmacol. 1998, **148**:65-70.

[64] Heiss GS, Gowan B, and Dabbs ER. Cloning of DNA from a *Rhodococcus* strain conferring the ability to decolorize sulfonated azo dyes. FEMS Microbiol. Lett. 1992, **99**:221-226.

[65] Holt JG, Krieg NR, Sneath PHA, Stanley JT, and Williams ST (Eds). Bergey's manual of determinitive bacteriology. 2000, 9ed, Lippicott Williams & Wilkens, Philadelphia, USA. pp. 573,583,619,625,626,634 and 636.

[66] Hughes MF. Arsenic toxicity and potential mechanisms of action. Toxicol. Lett. 2002, **133**:1-16.

[67] Inoue H, Nojima H, and Okayama H. High efficiency transformation of *Escherichia coli* with plasmids. Gene 1990, 96:23-28.

[68] Islam MR, Rodova M, and Calvet JP. A fast and efficient method of DNA fragment isolation from agarose gels without using commercial kits. Am. Biotechnol. Lab. 2002, **20**:18

[69] Izumi Y, Ohshiro T, Ogino H, Hine Y, and Shimao M. Selective desulfurization of dibenzothiophene by *Rhodococcus erythropolis* D1. Appl. Environ. Microbiol. 1994, **60**:223-226.

[70] Ji G and Silver S. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. Proc. Natl. Acad. Sci. USA. 1992, **89**:9474-9478.

[71] Ji G and Silver S. Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258. J. Bacteriol. 1992, **174**:3684-3694.

[72] Ji G, Garber EAE, Armes LG, Chen C-M, Fuchs JA, and Silver S. Arsenate reductase of *Staphylococcus aureus* plasmid pI258. Biochemistry 1994, **33**:7294-7299.

[73] Jones D and Collins MD. 1986. Irregular non-sporing Gram(+) rods, pp. 1288-1298. In Bergey's manual of systematic bacteriology Vol III, PHA Sneath (Ed), Williams & Wilkens, Baltimore, USA.

[74] Katsoyiannis IA and Zouboulis AI. Application of biological processes for the removal of arsenic from groundwaters. Water Res. 2004, **38**:17-26.

[75] Kaur P and Rosen BP. Plasmid-encoded resistance to arsenic and antimony. Plasmid 1992, **27**:29-40.

[76] Kägi JH. Overview of metallothionein. Methods. Enzymol. 1991, **205**:613-626.

[77] Krafft T and Macy JM. Purification and characterization of the respiratoty arsenate reductase of *Chrysiogenes arsenatis*. Eur. J. Biochem. 1998, **225**:647-653.

[78] Kulp TR, Hoeft SE, and Oremland RS. Redox transformations of arsenic in periphyton communities. Appl. Environ. Microbiol. 2004, **70**:6428-6434.

[79] Langhans D, Lord A, Lampshire D, Burbank A, and Baglin E. Biooxidation of an arsenic-bearing refractory gold ore. Min. Eng. 1995, **8**:147-158.

[80] Lechavalier HA. 1986. Nocardioforms, pp. 1458-1459,1472-81. In Bergey's manual of systematic bacteriology VolIII, PHA Sneath (Ed), Williams & Wilkens, Baltimore, USA.

[81] Le Duc DL and Terry N. Phytoremediation of toxic trace elements in soil and water. J. Ind. Microbiol. Biotechnol. 2005, **32**:514-520.

[82] Li J, Liu S, and Rosen BP. Interaction of ATP binding sites in the ArsA ATPase, the catalytic subunit of the Ars pump. J. Biol. Chem. 1996, **273**:25247-25252.

[83] Li S, Chen Y, and Rosen BP. Role of vicinal cysteine residues in metalloid sensing by the ArsD As(III)-responsive repressor. Mol. Microbiol. 2001, **41**:687-696.

[84] Li S, Rosen BP, Borges-Walmsley MI, and Walmsley AR. Evidence for cooperativity between the four binding sites of dimeric ArsD, an As(III)-responsive transcriptional regulator. J. Biol. Chem. 2002, **272**:25992-26002.

[85] Liu J and Rosen BP. Ligand interactions of the ArsC arsenate reductase. J. Biol. Chem. 1997, **272**(34):21084-21089.

[86] Liu J, Gladysheva TB, Lee L, and Rosen BP. Identification of an essential cysteinyl residue in the ArsC arsenate reductase of plasmid R773. Biochemistry 1995, **34**:13472-13476.

[87] Liu Z, Shen J, Carbrey, JM, Mukhopadhyay R, Agre P, and Rosen BP. Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. Proc. Natl. Acad. Sci. 2002, **99**:9053-6058.

[88] Lopez-Maury L, Florencio FJ, and Reyes JC. Arsenic sensing and resistance system in the cyanobacterium *Synechocystis* sp. strain PCC 6803. J. Bacteriol. 2003, **185**:5363-5371.

[89] McDonald CR, Cooper CG, and Zajic J. Surface active lipids from *Nocardia erythropolis* grown on hydrocarbons. Appl. Environ. Microbiol. 1981, **41**:117-123.

[90] Maciaszczyk E, Wysocki R, Golik P, Lazowska J, and Ulaszewski S. Arsenical resistance genes in *Saccharomyces douglasii* and other yeast species undergo rapid evolution involving genomic rearrangements and duplications. FEMS Yeast Res. 2004, **4**:821-831.

[91] Martin P, DeMel S, Shi J, Gladysheva TB, Gatti DL, Rosen BP and Edwards BFP. Insights into the stucture, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. Structure 2001, **9**:1071-1081.

[92] McNeil MM and Brown JM. The medically important aerobic actinomycetes: epidemiology and microbiology. Clin. Microbiol. Rev. 1994, **7**:357-417.

[93] Meagher RB and Heaton ACP. Strategies for the engineeered phytoremediation of toxic elements pollution: mercury and arsenic. J. Ind. Microbiol. Biotechnol. 2005, **32**:502-513.

[94] Meng Y-L, Liu Z, and Rosen BP. As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*, J. Biol. Chem. 2004, **279**:18334-18341.

[95] Messens J, Hayburn G, Desmyter A, Laus G, and Wyns L. The essential catalytic redox couple in arsenate reductase of *Staphylococcus aureus*. Biochemistry 1999, **38**:16857-16865.

[96] Miller WH Jr. Schipper HM, Lee JS, Singer J, and Waxman S. Mechanisms of action of arsenic trioxide. Cancer Res. 2002, **62**:3893-3903.

[97] Mukhopadyay R, and Rosen BP. Arsenate reductases in prokaryotes and eukaryotes. Environ. Health Perspect. 2002, **110**745-748.

[98] Mukhopadhyay R, and Rosen BP. The phosphatase $C(X)_5R$ motif is required for catalytic activity of the *Saccharomyces cerevisiae* Acr2p arsenate reductase. J. Biol. Chem. 2001, **276**:34738-34742.

[99] Mukhopadhyay R, Shi J, and Rosen BP. Purification and characterization of Acr2p, the *Saccharomyces cerevisiae* arsenate reductase. J. Biol. Chem. 2000, **275**:21145-21157.

[100] Mukhopadhyay R, Rosen BP, Phung LT, and Silver S. Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol Rev. 2002, **26**:311-325.

[101] Muller D, Lievremont D, Simeonova DD, Hubert J-C, and Lett M-C. Arsenite oxidase *aox* genes from a metal-resistant β- Proteobacterium. J. Bacteriol. 2003, **185**:135-141.

[102] Mulligan CN, Yang RN, and Gibbs BF. Remediation technologies for metal-contaminated soils and groundwater: an evaluation. Eng. Geology. 2001, **60**:193-207.

[103] Neyt C, Iriarte M, Thi V-H, and Cornelis GR. Virulence and arsenic resistance in *Yersiniae*. J. Bacteriol. 1997, **179**:612-619.

[104] Nies DH. Microbial heavy metal resistance. Appl. Microbiol. Biotechnol. 1999, **51**:730-750.

[105] Niggemyer A, Spring S, Strackenbrandt E, and Rosenzweig RF. Isolation and characterization of a novel As(V)- reducing bacterium - implications for arsenic mobilization and the genus *Desulfitobacterium*. Appl. Environ. Microbiol. 2001, **67**:5568-5580.

[106] Ordónez E, Letek M, Valbuena N, Gil JA, and Mateos LM. Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. Appl. Environ. Microbiol. 2005, 71:6205-6215.

[107] Oremland RM and Stolz JF. The ecology of arsenic. Science May 2003, **300**:939-944.

[108] Parekh NR, Walker A, Roberts SJ, and Welch SJ. Rapid degradation of the triazinone herbicide metamitron by a *Rhodococcus* sp. isolated from treated soil. J. Appl. Bacteriol. 1994, **77**:467-475.

[109] Pradosh R and Anupama S. Metabolism and toxicity of arsenic: a human carcinogen. Curr. Science. 2002, **82**:38-45.

[110] Prescott LM, Harley JP, and Klein DA. Microbiology, 5ed, 2002, WCB McGraw-Hill, Boston, pp.541-544.

[111] Qin J, Rosen BP, Zhang Y, Wang G, Franke S, and Rensing C. Arsenic detoxification and evolution of trimethylarsine gas by microbial arsenite S-adenosylmethionine methyltransferase. Proc. Nat. Acad. Sci. 2006, **103**:2075-2080.

[112] Quan S and Dabbs ER. Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. Plasmid 1993, **29**:74-79.

[113] Radabaugh TR and Aposhian HV. Enzymatic reduction of arsenic compounds in mammalian systems: Reduction of arsenate to arsenite by human liver arsenate reductase. Chem. Res. Toxicol. 2000, **13**:26-30.

[114] Radabaugh TR, Sampayo-Reyes A, Zakharyan RA, and Aposhian HV. Arsenate reductase II: Purine nucleoside phosphorylase in the presence of dihydrolipoic acid is a route fot reduction of arsenate to arsenite in mammalian systems. Chem. Res. Toxicol. 2002, **15**:692-698

[115] Rawlings DE, Dew D, and du Plessis C. Biomineralization of metal- containing ores and concentrates. Trends Biotechnol. Jan 2003, **21**:38-44.

[116] Rodríguez VM, Jiménez-Capdeville ME, and Giordana M. The effects of arsenic exposure on the nervous system. Toxicol. Lett. 2003, **145**:1-8.

[117] Rosen BP. Families of arsenic transporters. Trends Microbiol. 1999, **7**:207-212.

[118] Rosen BP. Biochemistry of arsenic detoxification. FEBS Lett. 2002, **529**:86-92.

[119] Rosen BP. Transport and detoxification systems for transition metals, heavy metals, and metalloids in eukaryotic and prokaryotic microbes. Comp. Biochem. Physiol. Part A 2002, **133**:689-693.

[120] Rosen BP. Bacterial resistance to heavy metals and metalloids. J. Biol. Inorg. Chem. 1996, 1:273-277.

[121] Rosen BP, Bhattacharjee H, Zhou T, and Walmsley AR. Mechanism of the ArsA ATPase. Biochim. Biophys. Acta 1999, **1461**:207-215.

[122] Rosenstein R, Peschel A, Wieland B, and Götz F. Expression and regulation of the antimonite, arsenite,and arsenate resistance operon of *Staphylococcus xylosus* plasmid pSX267. J. Bacteriol. 1992, **174**:3676-3683.

[123] Ryan D and Colleran E. Arsenical resistance in the IncHI2 plasmids. Plasmid 2002, **47**:234-240.

[124] Saltikov CW and Newman DK. Genetic identification of a respiratory arsenate reductase. Proc. Natl. Acad. Sci. USA. 2003, **100**:10983-10988.

[125] Sambrook J, Freitch EF, and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2ed, Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY, USA.

[126] Sanders OI, Rensing C, Kuroda M, Mitra B, and Rosen BP. Antimonite is accumulated by the glycerol facilitator GlpF in *Escherichia coli*. J. Bacteriol. 1997, **179**:3365-3367.

[127] Sampayo-Reyes A, Zakharyan RA, Healy SM, and Aposhian HV. Monomethyl arsonic acid reductase and monomethylarsonic acid in hamster tissue. Chem. Res. Toxicol. 2000, **13**:1181-1186.

[128] Santini JM and van den Hoven RN. Moybdenum containing arsenate reductase of the chemolithotrophic arsenite oxidizer NT-26. J. Bacteriol. 2004, **184**:1614-1619.

[129] Sato T and Kobayashi Y. The *ars* operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite. J. Bacteriol. 1998, **180**:1655-1661.

[130] Seto M, Kimbara K, Shimura M, Hatta T, Fukuda M, and Yodo K. A novel transformation of polychlorinated biphenyls by *Rhodococcus* sp. strain RHA1. Appl. Environ. Microbiol. 1995, **61**:3353-3358

[131] Shen J, Hsu C-H, Kong B-K, Rosen BP, and Bhattacharjee H. The *Saccharomyces cerevisiae* Arr4p is involved in metal and heat tolerance. Biometals 2003, **16**:369-378.

[132] Shi J, Mukhopadhyay R, and Rosen BP. Identification of a triad of arginine residues in the active site of the ArsC reductase of plasmid R773. FEMS Microbiol. Lett. 2003, **227**:295-301.

[133] Shi W, Dong J, Scott RA, Ksenzenko MY, and Rosen BP. The role of arsenic-thiol interactions in metalloregulation of the *ars* operon. J. Biol. Chem. 1996, **271**(16):9291-9297.

[134] Silver S. Genes for all metals - a bacteial review of the periodic table. The 1996 Thom Award lecture. J. Ind. Microbiol. Biotechnol. 1998, **20**:1-12.

[135] Silver S and Phung LT. Genes and enzymes in bacterial oxidation and reduction of inorganic arsenic. Appl. Environ. Microbiol. 2005, **71**:599-608.

[136] Silver S and Phung LT. Bacterial heavy metal resistance: new surprises. Annu. Rev. Microbiol. 1996, **50**:753-789.

[137] Silver S and Phung LT. A bacterial review of the periodic table: genes and proteins for toxic inorganic ions. J. Ind. Microbiol. Biotechnol. 2005 (Special biodiversity topic), **32**:587-605.

[138] Stolz JF and Oremland RS. Bacterial respiration of arsenic and selenium. FEMS Microbiol. Rev. 1999, **23**,615-627.

[139] Stolz JF, Basu P, and Oremland RS. Microbial transformation of elements: the case of arsenic and selenium. Int. Microbiol. 2002, **5**:201-207.

[140] Stecker C, Johann A, Herzberg C, Averhoff B, and Gottschalk G. Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. J. Bacteriol. 2003, **185**:5269-5274.

[141] Styblo M, Del Razo LM, Vega L, Germolec DR, LeCluyse EL, Hamilton GA, Wang C, Culleb WR, and Thomas DJ. Comparitive toxicity of trivalent and pentavalent inorganic and methylated arsenicals in human cells. Arch. Toxicol. 2000, **74**:289-299.

[142] Styblo M, Del Razo LM, LeCluyse EL, Hamilton GA, Wang C, Cullen WR, and Thomas DJ. Metabolism of arsenic in primary cultures of human and rat hepatocytes. Chem. Res. Toxicol. 1999, **12**:560-565.

[143] Suzuki K, Wakao N, Kimura T, Sakka K, and Ohmiya K. Expression and regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU plasmid pKW301 in *Escherichia coli*. Appl. Environ. Microbiol. 1998, **64**:411-418.

[144] Tamaki S and Frankenberger WT Jr.. Environmental biochemistry of arsenic. Rev. Environ. Contam. Toxicol. 1992, **124**:79-110.

[145] Tamas MJ and Wysocki R. Mechanisms involved in metalloid transport and tolerance acquisition. Curr. Genet. 2001, **40**:2-12.

[146] Thomas DJ, Waters SB, and Styblo M. Elucidating the pathway for arsenic methylation. Toxicol. Appl. Pharmacol. 2004, **198**:319-326

[147] Thomas DJ, Styblo M, and Lin S. The cellular metabolism and systematic toxicity of arsenic. Toxicol. Appl. Pharmacol. 2001, **176**:127-144.

[148] Tisa LS and Rosen BP. Molecular characterization of an anion pump. J. Biol. Chem. 1990, **265**:190-194.

[149] Tu S, Ma LQ, Kayiga AO, and Zilloux EJ. Phytoremediation of arsenic contaminated groundwater by the arsenic hyperaccumulating fern *Pteris vittata* L. Int. J. Phytoremediation 2004, **6**:35-47.

[150] Tuffin M, de Groot P, Deane SM, and Rawlings DE. Multiple sets of arsenic resistance genes are present within highly arsenic resistant industrial strains of the biomining bacterium *Acidothiobacillus caldus*. Int. Congress Series. 2004, **1275**:165-172.

[151] Turner JS, Robinson NJ, and Gupta A. Cyanobacterial metallothioneins : biochemistry and molecular genetics. J. Ind. Microbiol. 1995, **14**:119-125.

[152] Vahter M. Variation in human metabolism of arsenic, pp.267-279. In: Arsenic exposure and health effects, Chappell WR, Abernathy CO, and Calderon RL (Eds), Elsevier, NY, USA.

[153] van den Hoven RN and Santini JM. Arsenite oxidation by the heterotroph Hydrogenophaga NT-14: the arsenite oxidase and physiological electron acceptor. Biochim. Biophys. Acta 2004, **656**:48-55.

[154] Venter JC, Remimgton, Heidelberg JF, Haipern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson, Pfannkock C, Rogers YH, and Smith HO. Environmental genome shotgun sequencing of the Sargasso Sea. Science 2004, **304**:66-74.

[155] Wang G, Kennedy SP, Fasiludeen S, Rensing C, and DasSarma S. Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system. J. Bacteriol. May 2004, **186**:3187-3194.

[156] Wang L, Chen S, Xiao X, Huang X, You D, Zhou X, and Deng Z. *ars*RBOCT arsenic resistance system encoded by linear plasmid pHZ227 in *Streptomyces* sp. strain FR008. Appl. Environ. Microbiol. 2006, **72**:3738-3742.

[157] Waxman S and Andersen KC. History of the development of arsenic derivatives in cancer therapy. The Oncologist 2001, **6**:3-10.

[158] Wildfang E, Zakharyan RA, and Aposhian HV. Characterization of hamster liver arsenite and methylarsonic acid methyltransferase activity in vitro. Toxicol. Appl. Pharmacol. 2005, **152**:366-375

[159] Willsky GR and Malamy MH. Effect of arsenate on inorganic phosphate transport in *Escherichia coli*. J. Bacteriol. 1980, **144**:366-374.

[160] Wu J and Rosen BP. Metalloregulated expression of the *ars* operon. J. Biol. Chem. 1993, **268**:52-58.

[161] Wu J, Tisa LS, and Rosen BP. Membrane topology of the ArsB protein, the membrane subunit of an anion translocating ATPase. J. Biol. Chem. 1992, **267**:12570-12576.

[162] Wu J and Rosen BP. The ArsR protein is a trans-acting regulatory protein. Mol. Microbiol. 1991, **5**:1331-1336.

[163] Wu J and Rosen BP. The *ars*D gene encodes a second trans-acting regulatory protein of the plasmid-encoded arsenical resistance operon. Mol. Microbiol. 1993, **8**:615-623.

[164] Wysocki R, Fortier P-K, Maciaszczyk E, Thorsen M, Leduc A, Odhagen A, Owsianik G, Ulaszewski S, Ramotar D, and Tamas MJ. Transcriptional activation of metalloid tolerance genes in *Saccharomyces cerevisiae* requires the AP-1-like proteins Yap1p and Yap8p. Mol. Biol. of the cell. 2004, **15**:2049-2060.

[165] Wysocki R, Bobrowicz P, and Ulaszewski S. The *Saccharomyces cerevisiae acr*3 gene encodes a putative membrane protein in arsenite transport. J. Biol. Chem. 1997, **272**:30061-30066.

[166] Wysocki R, Chéry CC, Wawrzycka D, van Hulle M, Cornelis R, Thevelein JM, and Tamas MJ. The glycerol channel Fps1p mediates the uptake of arsenite and antimonite in *Saccharomyces cerevisiae*. Mol. Microbiol. 2001, **40**:1391-1401.

[167] Xu C and Rosen BP. Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of *Escherichia coli*. J. Biol. Chem. 1997, **272**:15734-15738.

[168] Yang C-H, Cheng J, Finan TM, Rosen BP, and Bhattacharjee H. Novel pathway for arsenic detoxification in the legume symbiont *Sinorhizobium meliloti*. J. Bacteriol. 2005. **185**:6991-6997.

[169] Zakharyan RA and Aposhian HV. Enzymatic reduction of arsenic compounds in mammalian systems: The rate limiting enzyme of rabbit liver arsenic transformation is MMA^V reductase. Chem. Res. Toxicol. 1999, **12**:1278:1283.

[170] Zakharyan RA, Sampayo-Reyes A, Healy SM, Tsaprailis G, Board PG, Liebler DC, and Aposhian HV. Human monomethylarsonic acid (MMA(V)) reductase is a member of the glutathione-S-transferases superfamily. Chem. Res. Toxicol. 2001, **14**:1051-1057.

[171] Zakharyan RA, Wu Y, Bogdan GM, and Aposhian HV. Enzymatic methylation of arsenic compunds I: Assay, partial purification, and properties of arsenite methyltransferase and monomethylarsonic acid methyltransferase of rabbit liver. Chem. Res. Toxicol. 1995, **8**:1029-1038.

[172] Zegers I, Martins JC, Willem R, Wyns L, and Messens J. Arsenate reductase from *S. aureus* pI258 is a phosphatase drafted for redox duty. Nature Struct. Biol. 2001, **8**:843-847.

[173] Zhou T and Rosen BP. Tryptophan fluorescence reports nucleotide induced conformational changes in α- domain of the ArsA ATPase. J. Biol. Chem 1997, **272**:19731-19737.

[174] Zhou T, Radaev S, Rosen BP, and Gatti DL. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. EMBO J. 2000, **19**:4838-4845.

[175] Zouboulis AI and Katsoyiannis IA. Recent advances in the bioremediation of arsenic contaminated groundwaters. Environ. Int. 2005, **31**:213-219.