Analysis of Arsenic Resistance in the Biomining Bacterium, *Acidithiobacillus caldus*

Andries Albertus Kotzé

Thesis presented in partial fulfillment of the requirements for the degree of Master of Sciences at the University of Stellenbosch

Supervisor: Professor Douglas E. Rawlings

December 2006

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

Signature:_____ Date:_____

(Andries A. Kotzé)

Abstract

In this study the chromosomal arsenic resistance (*ars*) genes shown to be present in all *Acidithiobacillus. caldus* isolates were cloned and sequenced from *At. caldus* #6. Ten open reading frames (ORFs) were identified on a clone conferring arsenic resistance, with three homologs to arsenic genes, *arsC* (arsenate reductase), *arsR* (regulator) and *arsB* (arsenite export). This *ars* operon is divergent, with the *arsRC* and *arsB* genes transcribed in opposite directions. Analysis of the putative amino acid sequences of these *arsRC* and *arsB* genes revealed that they are the most closely related to the *ars* genes of *Acidithiobacillus ferrooxidans*.

These *ars* genes were functional when transformed into an *Escherichia coli ars* deletion mutant ACSH50I^{q}, and conferred increased levels of resistance to arsenate and arsenite. ArsC was required for resistance to arsenate, but not for resistance to arsenite. None of the other ORFs enhanced arsenic resistance in *E. coli*. A transposon located arsenic resistance system (TnAtcArs) has been described for highly arsenic resistant strains of the moderately thermophilic, sulfur-oxidizing, biomining bacterium *At .caldus* #6. In the latter study it was shown that TnAtcArs confers higher levels of resistance to arsenate and arsenite than the chromosomal operon. TnAtcArs was conjugated into a weakly *ars* resistant *At. caldus* strain (C-SH12) and resulted in greatly increased arsenite resistance. RT-PCR analysis revealed that *arsR* and *arsC* are co-transcribed. Despite ORF1 (cadmium inducible-like protein) and ORF5 (putative integrase for prophage CP-933R) not being involved in resistance to arsenic, ORF1 was co-transcribed with *arsRC* and ORF5 with *arsB*. Using *arsR-lacZ* and *arsB-lacZ* fusions it was shown that the chromosomal ArsR-like regulator of *At. caldus* acts as a repressor of the *arsR* and *arsB* promoter expression. Induction of gene expression took place when either arsenate or arsenite was added. The chromosomal located ArsR was also able to repress TnAtcArs, but the transposon-located ArsR was unable to regulate the chromosomal system.

Opsomming

In hierdie studie is die chromosomale arseen weerstandbiedendheidsgene (*ars* gene), teenwoordig in alle *Acidithiobacillus caldus* isolate, gekloon en die DNA volgorde daarvan vanaf *At. caldus* #6 bepaal. Tien oopleesrame (ORFs) is geïdentifiseer op 'n kloon wat arseen weerstandbiedend is, met drie homoloog aan *ars* gene, nl. *arsC* (arsenaat reduktase), *arsR* (reguleerder) en *arsB* (membraan-geleë pomp wat arseniet uitpomp). Die *ars* operon is gerangskik met die *arsRC* en *arsB* gene wat in teenoorgestelde rigtings getranskribeer word. Analise van die afgeleide aminosuurvolgorde van dié *ars* gene het getoon hulle is naverwant aan die *ars* gene van *Acidithiobacillus ferrooxidans*.

Die *ars* gene was funksioneel na transformasie na 'n *E. coli ars* mutant (ACSH50I^{q}), en het 'n hoër vlak van weerstand teen arsenaat en arseniet gebied. ArsC was nodig vir weerstand teen arsenaat, maar nie vir weerstand teen arseniet nie. Geen van die ander ORFs het arseen weerstandbiedendheid in *E. coli* bevorder nie. Voorheen is 'n *ars* operon, geleë op 'n transposon (TnAtcArs), in 'n hoogs arseen-weerstandbiedende stam van die middelmatige termofiliese, swawel-oksiderende, bio-ontgunning ("biomining") bakterie *Acidithiobacillus caldus* #6 beskryf. In laasgenoemde studie is gevind dat TnAtcArs hoër vlakke van weerstand bied teen arsenaat en arseniet as die chromosomale operon. TnAtcArs is na 'n lae arseen-weerstandbiedende *At. caldus* stam (C-SH12) gekonjugeer. Die resultaat was 'n groot verhoging in arseen weerstandbiedendheid. RT-PCR analise het onthul dat *arsR* en *arsC* saam getranskribeer word. Benewens die feit dat ORF1 (kadmium induseerbare protein) en ORF5 (afgeleide integrase vir profaag CP-933R) nie betrokke is in weerstand teen arseniet and arsenaat nie, is ORF1 saam met *arsRC* getranskribeer en ORF5 saam met *arsB*. Deur gebruik te maak van die fusie-gene *arsR-lacZ* en *arsB-lacZ* is bewys dat die chromosomale ArsR reguleerder van *At. caldus* as 'n inhibeerder van die *arsR* en *arsB* promoter uitdrukking funksioneer. Indusering van geen uitdrukking vind plaas wanneer arseniet of arsenaat bygevoeg word. Die chromosomaal-geleë ArsR is ook in staat om TnAtcArs te inhibeer, terwyl die transposon geleë ArsR nie daartoe in staat is om die chromosomale *ars* sisteem te reguleer nie.

Acknowledgements

Most importantly, I would like to thank those who have provided me with support in many ways during this time, including:

My family and most importantly my parents, for giving me the opportunity to attend Stellenbosch University and for supporting me in all possible ways.

My supervisor, Professor Douglas Rawlings, who always had time in his busy schedule to give advice and support when problems appeared.

Dr. Marla Tuffin, who always watched over my shoulder as I learned the ropes in the lab. Thank you for all the friendship, guidance and patience and for always being willing to lend an ear and share ideas. Thank you also for proof-reading my thesis and correcting my many mistakes.

The rest of the Biomining Research Group for their friendship, support and for providing many a fun and amusing moment in the lab.

Finally, our Heavenly Father, for blessing me with the strength and opportunity to complete this study.

This work was supported by grants from the NRF (National Research Foundation, Pretoria, South Africa), BHP Billiton Process Research Laboratories (Randburg, South Africa) and the Biomine Project of the European Framework 6 program.

Table of contents

Abbreviations

aa	amino acids
ADP	adenosine 5’ diphisphate
Amp	ampicillin
At.	*Acidithiobacillus*
ATP	adenosine 5’ triphosphate
AsV	arsenate
AsIII	arsenite
bp	base pair (s)
^{o}C	degrees Celsius
C-terminal	carboxyl-terminus
Cys	cysteine
DNA	deoxyribonucleic acid
g	gram
h	hour
His	histidine
Inc	incompatibility group (s)
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pair (s) or 1000-bp
kDa	kilodalton (s) or 1000 daltons
Km	kanamycin
l	liter
LA	Luria agar
LB	Luria Bertani
LMW	low molecular weight
mg	milligram
min	minute (s)
ml	milliliter
µg	microgram
µl	microliter
NCBI	National Centre for Biotechnology Information
NDB	nucleotide binding domain

N-terminal	amino-terminus
ng	nanogram
nm	nanometer
ORF	open reading frame
p	plasmid
PCR	polymerase chain reaction
pH	potential of hydrogen
PTPase	protein-tyrosine-phosphatase
RBS	ribosome binding site
rpm	revolutions per minute
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
Trp	tryptophan
w/v	weight per volume

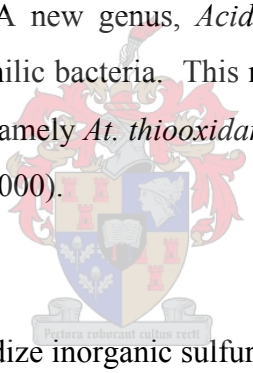
Chapter One

General Introduction

Contents

1.1. *Acidithiobacillus caldus* and its role in biomining

Acidithiobacillus caldus is a moderately thermophilic acidophile with optimum growth at pH 2.5 and a temperature of 45 ^{o}C. It can grow in a temperature range of 32 ^{o}C to 52 ^{o}C, and a pH range of 1.0 - 4.6. Cells are short rods, motile, Gram-negative and are capable of chemolithoautotrophic growth by the oxidation of reduced inorganic sulfur compounds. *At. caldus* is incapable of oxidizing ferrous iron or iron sulfides. It obtains its carbon by the reductive fixation of atmospheric CO_2. On solid tetrathionate medium the colonies are small, circular, convex and with precipitated sulfur in the centre (Hallberg *et al.*, 1994; Hallberg *et al.*, 1996a and Hallberg *et al.*, 1996b). This bacterium was previously classified as *Thiobacillus caldus,* but as a result of 16S rRNA sequence analysis, it became clear that the genus *Thiobacillus* included members of the α-, β- and γ-subclasses of *Proteobacteria*. A new genus, *Acidithiobacillus*, was then created to accommodate these highly acidophilic bacteria. This new genus contains four species of the γ-subclass of *Proteobacteria,* namely *At. thiooxidans, At. ferrooxidans, At. caldus* and *At. albertensis* (Kelly and Wood, 2000).



The ability of this organism to oxidize inorganic sulfur makes it very useful in converting insoluble metal sulfides into soluble metal sulfates which can be leached from their surroundings. *At. caldus* plays an important role in a commercial biooxidation process called biomining. This process uses the oxidizing properties of various acidophilic bacteria either to convert insoluble metal sulphides to water soluble metal sulphates or as a pretreatment process to break up the structure of the mineral, thus permitting other chemicals to penetrate and solubilize the metal. An example of the first process is the solubilization of copper from minerals such as covellite (CuS) or chalcocite (Cu_2S). An example of the second process is the removal of arsenic, iron and sulfur from gold-bearing arsenopyrite. The gold that remains in the mineral is then more easily accessible to cyanide for subsequent extraction (Van Aswegen *et al.*, 1991).

There are currently two main types of microbially assisted biomining processes for the recovery of metals. The first process is an irrigation-type or heap leaching process and involves the percolation of leaching solutions through crushed ore that have been stacked in columns, dumps or heaps (Rawlings *et al.*, 1999a). Each dump is irrigated with iron- and sulfate-rich wastewater from which copper has been removed. Microorganisms growing on the surface of the mineral in the heap catalyze the chemical reactions that result in the conversion of insoluble copper sulfides to soluble copper sulfate. The copper sulfate-containing leach solutions are collected at the bottom of the dump and sent for metal recovery by a process of solvent extraction and electrowinning. Although heap reactors are very cost effective, they are very difficult to aerate efficiently and the pH and nutrient levels are difficult to manage. These dumps are relatively cheap to construct and mainly consist of waste ore, and therefore it is considered to be a low technology process. The metal recovered in the largest quantity by this process is copper (Rawlings, 2002, 2005).

The second process is a stirred tank process and employs highly aerated, continuous-flow bioreactors. The bioreactors are arranged in a series, with finely milled ore or concentrate being added to the first tank together with inorganic nutrients in the form of fertilizer-grade $(NH4)_2SO_4$ and KH_2PO_4. The mineral suspension then flows through a series of highly aerated tanks that are pH and temperature controlled until biooxidation of the mineral concentrate is complete. Because mineral biooxidation is an exothermic process, cooling of the bioreactors is important. Large volumes of air are blown through each reactor, and a large agitator ensures that an even suspension of the solids is carried over to the next tank. A big advantage of this process is that mineral solubilization in these tanks takes place in days, while heap reactors can take up to months to solubilize the minerals. These reactors are extremely expensive to construct and maintain, and therefore their use is mainly restricted to high-value ores and concentrates. Stirred tank bioreactors are mostly used as a pretreatment process for the recovery of gold from ores where the gold is finely divided in a mixture of pyrite or arsenopyrite. These gold particles cannot easily be accessed by cyanide for subsequent solubilization of the mineral. Treatment of the ores with this process decomposes the arsenopyrite, which

allows the cyanide to come in contact with the gold (Rawlings, 2002, 2005; Rohwerder *et al.*, 2003).

These biomining organisms have several physiological features in common. They are all chemolithoautotrophic and use ferrous iron or reduced inorganic sulfur (or both) as an electron donor. A result of sulfur-oxidation is the production of sulfuric acid, and therefore these organisms are all acidophilic and have a pH optimum of 1.5 – 2.0. They can all fix CO_2 and generally are resistant to a wide range of metal ions. In these biomining processes, air provides the carbon source (CO_2) and the electron acceptor (O_2), water provides trace elements and is also the medium of growth. The mineral ore supplies the electron donor in the form of ferrous iron and/or reduced inorganic sulfur (Rawlings, 2002). In general, the same organisms are found in heap-leaching and stirred tank processes. The proportions of the organisms may vary, however, depending on the mineral being solubilized and conditions under which the heaps or tanks are operated (Rawlings, 2005). Unlike other fermentation processes, biomining processes are not sterile, and no attempt is made to maintain sterility. This is not necessary because these highly acidophilic organisms create an environment that is only suited for themselves and not for other organisms. Another reason why this process does not need to be sterile is because the microorganism that decomposes the mineral the most efficiently will out-compete those that are less efficient (Rawlings, 2002).

In processes that operate at moderate temperatures in the range of 35°C - 40°C the most important bacteria are considered to be the iron- and sulfur-oxidizing *At. ferrooxidans*, the sulfur-oxidizing *At. thiooxidans* and *At. caldus*, and the iron-oxidizing *Leptospirillum ferrooxidans* and *Leptospirillum ferriphilum* (Suzuki, 2001). For many years *At. ferrooxidans* was believed to be the most important biomining organism at 40°C, but with the development of techniques such as PCR amplification of 16S rRNA genes from total DNA extracted from environmental samples, it became clear that *At. ferrooxidans* plays a minor role in continuously operating stirred tank reactors operating under steady-state conditions. Here a combination of *Leptospirillum* and *At. thiooxidans* or *At. caldus* are the most important oxidizing bacteria. In a pilot scale, stirred tank operation in which a

polymetallic sulfide ore was treated at 45°C, it was found that *At. caldus*-like, *L. ferriphilum*-like and *Sulfobacillus*-like bacteria was dominant (Okibe *et al*., 2003). *At. ferrooxidans* may be the dominant organism in dump or heap leaching environments where there is a lower redox potential (Rawlings *et al.*, 1999a; Rawlings, 2002; Rawlings, 2005).

Because *At. caldus* and *At. thiooxidans* are closely related, initial studies indicated that *At. thiooxidans* dominated continuous-flow tanks that operated at 40°C. By using restriction enzyme analysis to identify the bacteria in biooxidation tanks, Rawlings *et al.* (1999b) have shown that these tanks used to treat gold-bearing arsenopyrite concentrates are in fact dominated by *At. caldus* instead of *At. thiooxidans*. *At. caldus* are the most dominant sulfur-oxidizing bacteria in pilot plants that treat copper concentrates at 40°C and nickel concentrates at 45-55°C and are considered to be the "weed" of biomining bacteria under these conditions (de Groot *et al.*, 2003). However, for the biooxidation of most ores to proceed efficiently, bacteria capable of oxidizing both sulfur and iron need to be present. *At. caldus* and *At. thiooxidans* cannot, like *At. ferrooxidans* and *L. ferrooxidans*, oxidize an ore such as pyrite when growing in pure culture (Okibe and Johnson, 2004). The reason for this is that the sulfur in the ore is only available to these bacteria after the iron has been oxidized. Therefore, they need to be grown in a mixed culture with iron-oxidizing bacteria. Oxidation of an ore by such a consortia of bacteria generally takes place at a higher rate than in pure culture (Rawlings *et al.*, 1999a).

Several roles of *At. caldus* in the biomining environment have been suggested. Dopson and Lindström (1999) showed in laboratory studies that *At. caldus* is able to enhance the ability of the iron-oxidizing bacterium *Sulfobacillus thermosulfidooxidans* to oxidize arsenopyrite ores at 45 °C. A result of the chemical leaching process is the formation of sulfur (S^0). When sulfur accumulates, it forms a layer on the surface of the mineral. It might be that the presence of *At. caldus* increases the leaching rate by removing the elemental sulfur that builds up on the mineral surface, thereby allowing bacterial and chemical (Fe^{3+}) access to the mineral (Dopson and Lindström, 1999). Bacteria need to attach to the mineral to cause leaching (Arredondo *et al.*, 1994). Another possible

mechanism by which *At. caldus* may affect the leaching rate of arsenopyrite is by the production of organic growth factors. These growth factors may stimulate heterotrophic and mixotrophic growth of bacteria. It is also possible that *At. caldus* might provide organic material in the form of cross-feeding that aids mixotrophic growth. This cross-feeding could also be in the form of a symbiotic relationship, where the iron oxidizing bacteria such as *Leptospirillum* spp. may reduce the concentration of organic chemicals that inhibits growth by accumulation in the cytoplasm. A last possible mechanism by which *At. caldus* might affect the arsenopyrite leaching rate is by production of surface-active agents to solubilize the sulfur (Dopson and Lindström, 1999).

During this process of mineral solubilization, arsenic compounds are leached from the arsenopyrite ore. These arsenic compounds are toxic to the bacteria used in this process. During early stages of development of the biooxidation process, solubilization of the arsenopyrite ore was inefficient and slow and bacteria required a retention time of over 12 days to oxidize the concentrate efficiently for gold recovery. The reason was that the bacteria were not tolerant to such high levels of arsenic. Initially the bacteria were sensitive to less than 1 g arsenic/l. At the Fairview mine in Barberton, South Africa, soluble As^{V} concentrations as high as 12 g/l were reported and at the Sao Bento mine in Brazil As^{III} concentrations of 3-6 g/l (Dew *et al.*, 1997). This arsenic inhibition was so severe that after a period of time the arsenic had to be removed from the tanks before the process could continue. This removal of the arsenic was time consuming and very uneconomic. However, in a period of 2 years, through a process of selection using a bacterial chemostat, the bacterial consortium became sufficiently resistant to these high levels of arsenic and no further removal of arsenic compounds was needed. The retention time needed to oxidize the concentrate efficiently was also reduced from 12 days to less than 7 days (Rawlings and Woods, 1995). After another 3 years of operation at the Fairview mine, the retention time was reduced to a little over 3 days, and the quantity of the suspended mineral concentrate had been increased from 10% to 19% w/v.

1.2. Properties of arsenic

It is believed that Albertus Magnus discovered arsenic around 1250. Arsenic is a semi-metal or metalloid that exists in the environment in two biologically important oxidation states, namely As^{V} (arsenate) and As^{III} (arsenite). It is one of the most prevalent toxic metals in the environment (although in very low concentrations) and is commonly associated with the ores of metals like gold, copper and lead. Arsenite absorbs less strongly and to fewer minerals than arsenate, which makes it a more mobile oxyanion than arsenate. Arsenate is the predominant form of arsenic in aerobic environments, while arsenite is more predominant in an anoxic environment (Oremland and Stolz, 2003; Rosen, 2002a).

Arsenite is said to be about 100 times more toxic than arsenate. In a study where the genotoxic activity of arsenite and arsenate was tested in TK6 human lymphoblastiod cell lines, arsenite was far more genotoxic than arsenate (Guillamet *et al.*, 2004). The toxicity of arsenite lies in its ability to bind to the sulfur groups of essential cysteines in proteins. Because arsenite binds two sulfur groups it can cross link proteins, altering their overall shape and thereby impeding their function. Because of the similarity in structure and solubility between arsenate and phosphate, arsenate can act as a phosphate analog and competes for the formation of ATP (Coddington, 1986). Kenney and Kaplan strengthened this theory by showing that arsenate substituted phosphate in both the sodium pump and the anion exchanger of human red blood cells (Kenney and Kaplan, 1988). The transport of arsenate into bacterial cells is via two phosphate transport membrane systems, the Pit and Pst system, while the Pst system seems to be the predominant system for arsenate uptake. Arsenite is transported into bacterial cells by aqua-glycerolporins (glycerol transport proteins) (Sanders *et al.*, 1997) (Figure 1.1). A mutation in the glycerol facilitator of *E. coli* (GlpF) converted the *E. coli* strain to antimony (Sb^{III}) resistance. Because the chemical properties of arsenite and antimony are similar, it is very likely that GlpF is also an As^{III} transporter (Mukhopadhyay *et al.*, 2002; Rosen, 2002a; Rosen, 2002b). Recently Meng *et al.* (2004) confirmed As^{III} uptake by GlpF in *E. coli*.

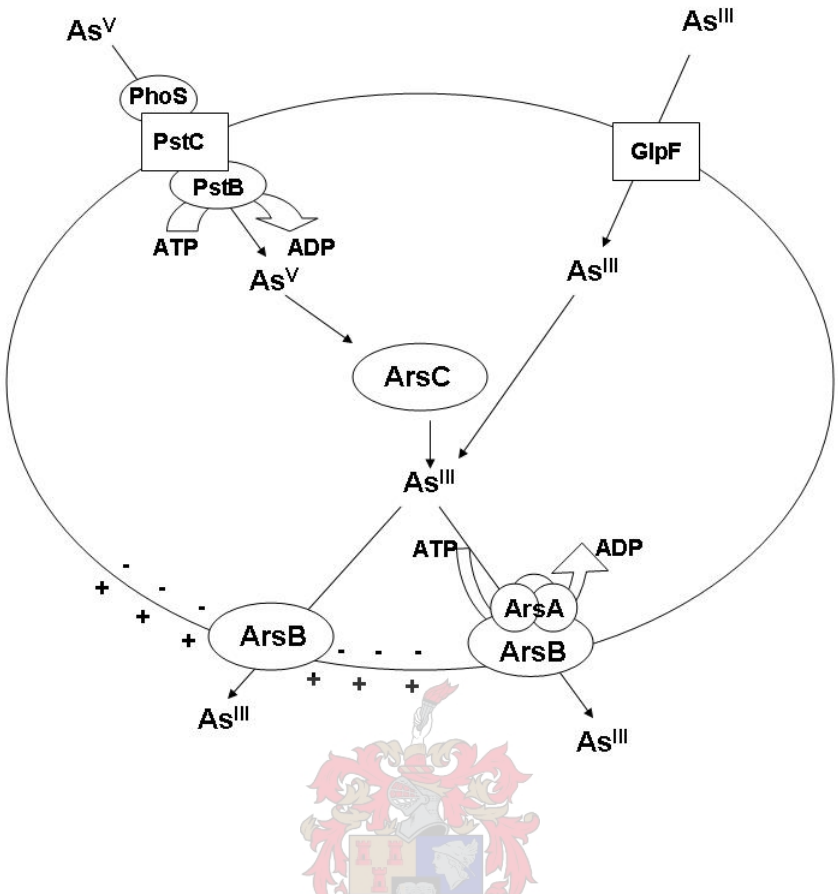


Figure 1.1: Uptake mechanisms of arsenate and arsenite in *E. coli*. Arsenate is brought into cells by phosphate transporters, the Pst and the Pit system (top left). Arsenate is then reduced to arsenite by the ArsC (arsenate reductase, middle). Arsenite uptake is facilitated by the aquaglycerolporin GlpF (top right). Arsenite is actively extruded by the ArsAB complex (As^{III}/Sb^{III}-translocating ATPase, bottom right). If ArsA is absent, ArsB acts as a secondary arsenite carrier protein and efflux is coupled to membrane potential (bottom left). Adapted from Rensing *et al.* (1999).

Arsenical compounds have been used for many purposes, mainly in medicine (mostly for treating protozoan diseases), in agriculture (as herbicides, fungicides, pesticides and animal feed additives) and as a poison. In 1908 Paul Ehrlich won the Nobel Prize in medicine for the use of arsenic compounds as chemotherapeutic agents (Mukhopadhyay *et al.*, 2002). Due to the fact that arsenic is so prevalent in the environment, microorganisms have developed various resistance mechanisms to counteract its toxicity.

1.3. Bacterial resistance to arsenic

Due to the abundance of arsenic in the environment, microorganisms come into contact with this metalloid on a regular basis. Therefore it is not surprising that chromosomal and plasmid located genes which confer resistance to arsenic have been isolated from bacteria. Different mechanisms of resistance to arsenic have evolved in microorganisms, such as the overproduction of intracellular thiols (e.g. the protozoan *Leishmania*), phosphate pathways that do not transport arsenate efficiently (e.g. cyanobacteria), the oxidation of arsenite to the less toxic arsenate (e.g. *Alcaligenes faecalis*), sequestration in a vacuole (e.g. fungi) and methylation (Rosen, 1999; Oremland and Stolz, 2003; Cervantes *et al.*, 1994; Wang *et al.*, 2004). The best characterized and most studied way of detoxification is a mechanism where intracellular As^{V} is converted to As^{III} and a specific efflux pump then extrudes As^{III} from the cytoplasm (*ars* operon). This system has been extensively studied and will be the focus of this study.

The *ars* operon has been isolated from Gram-positive and Gram-negative bacteria. There are two common forms of the *ars* operon. The first and most common form consists of three genes, *arsRBC*. This form of the *ars* operon is located on chromosomes and plasmids of many organisms, but the best studied is the chromosomal *ars* system of *E. coli* K-12 and the *ars* systems on plasmids pI258 and pSX267 of *Staphylococcus* (Diorio *et al.*, 1995). The second operon is found on plasmids of Gram-negative bacteria (e.g. plasmid R773 and R46 of *E. coli* and plasmid pKW301 of *Acidiphilium multivorum*) and, more recently, on transposons of *At. caldus* (AY821803), *L. ferriphilum* (DQ057986), *A. faecalis* (Ay297781) and *M. flagellatus* (NZ_AADX01000013) (Tuffin *et al.*, 2006). this operon consists of five genes, *arsRDABC* (Figure 1.2) (Suzuki *et al.*, 1998). Both these operons have been organized into a single transcriptional unit. The *arsR* and *arsD* genes encode two different regulatory proteins. ArsR is a trans-acting repressor that senses intracellular As^{III} and controls the basal level of expression of ArsB and ArsC (Yang *et al.*, 2005). ArsD is an inducer-independent protein and controls the upper level of expression (Saltikov and Olson, 2002). ArsA is an arsenite stimulated ATPase, energizing ArsB (a membrane-bound arsenite and antimony efflux pump) by ATP hydrolysis (Silver, 1996). In the presence of ArsA, arsenite and antimony are exported

actively out of the cell by using cellular ATP. If the ArsA is absent, these oxyanions can still be excreted by ArsB using membrane potential (Cai *et al.*, 1998). ArsC is an arsenate reductase that converts arsenate to arsenite to get exported by the efflux pump.

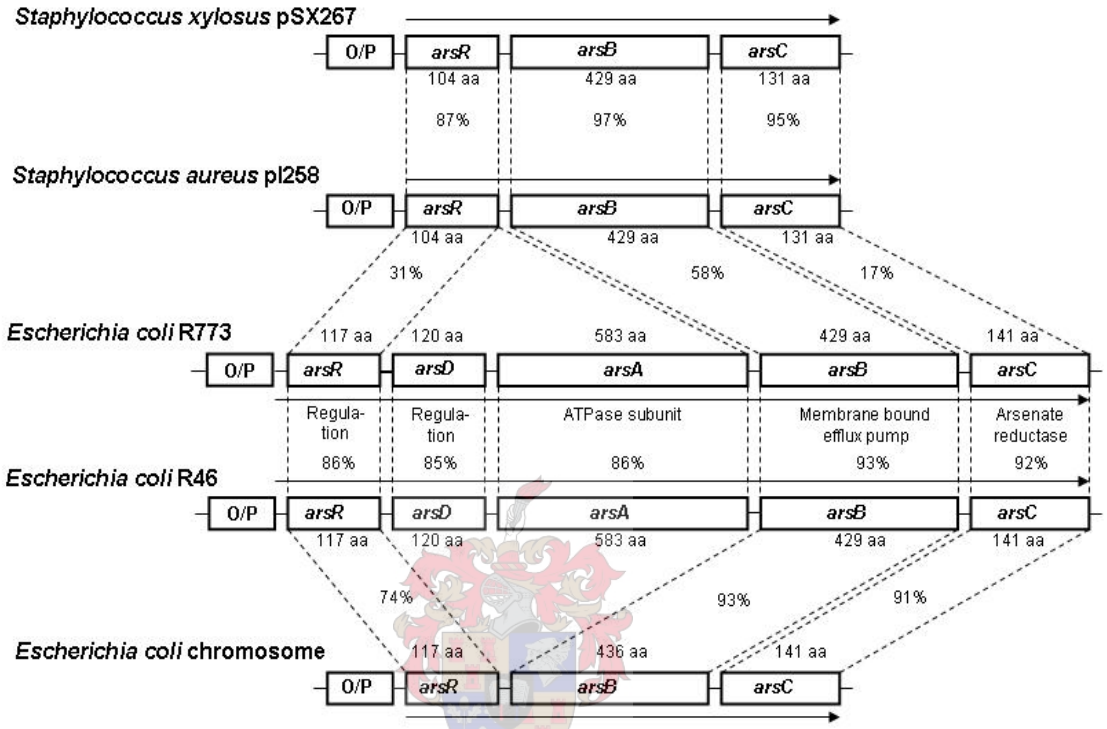


Figure 1.2: Homologies among the arsenic resistance determinants of *Staphylococcus* plasmids pSX267 and pI258, *E. coli* plasmids R773 and R46, and the chromosomal *ars* operon of *E. coli*. The sizes of putative gene products (in amino acids [aa]) are shown above or below the genes (boxes). The numbers between the dashed lines are percentages of similarity among the ArsR, ArsB and ArsC proteins. O/P, operator/promoter. Adapted from Silver and Phung (1996).

With new technology constantly developing and data generated from genome sequencing projects, it became clear that *ars* operons may differ from the two common forms described above. Southern blot hybridization experiments where different arsenic resistance genes were probed against total DNA isolated from bacteria revealed that these genes are very widespread among microorganisms (Dopson *et al.*, 2001). PCR experiments with degenerate primers to these resistance genes also revealed the prevalence of *ars* operons. There are however, other forms of arsenic operons that differ from the *arsRBC* and *arsRDABC* conformation. Other genes that have also been reported

to be associated with arsenic resistance are *arsH* and *arsM*. For example, an operon that is different in gene order and that contains an extra gene was identified in *At. ferrooxidans* (Butcher *et al.*, 2000). This operon consisted of *arsCRBH* genes with the *arsCR* and *arsBH* genes divergently transcribed. The function of the *arsH* gene is still unknown, but it was not required for and did not enhance arsenic resistance in *E. coli*. Cyanobacterium *Synechocystis* sp. strain PCC 6803 has an *arsBHC* arsenic resistance operon, where the *arsH* was also not required for resistance (Lopez-Maury *et al.*, 2003). Neyt *et al.* (1997) found an *arsHRBC* operon on plasmid pYV of *Yersinia enterocolitica* with the *arsH* gene also divergently transcribed. The presence of this gene however, either in *cis* or *trans* was essential for arsenic resistance. This was surprising, because *arsRBC* are sufficient to confer resistance in *E. coli* and staphylococci. The authors suggested that *arsH* might have a regulatory function similar to *arsD*. The same *arsHRBC* operon was also found on the IncH12 plasmid R478 (Ryan and Colleran, 2002). PCR amplification using primers to the *arsH* gene showed that this gene is present in many other arsenic resistant IncH12 plasmids. Removal of the *arsH*, like in *Y. enterocolitica,* rendered the cell sensitive to arsenic. The legume symbiont *Sinorhizobium meliloti* contains a very interesting *ars* operon. It consists of four genes, an *arsA*, followed by an *aqpS* (aquaglyceroporin) and then *arsCH*. A deletion of *arsH* also rendered the cell sensitive to arsenite and arsenate (Yang *et al*., 2005).

The archaeon, *Halobacterium* sp. strain NRC-1 contains a gene encoding a mammalian arsenite-methyltransferase homolog, named *arsM*. A deletion of this gene increased sensitivity to arsenite, indicating a novel mechanism of arsenic resistance, possibly by methylating intracellular arsenite and thereby creating a concentration gradient to the outside of the cell. It is also possible that arsenite could be methylated to a volatile trimethyl-arsine that would leave the cell by diffusion (Wang *et al.*, 2004; Tuffin *et al.*, 2005).

The *ars* operon of the acidophilic archaeon *Ferroplasma acidarmanus* consists of only two arsenic resistance genes, *arsRB* (Gihring *et al.*, 2003). An *arsC* gene was absent in the genome. Although no *arsC* gene is present, the organism was still resistant to

arsenate. It is suggested that *F. acidarmanus* employs phosphate specific transporters to reduce non-specific uptake of arsenate. *F. acidarmanus* also lacks a complete *arsA*-homologous open reading frame (ORF). The ArsB may thus not function as a subunit of a primary pump, but as a secondary carrier using membrane potential to export arsenite out of the cell. This may not be the case however, as acidophilic microorganisms commonly have a reversed membrane potential (Butcher *et al.*, 2000). Membrane potential may not be a suitable energy source to export arsenite out of the cell. The mechanism of arsenite efflux by *F. acidarmanus* is still unknown.

Another variation of the conventional *ars* operons is in *Mycobacterium tuberculosis*, where the *arsB* and *arsC* genes are fused into a single gene, encoding a 498-residue fusion protein (Mukhopadyay *et al.*, 2002). The reason for the fusion of these two genes is unknown. Another very interesting arsenic operon is that of the legume symbiont *Sinorhizobium meliloti*. The operon consists of an *arsR* gene, followed by an *aqpS* gene in place of *arsB*. A third ORF downstream of *aqpS* showed homology with an *arsC* and the fourth to an *arsH* gene. AqpS showed homology with the bacterial glycerol facilitator, GlpF, which facilitates transport of arsenite and antimonite into bacterial cells. *S. meliloti* utilizes a unique arsenic detoxification pathway where ArsC converts arsenate to arsenite, which is then transported down a concentration gradient through the AqpS channel to the outside of the cell (Yang *et al.*, 2005).

Sato and Kobayashi (1998) found an *ars* operon on the skin element of *Bacillus subtilis* that also differs from the conventional three or five gene operons. This operon contains the *arsRBC* genes, but also a fourth gene (*ORF2*) located between the *arsR* and *arsB* genes. This gene did not show homology to any known *ars* genes, but did show homology to an ORF of unknown function situated upstream of the *arsRBC* operon in *Mycobacterium tuberculosis*. These experiments conducted on the *ars* genes however, did not define the role of *ORF2* in the arsenic resistance mechanism of *Bacillus subtilis*.

1.4. Proteins involved in the *ars* operon

1.4.1. The ArsAB pump

The best studied ArsAB pump is that encoded by the *ars* operon of *E. coli* plasmid R773. ArsA normally is part of an aggregate with the membrane-bound ArsB protein, and together forms an ArsAB As^{III}/Sb^{III}-translocating ATPase, which extrudes As^{III} and Sb^{III} oxyanions out of the cell. The 63 kDa ArsA ATPase is the catalytic component of the ArsAB pump, with As^{III}/Sb^{III} stimulated ATPase activity. It hydrolyses ATP in the presence of antimonite and arsenite. The 45 kDa ArsB subunit (which forms the oxyanion-translocating pathway) uses the chemical energy released by ArsA to secrete arsenite and antimonite out of the cell (Tisa and Rosen, 1990). The *arsRDABC* operon confers higher levels of resistance than the *arsRBC* operon, indicating that the ATP-driven ArsAB pumps out arsenite more efficiently. This illustrates that ArsB in complex with ArsA can reduce the intracellular concentration of As^{III}/Sb^{III} to a greater extent than ArsB alone (Rosen, 1999).

1.4.1.1. ArsA (ATPase subunit)

When ArsA is expressed in the absence of ArsB, it is found in the cytosol and can be purified as a soluble protein (Walmsley *et al.*, 1999). The 583-residue ArsA has two homologous halves, a N-terminal A1 (residues 1-288) and a C-terminal A2 (residues 314-583), which are connected by a flexible linker of 25 residues (residues 289-313) (Bhattacharjee *et al.*, 2000) (Figure 1.3). These two homologous halves are most likely the result of ancestral gene duplication and fusion (Chen *et al.*, 1986). The crystal structure of the ArsA has been determined (Zhou *et al*., 2000), and three types of domains were found. Firstly, there are two nucleotide binding domains (NBDs) for ATP that contain residues from both A1 and A2. Secondly, there is a metalloid-binding domain (MBD). This site is allosterically activated and is positioned at the opposite end of the protein from the NBDs. The third domain is a signal transduction domain (DTAP) that connects the MBD to the two NBDs (Rosen, 2002a). There is also a linker that connects the A1 and A2 halves of ArsA.

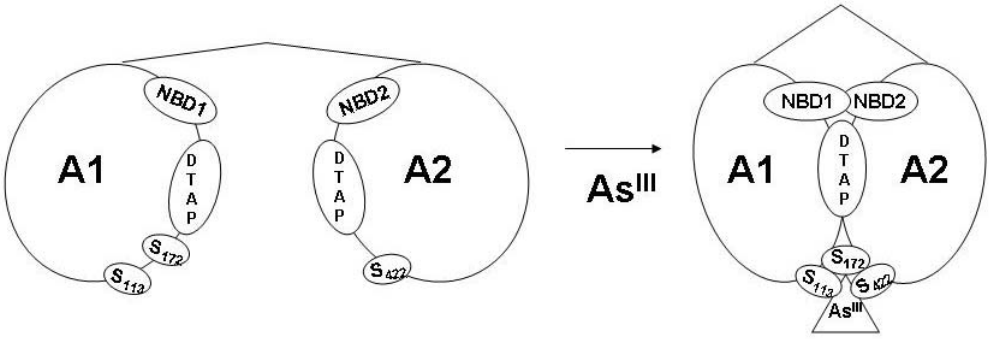


Figure 1.3: Model of the structure of the ArsA ATPase from plasmid R773. The protein consists of two homologous domains, a N-terminal A1 and a C-terminal A2, connected by a flexible linker region. Each domain contains a nucleotide binding domain (NBD), a signal-transduction domain (DTAP) and a metalloid binding domain. The cysteine residues have been found to comprise the metalloid binding domain (MBD). Binding of arsenite to the cysteine residues brings the A1 and A2 domains together, resulting in hydrolysis of ATP. Adapted from Rosen (2002a).

The Nucleotide Binding Domains (Catalytic Site)

The ArsA protein has two nucleotide binding domains, NBD1 and NBD2. Both these domains are composed of residues from the A1 and A2 halves of ArsA (Zhou *et al.*, 2000). Both NBDs are located at the interface between A1 and A2, in close proximity to each other (Zhou *et al.*, 2001). Mutations in either A1 or A2 NBD resulted in a sensitive strain with loss of resistance to arsenite, ATPase activity and transport (Karkaria *et al.*, 1990; Kaur and Rosen, 1992). This indicates that both the A1 and A2 NBDs are required for oxyanion-translocating ATPase activity. The sequence for the consensus binding sites of ATP (P-loop) is G^{15}KGGVGKTSIS^{25} and G^{334}KGGVGKTTMA^{344} in the A1 and A2 halves, respectively (Chen *et al.*, 1986). By using site-directed mutagenesis Li *et al.* (1996) altered the first glycine residues in A1 and A2 to cysteine residues. Cells expressing a $G^{15}C$ (A1) mutant and a $G^{15}C/G^{334}C$ double mutant exhibited moderate sensitivity to arsenite, whereas a $G^{334}C$ mutant in A2 retained arsenite resistance. Arsenite resistance was restored in the moderate sensitive mutants to the wild type with a single substitution of $A^{344}V$. Because $A^{344}V$ is located adjacent to the C-terminal NBD, it

suggests a spatial proximity of Gly^{15} and Ala^{344} and thus indicates that the A1 and A2 domains do interact to form a catalytic unit. Li and colleagues further studied this interaction of the A1 and A2 NBDs and found that the bigger the residue substitution at position 15 in A1, the weaker the resistance to arsenite. The larger the residue at position 344 in A2, the greater the suppression of the G^{15} mutants (Li *et al.*, 1998). This shows an inverse relationship between ATPase activity and residue volume at position 15, indicating that the geometry at the interface between A1 and A2 NBDs imposes steric constraints on residues allowed in the A1-A2 interface.

By looking at the structure of ArsA in complex with $Mg{\cdot}ADP$, it is clear that the A1 NBD (NBD1) is in a closed conformation and has ADP bound, while the A2 NBD (NBD2) has an open conformation, and ATP can be exchanged into this site. This indicates that NBD2 will be much more accessible than the NBD1 (Zhou *et al.*, 2000; Rosen, 2002b). These findings are consistent with studies done by Kaur (1999) which suggests that A2 is a low affinity, easily exchangeable site, while A1 is a high affinity, poorly exchangeable site. Further studies have also shown that the nature of the nucleotide bound at A2 is reflected in the conformation of the A1 domain (Zhou *et al.*, 2001). In the absence of antimonite or arsenite, ATP binding and hydrolysis is at a slow basal rate and occurs firstly in A1, which then results in a change in A2 so that A2 is now more competent to bind ATP. In the presence of As^{III}/Sb^{III} (multisite conditions), ATP binding and hydrolysis by A2 produces a change in A1, giving it a more open conformation. The result is the release of tightly bound ADP from A1 and the subsequent binding of another ATP molecule to A1 (Jia and Kaur, 2003).

In the presence of As^{III}/Sb^{III}, Walmsley *et al.* (2001) found that NBD1 hydrolyzed ATP 250-fold faster than NBD2. Although the two NBDs have overall structural similarities and similar evolutionary relationships, the two NBDs have intrinsic differences. These findings gave more weight to the possibility that the two NBDs have separate roles in catalysis (Zhou *et al.*, 2002). This hypothesis was further strengthened when a thrombin site was introduced into the linker that connects A1 and A2 to distinguish between events at NBD1 and NBD2. After labeling and thrombin digestion, A1 and A2 migrated at

different mobilities on SDS-PAGE. This showed that the two NBDs have different properties, and possibly different functions (Jiang *et al.*, 2005). They further showed that a) both nucleotides are catalytic and hydrolyze ATP in the absence and presence of Sb^{III}, b) the affinity for ATP is increased in both NBD1 and NBD2 by Sb^{III}, and c) NBD1 has a higher affinity for ATP than NBD2. These findings raised the question as to the function of NBD2. One possibility is that it could once have been an important evolutionary structure, but that it is not involved in ATP transport anymore. This possibility seems fairly unlikely, however, as Kaur and Rosen (1992) have shown that mutagenesis of NBD2 resulted in loss of ATPase activity and As^{III} resistance. Another possible function is that it could play a regulatory role. A third possibility is that the two NBDs could play equivalent roles in an intact ArsAB complex. The noted differences might be the result of the analysis of ArsA in the absence of ArsB (Jiang *et al.*, 2005).

Previously it was shown that Mg^{2+} is required for ArsA ATPase activity (Hsu and Rosen, 1989). An increase in intrinsic tryptophan fluorescence occurred only on the addition of MgATP, which indicated conformational changes at the NBDs. This effect was observed only in the presence of ATP, suggesting that Mg^{2+} binds to ArsA as a complex with ATP (Zhou *et al.*, 1995). From sequence alignment of ArsA homologs with enzymes such as NifH, RecA and GTP-binding proteins such as Ras p21, it was suggested that Asp^{45} might be a putative Mg^{2+} ligand. From the crystal structure of Ras p21 bound to GTP, the conserved Asp residue formed a portion of the Mg^{2+} binding site (Pai *et al.*, 1990). The Mg^{2+} ion bring together different components of the GTP binding core, resulting in information flow between domains. To examine the role of Asp^{45}, mutants were constructed in which Asp^{45} was changed to Asn, Glu or Ala. Cells expressing these mutated *arsA* genes almost completely lost arsenite resistance. These results supported the role of Asp^{45} as a Mg^{2+} ligand (Zhou and Rosen, 1999).

The Metalloid Binding Domain (Allosteric Site)

Arsenite and antimonite have been shown to allosterically activate ArsA ATPase activity (Rosen *et al.*, 1999). As^{III} stimulates ATPase activity 3-5 fold, while Sb^{III} as an activator stimulates ATPase activity 10-20 fold. In the absence of the metalloids, a low level of

ATPase activity is observed by ArsA. No other oxyanion tested had an effect on activity (Hsu and Rosen, 1989). One possibility was that the activation by Sb^{III} or As^{III} might be through the binding as oxyanions to anion binding sites. As metalloids however, they might react as soft metals through covalent binding with cysteine thiolates. In such a reaction, where arsenite reacts with thiolates such as dithiothreitol or glutathione, arsenite acts as a soft metal, forming direct metal-sulfur As-S bonds (Rosen *et al.*, 1999). If the activation of ArsA does involve direct metal-sulfur bonds, alteration of participating cysteine residues would greatly affect the activation of ArsA. The thiol-modifying reagent methyl methanethiosulphonate (MMTS) inhibits ArsA catalysis, suggesting that cysteine residues are involved in catalysis (Bhattacharjee *et al.*, 1995).

The ArsA of *E. coli* has four cysteine residues, Cys^{26}, Cys^{113}, Cys^{172} and Cys^{422} (Figure 1.4). For As^{III} or Sb^{III} to act as an allosteric activator, two or more cysteine residues must be in close proximity in the folded protein to interact with the metalloid. To investigate the role of the cysteine residues in the activation of ArsA, each cysteine was altered to a serine residue by site-directed mutagenesis (Bhattacharjee *et al.*, 1995). The purified $C^{26}S$ ArsA showed identical properties to those of the wild-type ArsA, with no change in resistance to As^{III} or Sb^{III}. This suggested that Cys^{26} is not involved in activation of ArsA. In contrast, cells expressing the other three mutants were sensitive to Sb^{III} and As^{III}. The purified $C^{113}S$, $C^{172}S$ and $C^{422}S$ enzymes each exhibited a similar affinity for ATP to the wild-type ArsA, but the concentration of Sb^{III} or As^{III} required for activation of ArsA was substantially increased, most likely reflecting a decrease in affinity for the metalloid. This resulted in reduced ATPase activity and rendered the mutants sensitive to As^{III} and Sb^{III}.

These three cysteines are distant from each other in the primary structure of ArsA, but as the protein folds it brings them into close proximity. As^{III} or Sb^{III} then interact with these cysteines to form a metal-thiol cage. To determine the distance between the cysteine residues, the wild-type ArsA and ArsA with Cys-Ser substitutions were treated with the homobifunctional crosslinker dibromobimane (bBBr) (Bhattacharjee and Rosen, 1996). bBBr has two bromomethyl groups that can crosslink a thiol pair located within 3-6 Å of

each other to form a fluorescent adduct. An ArsA protein with only one cysteine group altered by mutagenesis still formed fluorescent adducts. Proteins lacking any two of the three cysteines at residue 113, 172 or 422 did not form fluorescent adduct. These results demonstrate that Cys^{113}, Cys^{172} and Cys^{422} are in close proximity of each other in the tertiary structure so that As^{III} or Sb^{III} can interact with these three cysteine residues in a trigonal pyramidal geometry.

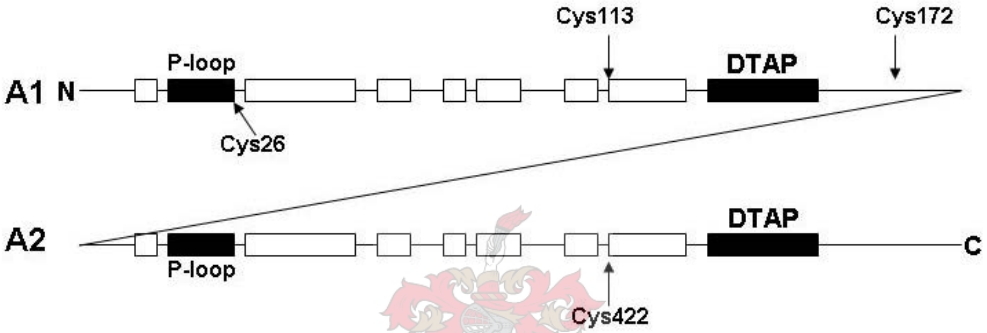


Figure 1.4: The primary sequence of ArsA is represented linearly, with the homologs A1 and A2 halves aligned and linked by the flexible linker region. Boxes indicate regions of greatest similarity. The black boxes indicate the location of the P-loop of the NBDs and the signal transduction domains (DTAP). The locations of the four cysteines are indicated. Cys^{113}, Cys^{172} and Cys^{422} form the metalloid binding domain (MBD). Adapted from Bhattacharjee *et al.* (2000)

Based on the structure of the ArsA protein, Zhou *et al.* (2000) suggested that it may be possible that a change in the affinity of ArsA for As^{III}/Sb^{III} might inject metal ions into ArsB as a result of conformational changes during ATP hydrolysis. This was contrary to previous ideas that the ions that activate ArsA are not transported by ArsB. To further investigate this hypothesis, ArsA was crystallized in complex with ATP, the non-hydrolyzable ATP analog AMP-PMP and the transition state analog of ATP hydrolysis, $ADP{\cdot}AlF_3$ (Zhou *et al.*, 2001). When ArsA crystals were formed or incubated in the presence of ATP, ATP was found at NBD2, and ADP at NBD1. Surprisingly, when

crystals were incubated with AMP-PNP, it was found only at NBD1. The inability to obtain crystals in the presence of AMP-PNP, suggested that crystals can only be formed if NBD1 contains ADP. A reason for this might be that conformational changes associated with ATP binding and hydrolysis at NBD1 is not allowed in the crystal. Using pre-steady-state kinetics, it has been shown that binding of ATP favors the uptake of As^{III}/Sb^{III}, while the release of ADP from NBD2 is associated with the release of these ions from the MBD (Walmsley *et al.*, 2001).

Changes in helices of A1 and A2 corresponded with ATP hydrolysis at NBD2. Helices H9-H10 of A1 and A2 forms the arms of a gate, alternating in the "open" and "closed" positions at the interface with ArsB (Figure 1.5). When As^{III}/Sb^{III} first reacts with ArsA, the H9-H10 region of A1 provides the ceiling of the cavity where the ion binds. When hydrolysis occurs at NBD2, the bound ion moves from the cytosolic side into a protected pocket at the interface with ArsB. Release of ADP from NBD2 triggers the release of the As^{III}/Sb^{III} ion inside this pocket. ATP hydrolysis at NBD1 is then required to bring ArsA back to the ground state. Based on the above, the catalytic cycle of ArsA functions similar to that of a reciprocating engine (Zhou *et al.*, 2001). Many elements of this model remain to be proven.

Signal Transduction Domain (DTAP)

The catalytic and allosteric sites of the ArsA ATPase are located distant from each other in the enzyme, requiring a mechanism that can transmit the information from the MBD (allosteric site) to the NBD (catalytic site) where ATP hydrolysis occurs. This flow of information has been shown to be mediated by a 12-residue consensus sequence, DTAPTGHTIRLL (Bhattacharjee *et al.*, 2000). This sequence is termed DTAP or the signal transduction domain and is found in each half of ArsA.

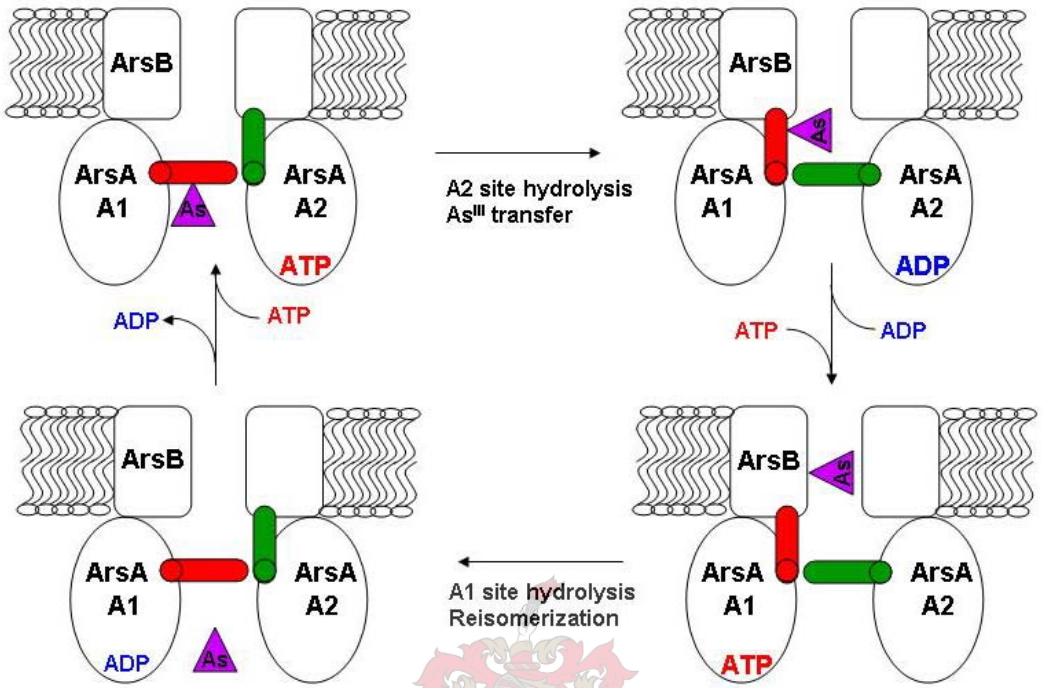


Figure 1.5: Schematic model of the proposed mechanism of the ArsA ATPase. Helices of the A1 (red) and A2 (green) domains form the arms of a gate that alternates between the open and closed conformations. An As^{III} ion is shown as a purple triangle. Adapted from Zhou *et al.* (2001). See text for details.

ArsA has four tryptophan residues located at positions 159, 253, 522 and 542. The function of the DTAP domain in ArsA was examined by using the effects of substrates and effectors on intrinsic tryptophan fluorescence (Zhou *et al.*, 1997). Using site-directed mutagenesis two single-tryptophan ArsAs were constructed, containing either Trp^{141} or Trp^{159} at the N- and C-terminal side of the A1 DTAP domain respectively. In the Trp^{141} protein all native tryptophan residues were changed to tyrosine, and Phe^{141} was changed to tryptophan. In the Trp^{159} ArsA the other three tryptophan residues were changed to tyrosine. These two tryptophan residues served as intrinsic probes of the environment of the DTAP domain. During ATP hydrolysis this domain undergoes conformational changes that may be involved in coupling the allosteric and catalytic sites. The fluorescence spectra indicated that the N-terminal Trp^{141} is located in a hydrophobic

region, while the C-terminal Trp^{159} is located in a relatively hydrophilic environment. Binding of Mg·ADP moved Trp^{141} into a more hydrophilic region, exposing the N-terminal end of the DTAP domain to a more polar environment. Upon binding of Mg·ATP, Trp^{159} moved into a more hydrophobic environment, leaving the C-terminal end in a less polar environment. The catalytic and allosteric domains are thus connected to each other by a "flipping" movement in the signal transduction domain during catalysis.

Linker Region

The sequence of the *E. coli* plasmid R773 ArsA indicated that A1 and A2 are connected by a flexible linker peptide. The function of this linker sequence was examined by extending it with five glycine residues or by the deletion of 5, 10, 15 or 23 residues (Li and Rosen, 2000). Cells expressing an *arsA* gene with the extended linker exhibited similar resistance and affinity for ATP and Sb^{III} to the wild type. Cells expressing the *arsA* genes with deleted linkers had increasing levels of As^{III} sensitivity and a decreased affinity for ATP and Sb^{III}. The authors proposed that the linkers have evolved to the shortest length that still allow the two halves of the protein to interact and that it was not the sequence of the linker that was important, but the length of the linker. Jia and Kaur (2001) however, studied the role of the *E. coli* R773 linker by creating point mutations and complementation experiments and found that the nature of the residues in the linker are important for correct conformation of the NBDs and for catalytic activity. The increase in sensitivity seen by Li and Rosen may not have been due to the shortening of the linker, but because of the role the linker plays in the function of ArsA. Jia and Kaur expressed an N-terminal A1 clone, that lacks the linker sequence, with two different C-terminal A2 clones, one without the linker and the other one containing the linker sequence. Because each domain of ArsA was on a separate polypeptide, deletion of the linker would not inhibit movement of the two domains. Unless the linker is required for the function of ArsA, the absence of the linker would have little effect on the ArsA function. The N- and C-terminal halves were able to complement each other only if the linker was present, indicating that the linker is essential for function. Point mutations of certain residues in the linker resulted in a loss of As^{III} resistance and ATPase activity. These mutations also affected the conformation of ArsA. The authors proposed that

residues in the linker region play an active role in the function of the protein and that there is an interaction between the linker and the NBDs.

1.4.1.2. ArsB (Membrane bound efflux pump)

The *arsB* gene of the *E. coli* R773 *ars* operon encodes a protein of approximately 45.5 kDa. Hydropathy plots demonstrated the hydrophobic character of the protein. At least 10 regions of 19 or more residues has a high average hydropathy, indicating these regions may be potential membrane-spanning α-helices. Using an ArsB-β-galactosidase hybrid protein, β-galactosidase activity of different extracts was measured. Results indicated that the fusion protein was localized to the inner membrane and it was postulated to be the anion channel component of the ArsAB pump (Chen *et al.*, 1986; San Francisco *et al.*, 1989).

The ArsB protein is difficult to study due to the poor expression of the *arsB* gene. It was suggested that the level of the *arsB* expression is controlled at a translational level. Analysis of the *arsB* translational initiation region (TIR) revealed a possible hairpin at the third codon. Another potential mRNA secondary structure was identified immediately upstream of the ribosome binding site (Dou *et al.*, 1992). These structures may interfere with the ribosome, causing it to pause at the start of the *arsB* sequence, thereby reducing the amounts of ArsB produced. The use of the T7 expression system has allowed San Francisco and colleagues to visualize the ArsB protein as a [^{35}S] methionine-labeled membrane protein (San Francisco *et al.*, 1989).

Tisa and Rosen (1990) performed binding studies with purified ArsA protein to membranes with and without the *arsB* gene product to determine if expression of the *arsB* gene is required to anchor the ArsA protein to the inner membrane. In cells expressing the *arsB* gene, the presence of the ArsA protein on the membrane was shown by immunoblotting of the membrane with antiserum prepared against the ArsA protein. In cells expressing an ArsA protein with either a truncated or no ArsB protein, ArsA protein was found only in the cytosol. This indicated that the expression of the *arsB* gene is essential to anchor the ArsA protein to the inner membrane. Membranes of cells

lacking an *arsA* gene were able to bind purified ArsA protein when added exogenously, suggesting that the ArsB protein is inserted into the membrane in a functional form in the absence of ArsA.

By fusing certain regions of the *arsB* gene to reporter genes, a more accurate prediction of the topology of certain regions in the ArsB could be determined compared with the hydropathy plots reported by Chen *et al.* (1986). Three types of gene fusions have been used. Fusions with the *phoA* gene (which encodes for alkaline phosphatase) shows high activity if the fusion is made in the coding region for the periplasmic section of the ArsB, while fusions with the *lacZ* gene (which encodes for β-galactosidase) shows high activity if the fusion is made in the coding region for the cytosolic region of the ArsB. Fusions with the *blaM* gene (which encodes for β-Lactamase) will provide resistance to ampicillin if the fusion is made in the coding region for the periplasmic part of the ArsB. Based on the results of these fusions, a model was proposed where the ArsB protein has 12 membrane-spanning α-helices joined by six periplasmic loops and five cytoplasmic loops. The N- and C-termini are suggested to be located in the cytosol (Wu *et al.*, 1992).

From studies of the energetics of the ArsAB pump it has been shown that this complex is an arsenite-translocating ATPase (Dey *et al.*, 1994; Karkaria *et al.*, 1990). Certain observations suggested that this hypothesis may be too simple. Firstly, a characteristic motif of secondary carrier proteins is 12 membrane-spanning α-helices. Thus, the ArsB has a similar structure to secondary carrier proteins. Secondly, staphylococcal *ars* operons have no *arsA* gene, but still confer resistance to arsenite and extrude arsenite in an energy-dependent manner, suggesting that the ArsB protein can extrude arsenite in the absence of ArsA (Diorio *et al.*, 1995). Thirdly, when the R773 *arsB* gene is expressed in the absence of the *arsA* gene, it still confers an intermediate level of arsenite resistance in *E. coli*. These observations raised the question as to whether energy was supplied by another ATPase elsewhere in the genome in the absence of *arsA*, or if another form of energy drives the pump in the absence of ArsA. Dey and Rosen (1995) investigated the possibility of the R773 ArsB utilizing two different kinds of energy. To do this they used an *unc* strain of *E. coli* to compare arsenite resistance and *in vivo* energetics of arsenite

transport in a strain expressing both *arsA* and *arsB* genes and in a strain only expressing the *arsB* gene. The *unc* strain of *E. coli* lacks the H^+-translocating ATPase and is thus unable to convert ATP and the electrochemical proton gradient. An *arsA* deletion mutant pBC101 (*arsRDBC*) and a clone that only contains *arsA* (pArsA) were transformed into the *unc* deletion strain of *E. coli*. ATP levels were controlled by growing the cells in glucose (generates ATP through substrate-level phosphorylation) or succinate (generates almost no ATP). Cells expressing both *arsA* and *arsB* genes only actively exclude arsenite in the presence of ATP (when growing in glucose). When fluoride (which inhibits glycolysis and thus prevents ATP synthesis) was added, the active exclusion of arsenite was inhibited. Cyanide, which prevents respiration, but does not alter ATP levels, failed to inhibit exclusion of arsenite (Dey and Rosen, 1995). This result indicated that the removal of arsenite in cells containing both *arsA* and *arsB* genes was ATP dependent. In cells expressing only the *arsB* gene, arsenite was excluded in the presence and absence of ATP. This exclusion was inhibited by the addition of cyanide, showing that respiration rather than ATP levels sufficiently excludes arsenite. When the uncoupler CCCP (carbonyl cyanide *m*-chlorophenylhydrazone) which destroys the pH and ion gradients was added to cells only expressing *arsB*, arsenite was immediately taken up by the cells. This indicated that a proton motive force is necessary for the transport of arsenite when only the ArsB protein is present (Figure 1.1).

These results indicated that ArsB on its own is significant for arsenite resistance. The question of why the *arsA* gene exists was then raised. Dey and Rosen (1995) hypothesized that the ArsAB complex is a more effective resistance mechanism than ArsB alone. Under conditions of stress, the ATP levels will drop more slowly than the membrane potential, making the ArsAB complex a much more effective mechanism in the environment where conditions can change rapidly. The specificity of this extrusion system is also of great importance. It only transports arsenite and antimonite. This specificity may be a mechanism by which the cell circumvents phosphate starvation, because if the ArsB system were able to transport arsenate, it might also transport the structurally similar phosphate oxyanion, leaving the cell depleted of phosphate (Messens *et al.*, 1999).

Researchers also wanted to determine the chemical nature of the arsenite that is being pumped through ArsB. Chen *et al.* (1996) suggested two possibilities. As^{III} or Sb^{III} may be electrophoretically transported in response to a positive exterior membrane potential. Another possibility is that As^{III} or Sb^{III} binds to the cysteine residue in the ArsB protein and is then transported as a soft metal. To investigate the role of the only cysteine in ArsB, Cys^{369} was changed to serine and alanine codons by site-directed mutagenesis (Chen *et al.*, 1996). Cells bearing both pBC101 (*arsRDBC*) and pArsA (*arsA*) showed high levels of resistance to arsenite. No change in resistance was detected when pBC101 was replaced with the mutant *arsB* $C^{369}S$ or *arsB* $C^{369}A$. Thus, Cys^{369} is not required for arsenite resistance, indicating that the transport of arsenite by ArsB alone does not involve metal thiol chemistry. The most likely alternative way of transport is suggested to be electrophoretic anion transport. These weaker interactions assist with easy release of As^{III} or Sb^{III} extracellularly (Rosen, 1999). This was confirmed when the uptake of labeled arsenite by everted membrane vesicles expressing only *arsB* was coupled to membrane potential (Kuroda *et al.*, 1997).

1.4.2. ArsC (Arsenate Reductase)

In the primordial anaerobic environment, arsenite would have been the major chemical species. This led to the evolution of transporters such as ArsB that could extrude arsenite out of the cell. As the atmosphere became more oxidizing, arsenate would have replaced arsenite, as arsenate is the thermodynamically favorable form under aerobic conditions. A mechanism of arsenate resistance is of great importance for survival. Proteins evolved that were able to reduce arsenate to arsenite, which could then be pumped out of the cell by existing transport systems, such as ArsB (Rosen, 1999; Jackson and Dugas, 2003).

Chen *et al.* (1986) identified a 16 kDa protein on plasmid R773 of *E. coli* that was essential for resistance to arsenate, but not arsenite. It was first thought that ArsC functioned as an intracellular substrate-binding protein, making arsenate accessible to the ArsAB membrane complex, which would then function as an arsenate efflux system. Ji and Silver (1992) demonstrated that this hypothesis was incorrect. They showed that

arsenite was exported from a cell when arsenate was added in the presence of ArsC from *Staphylococcus aureus* plasmid pI258. They concluded that the *S. aureus* ArsC protein (131 aa) functions as an arsenate reductase, converting intracellular arsenate to arsenite, which could then be extruded out of the cell by ArsB. They also found arsenate reductase activity in the cytoplasm. An important question was the mechanism of energy coupling for the reduction of arsenate to arsenite. Purified ArsC protein coupled *in vitro* with thioredoxin and dithiothreitol (a nonbiological thiol compound) reduced arsenate to arsenite. Reduced glutathione or 2-mercaptoethanol was unable to reduce arsenate to arsenite. Thioredoxin can regenerate reduced cysteine residues on intracellular enzymes. The involvement of cysteine residues found in the sequence of ArsC in the reduction of arsenate to arsenite was suggested and that thioredoxin is necessary to keep the cysteine residues in a reduced state (Ji and Silver, 1992).

The ArsC protein (141 aa) of *E. coli* plasmid R773 also functions as an arsenate reductase, even though it exhibits less that 20% identity to the ArsC of *S. aureus* plasmid pI258 (Oden *et al.*, 1994). Mutations in *gshA* (encodes γ-glutamylcysteine) and *gshB* (encodes glutathione synthetase), which form glutathione, showed wild-type levels of arsenite resistance when the R773 *ars* operon was present, but reduced levels of arsenate resistance. Strains with mutations in either *trxB* (thioredoxin reductase) or *trxA* (thioredoxin) had wild type resistance to arsenate and arsenite. These results indicated that unlike the ArsC of *S. aureus* that requires thioredoxin for conversion of arsenate to arsenite, the ArsC of *E. coli* requires glutathione for the detoxification of arsenate (Oden *et al.*, 1994).

Sequence homologies and amino acid sequence alignments suggested that arsenate reducing enzymes arose independently a number of times (Rosen, 1999). Recent X-ray crystallographic solutions of protein structure and reaction pathways also indicated on an independent evolvement of different arsenate reductases. Based on the latter findings, arsenate reductases are divided into three families. The first family includes *E. coli* and uses glutathione/glutaredoxin (GSH/Grx) to reduce arsenate. The second family includes *S. aureus* and *B. subtilis* and uses thioredoxin (Trx) to reduce arsenate. The third family

is the only eukaryotic family and is called the yeast Arr2p family (Mukhopadhyay *et al.*, 2002; Mukhopadhyay and Rosen, 2002). This family was not relevant to the work presented in our study and will therefore not be discussed.

The GSH/Grx clade

The main characteristic of this family of arsenate reductases is that arsenate reduction is dependent on the presence of reduced glutathione and glutaredoxin. Since these reducing equivalents derive from cysteine thiolates, Liu *et al.* (1995) investigated the possibility that cysteinyl residues within the protein are involved in catalysis. The ArsC protein of plasmid R773 has only two cysteine residues, Cys^{12} and Cys^{106}. By using site-directed mutagenesis four mutants in the Cys^{12} codon were constructed. Cells expressing all the mutant *arsC* genes were sensitive to arsenate. This indicated that Cys^{12} plays an essential role in catalysis and that it must be located at the active site of ArsC. Tsai *et al.* (1997) hypothesized that a single thiol group may be efficient for reductase activity. An enzyme complex forms by arsenylating Cys^{12}, with the interaction of glutaredoxin with the complex. The result is the transfer of electrons to reduce arsenate to arsenite.

The product of the *grxA* gene, glutaredoxin 1 (Grx1), is a glutathione-dependent dithiol hydrogen donor for enzymes such as arsenate reductases. In the search for alternate reductants of ribonucleotide reductase, two new glutaredoxins, Grx2 and Grx3 (each of which has a Cys-Pro-Tyr-Cys dithiol consensus sequence) were identified. By studying the efficiency of these glutaredoxins to serve as a hydrogen donor, Grx2 was identified as the predominant glutaredoxin in *E. coli* cells (Shi *et al.*, 1999). The general function of Grx is that it can either catalyze intraprotein disulfide bond reduction or catalyze the reduction of mixed disulfides between a Cys thiol and glutathione (Bushweller *et al.*, 1992). All three glutaredoxins of *E. coli* have two cysteine residues in their active site. It has been shown that the N-terminal cysteine is required for both protein disulfide reduction and reduction of mixed protein-glutathione disulfides. To determine the function of the glutaredoxins in arsenate reduction, single cysteine mutants of all three *E. coli* glutaredoxins were constructed. Mutants lacking the C-terminal cysteine had no effect on ArsC activity. When the N-terminal cysteine was substituted with serine,

neither Grx1, Grx2 or Grx3 could serve as a hydrogen donor for arsenate reduction. This was consistent with a reaction cycle where the ArsC forms a mixed disulfide with glutathione. The role of glutaredoxin is thus to regenerate the reduced arsenate reductase by reducing the mixed disulfide (Shi *et al.*, 1999).

Martin *et al.* (2001) reported X-ray crystal structures for three forms of ArsC, without bound arsenic and crystals complexed with arsenate and arsenite respectively. No relationship was found between the tertiary structures of *E. coli* R773 (GSH/Grx clade) and the structures of *S. aureus* and *B. subtilis* (Trx clade). While the Trx clade has a core of four parallel β-sheet regions, the GSH/Grx clade has one antiparallel β-sheet segment. The Trx clade ArsC also exhibits a low rate of phosphatase activity, while the ArsC of the GSH/Grx clade does not exhibit any phosphatase activity (Zegers *et al.*, 2001)

For arsenate reduction to occur in *E. coli*, ArsC must form an active quaternary complex with GSH, arsenate and Grx simultaneously (Liu and Rosen, 1997). The catalytic Cys^{12} in the active site of ArsC is surrounded by a group of five basic residues, His^{8}, Arg^{16}, Arg^{60}, Arg^{94} and Arg^{107}. Together these five residues lower the pKa value of Cys^{12} to 6.4 (Gladysheva *et al.*, 1996; Martin *et al.*, 2001). Three of these basic residues, Arg^{60}, Arg^{94} and Arg^{107} interact directly with the arsenate and arsenite intermediates. Arg^{60} and Arg^{94} temporarily exchange places in the presence of arsenate to enhance hydrogen-bonding and stabilization of the intermediate. Arg^{107} remains stationary and binds the O1 oxygen attached to arsenic throughout the reduction reaction (Shi *et al.*, 2003). The role of Arg^{60} in product formation was further evaluated by mutagenesis. Crystal structures of an ArsC protein with a substitution for Arg^{60} equilibrated with arsenite revealed that Arg^{60} plays an important role in the stability of the bound arsenite product (DeMel *et al.*, 2004).

Based on previous research and crystal structures by Martin *et al.* (2001) and DeMel *et al.* (2004) the following reaction mechanism is proposed (Figure 1.6):

Step1 involves the nucleophilic attack by Cys^{12} on an arsenate that is noncovalently bound at the sulfate ion in the active site, followed by the release of OH^-. The result is the formation of a thioarsenate binary adduct of ArsC (Intermediate I). The existence of

this adduct was demonstrated by a difference in electron density. The close proximity of Arg^{60}, Arg^{94} and Arg^{107} plays a very important role in the binding of arsenate. Upon reaction of arsenate with ArsC, Arg^{60} and Arg^{94} move to new orientations in the protein that allow both side chains to bind to the arsenate adduct.

Step 2 involves the nucleophilic attack of glutathione on the arsenate adduct, with the release of water. The result is the formation of a {ArsC Cys^{12}}S-As-S{glutathione} tertiary complex (Intermediate II). Glutathione only reacts after arsenate binds to the active site. The reaction also requires a free thiol on glutathione and ArsC to proceed. The structure of intermediate II has not yet been obtained.

Step 3 involves the binding of glutaredoxin to intermediate II, with the reduction of arsenate, producing a dihydroxy arsenite intermediate (Intermediate III). A mixed disulfide (GrxS-SG) is released that would be recycled by glutathione reductase using a second equivalent of GSH. This intermediate can only be observed in mutant structures because it is not stable in the native enzyme.

In step 4 a monohydroxy, positively charged arsenite adduct is formed (Intermediate IV). This arsenite adduct has an unusual structure because it has only two atoms linked to the arsenite atom. This thiarsahydroxyl complex is much more unstable that most cysteine-arsenite complexes. It is suggested that this unstable conformation is necessary to ensure that arsenite does not function as an inhibitor of ArsC.

In step 5 arsenite is released upon the addition of a free OH^-. ArsC now returns to its original conformation.

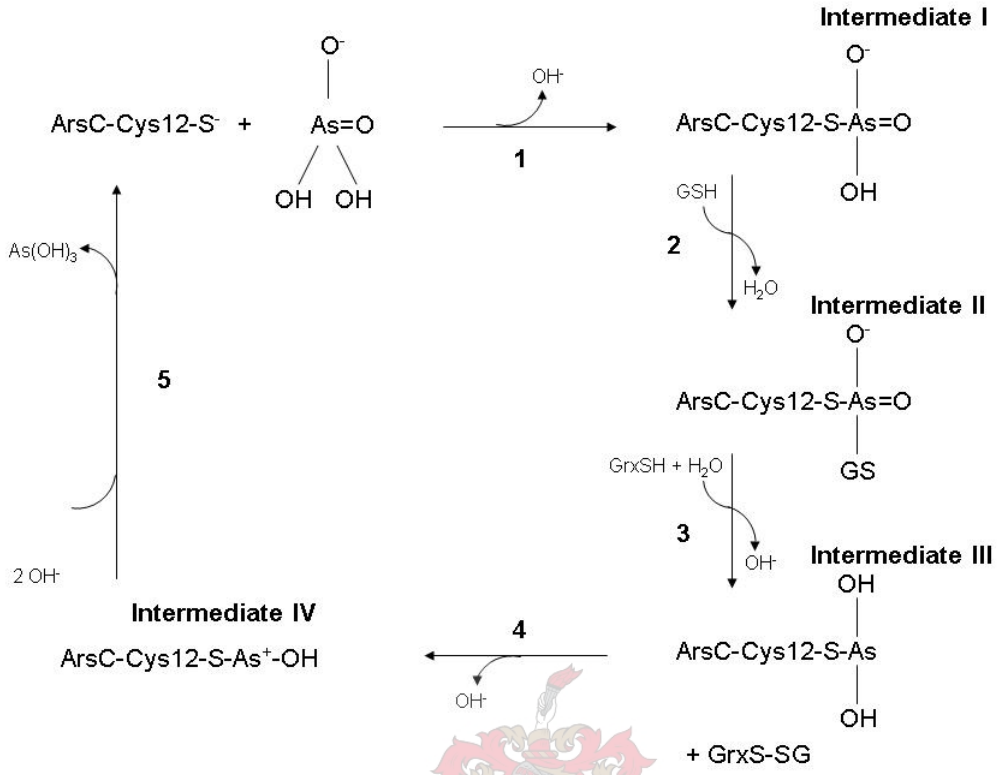


Figure 1.6: The proposed reaction mechanism of the *E. coli* R773 ArsC arsenate reductase (taken from DeMel *et al.*, 2004). See text for details.

The *ars* operon of *Synechocystis* sp. Strain PCC 6803 contains an ArsC homolog that exhibits 36 and 45% sequence identity to *S. aureus* and *B. subtilis*, respectively. This ArsC (SynArsC) exhibited weak phosphohydrolase activity, which is indicative of its descent from the LMW PTPase (low molecular weight protein tyrosine phosphatase) family of protein phosphatases. Like the ArsC of pI258, SynArsC contains three essential cysteine residues. All of the above indicate that SynArsC should belong to the Trx clade of arsenate reductases. SynArsC however, employs glutathione and glutaredoxin instead of thioredoxin as the source of reducing equivalents, making it part of the GSH/Grx clade of arsenate reductases. It appears that SynArsC and pI258 represent different branches in the evolution of their shared phosphohydrolytic ancestor (Li *et al.*, 2003)

The catalytic mechanism of SynArsC combines certain features from its *S. aureus* homolog with others from the *E. coli* R773 ArsC. Like R773, SynArsC forms a complex with glutathione. Both the active site cysteine residue and arsenate were needed for complex formation. Glutaredoxin also restored SynArsC to its reduced, catalytically ready state. A major difference between the mechanism of SynArsC and R773 is that the latter requires only a single active-site cysteine, while SynArsC needed two more cysteine residues (Cys^{80} and Cys^{82}) for catalysis. These two cysteines are similarly positioned to cysteine residues that play a key role in the reaction mechanism of pI258. Li *et al.* (2003) proposed a possible reaction scheme for SynArsC:

Step 1-3 is the same as for R773. An arsenoenzyme intermediate is formed that is attacked by glutathione to form a cysteinyl-arsenoglutathione tertiary complex. This complex is then broken down to yield arsenite by the binding of glutaredoxin. Step 4-6 is a modified version of the reaction mechanism of pI258 ArsC.

In step 4 a reaction takes place between the mixed disulfide bond of Cys^{8} and glutathione and the internal disulfide between Cys^{80} and Cys^{82}. The result is a shift of the mixed disulfide with glutathione to the surface of the enzyme.

In step 5 glutaredoxin reduces the mixed disulfide.

Step 6 involves the restoration of the active site cysteine to its reduced, reactive state. The original disulfide bond between Cys^{80} and Cys^{82} is also restored.

The Trx Clade

The most important characteristic of this family of arsenate reductases is that it requires the coupling to thioredoxin (Trx), thioredoxin reductase (TR) and NADPH to be enzymatically active (Ji *et al.*, 1994). The ArsC of *S. aureus* pI258 has a characteristic PTPase fold. It consists of a four-stranded parallel β-sheet and three α-helices. Unlike the R773 ArsC, the ArsC of pI258 shows phosphatase activity with the model substrate *p*-nitrophenyl phosphate (Zegers *et al.*, 2001). The pI258 ArsC sequence revealed that it has four cysteine residues, Cys^{10}, Cys^{15}, Cys^{82} and Cys^{89}. Cys^{82} and Cys^{89} are not found in the R773 ArsC. Messens *et al.* (1999) found that a substitution for any of Cys^{10}, Cys^{82} or Cys^{89} leads to an inactive enzyme. Cys^{10} is similarly located to Cys^{12} of *E. coli* R773 ArsC. Cys^{10} also functions as the reaction site of the arsenate reductase. By a series of

endoproteinase digestions after catalysis and peptide products identified by mass spectroscopy analysis, a Cys^{82}-Cys^{89} oxidized cysteine was assigned to be the redox couple responsible for electron donation to arsenate.

Zegers *et al.* (2001) determined the crystal structures of the reduced and oxidized forms of the *S. aureus* pI258 arsenate reductase. The secondary and tertiary structures showed great homology to those of LMW-PTPases from mammals, indicating that the pI258 ArsC may have evolved from LMW-PTPases. The homology between pI258 ArsC and LMW-PTPases is concentrated mainly in the catalytic motif of tyrosine phosphatases (P-loop), which has an identical CXGNXCR signature. The pI258 ArsC sequence has very little homology to R773, except for a common Cys-X_5-Arg anion binding motif.

Messens *et al.* (2002) provided further X-ray crystal structures for the pI258 arsenate reductase, providing evidence that Cys^{82} nucleophilically attacks Cys^{10} to form a Cys^{10}-Cys^{82} disulfide. This provides electrons to reduce arsenate to arsenite. The overall structure of the reduced (active) and oxidized (Cys^{82}-Cys^{89}, inactive) forms are basically the same, except for a major change in conformation, with Cys^{89} moving towards Cys^{82} to form a disulfide bond. In addition to the three essential cysteine residues, Arg^{16} in the P-loop was also essential for catalysis as a substitution for Arg^{16} resulted in a loss of ArsC activity. Mutational studies by Lah *et al.* (2003) revealed that potassium had a definite stabilizing effect on pI258 arsenate reductase.

Messens *et al.* (2004) also studied the mechanism by which thioredoxin regenerates the pI258 ArsC to its active form. A disulfide bridge between Cys^{10} and Cys^{15} in the P-loop is formed under oxidative stress. This Cys^{10}-Cys^{15} bridge is buried in the tertiary structure and is thus not accessible to thioredoxin. It is hypothesized that the protection of this bond may be a way to protect Cys^{10} against irreversible oxidation, such as the oxidation to sulfenic and sulfinic acids. Formation of a disulfide bridge between Cys^{10} and Cys^{15} is thus responsible for redox regulation. To reduce this buried Cys^{10}-Cys^{15} disulfide bridge, thioredoxin first reduces the surface exposed Cys^{82}-Cys^{89} disulfide bridge. Released Cys^{82} then attacks Cys^{10} in the buried disulfide bridge, forming a Cys^{10}-

Cys^{82} disulfide intermediate. Cys^{89} then attacks Cys^{82}, forming a Cys^{82}-Cys^{89} disulfide bridge, releasing Cys^{10}. A second molecule of thioredoxin reduces the newly formed Cys^{82}-Cys^{89} disulfide bridge, leaving a fully reduced ArsC.

A third X-ray crystal structure for an arsenate reductase was reported, this time for the ArsC in the SKIN element in the chromosome of *Bacillus subtilis* (Bennet *et al.*, 2001). The *B. subtilis* SKIN element ArsC contains 139 aa residues and shares about 65% identity to the ArsC of *S. aureus*. It exhibited phosphatase activity with *p*-nitrophenyl phosphate. It also uses thioredoxin to convert arsenate to arsenite and it contains the same essential cysteine residues (Cys^{10}, Cys^{82} and Cys^{89}) and Arg^{16} as the *S. aureus* pI258 ArsC. Arg^{16} is thought to stabilize the active site and to lower the pKa values of Cys^{10}, Cys^{82} and Cys^{89} in order for the reduction of arsenate to arsenite to occur. By using high resolution NMR spectroscopy, the solution structures of both the reduced and oxidized forms of the *B. subtilis* ArsC has been determined (Guo *et al.*, 2005). Backbone dynamics demonstrated that in the reduced form the Cys^{82}-Cys^{89} segment forms a helix that is not tightly structured and undergoes substantial conformational changes in the millisecond time scale. This may facilitate the formation of covalent intermediates and the subsequent reduction of arsenate to arsenite. In the oxidized form Cys^{82}-Cys^{89} shows flexibility on a sub-nanosecond time scale, which is favorable for interaction with thioredoxin, which regenerates the ArsC for the next cycle of arsenate reduction.

Based on work done by Zegers, Messens and Bennet a reaction mechanism for the *S. aureus* and *B. subtilis* ArsC is proposed (Figure 1.7).

Step 1: The reaction starts with ArsC in the reduced state. Arsenate triggers a nucleophilic attack of the thiol of Cys^{10}. A covalent Cys^{10}-$HAsO_3^-$ intermediate is formed with the release of a hydroxyl that is protonated by the acidic Asp^{105} to form a water molecule.

Step 2 involves the three essential cysteine residues in a triple redox relay system. Firstly, the thiol of Cys^{82} attacks Cys^{10} to form a Cys^{10}-Cys^{82} disulfide bond. Cys^{10} donates its electron pair to the arsenate, reducing it to arsenite, which then dissociates from the enzyme. Secondly, Cys^{89} attacks Cys^{82}, forming a Cys^{82}-Cys^{89} disulfide bond,

releasing Cys^{10}. The disulfide bridge formed by Cys^{82}-Cys^{89} becomes exposed to the enzyme surface. The positively charged Arg^{16} plays a very important role in stabilizing the P-loop and lowering the pKa values of the three cysteines involved to activate them for the above mentioned reactions.

In Step 3 thioredoxin reduces the exposed Cys^{82}-Cys^{89} disulfide, regenerating ArsC for the next cycle of reduction.

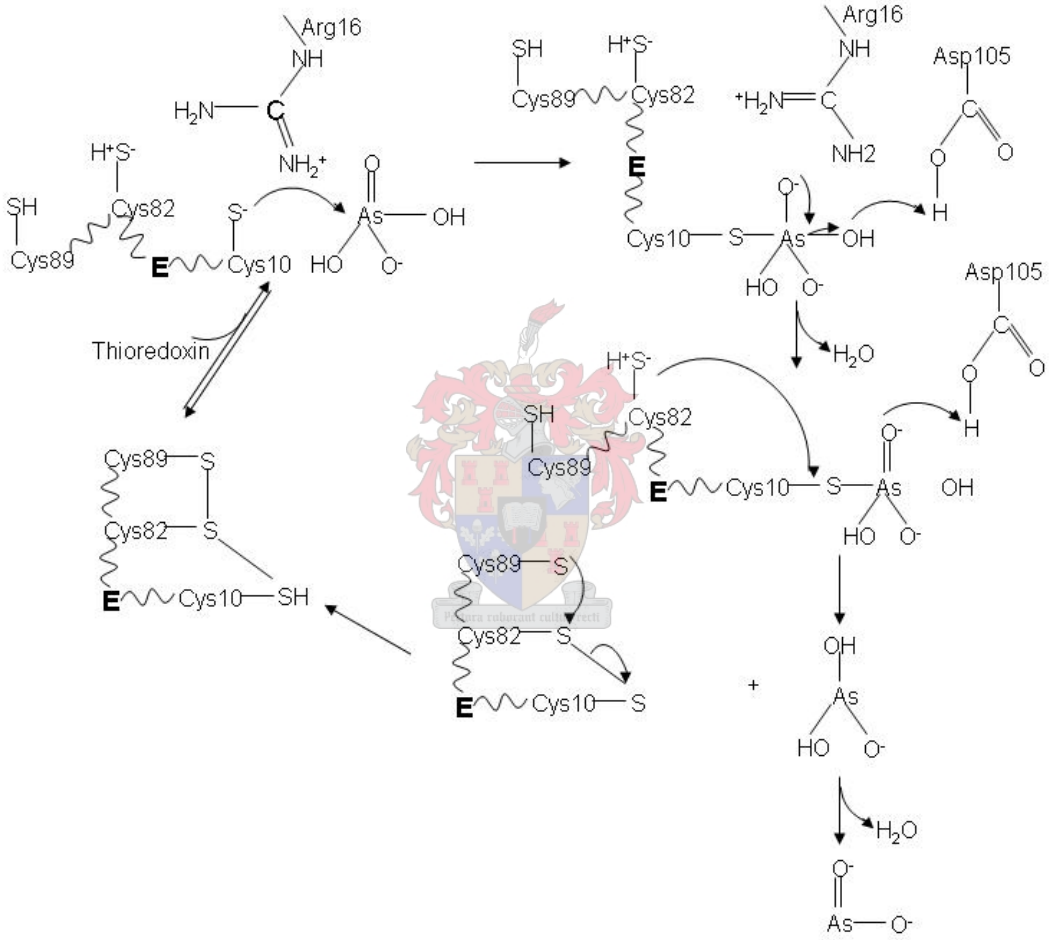


Figure 1.7: The proposed reaction mechanism of the *B. subtilis* ArsC arsenate reductase. The notation H^+S^- refers to the sulfhydryl group of a cysteine residue that will be protonated when placed in close proximity of a NH_2^+ group. The notation E refers to the ArsC enzyme. Adapted from Bennet *et al.* (2001). See text for details.

1.4.3. Regulation of the *ars* operon

1.4.3.1. ArsR

The small 13 kDa ArsR protein from the *E. coli* plasmid R773 has been intensively studied. The *arsR* gene encodes a trans-acting repressor that binds as a dimer to an operator and represses transcription. Wu and Rosen (1991) showed that the expression of this ArsR protein was autoregulated and that this expression was repressed in the absence of inducers (As^{III} and Sb^{III}). In the presence of inducers the expression was induced. Although the ArsR has been shown to be a trans-acting regulator, the *arsR* sequence contains no obvious DNA binding motif. Further studies by Wu and Rosen (1993a) led to the finding of the ArsR protein binding site. Using DNase I footprinting assays they found that the ArsR repressor binds to the DNA just upstream (from -64 to -40) of the -35 site. The ArsR would thus interfere with transcriptional initiation. The ArsR binding site contains an imperfectly symmetrical dyad that may be important for repressor recognition and binding. Hydroxyl radical footprinting revealed two small 4 bp regions of protection. Ten bp separated these two regions, indicating that the repressor only binds to one side of the DNA. The two short protected areas also indicated that the active form of ArsR must be a dimer. Gel retardation assays demonstrated the formation of DNA-ArsR protein complexes, with the release of the ArsR protein on interaction with As^{III} or Sb^{III} (Wu and Rosen, 1993a).

The ArsR protein belong to a new family of metalloregulatory proteins called the SmtB/ArsR family. This family functions exclusively as transcriptional repressors and includes the *Synechococcus* sp. SmtB repressor (activated by zinc) that represses the *smtA* gene (encodes for metallothionein), the *E. coli* ArsR repressor and the *S. aureus* pI258 CadC repressor (activated by cadmium). Cook *et al.* (1998) determined the crystal structure of the SmtB metallothionein repressor. It is suggested that the ArsR proteins will have a similar structure to SmtB. Bairoch (1993) proposed that the *E. coli* ArsR contains a DNA-binding helix-turn-helix structure at residues 32-52, containing two conserved cysteine residues. Sequence alignments of members belonging to the SmtB/ArsR family of regulators revealed a highly conserved ELCVCDL region (Shi *et al.*, 1994). However, not all the members of this family contain this consensus sequence,

e.g. the ArsR of *At. ferrooxidans* (Butcher and Rawlings, 2002). To identify the inducer binding domain in ArsR, three mutants with a noninducible phenotype ($C^{32}Y$, $C^{32}F$ and $C^{34}Y$) were isolated. The mutated ArsR proteins were still able to bind to the *ars* promoter, but they had reduced inducer response. This indicated that Cys^{32} and Cys^{34} forms part of the metal binding site of ArsR and that the cysteine residues are located outside of the DNA-binding helix-turn-helix motif. Binding of a metal induces a conformational change in the helix-turn-helix domain, resulting in the dissociation of the repressor from the DNA. A His^{50} mutation ($H^{50}Y$) located in the helix-turn-helix motif resulted in constitutive expression, suggesting a decrease in binding of the ArsR protein to DNA (Shi *et al.*, 1994).

In most molecule thiolates-As^{III} complexes, As^{III} is three-coordinated, therefore Shi *et al.* (1996) proposed that a third arsenic ligand (only Cys^{32} and Cys^{34} had been identified) may be involved in metal binding. In addition to Cys^{32} and Cys^{34}, the ArsR contains three more cysteine residues, Cys^{37}, Cys^{108} and Cys^{116}. Cys^{108} and Cys^{116} have been shown not to be involved in repressor function (Wu and Rosen, 1991). When Cys^{37} was altered to alanine, it had no effect on regulation or DNA binding. X-ray absorption spectroscopy of the As^{III}-ArsR complex showed that As^{III} is coordinated to three sulfur atoms. Shi *et al.* (1996) investigated the possibility that only two coordinates are needed for metal binding. Phenylarsine oxide (PAO) was used as it can form only two coordinates, but can still act as a good inducer. Interaction with PAO can be measured using matrix-bound PAO. ArsR proteins in which only Cy^{32} or Cys^{34} were altered to glycine bound strongly to the PAO resin, but a $C^{32}G/C^{34}G$ double mutant was unable to bind to POA. If binding requires two groups, an additional ligand must have been present when only Cys^{32} or Cys^{34} was involved. It was suggested that binding requires two cysteine residues, and that any combination of Cys^{32}, Cys^{34} and Cys^{37} will produce a conformational change that results in the release of the ArsR repressor from the DNA. Cys^{37} is thus dispensable for function, but participates in binding when present.

The chromosome of *E. coli* also contains an *arsRBC* operon (Carlin *et al.*, 1995). Xu *et al.* (1996) further studied the *E. coli* chromosomal ArsR. The chromosomal and R773

ArsRs exhibit 75% sequence similarity and therefore, as expected, most characteristics were shown to be the same. The chromosomal ArsR was able to regulate expression of the *ars* operon, functioned as a homodimer and bound to promoter DNA. Induction also resulted in the dissociation of the repressor from the DNA. Analysis of the ArsR binding site revealed that both the chromosomal and R773 *ars* sequence contained a TCAT and a TTTG region separated by 7 bp. It was thus possible that the two regulators could bind to each other's promoter. It was found that the chromosomal repressor and the repressor from R773 were interchangeable, as the chromosomally-encoded repressor was able to regulate expression of the R773 *ars* operon, and the R773 repressor cross-regulated expression of the chromosomal *ars* operon (Xu *et al.*, 1996).

It was proposed that members of the ArsR/SmtB family of repressor proteins have at least three domains, namely a metal binding domain, a DNA binding domain and a dimerization domain. While the first two domains have been identified, no evidence could locate the dimerization domain to a specific region in ArsR. Xu and Rosen (1997) investigated the dimerization properties of the *E. coli* chromosomal ArsR by using the yeast two-hybrid system. Results indicated that residues 1-8 and 90-117 are not required for the dimerization of ArsR. *β*-galactosidase assays and gel mobility shift assays showed that dimers, but not monomeric ArsRs were able to bind to the *ars* promoter. This led to the conclusion that a core of approximately 80 residues is sufficient for all the ArsR regulatory properties: metal binding, DNA binding and dimerization.

An evolutionary and structural study of six members of the SmtB/ArsR family produced new insights on the evolution of metal resistance in this family (Busenlehner *et al.*, 2003). The authors found that two distinct metal binding sites exist in this family. The first is present in the α3 helix [ELCV(C/G)D] and is termed the α3N metal binding site. The second metal binding site is present in the α5 helix and is called the α5C metal binding site. The α3N site is represented by *S. aureus* pI258 CadC, *Listeria monocytogenes* CadC and the *E. coli* ArsR, while the α5C binding site is represented by *Synechococcus* SmtB, *Synechocystis* ZiaR, *S. aureus* CzrA and *Mycobacterium tuberculosis* NmtR (Figure 1.8). The authors suggested that individual members of this

family have evolved distinct metal specificity profiles by the alteration of one or both of the α3N and α5C metal binding sites. The structures of all the metal-inducible operons contain one or two (in the case of the *smt* operon) imperfect 12-2-12 inverted repeats (Busenlehner *et al.*, 2003).

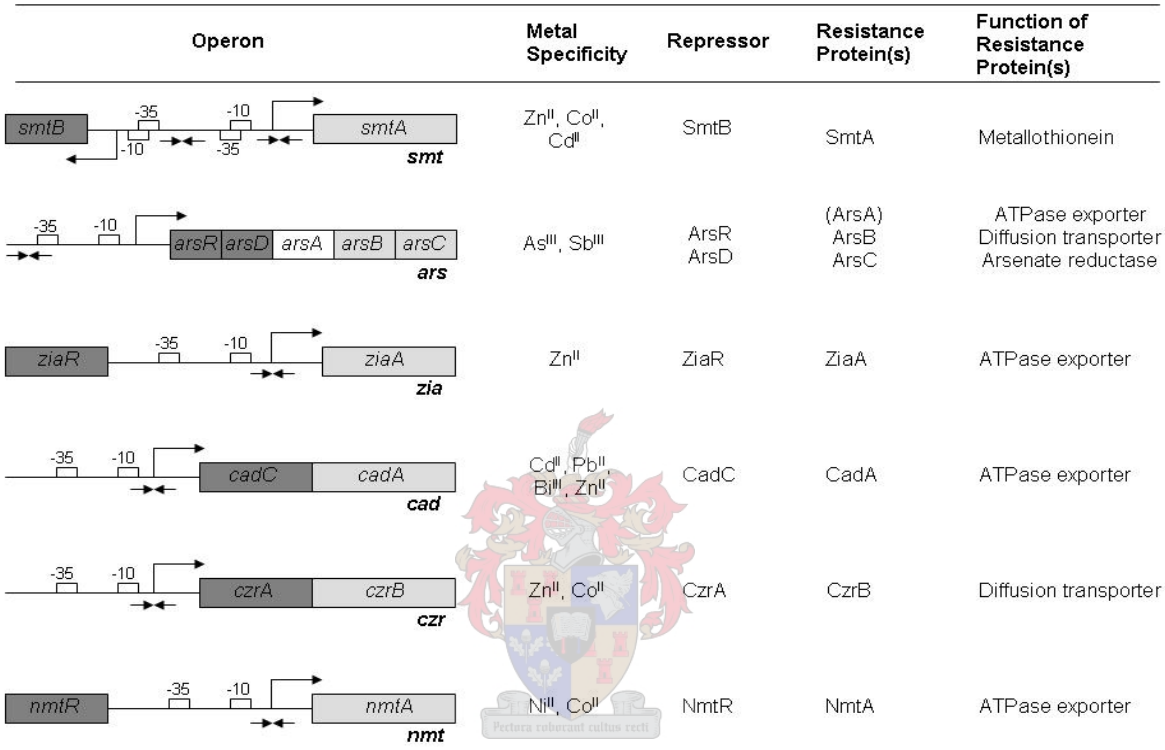


Figure 1.8: Different metal resistance operons regulated by the SmtB/ArsR family of transcriptional regulators. Dark grey boxes indicate repressor genes, while light grey boxes indicate resistance genes. Arrows indicate the positions of the imperfect 12-2-12 inverted repeats. Adapted from Busenlehner *et al.* (2003).

In an attempt to increase the bioaccumulation of arsenic, the chromosomal ArsR of *E. coli* was overexpressed (Kostal *et al.*, 2004). This overexpression resulted in elevated levels of arsenite in cells. High levels of ArsR could possibly compete with ArsB for intracellular As^{III}, thus rendering the bound As^{III} in a nontoxic form by sequestration. Another possibility is that the overproduction of ArsR may result in repression of the *ars* operon, reducing the efflux efficiency of ArsB and increasing the intracellular sequestration of As^{III}. This could be a promising alternative to chemical ligands for the treatment of contaminated water.

1.4.3.2. ArsD

A second regulatory protein, the ArsD, has been identified for the plasmid encoded R773 *ars* operon of *E. coli* (Wu and Rosen, 1993b). Like the ArsR protein, the ArsD also encodes a trans-acting regulatory protein. Although both the ArsR and ArsD proteins are 13kDa homodimers, they share no sequence similarity. A frameshift mutation introduced within the *arsD* gene resulted in elevated levels of expression of three downstream *ars* genes. When the *arsD* gene was co-expressed in trans the elevated expression of the *ars* operon was reduced to wild-type levels. It appeared that the down-regulation of the *ars* operon was dependent on the amount of ArsD present. This indicated that the *arsD* gene is required for resistance when the *arsABC* genes are controlled by the *ars* promoter. The authors suggested that a regulatory circuit is formed by ArsR and ArsD (Wu and Rosen, 1993b).

Chen and Rosen (1997) revealed that ArsD is a homodimer consisting of two monomeric 129 aa residue monomers. By using DNase I footprinting experiments it was shown that ArsD binds to the same promoter site occupied by ArsR before induction. The ArsD however, binds with two orders of magnitude lower affinity than ArsR. Repression of ArsR was relieved with 10μM arsenite, while ArsD repression could only be relieved with the addition of 100μM arsenite. This suggested that the ArsR has a higher affinity for inducer than ArsD.

ArsD has three vicinal cysteine pairs, Cys^{12}-Cys^{13}, Cys^{112}-Cys^{113} and Cys^{119}-Cys^{120}. Li *et al.* (2001) investigated the role of these cysteines in the repression of the *ars* operon. No effect on repressor or metalloid responsiveness was observed when the Cys^{119}-Cys^{120} pair were mutated or deleted. Mutations in either the Cys^{12}-Cys^{13} or the Cys^{112}-Cys^{113} pair had no effect on repression of the *ars* operon, but resulted in a loss of inducibility. This result suggested that both the Cys^{12}-Cys^{13} and Cys^{112}-Cys^{113} pair form metalloid-binding sites. Binding of both of the cysteine pairs to As^{III} or Sb^{III} is believed to be required for ArsD function. The fact that ArsD binds to the promoter as a dimer, suggests that ArsD appears to have four one-half metalloid-binding sites instead of two (Li *et al.*, 2001). Based on the above results experiments were designed to test whether these binding sites

act in tandem or independently to trigger a conformational change in ArsD in order to dissociate from the promoter (Li *et al.*, 2002). Results of tryptophan fluorescence experiments suggested cooperativity between the four binding sites of ArsD. The authors proposed that metalloid-binding sites are formed between equivalent residues of each subunit of the dimer. They also propose that the binding sites bind metalloids sequentially, starting with Cys^{113} and ending with Cys^{12}. Dissociation will occur when the concentration of metalloid bound to ArsD is sufficient.

Based on work done by Chen and Rosen (1997) a model for regulation of the *ars* operon by ArsR and ArsD was proposed (Figure 1.9): When no environmental arsenite or antimonite (inducer) is present, ArsR (which has a higher affinity than ArsD for the *ars* promoter) would bind to the operator and repress expression of the *ars* operon. ArsR also has a higher affinity for inducer than ArsD, and therefore a low level of inducer would result in the dissociation of the ArsR, resulting in the expression of the *ars* operon. Continued expression of the *ars* operon would result in increasing amounts of ArsD. When the ArsD concentration reaches a critical level it would bind to the *ars* operator site and repress expression. This is important as the overexpression of certain proteins, like ArsB, is toxic to the cell. Because ArsD has a lower affinity for inducer, a low concentration of inducer would not prevent the ArsD from binding to the operator site. If the inducer level reaches a certain concentration, ArsD would dissociate from the operator and expression would take place. These two regulators thus form a regulatory circuit with ArsR controlling the basal level of expression, while ArsD controls the upper level of expression.

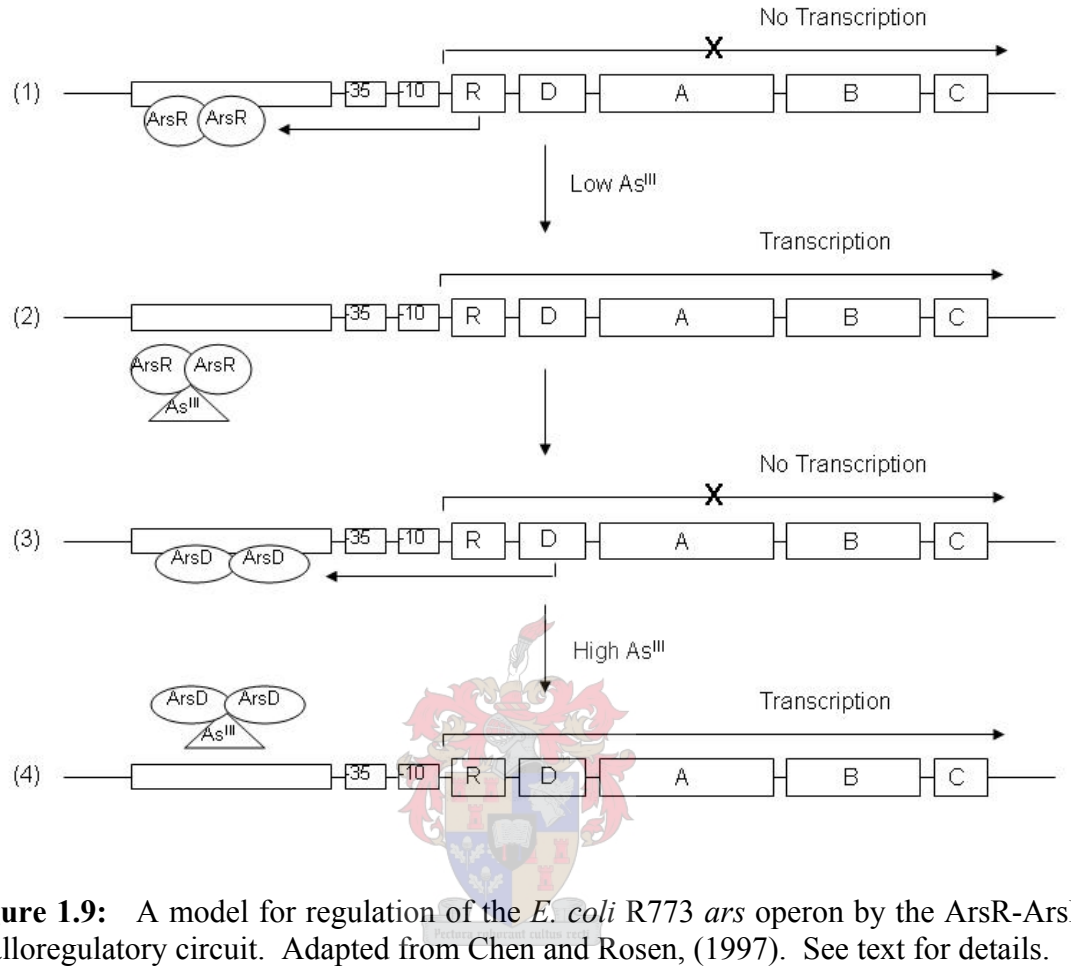


Figure 1.9: A model for regulation of the *E. coli* R773 *ars* operon by the ArsR-ArsD metalloregulatory circuit. Adapted from Chen and Rosen, (1997). See text for details.

1.5. Aim of thesis

At. caldus plays a very important role in a commercial biooxidation process called biomining. During this process various acidophilic bacteria either convert insoluble metal sulfides to water soluble metal sulphates or they break up the structure of the mineral, thus permitting other chemicals to penetrate the mineral and solubilize the desired metal. During this process of mineral solubilization, toxic arsenic compounds are leached into the environment. To survive in such an arsenic-rich environment, biomining bacteria evolved arsenic resistance mechanisms. The bacterial consortium used in the biomining process became more resistant to arsenic over a period of time. Research on

the *ars* systems of *At. caldus* revealed that *At*. *caldus* strain #6 is more resistant to arsenic than other *At. caldus* strains (Tuffin *et al.*, 2004).

Two copies of transposon-linked arsenic resistance operons have already been isolated from *At. caldus* #6 (Tuffin *et al.*, 2004) and evidence suggests that there may be a third *ars* operon. This study is aimed at isolating and characterizing the third *ars* operon *of At caldus* #6. Expression and regulatory studies will be performed on the *ars* operon and possible interaction between the three operons analyzed.

Chapter Two

Sequence analysis of the *Acidithiobacillus caldus* chromosomal *ars* operon

Contents

2.1. Introduction

Previously two copies of transposon-linked arsenic resistance operons were described for *At. caldus* strain #6 (Figure 2.1 a and b), with operon b a truncated version of operon a. These two operons were named family A and family B respectively. The *ars* genes in clone family A are located within a 12.2 kb transposon. This transposon has Tn*21*-like inverted repeats (IR) and contains a resolvase and transposase characteristic of the Tn*21*-like family. Located between the transposase and resolvase genes, arsenic resistance genes with high homology to *arsR*, *arsC*, duplicate copies of *arsD* and *arsA*, and *arsB* were identified. Two more ORFs were identified between the second copy of *arsA* and the *arsB*. ORF7 showed clear homology to the NADH-dependent dehydrogenase of *Thermoanaerobacter tencongensis*, while ORF8 showed homology to the cystathione-ß-synthase (CBS) domain of *Desulfitobacterium hafiense*. Clone family B contains a series of genes similar to those found in family A, except that only a single copy of *arsD* and *arsA* is present and no ORF8, *arsB* and *tnpA* or second IR have been reported (Tuffin *et al.*, 2004).

Probing of the DNA of *At. caldus* strain #6 with the transposon-borne *arsB* revealed a strong 15-kb signal, while a faint 2.5-kb signal was detected at low stringency conditions (Figure 2.2 a; de Groot *et al.*, 2003). The 2.5-kb signal represented a possible chromosomal *arsB* gene and the strong 15-kb signal the transposon *arsB*. Probing with the chromosomal *arsC* of *At. ferrooxidans* yielded a stronger 2.5 kb signal, and a weaker signal in the 13-15 kb region (Figure 2.2 b). These results suggested that there was a third arsenic operon in *At. caldus*. Based on the above evidence, and on the known *arsRBC* structure of most chromosomal arsenic operons like that of *E. coli*, it was thought that *At. caldus* might contain a chromosomal arsenic operon.

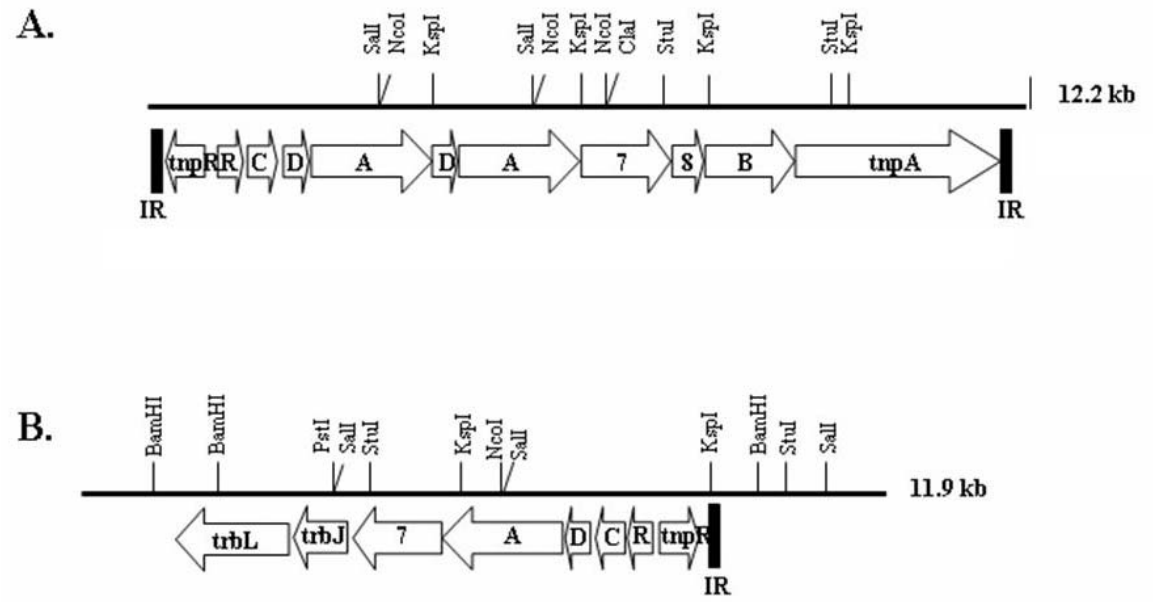


Figure 2.1: Two transposon-linked arsenic resistance operons identified in *At. caldus* strain #6. IR, inverted repeats; *tnpR* and *tnpA* are genes encoding for resolvase and transposase, respectively; R, C, D, A and B are arsenic resistance genes; 7 and 8 are ORF7 and ORF8, respectively; *trpL* and *trbJ* are homologs to conjugation genes from the Ti plasmid of *A. tumefaciens* (taken from Tuffin *et al.*, 2004).

A preliminary study involved the screening of an *At. caldus* genebank for the ability to confer arsenic resistance to the *E. coli* arsenic mutant, ACSH50I^{q}. Colony hybridizations were performed on clones that were able to confer resistance to arsenic. Plasmids that did not give a signal when probed with the transposon-borne *arsDA*, but hybridized with *arsB* and *arsC* of *At. ferrooxidans*, were believed to contain the *At. caldus* chromosomal *ars* operon. Two plasmids fitted these criteria, pTcC-#4 and pTcC-#6, and were selected for further studies. Restriction mapping of the inserts of the two plasmids revealed that pTcC-#4 contained the complete 8.6 kb insert of pTcC-#6, with an additional 1.1 kb on the one end. Analysis of partial sequences revealed homology to *arsR* and *arsB* from other *ars* operons (Figure 2.3). The sequences differed from the transposon *ars* genes that have already been isolated from *At. caldus*.

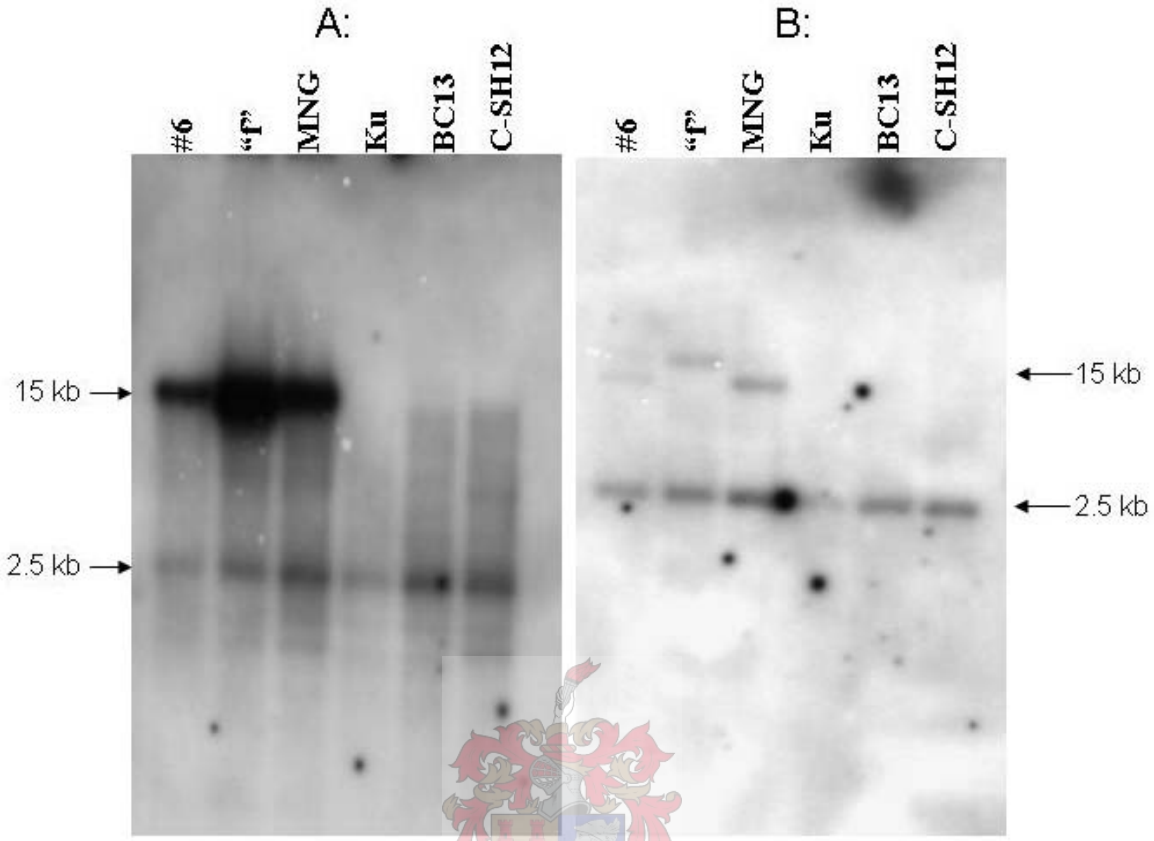


Figure 2.2: Southern hybridization of chromosomal DNA from six *At. caldus* strains, digested with *Bam*HI and probed with (A) the transposon-borne *arsB* from *At. caldus* strain #6 (less stringent conditions) and (B) the *arsC* from *At. ferrooxidans* (De Groot *et al.*, 2003).

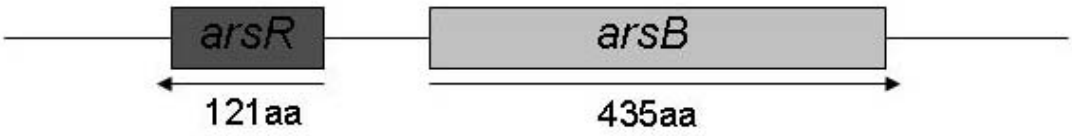


Figure 2.3: Arrangement of the *At. caldus* #6 chromosomal *ars* operon (based on preliminary results).

Here I report on the sequence analysis of the chromosomal *ars* operon of the arsenic tolerant *At. caldus* strain #6. I further show that this operon is functional in *E. coli* and confers arsenite and arsenate resistance in *E. coli* ACSH50I^{q}.

2.2. Methods

2.2.1. Media, bacterial strains and plasmids

Bacterial strains, plasmids and primers used in this study are listed in Table 2.1. *E. coli* strains were grown in Luria-Bertani (LB) broth medium (Sambrook *et al.*, 1989) with ampicillin (100 μg/ml) or kanamycin (50 μg/ml) added as required. *At. caldus* strains were grown at 37°C in tetrathionate medium (3 mM), autoclaved and adjusted to pH 2.5 (Rawlings *et al.*, 1999b)

Table 2.1: Bacterial strains, plasmids and primers used in this study

Strain, plasmid or primer	Description	Source or reference
Strains		
Acidithiobacillus caldus		
#6	Fairview mine, Barberton, S. Africa	Rawlings *et.al.* (1999b)
E. coli		
DH5α	Φ80d*lacZ*ΔM15, *endA*1, *recA*1, *gyrA*96, *thi*-1, *hsdR*17 (r_k^-, m_k^+), *relA*1, *supE*44, *deoR*,Δ(*lacZYA-argF*) U169	Promega
XL 1 Blue	*recA*1, *endA*1, *gyrA*96, *thi, hsdR*17 (r_k^-, m_k^+), *relA*1, *supE*44, *relA*1, lac [F' *proAB*+ *lacI*^{q}ZΔM15 ::Tn10(tet^r)]	Stratagene
ACSH50I^q	*rpsL*, Δ(*lac-pro*), Δ*ars*::*cam*	Butcher & Rawlings (2002)
BH5262	K-12 *araD*139?, *galU*, *galK*, *hsdR*, *rpsL*, *argH*1, *trxA*7004, *ilvC*::Tn10, *gshA*	Lim *et al.* (1986)
Plasmids		
p*Eco*R252	Ap^r, *Eco*RI inactivation cloning vector	Zabeau & Stanley (1982)
pBluescript SK	Ap^r, *LacZ*', ColE replicon, cloning vector	Stratagene
pGL10	Km^r, RK2/RP4 replicon, cloning vector	A. Toukdarian
pUCBM21	Ap^r, LacZ', ColE1 replicon, cloning vector	Boehringer-Mannheim

pTrx6	Ap^r, contains *At. ferrooxidans* ATCC 33020 *trxA* gene cloned in pBluescript	Powles *et al.* (1995)
pTcC-#4	Ap^r, 10kb *Sau*3A fragment of *At. caldus* #6 cloned into the *Bgl*II site of p*Eco*R252	This study
pTcC-#6	Ap^r, pTcC-#4 with *arsC* and *ORF1* deleted.	This study
pTcC-PH29	Ap^r, A blunted 2.9kb *Pvu*II-*Hind*III fragment from pTcC-#4 cloned into a blunted *Bgl*II site of p*Eco*R252	This study
pTcC-PH29GL	Km^r, A 2.9kb *Pvu*II-*Hin*dIII fragment from pTcC-#4 cloned into a *Hin*dIII-*Sma*I site of pGL10	This study
pEcoBlunt	Ap^r, p*Eco*R252 blunted at the *Bgl*II site and relegated to inactivate the *Eco*RI endonuclease	Tuffin *et al.* (2005)
pTcC-N1.3	Ap^r, A partial 1.3kb *Nco*I fragment from pTcC-SH3.9 cloned into pUCBM21	This study
pTcC-Bs2.6	Ap^r, A blunted 2.6kb *Bst*EII fragment from pTcC-#4 cloned into a blunted *Bgl*II site of p*Eco*R252	This study
pTcC-RB+5	Ap^r, pTcC-#6 cut with *Bst*EII-*Sph*I, blunted and ligated	This study
pTcC-RB-5	Ap^r, pTcC-#6 cut with *Pvu*II-*Sph*I, blunted and ligated	This study

2.2.2. DNA isolation and manipulations

Small-scale plasmid preparation, restriction endonuclease digestion, ligation, gel electrophoresis and Southern blot hybridization were performed using standard methods (Sambrook *et al.*, 1989). Large-scale preparation of plasmid DNA was done with Nucleobond AX100 (Macherey-Nagel) or by using the alkaline lysis method, followed by CsCl gradient centrifugation. Sequencing was done by the dideoxy chain termination method, using an ABI PRISM^{TM} 377 automated DNA sequencer. The sequence was analyzed using a variety of software programmes, but mainly by DNAMAN (version 4.1; Lynnon Biosoft). Comparison searches were performed using the gapped-BLAST program of the National Center for Biotechnology Information (Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST). The phylogenetic tree was constructed using the Multiple Sequence Alignment tool in DNAMAN.

2.2.3. Arsenic resistance assays

To test for arsenic resistance in *E. coli*, constructs were transformed into competent *E. coli* ACSH50I^{q} cells. Growth assays to determine resistance to arsenite (As^{III}) were carried out in Luria Bertani (LB) medium. Growth assays to determine the resistance to arsenate (As^{V}) were performed in low-phosphate medium (Oden *et al.*, 1994), supplemented with 2 mM K_2HPO_4. Overnight cultures were diluted 100-fold into fresh medium containing the appropriate antibiotics, 0.25 mM sodium arsenite and 0.5 mM sodium arsenate. Cultures were incubated at 37^{o}C and absorbance reading (600 nm) determined every hour until stationary phase.

2.2.4. Analysis of the arsenate reductase mechanism

To test for arsenate resistance in the *E. coli*, pTcC-PH29GL (contains the *At. caldus* chromosomal *ars* genes cloned in pGL10) and pTrx6 (contains the *At. ferrooxidans trxA* gene) were transformed in the *E. coli* BH5262 double mutant (which has mutations in the thioredoxin gene and the γ-glutamylcysteinyl gene). As controls pGL10 was transformed into *E. coli* BH5262 with co-resident plasmids pTrx6 and pKS respectively, and pTcC-PH29GL in *E. coli* BH5262 with co-resident pKS (Table 2.3). To determine resistance to arsenate, cells were plated on Luria agar containing 0, 1 and 2 mM sodium arsenate and were incubated overnight at 37^{0}C.

2.3. Results

2.3.1. Cloning of construct pTcC-#4

The inserts of the two plasmids, pTcC-#4 and pTcC-#6, that were isolated as described in the introduction, were mapped and revealed that pTcC-#4 contained the complete 8.6 kb insert of pTcC-#6, with an additional 1.1 kb on one end. The source of the pTcC-#6 insert was confirmed by Southern hybridization. An internal 1.3 kb *Nco*I fragment (Figure 2.4) was labeled and used to probe the chromosomal DNA of *At. caldus* strain #6 and pTcC-#6 plasmid DNA digested with *Stu*I, *Bam*HI, *Nco*I and *Sal*I. The following fragments were obtained for both the chromosomal DNA of *At. caldus* strain #6 and

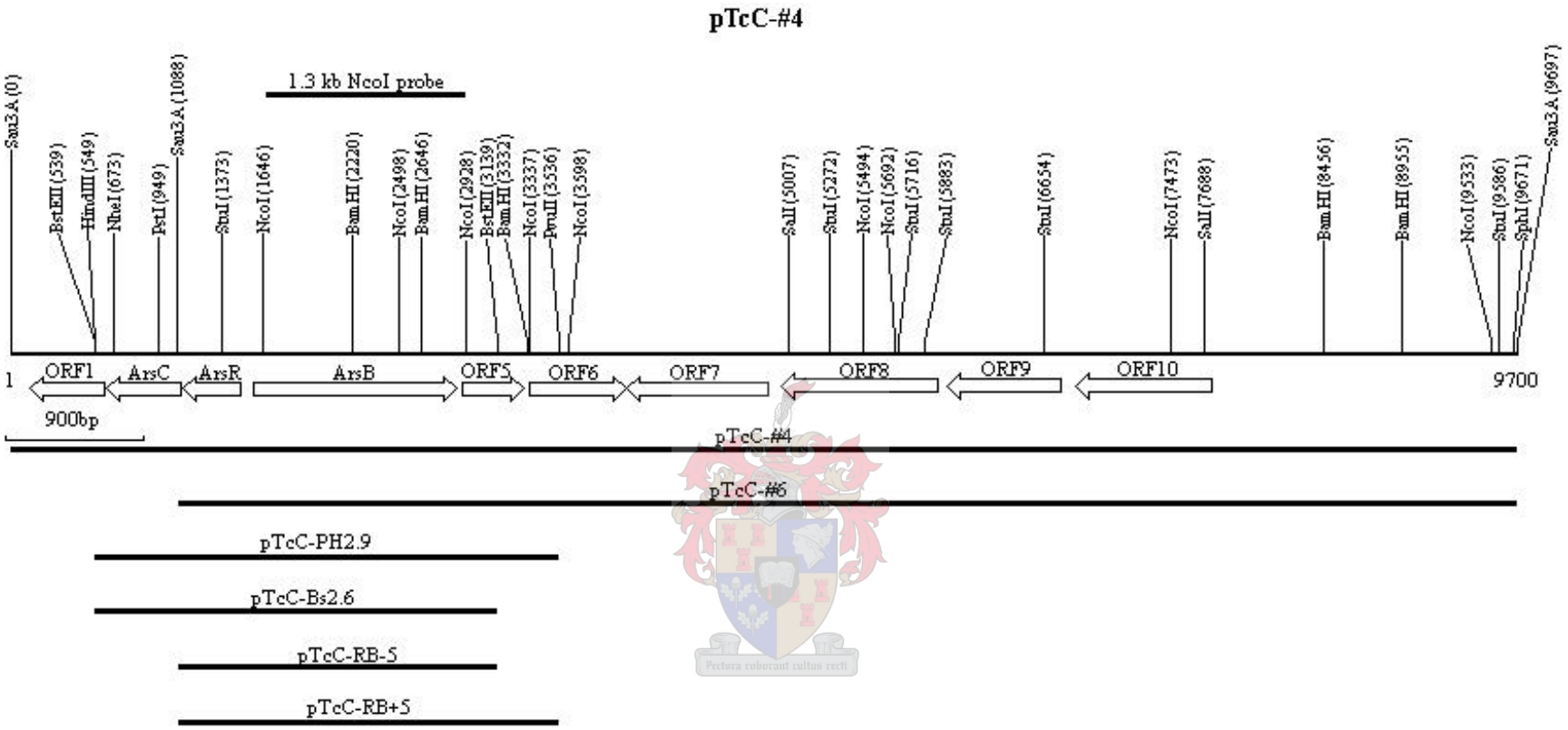


Figure 2.4: Diagram showing the restriction endonuclease map of pTcC-#4 and clones constructed in this study. Also shown are the position of the 1.3 kb *Nco*I probe used to determine the origin of pTcC-#6 and the ten ORFs identified.

pTcC-#6: a 0.4 kb and 0.7 kb *Bam*HI fragment, a 4.2 kb and 4.7 kb *Stu*I fragment (the reason why the 4.2 kb and the 4.7 kb bands are not identical is that the 4.7 kb *Bam*HI fragment contains part of the vector), a 0.85 kb and 0.4 kb *Nco*I fragment and a 9.3 kb *Sal*I fragment (Figure 2.5). This indicated that the insert DNA originated from *At. caldus* strain #6 was present in a single copy and that no rearrangements in the region of the probe and restriction fragments tested had occurred during cloning.

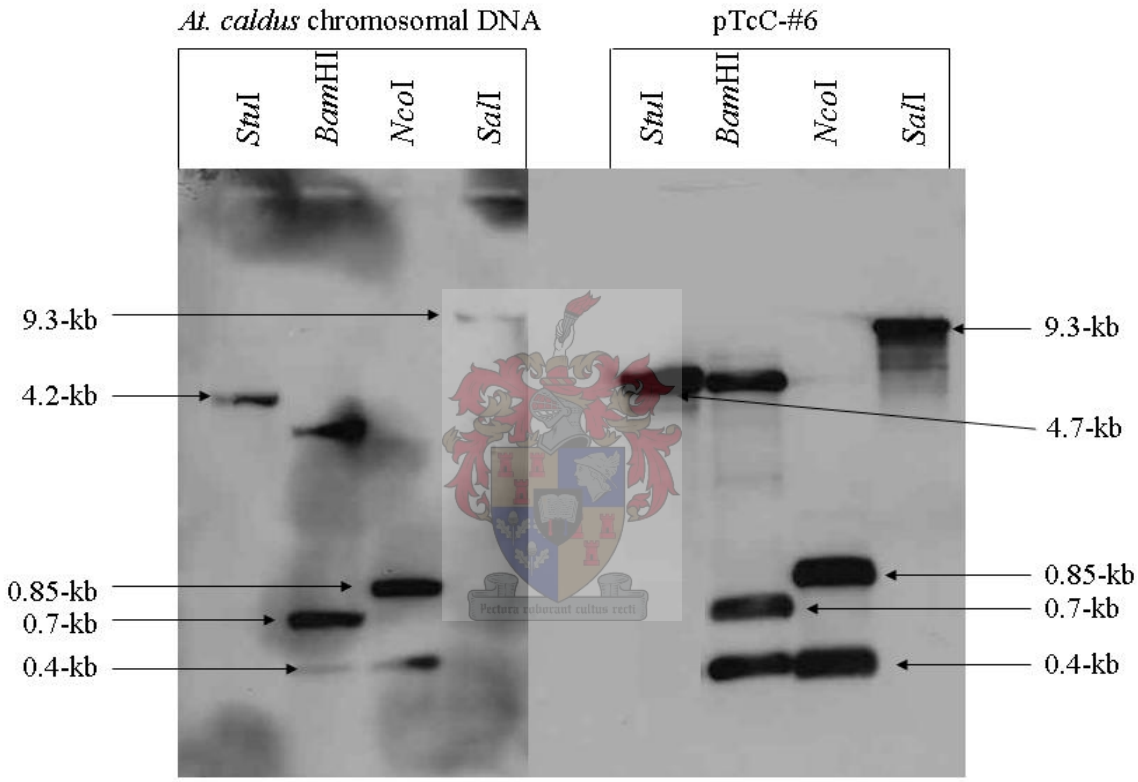


Figure 2.5: Southern hybridization analysis of *At. caldus* chromosomal DNA and pTcC-#6 plasmid DNA probed with an 1.3 kb internal *arsB* fragment and digested with *Stu*I, *Bam*HI, *Nco*I and *Sal*I.

2.3.2. Sequence analysis of pTcC-#4

The entire insert of pTcC-#4 was sequenced in both directions and ten open reading frames (ORFs) were identified based on sequence similarity with known genes in the NCBI database. The results and characteristics of the predicted products of the ten ORFs are shown in Table 2.2 and Figure 2.4. Homologs to the *arsB* and *arsC* genes of other

bacteria were identified. Between the *arsB* and *arsC* genes, a third ORF was identified that showed homology to many transcriptional regulators of the ArsR family. This *ars* operon however, differs from conventional chromosomal *arsRBC* operons like those of *E. coli* and *Staphylococcus* spp. Unlike other systems where the *arsC* gene is located downstream of the *arsB* gene, the *At. caldus* chromosomal *arsC* gene is located upstream of the *arsB* gene. The *arsC* and *arsR* genes are divergently transcribed to the *arsB* in a manner previously reported only for *At. ferrooxidans* (Butcher and Rawlings, 2000).

Table 2.2: Location and of ORFs identified in pTcC-#4

ORF	Position	No. of aa	Most related protein and predicted size	%identity/ similarity (part of protein)	Reference NCBI accession number
ORF1	604-131	157	Lactoylglutathione lyase [*Thermobifida fusca*] 143aa	34/52 (63)	ZP_00291564
			Cadmium inducible protein CadI [*Mycobacterium tuberculosis*] 152aa	24/39 (116)	NP_217157
arsC	1106-615	163	arsenate reductase [*Acidithiobacillus ferrooxidans*] 163aa	78/87 (159)	AAF69239
arsR	1477-1112	121	ArsR-like protein [*Acidithiobacillus ferrooxidans*] 118aa	74/84 (95)	AAF69241
arsB	1567-2874	435	arsenical membrane pump [*Acidithiobacillus ferrooxidans*] 436aa	85/90 (418)	AAF69238
ORF5	2911-3303	130	Putative integrase for prophage CP-933R [*Shigella flexneri*] 201aa	58/70 (96)	NP_837464
ORF6	3339-39671	210	Predicted nucleotidyltransferase [*Dechloromonas aromatica* RCB] 196aa	35/52 (176)	ZP_00203765

ORF7	4882-3962	306	Permeases of the drug/metabolite transporter (DMT) superfamily [*Methylobacillus flagellatus KT*] 305aa	36/49 (280)	ZP_00172424
ORF8	5975-4962	337	NADH:flavin oxidoreductases [*Geobacter metallireducens* GS-15] 355aa	63/76 (333)	ZP_00298761
ORF9	6763-6035	242	Dienelactone hydrolase [*Synechococcus elongatus* PCC 7942] 244aa	36/51 (233)	ZP_00163798
ORF10	7737-6862	291	Transcriptional regulator [*Magnetococcus* sp. MC-1] 205aa	37/53 (196)	ZP_00290374

Alignments with other ArsB proteins in the NCBI database revealed that the 435 amino acid *At. caldus* chromosomal ArsB is most closely related to the ArsB protein of the *At. ferrooxidans ars* operon. The predicted amino acid sequence of the *At. caldus* chromosomal ArsB exhibits 88% and 67% identity to the ArsB proteins of *At. ferrooxidans* and *Psychrobacter arcticus* 273-4 respectively, while only 60% identical to the *At. caldus* transposon ArsB protein. Phylogenetic analysis of ArsB proteins showed that the *At. caldus* chromosomal ArsB groups with ArsB proteins of other Gram-negative bacteria (Figure 2.6). Using the web-based program TMHMM (which predicts trans-membrane spanning domains; http://www.cbs.dtu.dk/services/TMHMM) the chromosomal ArsB of *At. caldus* is predicted to have 11 trans-membrane spanning domains (Figure 2.7).

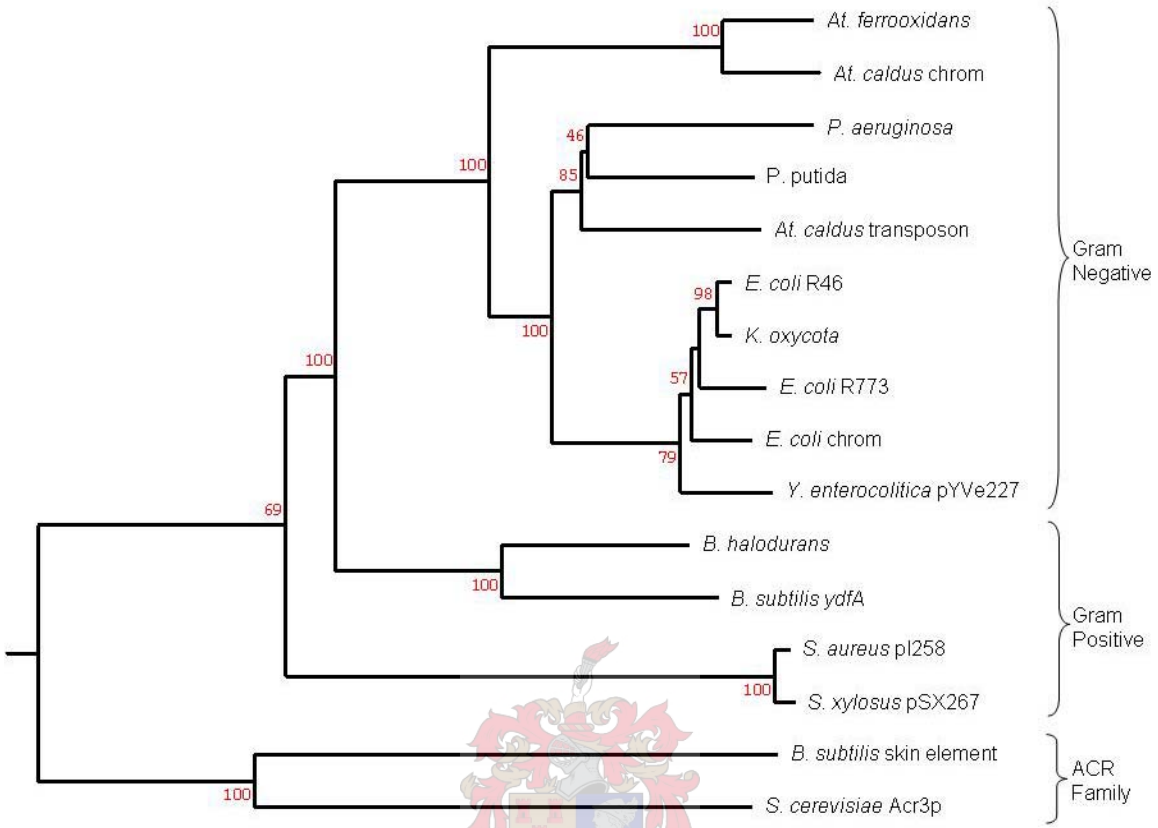


Figure 2.6: Phylogenetic tree of different ArsB proteins. The chromosomal *At. caldus* ArsB groups with other ArsB proteins of Gram-negative bacteria. Accession numbers: *At. ferrooxidans*, AAF69238; *P. aeruginosa*, NP_250968; *P. putida*, NP_744082; *At. caldus* transposon, AAX35679; *E. coli* R46, NP_511239; *K. oxycota* pMH12, AAF89641; *E. coli* R773, ARB1_ECOLI; *E. coli* chromosome, NP_417959; *Y. enterocolitica* pYVe227, NP_052439; *B. halodurans*, NP_243865; *B. subtilis* ydfA, NP_388415; *S. aureus* pI258, P30329; *S. xylosus* pSX267, Q01255; *B. subtilis* skin element, BAA06969; *S. cerevisiae* Acr3p, ACR3_YEAST.

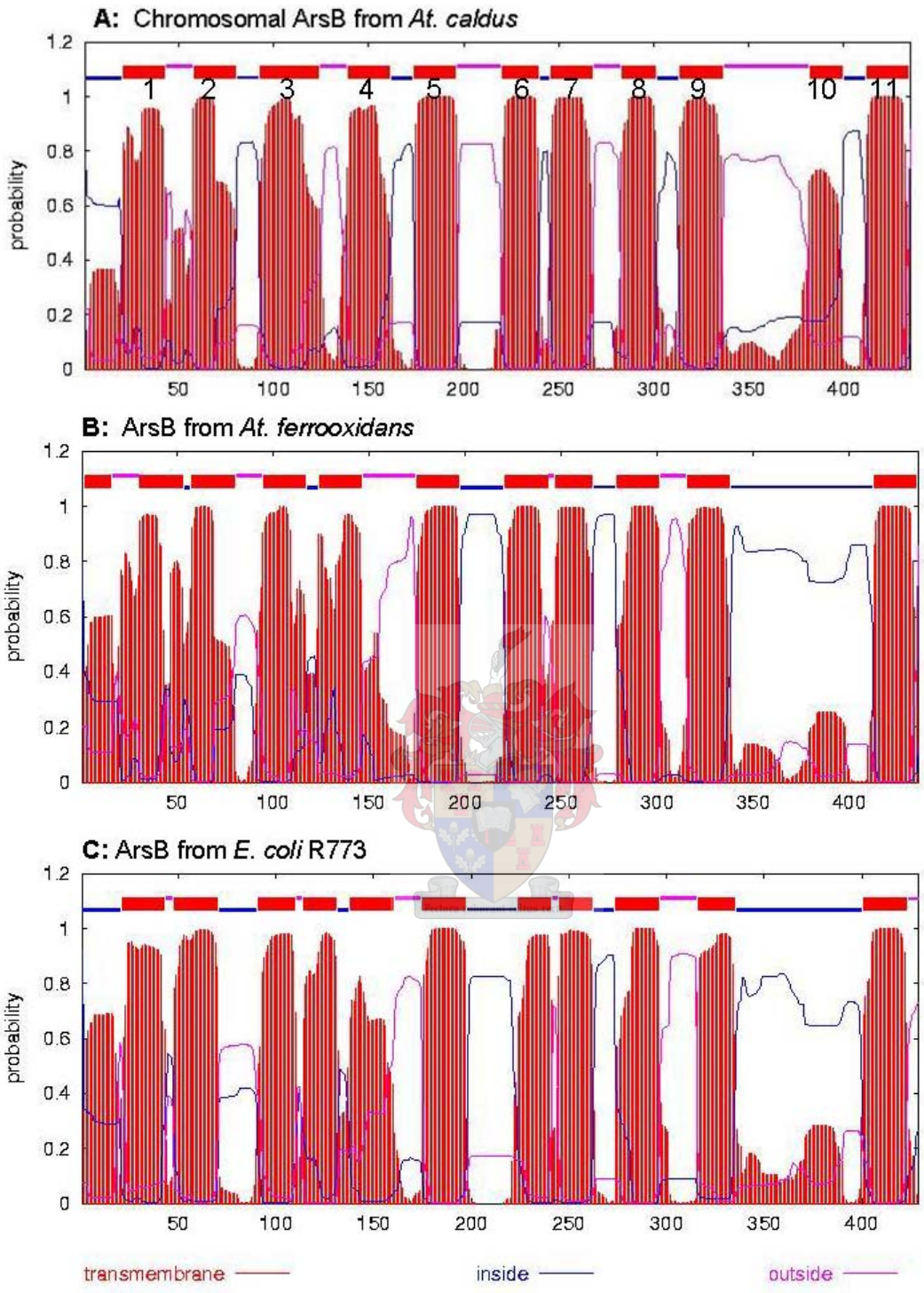


Figure 2.7: Prediction of trans-membrane spanning domains by the web-based TMHMM program for the ArsB proteins of *At. caldus* chromosome (A), *At. ferrooxidans* (B) and *E. coli* R773 (C).

As discussed in Chapter 1 (1.4.2) bacterial ArsC proteins can be divided into two families. The first family (GSH/Grx clade) requires glutathione (GSH) and glutaredoxin (Grx) to reduce arsenate to arsenite and includes the *E. coli* R773 ArsC. The second family of arsenate reductases (Trx clade) includes the *S. aureus* pI258 and *B. subtilis* ArsC proteins. Phylogenetic and alignment analysis showed that the *At. caldus* chromosomal ArsC groups with other ArsC proteins of the Trx clade (Figure 2.8 and 2.9). Members of the GSH/Grx clade of ArsC proteins are more closely related to each other than members of the Trx family with 93.52% sequence identity for the GSH/Grx family and only 53.3% sequence identity for the Trx family. Alignments with ArsC proteins in the NCBI database revealed that the *At. caldus* chromosomal ArsC is most similar to the ArsC protein of *At. ferrooxidans* with 78% identity. It is only 72% identical the *At. caldus* transposon ArsC protein. The Trx clade ArsC proteins use thioredoxin as a reducing power and contain four cysteine residues, Cys^{10}, Cys^{15}, Cys^{82} and Cys^{89}, of which only three are essential (Cys^{10}, Cys^{82} and Cys^{89}) for the conversion of arsenate to arsenite. Besides the three cysteine residues, Arg^{16} and Asp^{105} are also essential for catalysis. Sequence analysis of the *At. caldus* chromosomal ArsC revealed that it contains four cysteine residues (Cys^{10}, Cys^{15}, Cys^{82} and Cys^{90}) with spacing similar to the spacing in the ArsC proteins of the Trx clade (Figure 2.9). The third and fourth cysteines are separated by seven amino acids instead of six, as is the case with *At. ferrooxidans*, *P. putida* and *P. aeruginosa*. Arg^{16} and Asp^{105}, which are essential for catalysis by the Trx ArsC are also conserved in the *At. caldus* ArsC.

As discussed in Chapter 1 (1.4.3.1), the ArsR protein of *E. coli* R773 contains a possible DNA binding helix-turn-helix domain (Bairoch, 1993) and a highly conserved ^{30}ELCVCDL^{36} metal binding domain with conserved Cys^{32}, Cys^{34} and Cys^{37} residues. Binding requires only two cysteine residues, and any combination of Cys^{32}, Cys^{34} and Cys^{37} will result in the release of the repressor from the DNA (Shi *et al.*, 1994; Shi *et al.*, 1996). Busenlehner *et al.* (2003) reported that two distinct metal binding sites exist in the SmtB/ArsR family of transcriptional regulators. The first (α3N) metal binding site is present in the α3 helix and contains the conserved ELCVCDL domain. The second metal binding site (α5C) is present in the α5 helix (Figure 1.8). The predicted amino acid

sequence of the *At. caldus* chromosomal ArsR was most similar to the ArsR protein of *At. ferrooxidans* with 74% similarity, while it was only 45% identical to the transposon-linked ArsR protein of *At. caldus*. An alignment of the *At. caldus* chromosomal ArsR protein with its closest matches from the NCBI database revealed that none of the ArsR proteins contained the conserved ELCVCDL domain in the α3 helix region (Figure 2.10), indicating that the chromosomal ArsR and its closest homologs may use a different metal binding site than ArsR proteins with the ELCVCDL metal binding domain. The *At. caldus* chromosomal ArsR and its eight closest matches contains the conserved GXIA region (identified by Butcher and Rawlings, 2002) immediately downstream of the region corresponding to the ELCVCDL binding site. They all also contain another conserved LVAYLTENCC region near the C-terminal end of the protein. Secondary structure prediction of the *At. caldus* chromosomal ArsR (performed by the web-based program PsiPRED v2.4, http://www.predictprotein.org) revealed that it contains, like other ArsR proteins, two possible helix-turn-helix motifs (Figure 2.11). The position and spacing of these helix-turn-helix domains of the *At. caldus* chromosomal ArsR is very similar to those of the well studied *E. coli* R773 ArsR. All proteins that lacked the conserved metal binding domain have two extra residues before the αR helix (Figure 2.11). The significance of these extra residues in arsenic resistance has yet to be shown. Phylogenetic analysis showed that the chromosomal *At. caldus* ArsR, together with its closest matches form a subgroup of ArsR regulators (Figure 2.12).

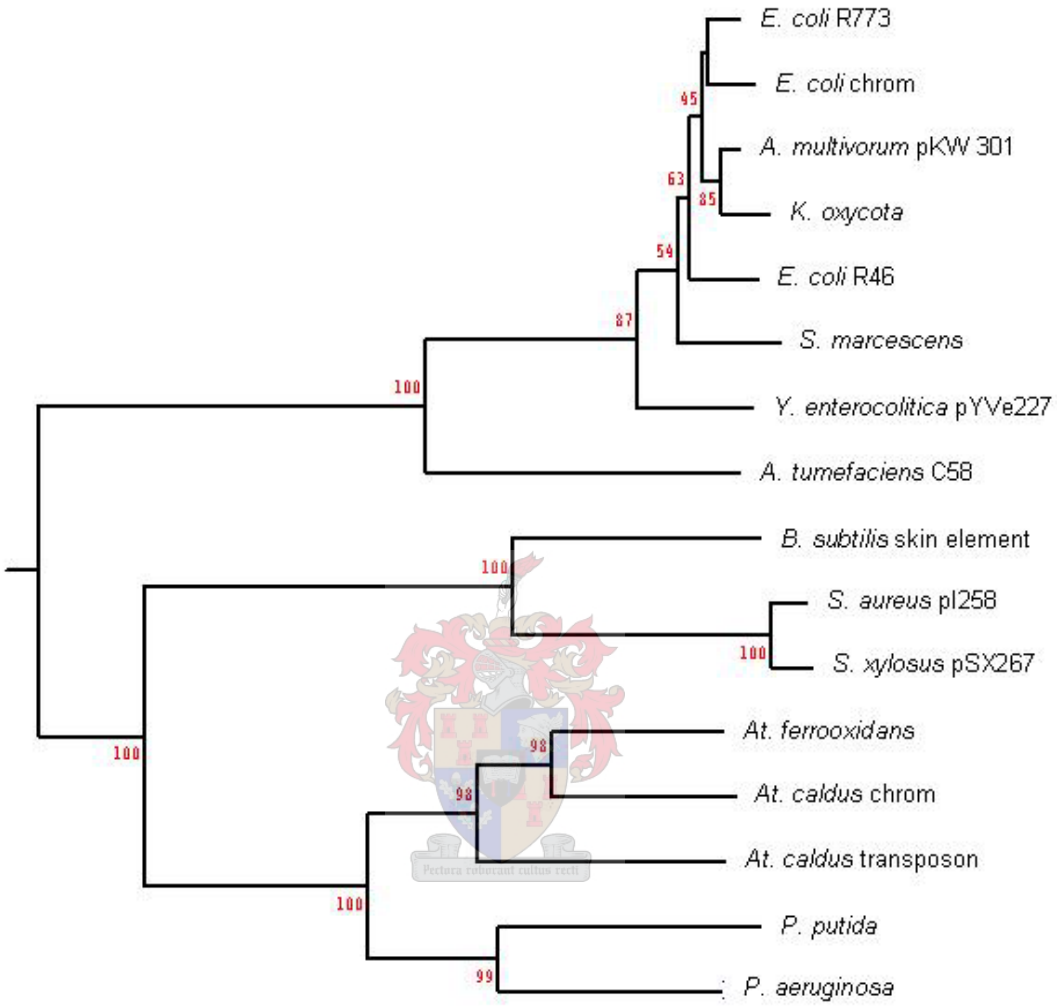


Figure 2.8: Phylogenetic tree of different ArsC proteins. The *At. caldus* chromosomal ArsC is most closely related to the transposon-located ArsC of *At. caldus* and the ArsC of *At. ferrooxidans*. Accession numbers: *E. coli* R773, AAA21096; *E. coli* chromosome, NP_417960; *A. multivorum* pKW301, BAA24824; *K. oxycota*, AF168737_5; *E. coli* R46, NP_511240; *S. marcescens*, CAB88404; *Y. enterocolitica* pYVe227, NP_052438; *A. tumefaciens* C58, NP_532180; *B. subtilis* skin element, ARSC_BACSU; *S. aureus* pI258, ARSC_STAAN; *S. xylosus* pSX267, ARSC_STAXY, *At. ferrooxidans*, AAF69239; *At. caldus* transposon, AAX35680; *P. putida*, CAC18654; *P aeruginosa*, NP_250969.

GSH/Grx clade

```
Ec R773       MSNITIYHNPACGTSRNTLEMIRNSGTEPTIILYLENPPSRDELVKLIADMGISV   55
Ec chrom      MSNITIYHNPACGTSRNTLEMIRNSGTEPTIIHYLETPPTRDELVKLIADMGISV   55
Ec R46        MSNITIYHNPHCGTSRNTLEMIRNSGIEPTVILYLETPPSRDELLKLIADMGISV   55
K oxytoca     MSNITIYHNPACGTSRNTLEMIRNSGNEPTVIHYLENPPSRDELVKLIADMGISV   55
S marces      MSNITIYHNPACGTSRNTLEMIRNSGTEPTVIHYLETPPSRAELVKLIADMGITV   55
Y ent pYV     MSNITIYHNPTCGTSRNTLEMIRNSGNEPTVIYYLETPPTHDELVKLIADMGITV   55
Am pKW301     MSNITIYHNPACGTSRNTLEMIRNSGNEPTVIHYLENPPSRDELVKLIADMGISV   55

Ec R773       RALLRKNVEPYEQLGLAEDKFTDDQLIDFMLQHPILINRPIVVTPLGTRLCRPSE  110
Ec chrom      RALLRKNVEPYEELGLAEDKFTDDRLIDFMLQHPILINRPIVVTPLGTRLCRPSE  110
Ec R46        RALLRKNVEPYEELGLAEDKFTDDQLIDFMLQHPILINRPIVVTPLGTKLCRPSE  110
K oxytoca     RALLRKNVEPYEELGLAEDKFTDGELIDFMLQYPILINRPIVVTPLGTRLCRPSE  110
S marces      RALLRKNVEPFEALGLAEDRFTDEQLIDFMLQHPVLINRPIVVTPLGTRLCRPSE  110
Y ent pYV     RALLRKNVEPYEELGLAEGTFSDEQLIGFMLEHPILINRPIVVTPLGTRLCRPSE  110
Am pKW301     RALLRKNVEPYEELGLEEDKFTDDQLIDFMLQHPILINRPIVVTPLGTRLCRPSE  110

Ec R773       VVLDILQDAQKGAFTKEDGEKVVDEAGKRL                           140
Ec chrom      VVLEILPDAQKGAFSKEDGEKVVDEAGKRL                           140
Ec R46        VVLDILPDAQKAAFTKEDGEKVVDDSGKRL                           140
K oxytoca     VVLDILPDAQKGAFAKEDGEKVVDETR...                           137
S marces      VVLDILYDAQKSAFTKEDGEKVVDEKGNRL                           140
Y ent pYV     VVLDILPEPQQGAFTKEDGEKITDESGKRL                           140
Am pKW301     VVLDILPDAQKGAFAKEDGEKVVDEAGKRL                           140
```

Trx clade

```
Bs skin       MENKIIYFLCTGNSCRSQMAEGWAKQYLGDEWKVYSAGIEAHGLNPNAVKAMKEV   55
At_ferro      MKTPEILFLCTGNSCRSILAEVTFNALAGPGMHATSAGSHPAGYVHTRSINLLER   55
Sx_pSX267     MDKKTIYFICTGNSCRSQMAEGWGREILGEDWNVYSAGIETHGVNPKAIEAMKEV   55
P_aerug__     ...MRVLFMCTANSCRSILSEAMFNHLAPPGFEACSAGSQPSGRVHPRSLATLEQ   52
P_putida      ...MKVLFMCTANSCRSILSEAMFNHLAPDGFEAISSGSFPKGLVLPRSLSTLQA   52
Sa_pI258      MDKKTIYFICTGNSCRSQMAEGWGKEILGEGWNVYSAGIETHGVNPKAIEAMKEV   55
At_c_Ch       MKTPEILFLCTGNSCRSILAEATLNGLAQGRLHASSAGSHPAGYVHPRSVALLQR   55

Bs skin       GIDISNQTSDIIDSDILNNADLVVTLCGDAADK..CPMTPPHVKREHWGFDDPAR  108
At_ferro      EGFRTDGLHSKSWEDLKET.PDIVITVCADAAGETCPAYLGPAIRTHWGVEDPAK  109
Sx_pSX267     DIDISNHTSDLIDNHILKQSDLVVTLCSDADDN..CPILPPNVKKEHWGLEDPAG  108
P_aerug__     AGIATHGLYSKGSEAFEGAPPDIVITVCDAAAGEACPLYLGAALKAHWGLADPSA  107
P_putida      AGIRTEGLYSKGNDVFEGSPPDVVITVCDKAAGEACPVYFGPAVKAHWGLEDPSD  107
Sa_pI258      DIDISNHTSDLIDNDILKQSDLVVTLCSDADNN..CPILPPNVKKEHWGFDDPAG  108
At_c_Ch       EGFSTEGLASKSWDDLPTT.PDIVITVCASAAGETCPAYLGPALRAHWGVDDPAK  109

Bs skin       AQGTEEEKWAFFQRVRDEIGNRLKEFAETGK.....................     139
At_ferro      VTGTEAQIEAAFDTAYHILRHRIEALLQLPVAELLEKDPAKLRQELERIGTLL     162
Sx_pSX267     KEWSE......FQRVRDEIKLAIENFKLR.......................     131
P_aerug__     LDGDEALRDAAFHATLARIEQRCRAFLGLPFATLDRDQLKR...ELERIGSL.     156
P_putida      VQGDDARVQAAFDATLKTIATRCRAFFILPFAELSPTELQA...ELARIAEL.     156
Sa_pI258      KEWSE......FQRVRDEIKLAIEKFKLR.......................     131
At_c_Ch       ATGSEAEIEAAFEKAYRILRHRIEALLQLPVAEWLAEDPSRLRTELERIGHSE     162
```

Figure 2.9: Multiple sequence alignments of different ArsC proteins compared with the *At. caldus* chromosomal ArsC. The *At. caldus* chromosomal ArsC protein groups with other ArsC proteins of the Trx clade. See figure 2.5 for accession numbers.

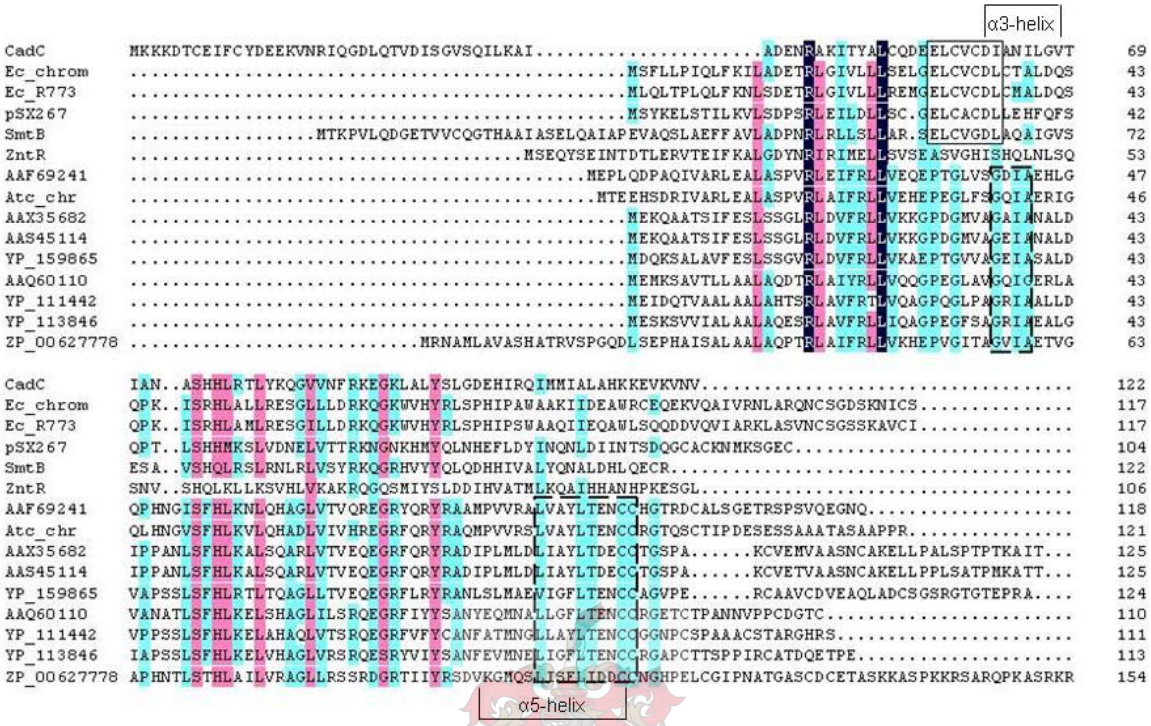


Figure 2.10: Multiple sequence alignments of different ArsR proteins (the eight closest matches obtained from a NCBI BLAST search) compared with the chromosomal ArsR of *At. caldus*. Also included are proteins from the ArsR/SmtB family of transcriptional regulators. Labels and accession numbers are as follows: CadC, *S. aureus* pI258 CadC (B32561); Ec_chrom, *E. coli* chromosome (AAS76526); Ec_R773, *E. coli* R773 (CAA34168); pSX267, *S. xylosus* pSX267 (AAA27587); SmtB, *Synechococcus* PCC7942 SmtB, (S31197), ZntR, *S. aureus* ZntR (AAC32484); Atc_chr, *At. caldus* chromosome. The nine closest matches are labeled by their accession numbers: AAF69241, *At. ferrooxidans*; AAX35682, *At. caldus* transposon; AAS45114, *Alcaligenes faecalis*; YP_159865, *Azoarcus* sp. EbN1; AAQ60110, *Chromobacterium violaceum*; YP_111442, *Burkholderia pseudomallei* K96243; YP_113846, *Methylococcus capsulatus* str. Bath; ZP_00627778, *Nitrobacter hamburgensis* X14. The conserved ELCVCDL metal binding motif found in members of the ArsR/SmtB family of transcriptional regulators has been boxed with solid lines, while conserved regions of an atypical ArsR group identified by Butcher and Rawlings (2002) are boxed in broken lines. The *At. caldus* chromosomal ArsR groups with the atypical ArsR group and also contains the two conserved domains. Also shown is the *At. caldus* chromosomal ArsR α3 and α5-helices.

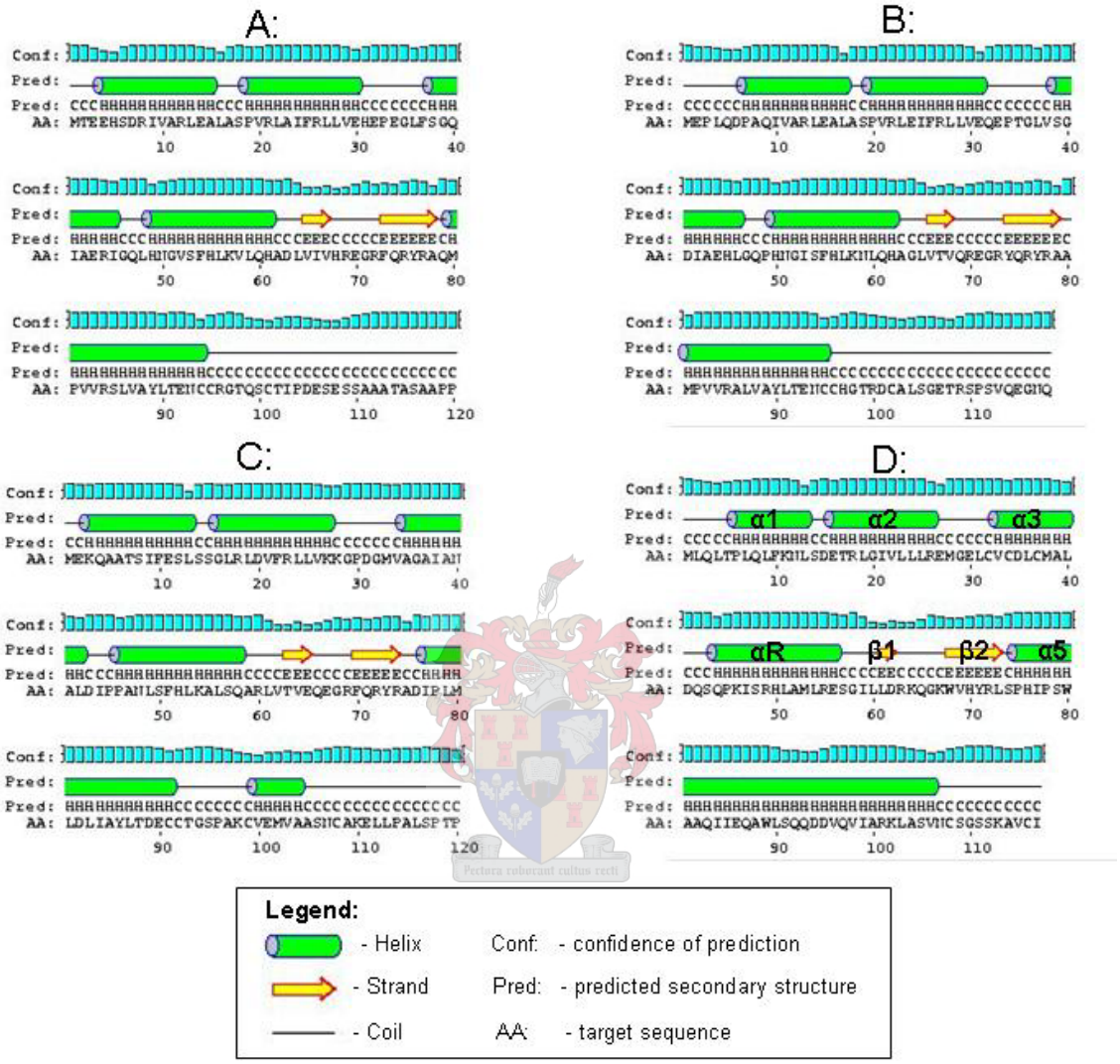


Figure 2.11: Secondary structure prediction of the (A) *At. caldus* chromosomal ArsR, (B) *At. ferrooxidans* ArsR, (C) *At. caldus* transposon ArsR and (D) *E. coli* R773 ArsR by the PsiPRED web-based program. Also included are the α-helices and β-sheets as indicated by Busenlehner *et al.* (2003). Proposed metal binding domains are located in the α3 and α5-helices, as described in the text.

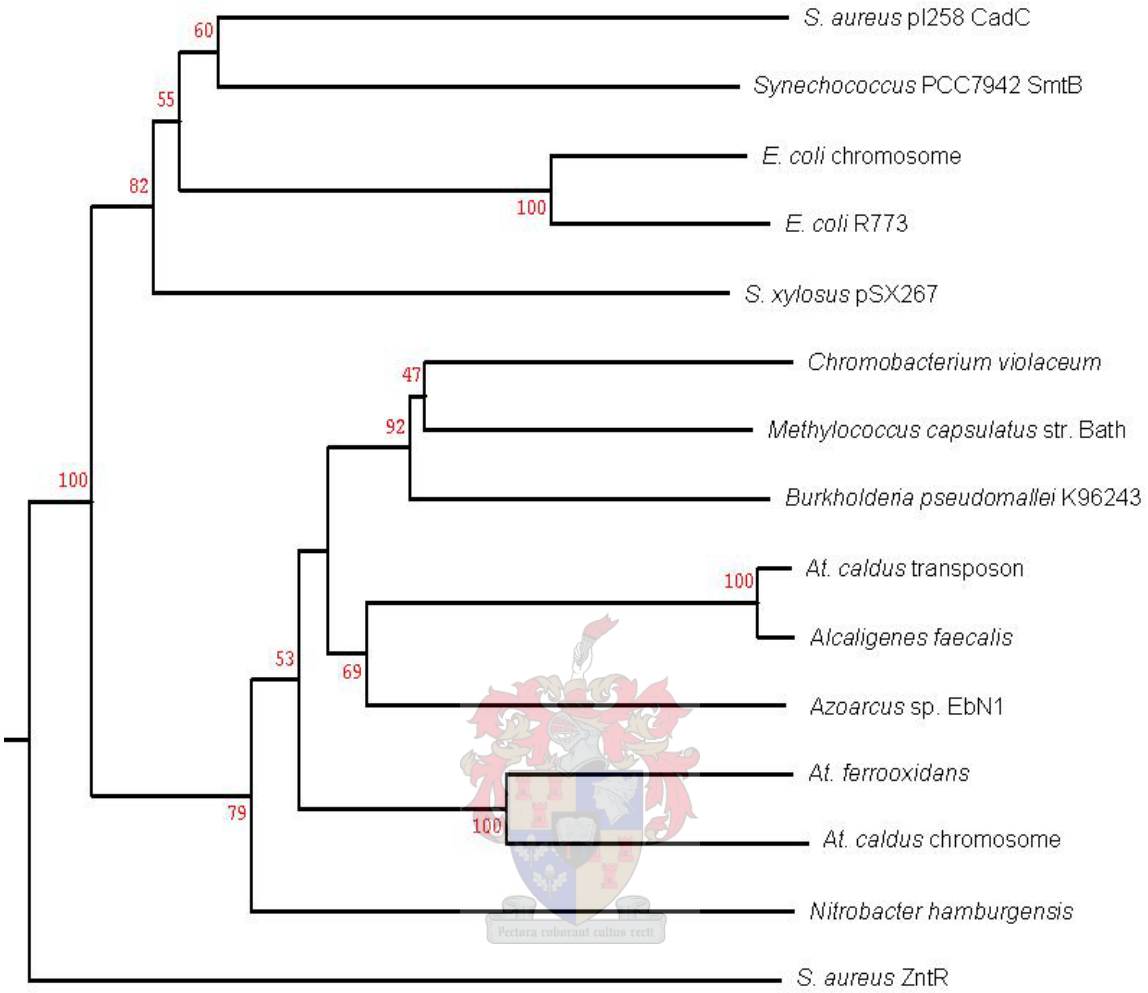


Figure 2.12: Phylogenetic tree of different ArsR proteins. The *At. caldus* chromosomal ArsR and its eight closest matches cluster as a second subgroup of ArsR regulators. See Figure 2.10 for accession numbers.

ORF1 showed sequence identity to a lactoylglutathione lyase (also known as glyoxalase I) of *Thermobifida fusca*. Glyoxalase I forms part of the glyoxalase pathway that protects against cellular damage caused by reactive aldehydes such as methylglyoxal with the help of glutathione hemithioacetal as substrate (Vickers *et al.*, 2004). Glyoxalase I and glyoxalase II (hydroxyacylglutathione hydrolase) act in concert to convert the hemithioacetal adduct formed between glutathione and methylglyoxal into D-lactate and free glutathione (Thornalley, 1996). Because ORF1 is located adjacent to *arsC*, the possibility exists that ORF1 may be involved in the conversion of arsenate to arsenite by the ArsC (if it uses glutathione as a reducing power). ORF1 also showed sequence identity to a cadmium inducible protein (CadI) of *Mycobacterium tuberculosis*. *cadI* homologs are often found adjacent to, or within *ars* operons (Hotter *at. al*., 2001). CadI is related to members of the βαββββ metalloenzyme superfamily, which has important detoxification functions (Bernat *et al.*, 1997). When looking at the sizes of the predicted lactoylglutathione lyase and CadI proteins, it seems more likely that ORF1 functions as a CadI protein. Furthermore, ORF1 is located adjacent to the chromosomal *ars* operon of *At. caldus*, as is the case with several CadI proteins of other bacteria (Hotter *et al.*, 2001). Cadmium induces the *ars* operon of *B. subtilis* (Moore *et al.*, 2005). Experiments were conducted to determine if ORF1 may be involved in the conversion of arsenate to arsenite (2.3.3) and to determine if the *At. caldus* chromosomal *ars* operon can be induced by cadmium, as is the case of *B. subtilis* (3.3.1).

2.3.3. The requirement of glutathione/thioredoxin for arsenate reductase

It is well known that ArsC proteins are grouped on their source of reducing power that is used for the reduction of arsenate to arsenite. The first family (GSH/Grx clade) uses glutathione and glutaredoxin to reduce arsenate, while the second family (Trx clade) uses thioredoxin for the reduction of arsenate to arsenite (Mukhopadhyay *et al.*, 2002; Mukhopadhyay and Rosen, 2002). Based on the observation that the *At. caldus* chromosomal ArsC grouped with other ArsC proteins of the Trx clade (Figure 2.8 and 2.9) and that it was 78% identical to the *At. ferrooxidans* ArsC (which has been shown to use thioredoxin to reduce arsenate to arsenite, Butcher *et al.*, 2000), tests were conducted

to determine whether the *At. caldus* chromosomal ArsC protein would use thioredoxin rather than glutathione and glutaredoxin as a reducing power for the reduction of arsenate to arsenite. To investigate this, an *E. coli* strain (BH5262) with mutations in the thioredoxin gene (*trxA*) and the *γ*-glutamylcysteinyl synthetase gene (*gshA*, responsible for the synthesis of glutathione) was transformed with various plasmids. *E. coli* BH5262 transformed with the *At. caldus* chromosomal *ars* genes (pTcC-PH29GL) and Trx6 (contains the *At. ferrooxidans trxA* gene) were unable to confer resistance to 1 mM arsenate. The *ars* mutant ACSH50I^{q} transformed with pTcC-PH29GL did confer resistance to 1 mM arsenate, indicating that the *At. caldus* chromosomal *ars* genes were sufficient for resistance to 1 mM arsenate (Table 2.3). Surprisingly, the *At. ferrooxidans* thioredoxin containing plasmid, pTrx6, did not permit the *E. coli* double mutant BH5262 containing the cloned *At. caldus* chromosomal *ars* genes (pTcC-PH29GL) to grow on arsenate, as reported when the *At. ferrooxidans ars* genes were tested in a similar experiment (Butcher and Rawlings, 2000). Because the *At. caldus* chromosomal ArsC groups with the ArsC proteins of the Trx clade, it seems very unlikely that it will use glutathione as a reducing power. It is possible that the *At. ferrooxidans trxA* gene may not complement the *At. caldus* chromosomal ArsC in the reduction of arsenate to arsenite. This experiment could be repeated using a cloned *trxA* gene from *At. caldus* to confirm if the *At. caldus* chromosomal ArsC uses thioredoxin in the conversion of arsenate to arsenite. Despite grouping with other ArsC proteins of the Trx clade (Figure 2.9), it is possible that the *At. caldus* chromosomal ArsC may use glutathione to reduce arsenate to arsenite and that ORF1 may play a role in this reduction. To investigate this, *E. coli* BH5262 was transformed with pTcC-#4. If ORF1 aids the ArsC in reducing arsenate to arsenite, arsenate resistance would be acquired. This was not the case however, as *E. coli* BH5262 transformed with pTcC-#4 was sensitive to 1mM arsenate (Table 2.3). This also indicated that none of the other ORFs on pTcC-#4 aid in the reduction of arsenate to arsenite.

Table 2.3: Arsenate resistance in *E. coli* mutant strains

***E. coli* strain**	**Transformed with plasmid**	**Resistance to 1mM arsenate** [a]
BH5262	pGL10 + Trx6	–
BH5262	pGL10 + pKS	–
BH5262	pTcC-PH29GL + pKS	–
BH5262	pTcC-PH29GL + Trx6	–
ACSH50I^{q}	pTcC-PH29GL	+
BH5262	pTcC-#4	–

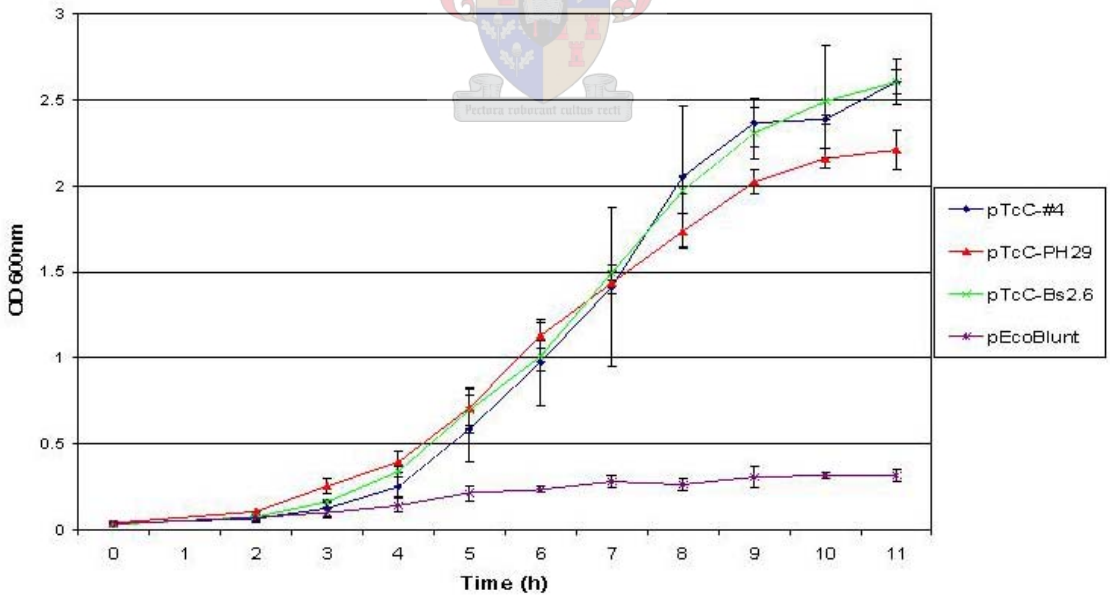
[a] – indicates no growth on Luria Agar with 1mM sodium arsenate; + indicates growth on Luria Agar with 1mM sodium arsenate

2.3.4. ORF1 and ORF5 are not required for arsenic resistance in *E. coli*

Arsenic assays were performed in order to determine (a) the resistance to arsenic that construct pTcC-#4 isolated from *At. caldus* confers on the *E. coli* arsenic mutant ACSH50I^{q}, (b) to determine if ORFs 5-10 (Figure 2.4; Table 2.2) play a role in arsenic resistance and (c) to determine if ORF1 aids the *arsC* gene product in converting arsenate to arsenite (see section 2.3.3). Growth assays were performed in 0.25 mM arsenite (As^{III}) and 0.5 mM arsenate (As^{V}). Plasmid pTcC-#4 conferred greater resistance to the *E. coli ars* mutant ACSH50I^{q} in both arsenate and arsenite in comparison to the negative control p*Eco*Blunt (Figure 2.13 a and b). Construct pTcC-#6 conferred lower resistance to arsenate, which was expected as it does not contain a functional *arsC* that is required for the conversion of arsenate to arsenite. Arsenic resistance did not change significantly in

cells harboring only *arsCRB*ORF5 (pTcC-PH29), suggesting that ORFs 6-10 were not required for arsenic resistance in *E. coli* ACSH50I^{q}. Since ORFs 6-10 were not needed for arsenic resistance, we could directly compare pTcC-#4 and pTcC-PH29. No differences in levels of resistance to arsenate were detectable between pTcC-#4 and pTcC-PH29. This indicated that ORF1 is not required for the reduction of arsenate to arsenite in *E. coli*.

Construct pTcC-Bs2.6 (which contains *arsCRB*) also showed high levels of resistance to arsenate and arsenite (Figure 2.13 a and b). As it does not contain ORF5, it was suggested that ORF5 is not necessary for arsenic resistance in *E. coli*. To determine the role ORF5 plays in arsenate resistance, pTcC-RB+5 and pTcC-RB-5 were constructed. Growth curves done with these constructs showed the same levels of resistance as pTcC-#6, indicating that ORF5 does not play a role in arsenate resistance in *E. coli* ACSH50I^{q} in the absence of ArsC.



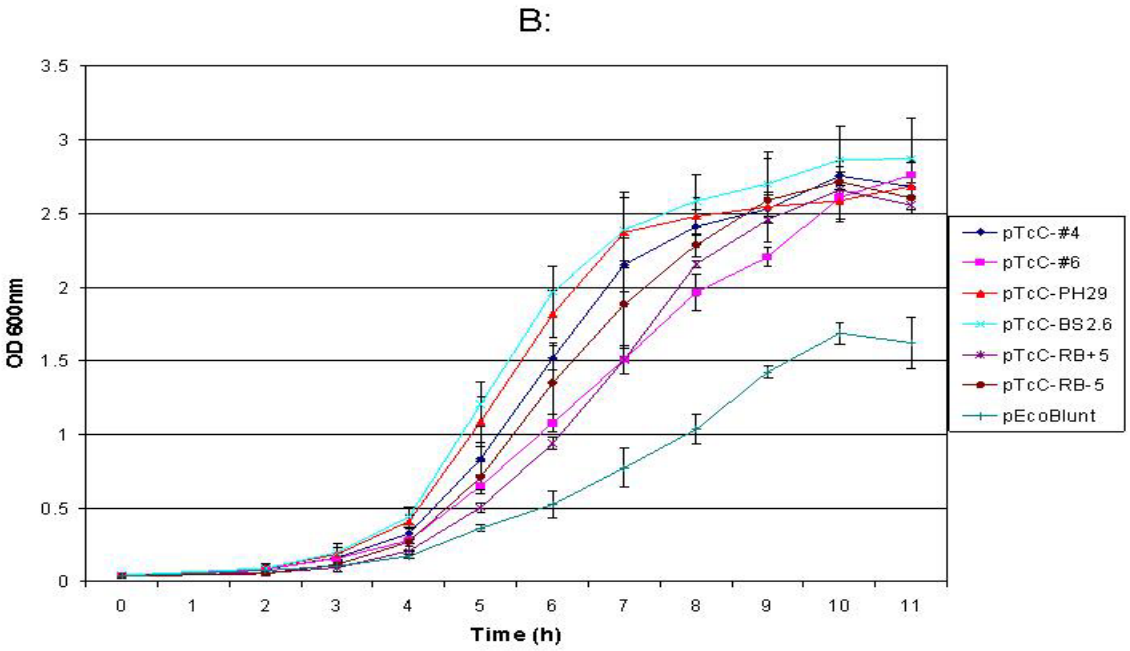


Figure 2.13: Growth curves performed with the *E. coli ars* mutant ACSH50I^q, in the presence of 0.25 mM arsenite (A) and 0.5 mM arsenate (B). Each data point represents the results of three assays of three independent experiments. Error bars indicate standard deviations. See figure 2.4 for location of the constructs. Optical density (OD) readings were recorded at 600 nm.

2.4. Discussion

Acidithiobacillus caldus is a member of a consortium of bacteria used in a commercial biooxidation process called biomining. During this process, toxic arsenic compounds are released into the environment. Due to the fact that *At. caldus* is able to live in such a harsh environment, it was not surprising to find arsenic resistance genes in its genome. In this study we identified a chromosomal *ars* operon that conferred resistance to arsenate and arsenite. It consists of three *ars* genes, *arsCRB*, with no *arsA* or *arsD* genes present. This is not surprising, as the conventional chromosomal *ars* operon of bacteria like *E. coli* also contains an *arsRBC* operon (Diorio *et al.*, 1995). However, the orientation of the *At. caldus* chromosomal *ars* operon was atypical. Where the conventional chromosomal *ars* operons are transcribed as a single transcriptional unit, the *At. caldus*

chromosomal *ars* operon is divergently transcribed, with the *arsC* and *arsR* genes divergently transcribed to the *arsB* gene, as has also been shown for *At. ferrooxidans*.

Results of BLAST searches in the NCBI database revealed that all three chromosomal arsenic resistance genes from *At. caldus* are most homologous to the arsenic resistance genes of another biomining bacterium, *Acidithiobacillus ferrooxidans*. As described earlier, *At. caldus* is a moderately thermophilic sulfur oxidizer and *At. ferrooxidans* a mesophilic iron- and sulfur oxidizer. The fact that such high homology was found (78%, 74%, and 85% identity for the *arsC*, *arsR* and *arsB* genes, respectively) between the *ars* genes of *At. caldus* and *At. ferrooxidans* strengthens the hypothesis that they may have a relatively recent common ancestor. The chromosomal ArsB protein of *At. caldus* grouped with ArsB proteins of other Gram-negative bacteria (Figure 2.6). Using a web-based program that predicts transmembrane spanning domains we predicted that the chromosomal ArsB has 11 membrane spanning domains. In *ars* systems where an ArsA is absent, it is believed that the ArsB cannot function as a primary pump using ATP hydrolysis as an energy source, but as a secondary carrier using membrane potential to export arsenite out of the cell. However, acidophilic microorganisms commonly have a reversed membrane potential, with a positive rather than a negative inside membrane potential (Martin, 1990). Despite originating from an organism with a positive inside membrane potential, the chromosomal *At. caldus* ArsB efflux pump was functional in *E. coli*, as resistance to arsenite was observed.

The *At. caldus* chromosomal ArsC grouped with ArsC proteins of the Trx family (Figure 2.9), suggesting that it will be more likely to use thioredoxin than glutathione and glutaredoxin as a reducing power for the conversion of arsenate to arsenite. To investigate this, *E. coli* BH5262 was transformed with the *At. ferrooxidans* thioredoxin gene (*trxA*) and pTcC-#4 in trans. It was surprising to find that this gene was unable to confer arsenate resistance to *E. coli* BH5262. It is possible that the *At. ferrooxidans trxA* gene may not complement the *At. caldus* chromosomal ArsC to reduce arsenate to arsenite, and therefore results in a arsenate sensitive phenotype. *E. coli* BH5262 transformed with pTcC-#4 was also sensitive to arsenate, indicating that none of the

ORFs on pTcC-#4 were able to function as a reducing power to help ArsC in the conversion of arsenate to arsenite. Sequence analysis of the *At. caldus* chromosomal ArsC revealed that it contains four cysteine residues (Cys^{10}, Cys^{15}, Cys^{82} and Cys^{90}), and that it is spaced similar to the spacing in other ArsC proteins of the Trx family. It also contains Arg^{16} and Asp^{105}, which have been shown to be essential for catalysis of the Trx family ArsC. Based on this sequence analysis, we proposed that the *At. caldus* chromosomal ArsC will function similar to the *B. subtilis* ArsC (Figure 1.7).

Sequence alignments of members of the ArsR family of transcriptional regulator proteins showed that these proteins show low similarity to each other. Alignments of the *At. caldus* chromosomal ArsR with eight of its closest matches from the NCBI database revealed that none of the ArsR proteins contained the conserved ELCVCDL metal-binding domain (Figure 2.10). Phylogenetic analyses of these ArsR proteins show that they form a subgroup of ArsR regulators (Figure 2.12). Using a web-based program (PsiPRED v. 2.4), we predicted that the *At. caldus* chromosomal ArsR contains, like other ArsR proteins, two helix-turn-helix domains with 5 α-helices and 2 β-sheets (Figure 2.11). The *At. caldus* ArsR and its eight closest matches from the NCBI database all have two extra residues before the first helix of the identified DNA-binding helix-turn-helix domain.

The SmtB/ArsR family of transcriptional regulators have two metal binding sites, the α3N in the α3 helix and the α5C metal-binding site in the α5 helix. They have shown that ArsR proteins contain the α3N metal binding site, with the ELCVCDL domain located in the α3 helix. The *At. caldus* chromosomal ArsR and its closest matches all contain a very conserved LVAYLTENCC domain near the C-terminal end of the protein (Figure 2.10). This conserved domain is not present in ArsR proteins with the consensus ELCVCDL domain. Five to seven residues separate another cysteine from the cysteine doublet at the end of the LVAYLTENCC domain of *At. caldus* and all of its closest matches. These three cysteine residues may form the metal-binding domain of these atypical ArsR proteins. This conserved region identified in the second ArsR group is present in the α5 helix (Figure 2.10). It is thus possible that the ArsR proteins may have two different

metal-binding domains, where one is located in the α3 helix and the other in the α5 helix. Analysis of a possible *At. caldus* chromosomal ArsR binding site did not reveal the TTTG and TCAT region identified in both the chromosomal and R773 *ars* sequences of *E. coli*.

Along with the three arsenic resistance genes, seven other ORFs were identified on pTcC-#4 (Table 2.2). Based on the proximity of ORF7-10 to one another and the direction of transcription, it is possible that these ORFs form an operon separate form the *ars* operon. The function of this possible operon is unknown, but it may play a role in the transport of certain molecules in and out of the cell (Figure 2.4; Table 2.2).

Both the *arsC* and *arsB* genes were expressed and were functional in the *E. coli ars* mutant ACSH50I^{q}. Growth assays in the presence of arsenite and arsenate were conducted to determine if the other ORFs play a role in arsenic resistance. It was found that ArsC is needed for resistance to arsenate, but not arsenite, while ArsB was essential for resistance to arsenite. Comparing arsenic resistance conferred by the *At. caldus* chromosomal *ars* operon to arsenic resistance conferred by the *At. caldus* transposon *ars* operon (Tuffin *et al.*, 2005), it seems that the transposon operon conferred much higher resistance to *E. coli* than the chromosomal *ars* operon. We did expect constructs not containing ArsC (pTcC-#6, pTcC-RB-5 and pTcC-RB+5) to be more sensitive to arsenate than was observed (Figure 2.13 a and b). A reason for increased arsenate resistance by these constructs may be that the *ars* mutant ACSH50I^{q} contains an element on the chromosome that can help with the conversion of arsenate to arsenite, which can then be transported out of the cell by the *At. caldus* ArsB.

This studies of the *At. caldus* chromosomal arsenic resistance system was carried out in *E. coli*. It would have been ideal if this studies could be repeated in *At. caldus*, as this system may not function the same way in *E. coli* as in *At. caldus*. Due to a lack of mutants, genetic tools for this organism and the difficulty to obtain large quantities of cells, these studies are limited to *E. coli* at present.

Chapter Three

Regulation and cross-regulation by the *Acidithiobacillus caldus* chromosomal ArsR

Contents

3.1. Introduction

Together with the work presented in Chapter 2, it is now clear that *At. caldus* #6 has three arsenic resistance operons, with two being located on transposons and a third on the chromosome. This raises interesting questions about how each of the systems interact to give arsenic resistance, and how each of these systems are regulated to give optimal resistance. In Chapter 4 work will be presented on the interaction of the transposon and chromosomal system to give arsenic resistance, while this chapter will focus on the regulatory mechanisms of the chromosomal operon, and whether the chromosomal and transposon resistance determinants can cross-regulate each other.

Sequence analysis (Chapter 2) revealed that the *At. caldus* chromosomal *ars* genes are arranged in the order *arsCRB*, with the *arsC* and *arsR* genes reading divergently to the *arsB* gene. This could suggest that *arsCR* and *arsB* may be transcribed as two separate transcriptional units. This is different to the typical *arsRBC* systems, which are transcribed as a single transcriptional unit. This divergent arrangement of *ars* genes has also been identified in *At. ferrooxidans,* where the *arsCR* were transcribed as one unit, separate to *arsBH* (Butcher *et al.*, 2000). We have shown in Chapter 2 that ORF1 and ORF5 were not required for, and did not enhance, arsenic resistance in *E. coli*. However, the gene layout (with stop and start codons very close together) could suggest co-transcription with the *ars* genes.

In this chapter expression of the *At. caldus* chromosomal *ars* system will be analyzed. Would this system be regulated by ArsR at a transcriptional level, binding to an operator sequence and thereby blocking transcription? It has been shown that with single unit operons like the chromosomal *ars* operon of *E. coli* K-12, the ArsR binds upstream of the *ars* operon (Diorio *et al.*, 1995). However, with the divergent operon, as has been shown in *At. ferrooxidans*, the ArsR autoregulates itself as well as *arsB* expression (Butcher and Rawlings, 2002). Sequence analysis of the *At. caldus* chromosomal ArsR revealed that it does not contain the highly conserved ELCVCDL region that is found in members belonging to the SmtB/ArsR family of transcriptional regulators (Shi *et al.*, 1994). As previously described, this domain forms part of metal binding site where arsenite binds in

order to induce a conformational change in ArsR, with the dissociation of the repressor from the DNA. The *At. caldus* chromosomal ArsR was most similar to the ArsR protein of *At. ferrooxidans* with 74% identity. The *At. ferrooxidans* ArsR and the *At. caldus* transposon-located ArsR also lack the conserved metal binding domain, but have been able to regulate their *ars* operons respectively (Butcher and Rawlings, 2002; Tuffin *et al.*, 2005). Based on the above, would ArsR function as a negative transcriptional regulator, and to what inducers would it respond? Does the *At. caldus* chromosomal ArsR regulate both itself and *arsB*? If it does regulate both, to what extent are they being regulated? Due to the presence of two different ArsR proteins in one organism, would there be any cross-regulation? The *At. caldus* transposon ArsR has been shown to function as a typical ArsR (Tuffin *et al.*, 2005). Despite the fact that the two ArsRs only share 45% identity at protein level, there is a possibility of cross-regulation between the two ArsR proteins.

3.2. Methods

3.2.1. Media, bacterial strains and plasmids

Bacterial strains, plasmids and primers used in this study are shown in Table 3.1. *E. coli* strains were grown in Luria-Bertani (LB) broth medium (Sambrook *et al.*, 1989) with ampicillin (100 μg/ml) or kanamycin (50 μg/ml) added as required. *At. caldus* strains were grown at 37^0C in tetrathionate medium (3 mM), sterilized and adjusted to pH 2.5 (Rawlings *et al.*, 1999b).

Table 3.1: Bacterial strains, plasmids and primers used in this study

Strain, plasmid or Primer	Description	Source or reference
Strains		
E. coli		
DH5α	Φ80d*lacZ*ΔM15, *endA*1, *recA*1, *gyrA*96, *thi*-1, *hsdR*17 (r_k^-, m_k^+), *relA*1, *supE*44, *deoR*,Δ(*lacZYA-argF*) U169	Promega
XL 1 Blue	*recA*1, *endA*1, *gyrA*96, *thi, hsdR*17 (r_k^-, m_k^+), *relA*1,	Stratagene

	*supE*44, *relA*1, lac [F' *proAB*+ *lacI*^{q}ZΔM15 ::Tn10(tet^r)]	
ACSH50I^q	*rpsL*, Δ(*lac-pro*), Δ*ars*::*cam*	Butcher & Rawlings (2002)
Plasmids		
p*Eco*R252	Ap^r, *Eco*RI inactivation cloning vector	Zabeau & Stanley (1982)
pBluescript SK	Ap^r, *LacZ*', ColE replicon, cloning vector	Stratagene
pGem-T®	Ap^r, T-tailed PCR product cloning vector	Promega
pGL10	Km^r, RK2/RP4 replicon, cloning vector	A. Toukdarian
pMC1403	Ap^r, promoterless *lacZY* operon, ColE1 replicon	Casadaban *et al.* (1983)
pKK223-3	Ap^r, p$_{tac}$, ColE1 replicon, cloning vector	Pharmacia
pTcC-#4	Ap^r, 10kb *Sau*3A fragment of *At. caldus* #6 cloned into the *Bgl*II site of p*Eco*R252	This study
ptacGL	Km^r, 1.4 kb *Sph*I-*Pvu*I fragment from pKK223.3 in pGL10	Tuffin *et al.* (2005)
pGEM-ChArsR	Ap^r, PCR product of arsR (bp 1049-1544) obtained using ChArsRF/ ChArsRR primers, cloned in pGEM-T	This study
pKKChArsR	Ap^r, 460 bp *Eco*RI-*Hin*dIII fragment from pGEM-arsR cloned in pKK223.3	This study
ptacChArsR	Km^r, An *arsR* containing blunted 1.58 kb *Bam*HI-*Sca*I fragment from pKKArsR cloned into a blunted *Sma*I site in pGL10	This study
pChArsRLacZ	Ap^r, PCR product of *arsR* promoter (bp 1455-1955) obtained with ChArsRLacZF/ ChArsRLacZR primers in pMC1403.	This study
pChArsBLacZ	Ap^r, PCR product of *arsB* promoter (bp 1383-1586) obtained with ChArsBLacZF/ ChArsBLacZR primers in pMC1403.	This study
pTnArs1GL	Km^r, The "transposon-like" arsenic operon isolated from *At. caldus* #6, truncated in the *tnpA* gene, in the cloning vector pGL10.	Tuffin *et al.* (2005)
TnArsRLacZ	Ap^r, *At. caldus* transposon PCR product of *arsR* promoter in pMC1403 (also named ArsRLacZ , Tuffin *et al.*, 2005)	Tuffin *et al.* (2005)
ptacTnArsR	Km^r, *At. caldus* transposon *arsR* in pGL10 (also named ptacArsR, Tuffin *et al.*, 2005)	Tuffin *et al.* (2005)
Primers*	**Sequence (5'-3')**	
ArsR-RT_Fw	CGAAGAGCACAGCGATCG	This study
ArsC-RT_Rev	CCAGGATGGAACGACAGG	This study

ORF1-RT_Rev	GATCGCGCAGCCAGAGTT	This study
ArsC-RT_Fw	CGGAAATCGAGGCCGCTT	This study
ORF5-RT_Rev	GTTTGGCAGGGATTGCGG	This study
ArsB-RT_Fw	ACCGTTCTGGCGGCGATT	This study
ORF6-RT-Rev	ATCTTCCTGGGTGCTGCC	This study
ORF5-RT_Fw	CGGGGAAGATCCTAGGTG	This study
ChArsBLacZF (*Eco*RI)	CG**GAATTC**TCGTGCTCCACCAGTAGG	This study
ChArsBLacZR (*Bam*HI)	CC**GGATCC**AAGATGGCCAGGGCAAGC	This study
ChArsRLacZF (*Eco*RI)	CG**GAATTC**AAGTCGAGCCGCAACAGG	This study
ChArsRLacZR (*Bam*HI)	CC**GGATCC**CGATCGCTGTGCTCTTC	This study
ChArsRF (*Eco*RI)	C**GAATTC**GCCGGTAGTCTTGACGAG	This study
ChArsRR (*Hin*dIII)	CCC**AAGCTT**CCAGGATGGAACGACAGG	This study
ArsRLacZR2 (*Bam*HI)	CGC**GGATCC**AGCGCGTTAGCAATCGC	Tuffin *et al.* (2005)
TnpRLacZR (*Bam*HI)	CGC**GGATCC**AGGATCGAACGGCGTG	Tuffin *et al.* (2005)
ArsRF2 (*Eco*RI)	C**GAATTC**TGATGGTTGTAGTATTATTTAA	Tuffin *et al.* (2005)
ArsRR (*Hin*dIII)	CCC**AAGCTT**CCGGTACAAAGAAACAGG	Tuffin *et al.* (2005)

*Restriction enzyme sites incorporated into primers are indicated in parenthesis and are shown in **bold** in the primer sequence.

3.2.2. DNA isolation and manipulations

Small-scale plasmid preparation, restriction endonuclease digestion, ligation and gel electrophoresis were performed using standard methods described previously (Sambrook *et al.*, 1989). Large-scale preparation of plasmid DNA was done with the Nucleobond AX100 (Macherey-Nagel) or by the alkaline lysis method followed by CsCl gradient centrifugation (Current protocols). Sequencing was done by the dideoxy chain termination method, using an ABI PRISM^{TM} 377 automated DNA sequencer and the sequence was analyzed using a variety of software programmes, but mainly by the PC based DNAMAN (version 4.1) package from Lynnon Biosoft.

3.2.3. PCR

The polymerase chain reaction was performed with primers described in Table 3.1, using 50 ng of plasmid DNA in a 50 µl volume containing 2 µM $MgCl_2$, 0.25 µM of each primer, 200 µM each of dNTP's and 1 U *Taq*I polymerase. The reaction was carried out in a Biometra thermocycler with an initial denaturation temperature of 94^0C for 60s, followed by 25 cycles of denaturation (30s at 94^0C), an annealing step of 30s, and a variable elongation step at 72^0C. Annealing temperatures and elongation times were altered as required.

3.2.4. Construction of the promoter-*lacZ* fusions

The putative promoter regions for the chromosomal *arsR* and *arsB* were amplified by PCR using the following primer pairs: ChArsBLacZF/ChArsBLacZR (bp 1417-1938) for *arsB* and ChArsRLacZF/ChArsRLacZR (bp 1400-1569) for *arsR* (Figure 3.1). The cloning of the putative promoter regions for the transposon-located *arsR* is described by Tuffin *et al.* (2005). The PCR products were digested with either *Bam*HI or *Eco*RI-*Bam*HI and ligated to the promoterless *lacZ* reporter gene of pMC1403. Fusions were confirmed by DNA sequencing.

3.2.5. β-galactosidase assays

Overnight cultures were diluted 1:200 into fresh medium containing the appropriate antibiotics, 0.4 mM IPTG, sodium arsenate or sodium arsenite (25 µM) when indicated, and incubated at 30^0C to an OD_{600} of 0.5. The β-galactosidase activities were measured using the method of Miller (1972).

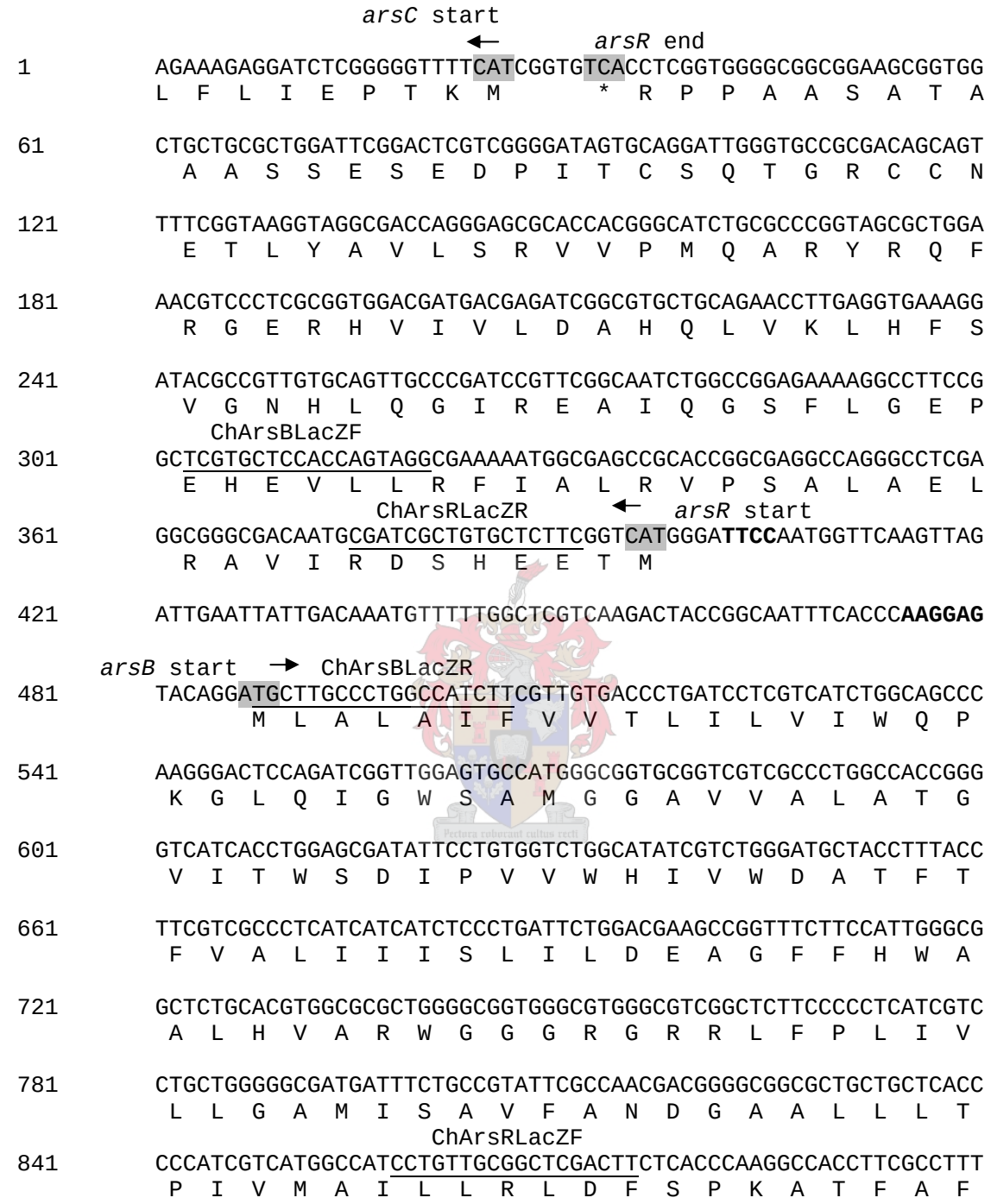


Figure 3.1: Sequence of the *arsR-arsB* intergenic region. The primers used to create the *arsR-lacZ* and *arsB-lacZ* fusions (underlined), as well as the *arsR* start and end codons and the *arsB* start codon (grey boxes) are shown. Arrows indicate the direction of transcription. Putative ribosome binding sites for the *arsR* and *arsB* are marked in bold.

3.2.6. RNA analysis and RT-PCR

Total RNA was isolated as described previously (Trindade *et al.*, 2003) from 50 ml of mid-exponential phase cultures of *E. coli* ACSH50I^q carrying various plasmids, grown in LB containing 25μM arsenate or μM arsenite or no arsenic, and with antibiotic selection. RNA was also isolated from *At. caldus* #6. Cells were first grown in tetrathionate medium without arsenite, and were then diluted 1:100 into fresh medium containing 25μM arsenite or no arsenite. For RT-PCR, the 1st Strand cDNA synthesis kit (AMV; Roche) was used for cDNA synthesis and cDNA product detection. The protocol of the manufacturer was used for the reverse-transcriptase reaction. The PCR was performed as described above using 1.5μl of the 20μl (total volume) reverse-transcriptase reaction, and the extension times were altered as required for the different primer pairs. The ORF1-RT_Rev, ArsC-RT_Rev, ORF5-RT_Rev, and ORF6-RT-Rev primers were used for cDNA synthesis from ORF1, *arsC*, ORF5 and ORF6 mRNA respectively (Table 3.1; Figure 3.5). For the PCR, the following primers were used in combination: ORF1-RT_Rev/ArsC-RT_Fw; ArsC-RT_Rev/ArsR-RT_Fw; ORF1-RT_Rev/ArsR-RT_Fw; ORF5-RT_Rev/ArsB-RT_Fw; ORF6-RT-Rev/ORT5-RT_Fw and ORF6-RT-Rev/ArsB-RT_Fw. To detect DNA contamination in the mRNA extracts, reactions were performed with each primer pair without any AMV reverse transcriptase.

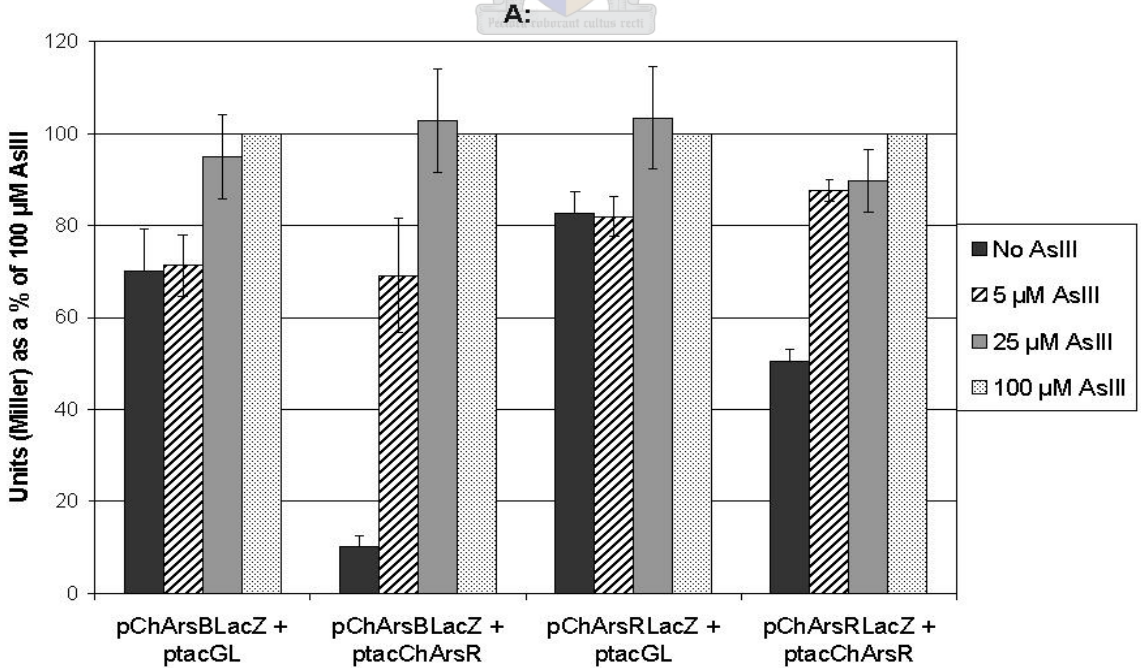
3.3. Results

3.3.1. Reporter-gene studies to investigate regulation of putative promoter regions

ArsR and ArsB promoter studies were performed in the *E. coli ars* mutant ACSH50I^q using a promoterless *lacZ* fusion construct (pMC1403) with each of the promoter fusions (pChArsRLacZ and pChArsBLacZ). Since only the first few amino acids are used when constructing an arsR-lacZ fusion, a functional ArsR would be absent. Therefore, ArsR expressed from a *tac* promoter (ptacChArsR; Table 3.1) was added in trans. β-

galactosidase assays were performed with no arsenic added, or 5, 25 or 100 μM arsenate or arsenite added. It is known that the transport of arsenate into bacterial cells is via phosphate transport systems (Sanders *et al.*, 1997). To determine if phosphate would compete with the uptake of arsenate into the cell, β-galactosidase assays with arsenate were done in LB (Luria Bertani) media and in a low phosphate media.

When transformed into the *E. coli ars* mutant ACSH50I^{q}, pChArsRLacZ and pChArsBLacZ gave β-galactosidase activity of 10-20 and 50-70 Miller units respectively. This suggested that transcription of the cloned genes from the promoter region was stronger in the direction of *arsB* than *arsRC*. In the presence of ArsR, *arsB* and *arsR* expression was repressed two-fold and twelve-fold respectively. Addition of arsenite and arsenate lifted the repression, and it was found that 25 μM of arsenite or arsenate is the best concentration to use for induction, as 5 μM did not fully induce the *ars* operon and 100 μM affected cell growth. No significant difference in induction between arsenite and arsenate was observed. Arsenate was able to induce the *ars* operon more efficiently in low phosphate media than in LB media (Figure 3.2).



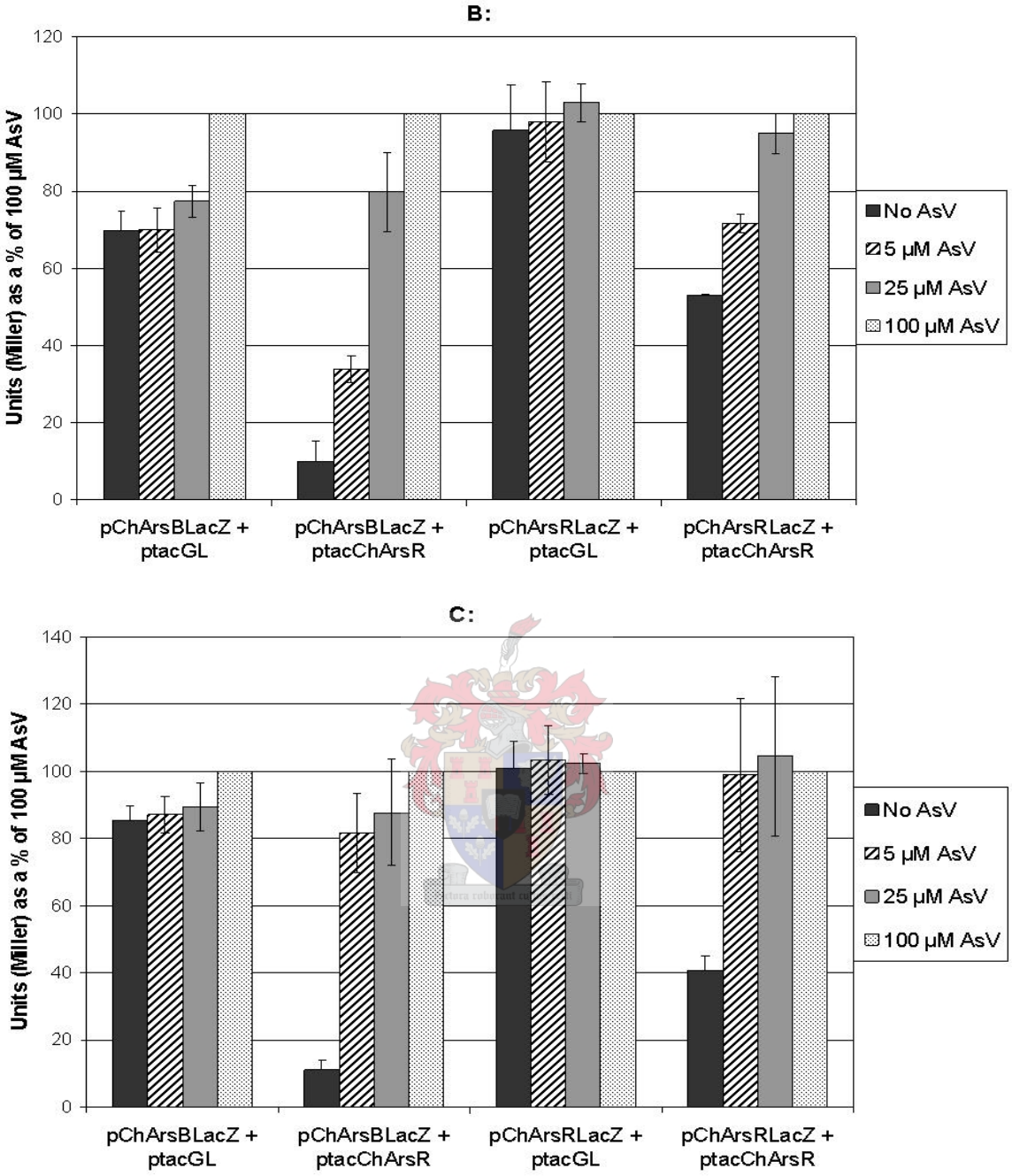


Figure 3.2: β-galactosidase activity measured in *E. coli* ACSH50I^{q} carrying the *arsR* (pChArsRLacZ) and *arsB* (pChArsBLacZ) promoter constructs in the presence of no arsenic and 5, 25 and 100 µM arsenate and arsenite. *arsR* expressed from a *tac* promoter (ptacChArsR) and a control plasmid (ptacGL10) were added in trans. Assays in arsenite were done in LB media (A), while assays in arsenate were done in LB media (B) and low phosphate media (C). Units are described as a percentage activity compared to 100 µM As^{III} and As^{V}. All assays were carried out in triplicate and error bars represent standard deviations of two or more experiments.

Sequence analysis revealed that ORF1 also showed homology to a cadmium inducible protein (CadI) of *Mycobacterium tuberculosis*. CadI proteins have been found in several different organisms and the genes encoding it are often found between *ars* genes of organisms like *Mycobacterium bovis* and *Mycobacterium tuberculosis* (Hotter *et al.*, 2001). *Moore et al.* (2005) found that 10 µM cadmium (Cd^{II}) was able to induce the *ars* operon of *Bacillus subtilis* by almost the same amount as 10 µM arsenate and arsenite. Based on these findings it was of interest to see if Cd^{II} could induce the *At. caldus* chromosomal *ars* operon.

In the presence of ArsR, *arsB* and *arsR* expression was 7.3 and 8.7 β-galactosidase units respectively. When the same constructs were expressed in the presence of Cd^{II}, β-galactosidase units did not increase significantly, indicating that ArsR still repressed *arsB* and *arsR* (Figure 3.3). These results indicated that the *At. caldus* chromosomal *ars* operon is not induced by Cd^{II}.

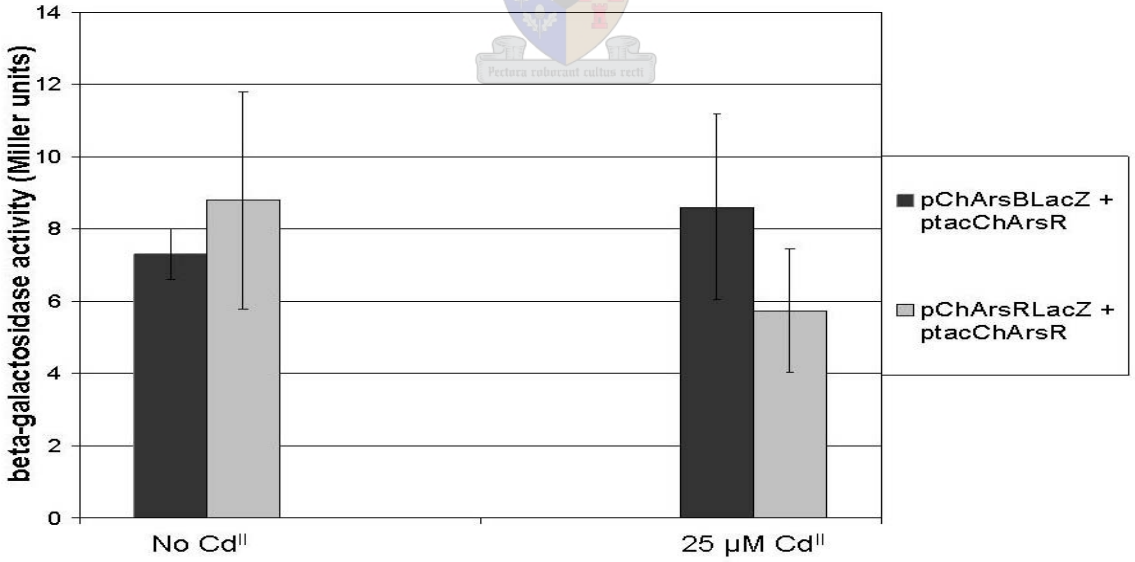


Figure 3.3: β-galactosidase activity measured in *E. coli* ACSH50I^{q} carrying the *arsR* (pChArsRLacZ) and *arsB* (pChArsBLacZ) promoter constructs in the presence of no cadmium and 25 µM cadmium. pGL10 (ptacGL10) and the *arsR* expressed from a *tac* promoter (ptacChArsR) are added in trans.

3.3.2. ORF1 and ORF5 are co-expressed with *arsCRB*

Structural gene arrangement suggests that *arsCR* may be transcribed as one transcriptional unit, separate to *arsB*. Although deletion experiments had indicated that they were not involved in arsenic resistance, because of their proximity to arsC and arsB, as well as their direction of transcription (Figure 3.5), we wanted to see if ORF1, ORF5 and ORF6 were transcribed with the *ars* genes. RT-PCR analysis was performed to determine if the *ars* genes were transcribed as a single operon and if ORF1, ORF5 and ORF6 were co-transcribed with the *ars* genes. RNA was isolated from *E. coli* ACSH50I^q transformed with pTcC-#4 grown in medium in the presence and absence of arsenite and arsenate. RNA was also extracted from *At. caldus* strain #6 grown in the presence of arsenite. The positions of the primers used in this study are shown in Figure 3.5. The RT-PCR experiment with the RNA extracted from *At. caldus* #6 was inconclusive, as some non-specific but no *ars* specific products could be obtained (data not shown). A possible reason for the non-specific products may be the degradation of the mRNA as a result of the slow growth rate of *At. caldus*. When using RNA samples extracted from *E. coli* ACSH50I^q in the presence and absence of arsenite and arsenate with the ORF1-RT_Rev/ArsC-RT_Fw primer combination, a 522 bp fragment (Figure 3.4 a) was obtained. This indicated that ORF1 is co-transcribed with *arsC*. When using the ArsC-RT_Rev/ArsR-RT_Fw primer combination, a 423 bp fragment was obtained, indicating that *arsC* is co-transcribed with *arsR* (Figure 3.4 a). A 451 bp product was obtained when primers ArsB-RT_Fw/ORF5-RT_Rev were used, showing that *arsB* and ORF5 are co-transcribed (Figure 3.4 b). If ORF5 and ORF6 were co-transcribed, a 473 bp product would have been obtained using primers ORF5-RT_Fw/ORF6-RT_Rev. No such product was visible, indicating that co-transcription of ORF5 and ORF6 did not occur (data not shown). In the absence of AMV no products were obtained, indicating that there was no DNA in the RNA preparations.

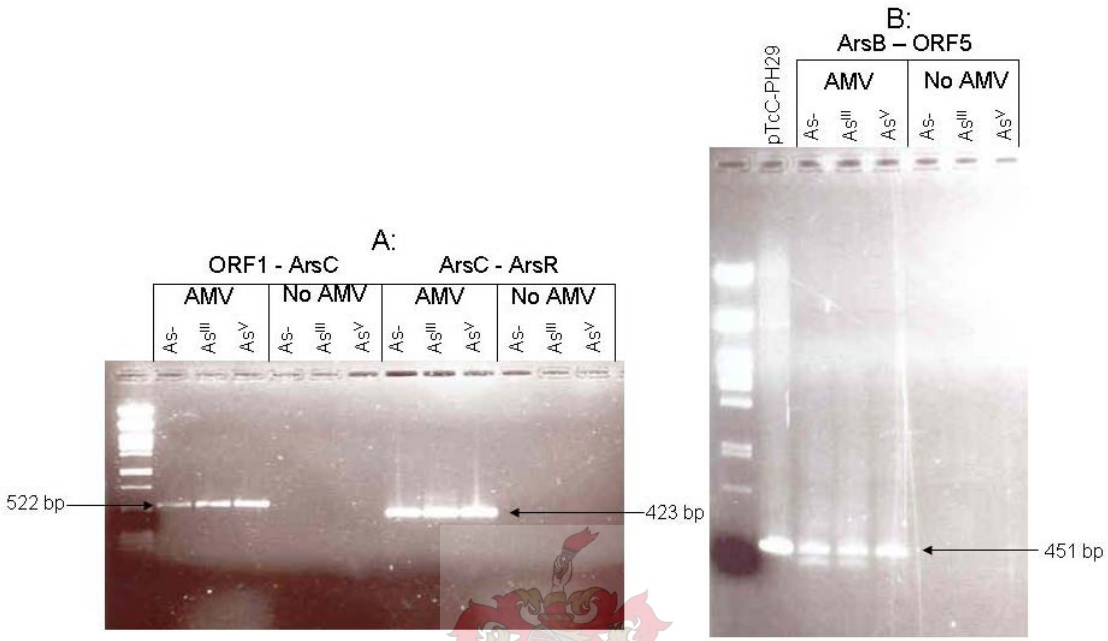


Figure 3.4: Ethidium bromide-stained agarose gels of RT-PCR amplification products. **A:** Co-transcription of ORF1-ArsC shown by the presence of a 522 bp fragment in the presence of AMV when using the ORF1-RT_Rev/ArsC-RT_FW primer combination, while co-transcription of ArsC-ArsR is shown by the presence of a 423 bp fragment when using the ArsC-RT_Rev/ArsR-RT_Fw primer combination. **B:** When using the ArsB-RT_Fw/ORF5-RT_Rev primer combination co-transcription of ArsB-ORF5 was shown by the presence of a 451 bp fragment. pTcC-PH29 plasmid DNA was included as a positive control.

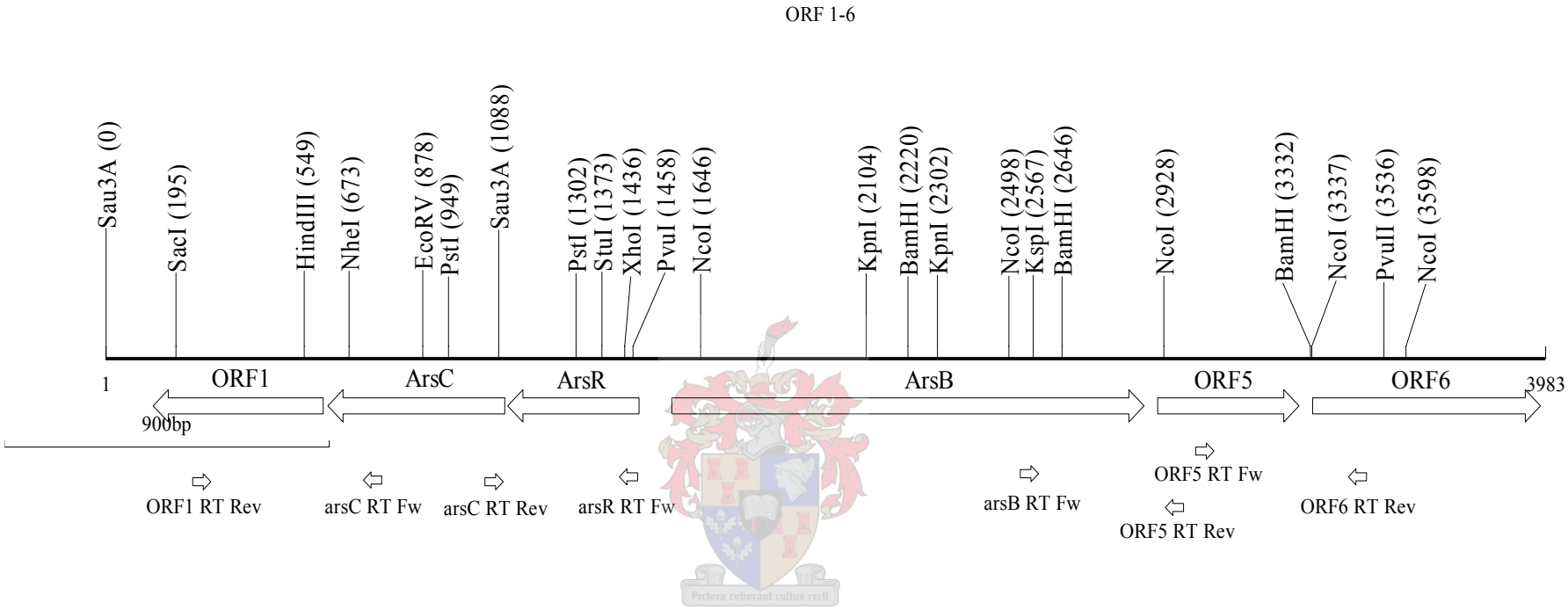


Figure 3.5: Diagram showing the restriction endonuclease map of ORF1-ORF6 as well as the directions of the ORFs. Also indicated are the positions of the primers used for the reverse transcriptase PCR experiment.

3.3.3. Cross-regulation of the transposon and chromosomal *ars* systems

To determine whether the chromosomal ArsR can regulate the transposon *ars* system and whether the transposon ArsR can regulate the chromosomal *ars* system, the transposon ArsR expressed from a *tac* promoter (ptacTnArsR) was placed in trans with the promoter constructs pChArsRLacZ and pChArsBLacZ. It was noted that the unregulated transposon *ars* operon *lacZ* fusion was expressed at much higher levels (1250-1450 Miller units) than either the chromosomal *arsB* (15-24 u) and *arsR* (40-80 u) fusions. ptacTnArsR did not affect the chromosomal *arsR* (Figure 3.6 A) nor the chromosomal *arsB* (Figure 3.6 B) expression, as β-galactosidase activities were unchanged whether ptacTnArsR was present or absent. However, when ptacChArsR and ptacTnArsR were added to their respective *arsR*-fusion constructs (pChArsRLacZ and pTnArsRLacZ) in trans, a two-fold (Figure 3.6 A) and four-fold (Figure 3.6 C) repression was observed, which was completely alleviated in the presence of arsenate and arsenite. Interestingly, when ptacChArsR was added to pTnArsRLacZ, a two-fold repression was observed, which once again was alleviated with the addition of arsenate and arsenite (Figure 3.6 C). This indicated that the transposon *arsR* gene product (ptacTnArsR) was unable to regulate the chromosomal arsenic-resistance operon, but the chromosomal *arsR* gene product (ptacChArsR) could regulate the chromosomal and transposon *ars* systems. However, the transposon *arsR* gene product was a much more effective repressor of the transposon *ars* system (four-fold repression) than the chromosomal *arsR* gene product (only two-fold repression). This indicates that the transposon ArsR is better equipped to regulate the transposon *ars* system than the chromosomal ArsR. It was noticeable that activity was higher when ptacTnArsR was added in trans to pChArsRLacZ (Figure 3.6 A). This higher activity however, was not affected by the addition of arsenate or arsenite, and is thus not relevant to this regulatory study.

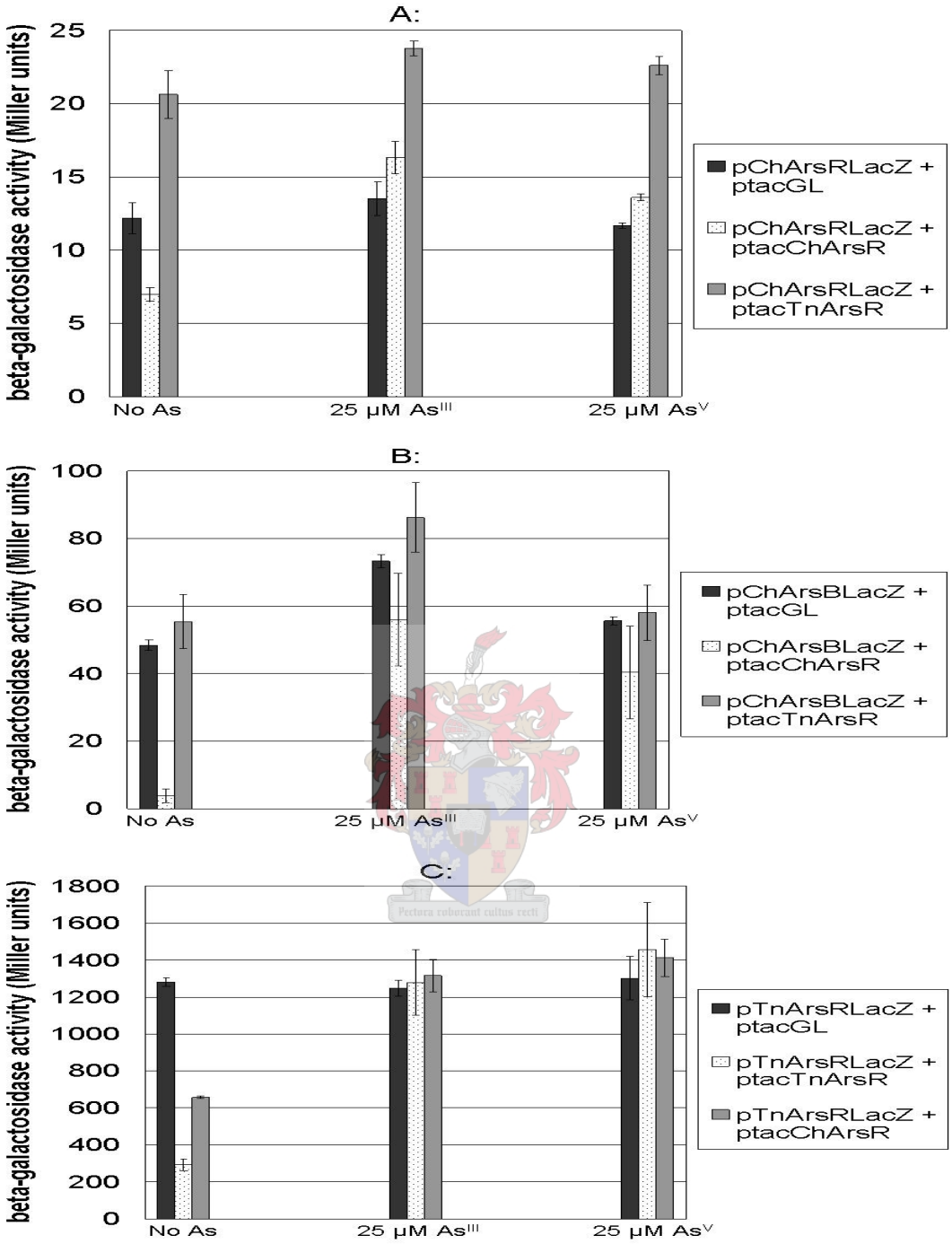


Figure 3.6: Expression of β-galactosidase of the following promoter constructs: A: the chromosomal *arsR* (pChArsRLacZ), B: the chromosomal *arsB* (pChArsBLacZ) and C: the transposon *arsR* (pTnArsRLacZ) in the presence and absence of 25 µM arsenite and arsenate. ptacGL (control), ptacChArsR (chromosomal *arsR* expressed from a *tac* promoter) and ptacTnArsR (transposon *arsR* expressed from a *tac* promoter) were added in trans. All assays were carried out in triplicate and bars represent standard deviations of three experiments.

3.4. Discussion

The chromosomal arsenic resistance operon of *At. caldus* contains the same *ars* genes as conventional chromosomal *ars* operons of e.g. *E. coli* and *Staphylococcus*, but it has an atypical arrangement with the *arsC* and *arsR* genes divergently transcribed to the *arsB* gene (Figure 2.2), although a similar arrangement was also found in *At. ferrooxidans.* In such a divergent arrangement, it is expected that the *arsRC* and *arsB* genes would be regulated separately.

RT-PCR analysis carried out with RNA isolated from *E. coli* ACSH50I^{q} transformed with pTcC-#4 revealed that ORF1, *arsC* and *arsR* are transcribed as a single transcriptional unit and that *arsB* and ORF5 are transcribed as one transcriptional unit. ORF1 showed homology to a cadmium inducible protein (CadI) of *Mycobacterium tuberculosis*. CadI proteins are often found between or adjacent to arsenic resistant genes, like in *Mycobacterium bovis*, *Bacillus subtilis* and *Mycobacterium tuberculosis* (Hotter *et al.*, 2001). It was therefore not surprising to find ORF1 transcribed with *arsC* and *arsR*. Moore *et al.* (2005) found that cadmium was able to induce the *ars* operon of *B. subtilis*. β-galactosidase analysis of ArsR-LacZ promoter fusions showed that cadmium was unable to induce the *At. caldus* chromosomal *ars* operon. This indicated that the *At. caldus* chromosomal ArsR is unable to bind cadmium and dissociate from the promoter region of the DNA like the ArsR of *B. subtilis* does. ORF5 did not show any identity to *ars* genes and therefore it was very surprising that it was transcribed with *arsB*, as it shows no connection to arsenic systems of other organisms. We have also shown in chapter 2 that it was not needed for and did not enhance resistance to arsenate or arsenite. However, these analysis were conducted in the heterologous host *E. coli* ACSH50I^{q}. No conclusions could be deduced from RT-PCR analysis performed with RNA isolated from *At. caldus* #6, possibly due to the inability to isolate mRNA as a result of the slow growth rate of *At. caldus*. Different results may be obtained when this experiment is repeated in *At. caldus*. As can be seen in Figure 3.4 a and b, ethidium bromide-stained agarose gels of RT-PCR amplification products showed the same intensity of the PCR product with RNA isolated from *E. coli* ACSH50I^{q} in the absence of arsenic as with RNA isolated when arsenate or arsenite was present. The RT-PCR

experiments however, were not quantitative and therefore do not give an indication of induction levels. Quantitative RT-PCR experiments as well as mRNA transcriptional studies using Northern hybridization will have to be performed in order to determine to what extent arsenate and arsenite induces the *At. caldus* chromosomal *ars* operon.

Using promoter fusions to a β-galactosidase reporter gene, we showed that 25 μM is the optimum concentration of arsenate and arsenite to use for induction of the *At. caldus* chromosomal *ars* operon. Five μM arsenate/arsenite was unable to fully induce the *ars* operon and 100 μM had an inhibitory effect on cell growth. Arsenate was more efficient in inducing the *ars* operon in low phosphate media than in LB media (Figure 3.2 b and c). This was expected, as it is well known that arsenate uptake is inhibited in the presence of phosphate. In the LB medium phosphate would compete with the arsenate for uptake into the cell. Less arsenate would thus be taken up when cells are grown in LB media and a true reflection of arsenic uptake and induction would not be given.

β-galactosidase assays also showed that the *At. caldus* chromosomal *ars* operon was regulated by ArsR. The *ars* operon of *E. coli* plasmid R773 has been shown to be induced by arsenite, but not arsenate (Wu and Rosen, 1993). Arsenate is first reduced to arsenite by the ArsC reductase before induction takes place. We showed that the *At. caldus* chromosomal ArsR was induced by arsenate and arsenite in the *E. coli ars* mutant ACSH50I^{q} which contained no ArsC. This suggests that the *At. caldus* chromosomal ArsR can either bind arsenate and arsenite, or that another mechanism exists that can convert arsenate to arsenite, which then binds to the ArsR to induce the *ars* operon. Results from Chapter 2 also suggested the possibility that another mechanism may exists that can convert arsenate to arsenite. Two other ArsR proteins that group with the *At. caldus* chromosomal ArsR, the *At. caldus* transposon ArsR (Tuffin *et al.*, 2005) and the *At. ferrooxidans* ArsR (Butcher and Rawlings, 2002) have also been shown to be induced by arsenate and arsenite. It may be that the metal-binding sites of these atypical ArsRs (which contains a conserved LVAYLTENCC domain) are able to bind arsenate and arsenite. The *At. caldus* chromosomal ArsR was also able to cross-regulate the *At. caldus* transposon *ars* system, but the transposon system was unable to regulate the chromosomal *ars* system. This suggested that, in spite of having amino acid sequence

identities of 45%, the chromosomal ArsR was able to bind to the DNA of the chromosomal and transposon *ars* systems, while the transposon ArsR was only able to bind to the DNA of the transposon system. It has been shown that the ArsR of *E. coli* plasmid R773 was able to regulate the *E. coli* chromosomal *ars* system and that the chromosomal ArsR was able to regulate the *ars* system of plasmid R773 (Xu *et al.*, 1996). It was thus not surprising to find that the *At. caldus* chromosomal ArsR was able to regulate the two different *ars* systems. However, it was surprising to find that the transposon ArsR was unable to regulate the chromosomal *ars* operon. Due to the fact that *arsR* sequences contain no obvious DNA binding motif (Wu and Rosen, 1991), it is difficult to speculate as to why the chromosomal ArsR can regulate the transposon *ars* system, but not visa versa. The chromosomal ArsR however, was not able to regulate the transposon *ars* system as well as the transposon ArsR, suggesting that the chromosomal ArsR is not optimally designed for the regulation of the transposon *ars* system.

By binding to the promoter sequence, ArsR functions at a transcription level. Regulatory studies were done by using *lacZ* fusions, which is both a transcriptional as well as a translational fusion. It is possible that there may be another level of regulation at the translational level, which could be masking the full regulatory potential of ArsR. This could be a reason why the fusion to *arsR* gave much lower activity than the fusion to *arsB*. Another reason for the lower activity of the fusion to *arsR* may be that the RBS of the *arsR* is much weaker than the RBS of the *arsB* (Figure 3.1). A third possible reason for the lower *arsR* expression may be that ArsR must always be made to a certain extent in order to repress the *ars* operon in the absence of arsenic. Because *arsR* and *arsC* are transcribed as one transcriptional unit, ArsC will also be made when it is not required. This low level of expression will ensure that only small amounts of ArsR and ArsC will be made in the absence of arsenic. To check for regulation on just a transcriptional level, Northern blots analysis will have to be performed.

Chapter Four

The *Acidithiobacillus caldus* transposon *ars* operon enhances arsenic resistance in strains only harboring a chromosomal copy

Contents

4.1. Introduction

As a result of the acidophilic, chemoautolithotrophic mode of life and its inhibition by organic compounds, *At. caldus* occupies an environmental niche with relatively few other bacterial species. In such an isolated niche, the amount of genetic transfer that would have taken place between *At. caldus* and other soil microorganisms, is uncertain. It is more likely that exchange would have occurred between *At. caldus* and other acidophilic organisms living in the same environment. However, the finding of a broad-host-range mobilizable plasmid (pTF-FC2) in *At. ferrooxidans* (which occupies the same ecological niche as *At. caldus*) may indicate that these microorganisms are not as genetically isolated as was initially thought (Rawlings *et al.*, 1993). Research into the molecular biology of these biomining bacteria is of great importance in order to optimize the biomining process. This research however, is inhibited greatly by a lack of an easily usable genetic system.

Currently two mechanisms exist to transfer genetic material into *At. ferrooxidans*. The first mechanism, electroporation, was achieved by Kusano *et al.* (1992) using natural plasmids from *At. ferrooxidans* as the vector and the *merA* gene (mercury reductase) as a selectable marker. Researchers were able to reisolate recombinant plasmids, and restriction enzyme cleavage patterns indicated that no plasmid rearrangements had occurred. They also showed that an IncQ-type of plasmid replicon was able to reproduce in *At. ferrooxidans*. However, this transformation procedure had several difficulties. Of 30 strains electroporated, only one gave transformants, and at a very low frequency. Selection for mercury resistance was also problematic because *At. ferrooxidans* cells took a long time to form colonies and mercury is naturally very volatile.

The second mechanism of gene transfer into *At. ferrooxidans* concerns the use of conjugation, using a procedure described by Jin *et al.* (1992). Matings were carried out using filters on plates containing an inorganic thiosulphate medium with 0.05% yeast extract added. Using this method, IncP, IncQ and IncW plasmids could be conjugated into *At. ferrooxidans* recipients (Peng *et al.*, 1994). Conjugation frequency was improved

by optimizing donor to recipient ratios as well as pH and medium composition (Liu *et al.*, 1999).

No gene transfer systems for *At. caldus* have been reported. This limits research into the molecular biology of *At. caldus* considerably. Although the external pH environments of *At. caldus* and *E. coli* are very different, the internal pH and cytoplasmic environment of the two bacteria are similar. There is therefore a good probability that genes from other bacteria will function in *At. caldus*. Based on evidence presented by de Groot *et al.* (2003) it is known that three *At. caldus* strains (#6, "f" and MNG) contain *ars* genes carried on a transposon as well as a chromosomal copy of *ars* genes. Three *At. caldus* strains (BC13, KU and C-SH12) contain only a chromosomal set of *ars* genes. It has been shown that the *ars* genes located on the transposon confer high levels of resistance to arsenate and arsenite (Tuffin *et al.*, 2004). It is known that, due to selective pressure, the biomining bacterial population became more resistant to high levels of arsenic (Rawlings and Woods, 1995). It is possible that an increase in resistance of *At. caldus* was obtained by the acquisition of the transposon systems. Based on the above, it was of interest to see if the acquisition of a transposon operon by the *E. coli ars* mutant ACSH50I^{q} containing only the chromosomal *ars* genes would increase arsenic resistance in *E. coli*. A gene transfer system for *At. caldus* will enable the introduction of the transposon *ars* operon to *At. caldus* strains harboring only the chromosomal *ars* genes. This will enable the effect the transposon *ars* system has on *ars* resistance to be determined and might shed light as to why certain strains of *At. caldus* acquired the transposon system.

Here I report on the levels of arsenic resistance conferred to *E. coli* by the *At. caldus* chromosomal and transposon *ars* systems individually and in combination. I also report on the introduction of the *At. caldus* #6 transposon *ars* system to a strain of *At. caldus* harboring only a chromosomal copy of *ars* genes by conjugation and on changes in levels of arsenite resistance after the acquisition of the transposon operon.

4.2. Methods

4.2.1. Media, bacterial strains and plasmids

Bacterial strains, plasmids and primers used in this study are shown in Table 4.1. *E. coli* strains were grown in Luria-Bertani (LB) broth medium (Sambrook *et al.*, 1989) with ampicillin (100 μg/ml) or kanamycin (50 μg/ml) added as required. *At. caldus* strains were grown at 37^0C in three different types of media: (a) tetrathionate medium (3 mM), sterilized and adjusted to pH 2.5 (Rawlings *et al.*, 1999b), (b) sodium thiosulfate medium, sterilized and adjusted to pH 4.6, or (c) elemental sulfur (S^0) medium (Jin *et al.* 1992), sterilized and adjusted to pH 4.6. For the sodium thiosulfate medium, the same basal salts and trace elements as the tetrathionate medium were used, but it contained 50 mM $Na_2S_2O_3$ instead of 50 mM $K_2S_4O_6$. For the S^0 medium, elemental sulfur was sterilized by adding 0.5 g S^0 to 5 ml water and heated to 105°C for 1 h on two successive days. The sterilized sulfur was then added to the basal medium (0.5 g/l).

Table 4.1. Bacterial strains, plasmids and primers used in this study

Strain, plasmid or primer	Description	Source or reference
Strains		
At. caldus		
#6	Fairview mine, Barberton, S. Africa	Rawlings *et.al*. (1999b)
BC13 (ATCC 51756)	Birch coppice, Warwickshire, UK	Hallberg & Lindström (1994)
KU (DSM 8584)	Kingsbury coal spoil, UK	Hallberg & Lindström (1994)
C-SH12 (DSM 9466)	Continuous bioreaktor, Brisbane, Australia	Goebel & Stackebrandt (1994)
E. coli		
DH5α	Φ80d*lacZ*ΔM15, *endA*1, *recA*1, *gyrA*96, *thi*-1, *hsdR*17 (r_k^-, m_k^+), *relA*1, *supE*44, *deoR*,Δ(*lacZYA-argF*) U169	Promega
XL 1 Blue	*recA*1, *endA*1, *gyrA*96, *thi, hsdR*17 (r_k^-, m_k^+), *relA*1, *supE*44, *relA*1, lac [F' *proAB*+ *lacI*^{q}ZΔM15	Stratagene

	::Tn10(tet^r)]	
ACSH50I^q	*rpsL*, Δ(*lac-pro*), Δ*ars*::*cam*	Butcher & Rawlings (2002)
HB101	F$^-$, Δ(*mcrC-mrr*), *hsdS20*, *recA13*, *ara-14*, *proA2*, *lacY1* λ$^-$, *galK2*, *rpsL20*(Sm^r), *xyl-5*, *mtl-1*, *supE44*	Boyer and Roulland-Dussoix (1969)
CSH56	F$^-$, *ara*, Δ(*lac pro*) *supD*, *nalA*, *thi*	Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Plasmids		
p*Eco*R252	Ap^r, *Eco*RI inactivation cloning vector	Zabeau & Stanley (1982)
pBluescript SK	Ap^r, *LacZ*', ColE replicon, cloning vector	Stratagene
pGL10	Km^r, RK2/RP4 replicon, cloning vector	A. Toukdarian
pSa	Km^r, Cmr, Spr, IncW replicon, mobilizing plasmid	Tait *et al.* (1982)
pTcC-PH29	Ap^r, A blunted 2.9kb *Pvu*II-*Hind*III fragment from pTcC-#4 cloned into a blunted *Bgl*II site of p*Eco*R252	This study
pEcoBlunt	Ap^r, p*Eco*R252 blunted at the *Bgl*II site and relegated to inactivate the *Eco*RI endonuclease	Tuffin *et al.* (2005)
pTnArs1GL	Km^r, The "transposon-like" arsenic operon isolated from *At. caldus* #6, truncated in the *tnpA* gene, in the cloning vector pGL10.	Tuffin *et al.* (2005)
pTn2	As^r Km^r, transposon transconjugant isolated from a mating experiment performed in *E. coli*	Tuffin *et al.* (2005)
Primers*	**Sequence (5'-3')**	
RNAarsAF	AATACCGCGAACGCATCG	Tuffin *et al.* (2005)
RNAarsAR	AGAAGAGGCTGCTCAGC	Tuffin *et al.* (2005)
fDD2 (*Bam*HI)(*Sal*I)	CC**GGATCCGTCGAC**AGAGTTTGATCITGGCTCAG	Rawlings *et al*. (1995)
rPP2 (*Hin*dIII)(*Xba*I)	CC**AAGCTTCTA**GACGGITACCTTGTTACGACTT	Rawlings *et al*. (1995)

*Restriction enzyme sites incorporated into primers are indicated in parenthesis and are shown in **bold** in the primer sequence

4.2.2. DNA isolation and manipulations

Small-scale plasmid preparation, restriction endonuclease digestion, ligation, gel electrophoresis and Southern blot hybridization were performed using standard methods described previously (Sambrook *et al.*, 1989). Large-scale preparation of plasmid DNA was done with the Nucleobond AX100 (Macherey-Nagel) or by the alkaline lysis method followed by CsCl gradient centrifugation (Current protocols). Labeling of probes, hybridization and detection was performed by using a dioxigenin-dUTP nonradioactive DNA labeling and detection kit (Roche). Total DNA was extracted from various *At. caldus* strains as follows: *At. caldus* cells were harvested by centrifugation, washed 3 times in acidified water (pH 1.8) and resuspended in TE buffer (pH 7.6). Cell lysis was done with 10% SDS in the presence of proteinase K (20 mg/ml) and lysozyme (2 mg/ml) at 37^0C for 1h and then at 50^0C for 1h. 7.5 M NH_4Ac was added and to the cell lysates and it was left at room temperature for 45 min. Precipitation of DNA was done with ethanol, washed twice in 70% ethanol and resuspended in TE buffer (pH 7.6).

4.2.3. PCR

The polymerase chain reaction was performed with primers described in Table 1, using 50 ng of chromosomal DNA in a 50 µl volume containing 2 µM $MgCl_2$, 0.25 µM of each primer, 200 µM each of dNTP's and 1 u *Taq*I polymerase. The reaction was carried out in a Biometra thermocycler with an initial denaturation temperature of 94^0C for 60 s, followed by 25 cycles of denaturation (30s at 94^0C), an annealing step of 30 s, and a variable elongation step at 72^0C. Annealing temperatures and elongation times were altered as required.

4.2.4. Arsenic resistance assays

To determine the growth of *At. caldus* in the presence of arsenite, cells were cultured in tetrathionate medium containing different concentrations of arsenite. Actively growing cultures were diluted 100-fold into fresh media and incubated for 19 days, and the absorbance at OD_{600nm} was determined every second day. The growth of *At. caldus* in the presence of arsenate was not tested, as the phosphate concentration in the growth medium would contribute to arsenate resistance (Silver *et al.* 1981). To test for arsenic resistance

in *E. coli,* constructs were transformed into competent *E. coli* ACSH50I^{q} cells. Growth assays to determine the resistance to arsenate (As^{V}) were performed in low-phosphate medium (Oden *et al.*, 1994) supplemented with 2 mM K_2HPO_4. Overnight cultures were diluted 100-fold into fresh medium containing the appropriate antibiotics and 1.5 mM sodium arsenate. The cultures were incubated at 37^{0}C and the absorbance at $OD_{600\ nm}$ was determined every hour until stationary phase was reached.

4.2.5. Mating assays

Donor (*E. coli* HB101 with pSa or pTn2) and recipient (*E. coli* CSH56 and *At. caldus* BC13, KU and C-SH12) strains were cultured separately in sodium thiosulfate medium (pH 4.6) with appropriate antibiotic selection. For *E. coli* HB101 and CSH56 the medium contained 0.05% (m/v) yeast extract. Cells were washed three times in 0.8% (w/v) NaCl solution and mixed in a donor/recipient ratio of 1:5. The solid medium was prepared in two parts, A (double-strength basal salts at pH 4.8 [Rawlings *et al.*, 1999b] and agar) and B (2% [w/v] sodium thiosulfate [$Na_2S_2O_3$]) which were combined in equal volumes after autoclaving and cooling to 50^{o}C. The solid medium was supplemented with 0.05% (w/v) yeast extract when used as a mating medium for *E. coli* and *At. caldus*. A 0.22 micron filter (Osmonics Inc. A02SP02500) was placed on the solid media and 200 µl of the donor/recipient mixture was spotted on the filter. The plate containing the filter and mixture was incubated at 37 ^{o}C for 11 days. The filter was washed in 5 ml of 0.8% NaCl solution, and vigorously shaken to dislodge mating cells. Cells were pelleted by a 2-min spin in a mirocentrifuge and resuspended in 1 ml of 0.8% NaCl solution. For the *E. coli*/*E. coli* matings the solution was plated onto Luria agar (LA) media that selected for the transconjugant cells, while the *E. coli*/*At. caldus* mating solutions were transferred to the sodium thiosulfate solid and liquid media without yeast extract, with selection for transconjugant cells.

4.3. Results

4.3.1. Difference in effectiveness of the chromosomal and transposon arsenic operons

To compare arsenic resistance of the transposon and chromosomal *ars* operons, assays were performed with *E. coli* ACSH50I^q (pTnArs1GL + pEcoBlunt) and *E. coli* ACSH50I^q (pTcC-PH29 + pGL10). Resistance conferred by the individual operons was also compared to that conferred by the two operons when present on compatible co-resident plasmids in *E. coli* ACSH50I^q (pTnArs1GL + pTcC-PH29). Growth curves were performed in the presence of 1.5 mM arsenate and in the absence of arsenate. The growth curves without arsenate served as a control to show that *E. coli* ACSH50I^q transformed with the different plasmids grew at the same rate. Growth curves in the presence of arsenite was not tested as the presence of arsenate tested for the functionality of both ArsC (converting arsenate to arsenite) and ArsB (exporting the arsenite out of the cell).

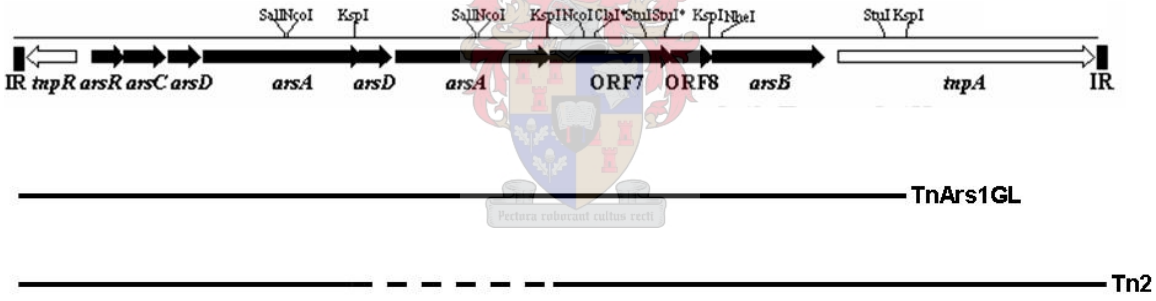


Figure 4.1: Genetic map of the *At. caldus* #6 transposon copy of arsenic resistance genes and subclones used in this study. The dotted line indicates a spontaneous *arsDA* deletion. IR, inverted repeats. The * indicates a methylated restriction enzyme site.

Results shown in Figure 4.2 indicate that in the absence of arsenate growth of all three strains was approximately the same irrespective of the plasmids that it contained or their selection. The transposon copy of arsenic resistance genes conferred much more resistance to arsenate than the chromosomal copy. The difference in resistance is not a true reflection however, due to the fact that pTnArs1GL was carried on the low-copy vector pGL10, while pTcC-PH29 (contains the chromosomal *arsRBC5*) was carried on the moderate-copy vector pEcoR252. The difference would have been bigger if both *ars*

operons were on the same copy number vector. When both operons were placed together in the *E. coli* ACSH50I^{q} *ars* mutant, no significant difference was seen between cells harboring both operons and cells harboring only the transposon operon (Figure 4.2).

4.3.2. Addition of the transposon *ars* system to strains only carrying the chromosomal *ars* genes

It has been shown that all six strains of *At. caldus* have a basic chromosomal arsenic resistance system, possibly an *arsRBC* system. The three strains isolated in South Africa (#6, MNG and "f") however, have acquired a transposon-located arsenic resistance mechanism, as all three isolates gave a positive hybridization signal using the transposon-borne *arsDA*, isolated from *At. caldus* #6, as a probe. The non-South African strains (KU, BC13 and C-SH12) did not give a signal when probed with *arsDA* (de Groot *et al.*, 2003; Tuffin *et al.*, 2004). It is also known that the transposon *ars* system confers higher resistance to arsenic than the chromosomal system. It was of interest to see if we could transfer the transposon *ars* system to one of the *At. caldus* strains that only contains the chromosomal *ars* system, and to determine whether an increase in arsenic resistance would be acquired.

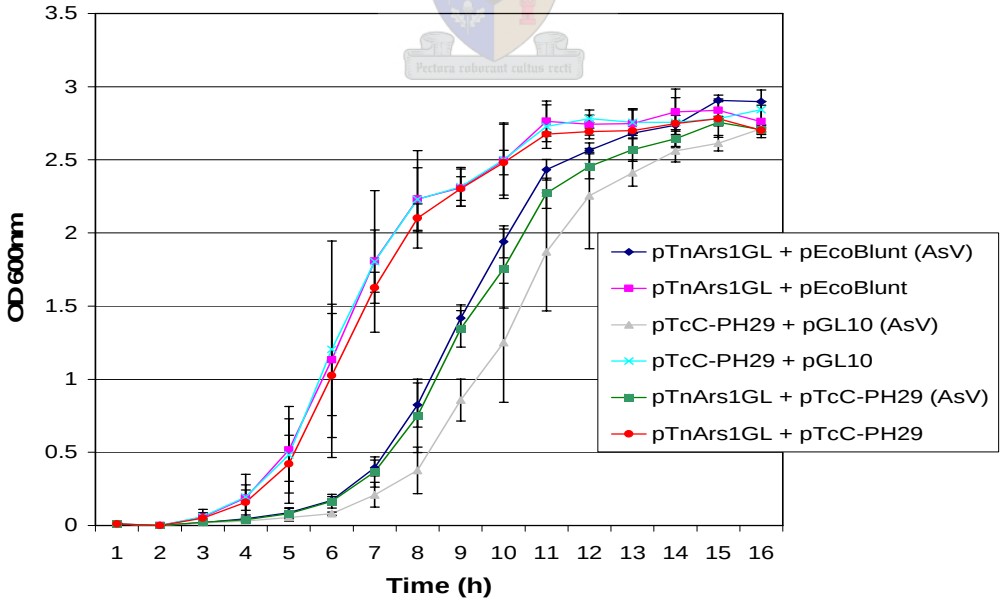


Figure 4.2: Growth curves performed in the *E. coli ars* mutant ACSH50I^{q} in the presence and absence of 1.5 mM arsenate. Each data point represents the results of two assays of two independent experiments. Error bars indicate standard deviations. Optical density (OD) readings were recorded at 600 nm.

E. coli HB101 containing either the broad host range conjugative plasmid pSa or pTn2 (pSa into which TnAtcArs had been transposed, Tuffin *et al*., 2005) was mated with *At. caldus* strains BC13, KU and C-SH12 for 11 days. *At. caldus* transconjugants were then selected in inorganic S_2O_3 media containing Km at pH 4.6. All of the *At. caldus* strains were able to grow on 100 mg/ml kanamycin (which selects for pSa). Difficulty was experienced in getting *At. caldus* to form well-isolated colonies on plates. Therefore total DNA was extracted and a PCR experiment using the transposon *arsA* primers (RNAarsAF/RNAarsAR) was conducted (Table 4.1, Figure 4.3). A 600 bp product was observed when using C-SH12-pTn2 total DNA as template, which corresponds to the size of the product the ArsA primer pair would form. This indicated that the transposon copy of *ars* genes (pTn2) did transfer from HB101 to C-SH12. No products were formed when using total DNA from BC13-pTn2 and KU-pTn2 as template. It is possible that the Km resistance observed in *At. caldus* BC13-pTn2 and KU-pTn2 was due to the breakdown of Km in the media.

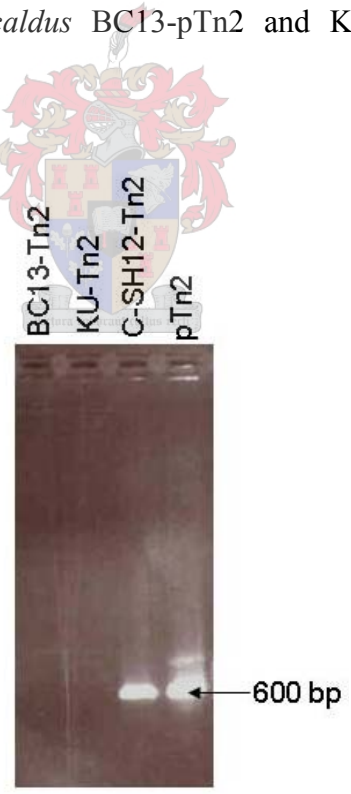


Figure 4.3: An ethidium bromide-stained agarose gel showing PCR amplification products when using an RNAarsAF/RNAarsAR primer combination. A product was only obtained for C-SH12-Tn2 and the positive control pTn2.

To confirm that *At. caldus* was the organism that grew in the S_2O_3 media and that no contamination occurred, the chromosomal DNA that was extracted from *At. caldus* strains C-SH12, BC13 and KU grown in the presence of 100 mg/ml kanamycin was used and a 16S rDNA PCR was performed. The PCR product was digested with *Bam*HI and *Stu*I (Figure 4.4). The fragments obtained indicated that the 16S rDNA was that of *At. caldus* and not a possible contaminating organism, or the *E. coli* HB101 donor strain.

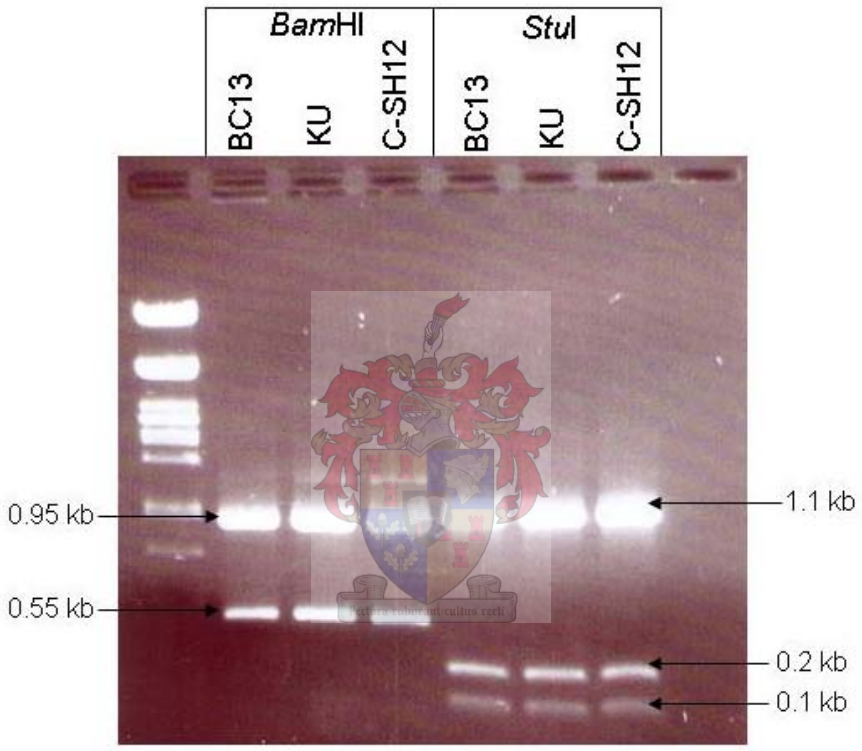


Figure 4.4: An ethidium bromide-stained agarose gel showing the 16S rDNA PCR amplification products when using chromosomal DNA isolated from *At. caldus* strains BC13, KU and C-SH12, digested with *Bam*HI and *Stu*I.

A Southern blot experiment was carried out where chromosomal DNA from *At. caldus* C-SH12 (pSa) and *At. caldus* C-SH12 (pTn2) transconjugants as well as pSa and pTn2 plasmid DNA were digested with *Ksp*I and probed with pSa and the transposon-borne *arsD*. When probing with *arsD* a signal was obtained only for lanes containing *At. caldus* C-SH12 (pTn2) and pTn2, indicating that the transposon copy did transfer and that it is not present in untransformed *At. caldus* C-SH12 and in *At. caldus* C-SH12 (pSa)

transconjugants (Figure 4.5 a). When probed with total pSa (Figure 4.5 b), several fragments were obtained for *At. caldus* C-SH12 (pSa) and *At. caldus* C-SH12 (pTn2) transconjugants, as well as for pSa and pTn2 plasmid DNA, and as expected no signal for untransformed *At. caldus* C-SH12 DNA. With *At. caldus* C-SH12 (pTn2) and pTn2 DNA an additional 5 kb *Ksp*I signal was obtained that was not obtained with *At. caldus* C-SH12 (pSa) and pSa DNA. This 5 kb signal represents a *Ksp*I fragment stretching from a *Ksp*I site in TnAtcArs to a *Ksp*I site in pSa, and therefore why it was not obtained in DNA extracts containing pSa only.

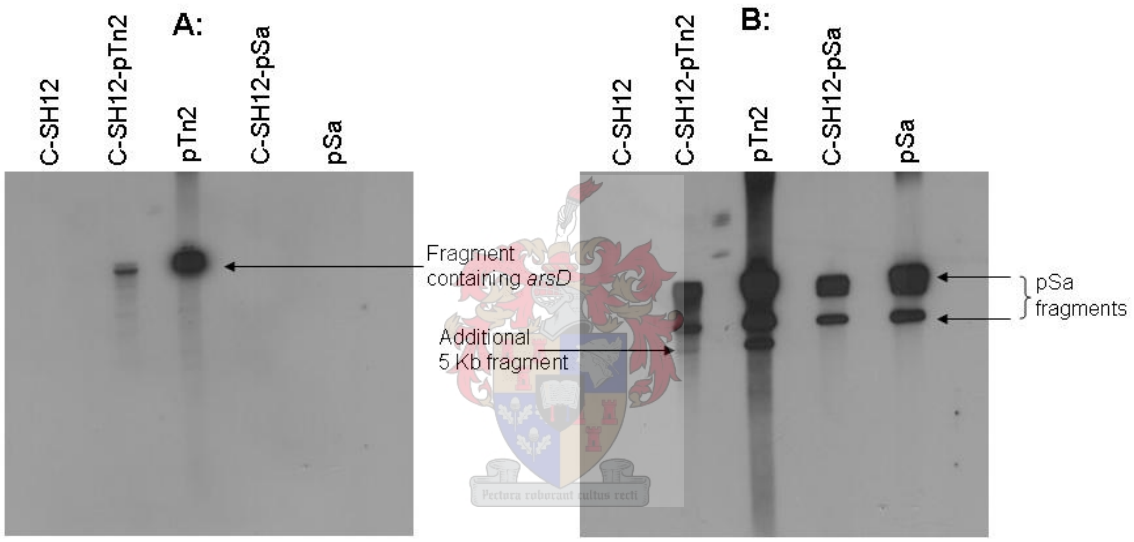


Figure 4.5: Southern blot experiment where *At. caldus* C-SH12 (pTn2) and *At. caldus* C-SH12 (pSa) transconjugants as well as untransformed *At. caldus* C-SH12 and pTn2 and pSa DNA were digested with *Ksp*I and probed with (A) the transposon *arsD* and (B) pSa.

Next we tested whether *At. caldus* C-SH12 (pTn2) transconjugants would be more resistant to arsenic than *At. caldus* C-SH12 transformed with just pSa and also to show that the addition of pSa does not have an effect on arsenic resistance. Growth curve experiments were done where *At. caldus* C-SH12 (pSa) and *At. caldus* C-SH12 (pTn2) transconjugants were grown in the presence of 0, 20, 30, 40 and 50 mM arsenite. After eight days, *At. caldus* cells containing the transposon copy of arsenic resistance genes (C-SH12-pTn2) were able to grow on 30 mM arsenite, while the *At. caldus* cells harboring only the chromosomal *ars* genes (C-SH12-pSa) showed weaker growth, even at 20 mM

arsenite. After 13 days, *At. caldus* C-SH12 (pTn2) was able to grow on 50 mM arsenite, while *At. caldus* C-SH12 (pSa) was sensitive to 30 mM arsenite. After 19 days C-SH12-pSa was still sensitive to 30 mM arsenite (Figure 4.6). These results indicate that the addition of the transposon *ars* operon of *At. caldus* strain #6 to *At. caldus* C-SH12 increased the resistance to arsenite more than two-fold.

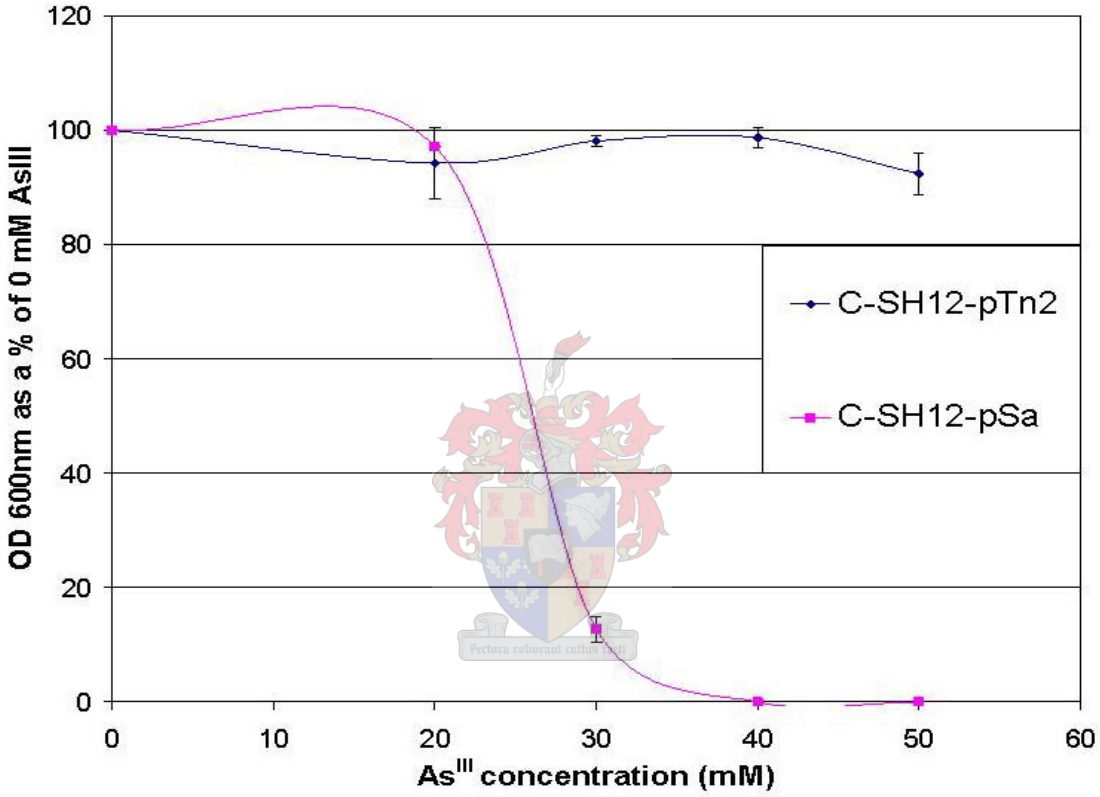


Figure 4.6: Arsenite growth curves performed with *At. caldus* C-SH12 (pTn2) and *At. caldus* C-SH12 (pSa) transconjugants in the presence of various concentrations of arsenite. Cell densities were determined (OD 600nm) and a percentage resistance was calculated (growth in the absence of arsenite was equivalent to 100%) after 19 days. Each data point represents duplicate results of at least two experiments. Error bars indicate standard deviations.

4.4. Discussion

It is known that *At. caldus* #6 has an *ars* operon carried on a transposon (de Groot *et al.*, 2003; Figure 4.1) and in this study we identified the chromosomal *ars* operon of this organism. Tuffin *et al.* (2004) showed that the *ars* genes on the transposon confer high levels of resistance to arsenate and arsenite. Growth curve experiments conducted with the transposon and chromosomal copy of *ars* genes in *E. coli* ACSH50I^{q} revealed that the transposon system conferred higher levels of resistance to arsenate than the chromosomal system. To determine arsenate resistance levels when both systems were in the same organism, growth curves were conducted where the chromosomal and transposon systems were placed together in trans in *E. coli* ACSH50I^{q}. No significant difference was observed between cells harboring both systems and cells harboring only the transposon system, indicating that the two operons did not have a synergistic effect on arsenate resistance and that resistance levels were similar to those of the transposon. However, this study was done in *E. coli*, and these results may not be a true reflection on the levels of resistance in *At. caldus*.

It is well known that the bacterial biomining population evolved to become more resistant to high levels of arsenic over a short period of time (Rawlings and Woods, 1995). In order to demonstrate that the acquisition of the transposon *ars* system resulted in the increase of arsenic resistance in *At. caldus*, a system had to be developed where the transposon copy of *ars* genes could be introduced into *At. caldus* strains only having the chromosomal *ars* operon (BC13, KU and C-SH12). Using a conjugational system, the transposon *ars* system (carried on the conjugative pSa) was transferred from *E. coli* HB101 to *At. caldus* C-SH12. Transformants were selected with kanamycin. Southern bolt hybridization (Figure 4.5) and PCR analysis (Figure 4.3) confirmed the transfer of the transposon *ars* system from *E. coli* HB101 to *At. caldus* C-SH12. The banding pattern of the Southern blot experiment also indicated that no plasmid rearrangements occurred during the conjugation process. We were unable to transfer the transposon *ars* system to *At. caldus* BC13 and KU. Although kanamycin resistance was also observed for *At. caldus* KU and BC13, it is most likely due to the breakdown of the kanamycin following extended aeration at low pH. Studies conducted by Hallberg and Lindstrom

(1996) showed that antiserum generated against whole cells from *At. caldus* KU recognized whole cells of isolates BC13 and KU, but not C-SH12. The authors propose that this lack of recognition by the antiserum is most likely due to the fact that the lipopolysaccharide (LPS) of isolate C-SH12 differs from the LPS of BC13 and KU. This means that *At. caldus* C-SH12 can be considered to belong to a different serotype from that of BC13 and KU. A difference in cell structure may thus be the reason why we were unable to transfer DNA to strain KU and BC13.

Growth curve experiments with *At. caldus* C-SH12 (pSa) and *At. caldus* C-SH12 with the introduced transposon operon (pTn2) were conducted to determine if the acquisition of the transposon operon could have resulted in the increase in arsenic resistance of certain *At. caldus* strains. Introduction of the transposon system to *At. caldus* C-SH12 resulted in an increase of arsenite resistance of more than two-fold. In the presence of arsenite the growth rate of the culture of transconjugants was also much faster than C-SH12 cells with only the chromosomal *ars* system. The increased levels of resistance and rate of growth had a dramatic effect on growth in the presence of arsenite (Figure 4.6). Based on these results we speculate that *At. caldus* #6 became more resistant to arsenic most likely by the acquisition of the transposon *ars* system.

The development of a conjugation system to transfer genes from *E. coli* to *At. caldus* C-SH12 is a major breakthrough in the study of the molecular biology of *At. caldus*. Optimization of this system may be achieved by optimizing donor to recipient ratios, pH and medium composition. Researchers in our laboratory are currently developing a mechanism to introduce gene knockouts in *At. caldus*. This system, together with the conjugation system, will enable us to study the molecular biology of biomining bacteria in greater depth than is currently possible. This could enable us to genetically engineer strains that could grow faster, at lower pH, at higher temperatures, at higher cell densities and with smaller quantities of added nutrients (Rawlings and Kusano, 1994) in order to achieve great economical savings in the biomining industry.

Chapter 5
General Discussion

Arsenic resistance (*ars*) systems had been identified in many different bacteria. Two common forms of the *ars* operon have been identified. The most common form consists of three genes (*arsRBC*) and is found on plasmids of *S. xylosus* (pSX267) and *S. aureus* (pI258). The second form consists of five genes (*arsRDABC*) and was found on plasmids R773 and R46 of *E. coli* and more recently on transposons of *At. caldus* (AY821803), *L. ferriphilum* (DQ057986), *A. faecalis* (Ay297781) and M. *flagellatus* (NZ_AADX01000013) (Tuffin *et al.*, 2006). During the sequencing of the *E. coli* chromosome, the first chromosomally encoded arsenic resistance genes were identified (Diorio *et al.*, 1995). Since then, with more chromosomal sequencing projects underway, arsenic resistance genes have been identified on many more bacterial chromosomes, and generally take the form *arsRBC*.

In a previous study two copies of transposon-linked *ars* operons had been identified in *At. caldus* strain #6 (Figure 2.1 a and b; Tuffin *et al.*, 2004). Using Southern blot analysis de Groot *et al.* (2003) identified a possible chromosomal *ars* operon in *At. caldus* strain #6 that was different to the *ars* genes in the transposon. The aim of this study was to isolate and sequence the third *ars* operon. Once characterized, expression and regulatory studies were performed and possible interaction between the transposon-located and chromosomal *ars* operons was analyzed.

Sequence analysis of the *At. caldus* #6 chromosomal *ars* operon revealed an *arsCRB* operon. This *ars* operon however, differs from the conventional chromosomal *arsRBC* operons of *E. coli* and *Staphylococcus* spp. where the operon is transcribed as one transcriptional unit (Diorio *et al.*, 1995). The *At. caldus* chromosomal *ars* operon is bi-directional, with the *arsC* gene located upstream of the *arsB* gene and the *arsC* and *arsR* genes divergently transcribed to *arsB*. This divergent arrangement of the *ars* genes raised the possibility that transcription in the two directions might be regulated differently.

In *ars* systems where an ArsA is absent (as in the *At. caldus* chromosomal *ars* operon), it is believed that the ArsB uses membrane potential rather that ATP hydrolysis as an energy source to export arsenite out of the cell. However, with acidophilic microorganisms like *At. caldus* this is likely to be different, as they have a reversed membrane potential, with a positive rather than negative inside membrane potential (Martin, 1990). Nevertheless, the *At. caldus ars* genes conferred enhanced resistance to arsenate and arsenite when cloned in *E. coli*. The mechanism how arsenite is exported by ArsB in this acidophilic organism lacking an ArsA is still unknown.

Alignment studies showed that the *At. caldus* chromosomal ArsC protein grouped with ArsC proteins of the Trx family (Figure 2.9). This suggested that this ArsC would use thioredoxin rather than glutathione as a reducing power to convert arsenate to arsenite. When tested to see if thioredoxin aids the ArsC in converting arsenate to arsenite, it was surprising to find that the *At. ferrooxidans* thioredoxin gene (*trxA*) was unable to help ArsC convert arsenate to arsenite. Based on the alignment studies and the high identity to the ArsC of *At. ferrooxidans*, it is likely that thioredoxin rather than glutathione is used by the *At. caldus* chromosomal ArsC as a reducing power. A possible explanation is that the *At. ferrooxidans trxA* gene may not complement the *At. caldus* chromosomal ArsC to reduce arsenate to arsenite.

Recently the crystal structures of ArsC proteins of the Trx and GSH/Grx clades have been determined and mechanisms of arsenate reduction predicted (Martin *et al.*, 2001; Zegers *et al.*, 2001). Sequence analysis of the *At. caldus* chromosomal ArsC revealed that it contains, like other ArsC proteins of the Trx clade, four cysteine residues (Cys^{10}, Cys^{15}, Cys^{82} and Cys^{90}) with spacing similar to those of other ArsC proteins of the Trx clade. Based on the above mentioned evidence, we propose that the *At. caldus* chromosomal ArsC will function similar to the *B. subtilis* ArsC (Figure 1.7).

RT-PCR analysis done in *E. coli* revealed that ORF1, *arsC* and *arsR* are transcribed as one transcriptional unit and that *arsB* and ORF5 are also transcribed as a single

transcriptional unit. Reporter gene studies in the *E. coli ars* mutant ACSH50I^q indicated that the level of expression in the direction of *arsB* was about five-fold higher than in the direction of *arsRC*. The *arsB* expression was also repressed to a greater extent by ArsR than *arsRC* expression. A divergent chromosomal *ars* operon was also identified in *At. ferrooxidans* (Butcher *et al.*, 2000). The *At. caldus* chromosomal *arsC*, *arsR* and *arsB* genes exhibited 78%, 74% and 85% identity, respectively, to the *At. ferrooxidans* chromosomal *ars* genes. This high similarity between the *ars* genes of the two organisms suggests that they may have a common ancestor. Although the genome of *At. ferrooxidans* has been sequenced, the *ars* genes are the first from the chromosome of *At. caldus* to be reported. It is therefore too early to know whether the relatedness between *ars* genes reflects a general relationship between *At. ferrooxidans* and *At. caldus* or whether it is a feature of *ars* genes.

Sequence alignments of the *At. caldus* chromosomal ArsR with its closest matches from the NCBI database showed that none of these ArsR proteins contain the conserved metal binding motif (ELCVCDL) known to be involved in the binding of arsenite. Arsenite binds to the cysteine residues in this domain, resulting in a conformational change to ArsR that causes the repressor to dissociate from the promoter, which then results in the expression of the *ars* genes (Shi *et al.*, 1994). The SmtB/ArsR family of transcriptional regulators have two metal binding sites, the α3N in the α3 helix and the α5C metal binding site in the α5 helix (Busenlehner *et al.*, 2003). It is known that the ELCVCDL motif is located in the α3 helix. The *At. caldus* chromosomal ArsR and its closest matches however, contain a conserved LVAYLTENCC domain near the C-terminal end of the protein (Figure 2.10). The two cysteine residues in this domain, together with another cysteine five to seven residues towards the C-terminal, may form the metal binding domain of these atypical ArsR proteins. The fact that this possible metal binding domain is located in the α5 helix strengthens the possibility that this could be the *At. caldus* chromosomal ArsR's metal binding domain, as Busenlehner *et al.* (2003) showed that this family of transcriptional regulators has metal binding domains in the α3 and/or α5 helices.

Despite the lack of the conventional ELCVCDL metal binding motif on the *At. caldus* chromosomal ArsR, translational fusions of the predicted promoter regions to a promoterless reporter gene (*lacZ*) showed that the expression from the *arsB* and *arsR* promoters were repressed in the presence of an intact *arsR* gene in trans. This repression was relieved with the addition of arsenite and arsenate (Figure 3.2). This indicated that the atypical ArsR does regulate expression from the *arsB* and *arsR* promoters. No detectable difference in the ability of arsenate and arsenite to derepress the transcription of the *ars* genes was observed. However, it was very interesting to find that ArsR was induced by arsenate, as most ArsR regulators respond to arsenite but not arsenate. The ArsR protein of *At. ferrooxidans* (with which the *At. caldus* chromosomal ArsR has 74% identity) has also been shown to be induced by arsenate (Butcher *et al.*, 2000). The ArsR protein of *At. caldus* is only the third example (all from *Acidithiobacillus*) of an arsenic regulator that is derepressed by both forms of arsenic. As no ArsC was present when ArsR was induced by arsenate and arsenite, it suggests that these atypical ArsR proteins can bind either arsenate or arsenite, or that a mechanism other that ArsC exists that can convert arsenate to arsenite prior to binding to the ArsR. Metal binding sites for both arsenate and arsenite might be a common characteristic for this atypical ArsR family of transcriptional regulators.

In future work it will be important to show that the ArsR binds to the predicted promoter regions. Shi *et al.* (1994) showed, by using gel mobility shift assays, that the *E. coli* R773 ArsR was able to bind to the predicted promoter region. Gel mobility shift assays can similarly be used to show that the *At. caldus* chromosomal ArsR binds to the promoter sequence in the absence of an inducer. If the ArsR does not bind in the presence of arsenite and arsenate, it will confirm our results that arsenate and arsenite can act as inducers of the *At. caldus* chromosomal ArsR. If mutations of Cys^{94}, Cys^{95} and Cys^{101} by site-directed mutagenesis result in the loss of induction of ArsR by arsenate and arsenite, it will indicate if the predicted LVAYLTENCC domain is indeed the metal binding domain and that these three cysteines are involved in arsenic binding.

An investigation of how the chromosomal and transposon *ars* systems were likely to interact when both were present in the same bacterium was carried out, as TnAtcArs is also found in *At. caldus* #6. Although the transposon and chromosomal *ars* operons are quite different and although the two ArsR proteins share only 45% sequence identity, promoter fusion studies in *E. coli* indicated that the *At. caldus* chromosomal ArsR was able to regulate the promoters (both directions) of its own operon as well as cross-regulate the transposon *ars* promoter. However, regulation of the transposon promoter by the chromosomal ArsR was not as effective as regulation by its own ArsR protein, indicating that the chromosomal ArsR is not optimally designed for the regulation of the transposon *ars* operon. Growth curve experiments conducted with the transposon and chromosomal copies of *ars* genes revealed that the transposon system conferred much higher levels of arsenic resistance than the chromosomal system in a heterologous host (Figure 4.2). The very much higher levels of gene expression from the transposon operon might be the reason for these higher levels of arsenic resistance conferred by the transposon operon in *E. coli*.

However, these experiments were done in a heterologous host, *E. coli*, and may therefore not directly reflect what happens in *At. caldus*. The fact that co-transcription of accessory genes (ORF1 and ORF5) was observed, but had no effect on arsenic resistance in *E. coli* is something of an anomaly. It could be that these ORFs confer some property that was not examined or that co-transcription in *E. coli* was artefactual. mRNA studies with RNA isolated from *At. caldus* would be able to confirm if the results obtained in the heterologous host, *E. coli*, were a true reflection of what happens in *At. caldus*. If the accessory genes are co-transcribed with the *ars* genes in *At. caldus*, knockout mutants in *At. caldus* would then shed more light on the functions of these genes.

It is known that three *At. caldus* strains (#6, "f" and MNG) that were previously exposed to arsenic contain two sets of arsenic resistance genes, a set on the chromosome as well as a set carried on a transposon, while strains BC13, KU and C-SH12 only contain a chromosomal copy of arsenic resistance genes (de Groot *et al.*, 2003). In this study it was shown that in *E. coli*, the transposon copy of *ars* genes confers much higher levels of

arsenic resistance than the chromosomal copy. We wanted to check on a possible difference in levels of *ars* resistance if the transposon arsenic system was introduced to a strain of *At. caldus* only harboring the chromosomal *ars* genes. A major problem was that no easy gene transfer system exists for these acidophilic organisms. No success has been achieved by chemical transformation and Kusano *et al.* (1992) were able to transform DNA to *At. ferrooxidans* by electroporation to only one strain of 30 strains electroporated. A more successful conjugational system was used by Jin *et al.* (1992) and was further improved by Liu *et al*. (1999). No DNA transformation systems for *At. caldus* had been reported prior to this study.

It is well known that the bacterial population used in the biomining process became highly resistant to arsenic over a period of three years (Rawlings and Woods, 1995). In order to demonstrate that the acquisition of the transposon *ars* system resulted in an increase in arsenic resistance in *At. caldus*, a system had to be developed where the transposon copy of *ars* genes could be introduced to strains harboring only a chromosomal copy. Using a conjugative system, the transposon *ars* system was introduced to *At. caldus* C-SH12. This was the first reported study where DNA was successfully transferred to *At. caldus*. As expected, the introduction of the transposon *ars* operon to *At. caldus* C-SH12 resulted in an dramatic increase of *ars* resistance (Figure 4.6). It was not possible to determine by how much arsenic resistance was increased, because growth of the transformed cultures occurred at concentrations above at which arsenite precipitated. Based on these results we speculate that *At. caldus* #6 obtained higher levels of arsenic resistance most likely by the acquisition of the transposon *ars* system from the gene pool.

The development of a genetic transfer system for *At. caldus* is of great importance in the future study of the molecular biology of this organism and possibly other acidophilic organisms. Researchers in our laboratory are currently developing a system to introduce gene knockouts in *At. caldus*. These two systems will enable us to study the molecular biology of these biomining bacteria in greater depth than is currently possible. *At. caldus* C-SH12 formed indistinct colonies on solid medium, and it was therefore not possible to

calculate the conjugation frequency. A means of obtaining more distinct *At. caldus* C-SH12 colonies on solid media has recently been achieved by replacing agarose with phytogel as a gelling agent. As we are now able to calculate the transformation frequency, experiments are currently in progress using this plasmid transfer system to optimize a conjugation system for *At. caldus*.

As the dominant iron-oxidizing bacterium in arsenopyrite biooxidation tanks, it was believed that a strain of *Leptospirillum ferriphilum* (isolated from a commercial biooxidation tank treating a gold-bearing arsenopyrite concentrate at the Fairview mine in South Africa) would also possess an effective arsenic resistance mechanism. This was the case, as Tuffin *et al.* (2006) showed that *L. ferriphilum* does, like *At. caldus*, carry a set of arsenic resistance genes on a transposon. This was not surprising, as *L. ferriphilum*, like *At. caldus*, was isolated from a very arsenic rich environment. *The L. ferriphilum* transposon *ars* operon had a very similar gene layout as the transposon *ars* operon of *At. caldus*, except for the absence of ORF7 and the *arsDA* duplication. Although these two *ars* operons originated from bacteria from the same environment, the amino acid sequence of their gene products were not necessarily the most closely related, with high sequence identity to genes on transposons from *A. faecalis* and *M. flagellatus* as well (Tuffin *et al.*, 2006; Figure 5.1). The differences in DNA sequence between the transposon *ars* genes of *At. caldus* and *L. ferriphilum* are far to substantial for the transposons to have passed from *At caldus* to *L. ferriphilum* or vice versa in the biooxidation environment. This suggests that these two types of bacteria have acquired the *ars* transposons independently of each other.

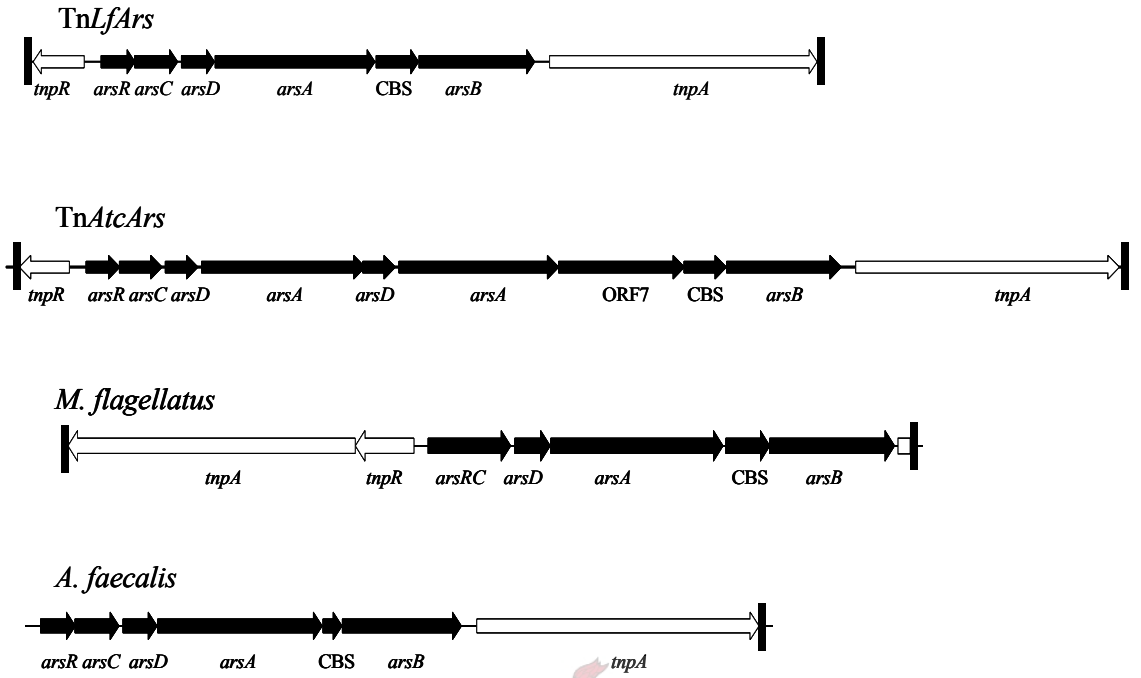


Figure 5.1: Diagrammatic representation of transposon located *ars* operons isolated from *Leptospirillum ferriphilum* Fairview (TnLfArs), *Acidithiobacillus caldus* #6 (TnAtcArs), *Methylobacillus flagellatus* and *Alcaligenes faecalis*. Inverted repeats are shown as vertical bars. The white box downstream of *M. flagellatus arsB* represents a 39 aa TnpA remnant. Taken from Tuffin *et al*. (2006).

Like in *At. caldus*, the *L. ferriphilum* transposon located *ars* genes conferred higher levels of arsenic resistance than the chromosomal *ars* genes in *E. coli*. The chromosomal *ars* genes were also very poorly expressed. Because the overexpression of *arsB* is lethal to cells, organisms do not necessarily want high expression levels of *ars* genes. If they live in an environment with low levels of arsenic, the low levels of expression from the chromosomal *ars* operons will probably be sufficient to confer arsenic resistance. Only when they are placed under selective pressure would they acquire an additional system that would confer higher levels of arsenic resistance. Tuffin *et al.* (2006) also showed that the *L. ferriphilum* strain containing the transposon *ars* operon was more resistant to arsenic than *L. ferriphilum* strains harboring only a chromosomal copy of *ars* genes. It thus seems that the quickest or easiest way for these bacteria to have survived these arsenic rich environments, was by the acquisition of another arsenic operon, as the time to mutate an existing operon (like the chromosomal *ars* operon) may take thousands of

years. The fact that these biooxidation environments are not sterile opens many options as to where these arsenic transposons originated from. The findings in this study are consistent with the prediction, reported by Woese (2004), that "cosmopolitan genes", of which arsenic resistance genes are a typical example, are more likely to be a characteristic of a certain environment than of particular lineages. If these "cosmopolitan genes" are not restricted to certain lineages, they will most likely pass between different types of bacteria within an environment that selects for them. Kondratyeva *et al.* (1995) suggested that both arsenic and zinc resistant strains of *At. ferrooxidans* have increased copy numbers of chromosomal genes encoding resistance. In this study we answered the question of how these biomining organisms became more resistant to arsenic in a relative short period of time. This increase in resistance was acquired probably not by an increase in copy number of arsenic resistance genes, but by the acquisition of a highly resistant arsenic operon carried on a transposon.

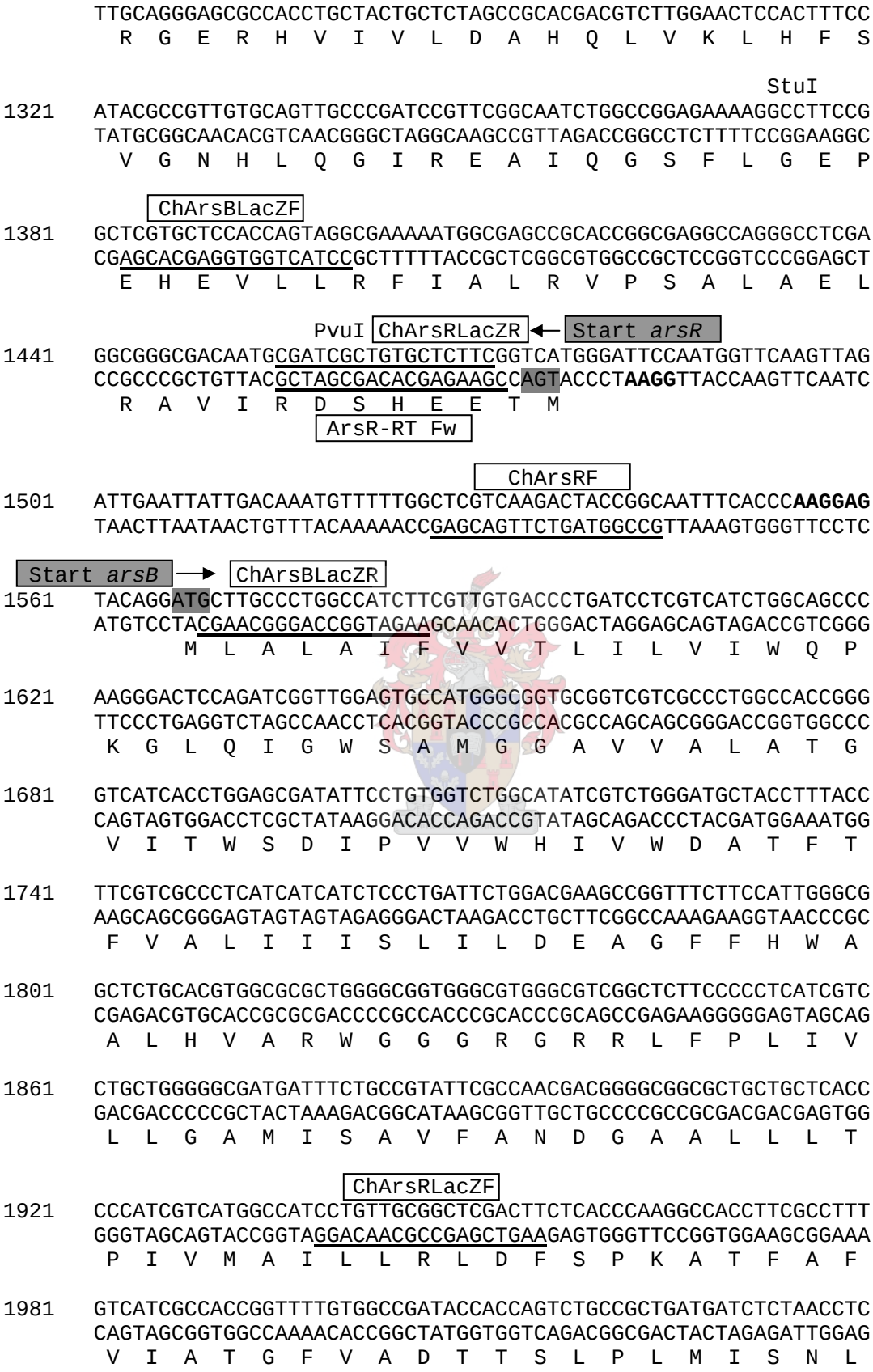
Appendix One: Annotated sequence obtained from pTcC-#4

Shown below is the double stranded sequence of the *At. caldus* chromosomal fragment obtained in pTcC-#4 during the course of this study. Protein translations of the ten open reading frames encoded on this fragment have been shown below the DNA sequence. Restriction endonuclease sites are labeled above the DNA sequence. The sequence to which primers were designed is underlined and labeled in white boxes. The start codons of open reading frames are labeled in grey boxes and the putative ribosome binding sites of *arsR* and *arsB* are shown in bold. Arrows indicate the direction of transcription.

```
1       GATCCCGACACCACCGAGATAAGCCCTCACTGCCTACGGCTCGCCGTCCGCAAACCGTTA
        CTAGGGCTGTGGTGGCTCTATTCGGGAGTGACGGATGCCGAGCGGCAGGCGTTTGGCAAT

61      GGCGCGACACGCTTCCAGGGTGGCTGTGCCTGGTGCTGGGCGAGCGCTTATCCCGAGAAA
        CCGCGCTGTGCGAAGGTCCCACCGACACGGACCACGACCCGCTCGCGAATAGGGCTCTTT

121     GTCGCACGGTTTACCTCGACTCCGTACCGGGCACCAGAAATACCGGCTCTTGGTCGGCGG
        CAGCGTGCCAAATGGAGCTGAGGCATGGCCCGTGGTCTTTATGGCCGAGAACCAGCCGCC
                  *  R  S  E  T  G  P  V  L  F  V  P  E  Q  D  A  P

181     GCATGGTGGCGAGCTCTTCTTCCGTCATCCGAGCGTAAATTTCGAGGAGATAACCCGTGG
        CGTACCACCGCTCGAGAAGAAGGCAGTAGGCTCGCATTTAAAGCTCCTCTATTGGGCACC
          M  T  A  L  E  E  E  T  M  R  A  Y  I  E  L  L  Y  G  T  P

          ORF1-RT_Rev
241     GATCGCGCAGCCAGAGTTCGTGCAGAGGTGTGCCGCGCCAGGTGCTGCGCGGCGGCTTGA
        CTAGCGCGTCGGTCTCAAGCACGTCTCCACACGGCGCGGTCCACGACGCGCCGCCGAACT
          D  R  L  W  L  E  H  L  P  T  G  R  W  T  S  R  P  P  K  V

301     CCACCTCGGCACCGGCGGCCACGGCGCGATGGTAGGCGGCGATCACCGCTTCCTTGCTCC
        GGTGGAGCCGTGGCCGCCGGTGCCGCGCTACCATCCGCCGCTAGTGGCGAAGGAACGAGG
          V  E  A  G  A  A  V  A  R  H  Y  A  A  I  V  A  E  K  S  R

361     GCACTCCCAGGCCCACGTGGTAGAGGCTGTAGGTGTCGAGGGGCTCGTCCCGGTCGTTGA
        CGTGAGGGTCCGGGTGCACCATCTCCGACATCCACAGCTCCCCGAGCAGGGCCAGCAACT
          V  G  L  G  V  H  Y  L  S  Y  T  D  L  P  E  D  R  D  N  V

421     CGAGAATCACCAGATTCAGGCAGAGATCGGGGACGATGAAGGTGCTGAAGCGCGCCGTCT
        GCTCTTAGTGGTCTAAGTCCGTCTCTAGCCCCTGCTACTTCCACGACTTCGCGCGGCAGA
          L  I  V  L  N  L  C  L  D  P  V  I  F  T  S  F  R  A  T  Q

                                                                BstEII
481     GTTGTCTGGGCTCAACCCCGAAAAAGTCCCGATAGAAGGCCGTGCTGCGGGCAAGATCGG
        CAACAGACCCGAGTTGGGGCTTTTTCAGGGCTATCTTCCGGCACGACGCCCGTTCTAGCC
          Q  R  P  E  V  G  F  F  D  R  Y  F  A  T  S  R  A  L  D  T
```

HindIII

```
541    TTACCCGGAAGCTTAGATGAAATTCCTGGCGAGCGGGCACCAGGGGAGACAACTCGGTGG
       AATGGGCCTTCGAATCTACTTTAAGGACCGCTCGCCCGTGGTCCCCTCTGTTGAGCCACC
         V  R  F  S  L  H  F  E  Q  R  A  P  V  L  P  S  L  E  T  S

       ←Start ORF1
601    ACATGCAGTAATCCTCAGGGTTCACTGTGACCAATGCGTTCCAGTTCCGTGCGCAGGCGA
       TGTACGTCATTAGGAGTCCCAAGTGACACTGGTTACGCAAGGTCAAGGCACGCGTCCGCT
         M           *  P  E  S  H  G  I  R  E  L  E  T  R  L  R

661    CTGGGATCTTCGGCTAGCCATTCGGCGACGGGCAGTTGCAAAAGCGCCTCGATGCGATGG
       GACCCTAGAAGCCGATCGGTAAGCCGCTGCCCGTCAACGTTTTCGCGGAGCTACGCTACC
       S  P  D  E  A  L  W  E  A  V  P  L  Q  L  L  A  E  I  R  H

                                    ArsC-RT_Fw
721    CGCAGGATTCGATAGGCTTTCTCGAAAGCGGCCTCGATTTCCGCCTCGCTGCCGGTCGCC
       GCGTCCTAAGCTATCCGAAAGAGCTTTCGCCGGAGCTAAAGGCGGAGCGACGGCCAGCGG
       R  L  I  R  Y  A  K  E  F  A  A  E  I  E  A  E  S  G  T  A

781    TTGGCGGGATCGTCTACCCCCCAATGGGCGCGCAGGGCAGGGCCGAGATAGGCCGGGCAG
       AACCGCCCTAGCAGATGGGGGGTTACCCGCGCGTCCCGTCCCGGCTCTATCCGGCCCGTC
       K  A  P  D  D  V  G  W  H  A  R  L  A  P  G  L  Y  A  P  C

                                                 EcoRV
841    GTCTCGCCAGCGGCACTGGCGCAGACGGTGATGACGATATCCGGGGTAGTGGGCAGATCG
       CAGAGCGGTCGCCGTGACCGCGTCTGCCACTACTGCTATAGGCCCCATCACCCGTCTAGC
       T  E  G  A  A  S  A  C  V  T  I  V  I  D  P  T  T  P  L  D

901    TCCCAGGACTTGCTGGCGAGACCTTCGGTAGAGAAGCCTTCCCGCTGCAGGAGGGCAACA
       AGGGTCCTGAACGACCGCTCTGGAAGCCATCTCTTCGGAAGGGCGACGTCCTCCCGTTGT
       D  W  S  K  S  A  L  G  E  T  S  F  G  E  R  Q  L  L  A  V

961    CTGCGCGGATGCACATAACCTGCCGGATGGCTGCCGGCGCTGCTGGCGTGGAGCCTGCCC
       GACGCGCCTACGTGTATTGGACGGCCTACCGACGGCCGCGACGACCGCACCTCGGACGGG
       S  R  P  H  V  Y  G  A  P  H  S  G  A  S  S  A  H  L  R  G

                                    ArsC-RT_Rev and ChArsRR
1021   TGGGCAAGGCCGTTGAGGGTGGCTTCGGCCAGGATGGAACGACAGGAGTTGCCGGTGCAC
       ACCCGTTCCGGCAACTCCCACCGAAGCCGGTCCTACCTTGCTGTCCTCAACGGCCACGTG
       Q  A  L  G  N  L  T  A  E  A  L  I  S  R  C  S  N  G  T  C

                                 ←Start arsC
1081   AGAAAGAGGATCTCGGGGGTTTTCATCGGTGTCACCTCGGTGGGGCGGCGGAAGCGGTGG
       TCTTTCTCCTAGAGCCCCCAAAAGTAGCCACAGTGGAGCCACCCCGCCGCCTTCGCCACC
       L  F  L  I  E  P  T  K  M        *  R  P  P  A  A  S  A  T  A

1141   CTGCTGCGCTGGATTCGGACTCGTCGGGGATAGTGCAGGATTGGGTGCCGCGACAGCAGT
       GACGACGCGACCTAAGCCTGAGCAGCCCCTATCACGTCCTAACCCACGGCGCTGTCGTCA
         A  A  S  S  E  S  E  D  P  I  T  C  S  Q  T  G  R  C  C  N

1201   TTTCGGTAAGGTAGGCGACCAGGGAGCGCACCACGGGCATCTGCGCCCGGTAGCGCTGGA
       AAAGCCATTCCATCCGCTGGTCCCTCGCGTGGTGCCCGTAGACGCGGGCCATCGCGACCT
         E  T  L  Y  A  V  L  S  R  V  V  P  M  Q  A  R  Y  R  Q  F

1261   AACGTCCCTCGCGGTGGACGATGACGAGATCGGCGTGCTGCAGAACCTTGAGGTGAAAGG
```


TTGCAGGGAGCGCCACCTGCTACTGCTCTAGCCGCACGACGTCTTGGAACTCCACTTTCC
R G E R H V I V L D A H Q L V K L H F S
StuI
1321 ATACGCCGTTGTGCAGTTGCCCGATCCGTTCGGCAATCTGGCCGGAGAAAAGGCCTTCCG
TATGCGGCAACACGTCAACGGGCTAGGCAAGCCGTTAGACCGGCCTCTTTTCCGGAAGGC
V G N H L Q G I R E A I Q G S F L G E P
ChArsBLacZF
1381 GCTCGTGCTCCACCAGTAGGCGAAAAATGGCGAGCCGCACCGGCGAGGCCAGGGCCTCGA
CGAGCACGAGGTGGTCATCCGCTTTTTACCGCTCGGCGTGGCCGCTCCGGTCCCGGAGCT
E H E V L L R F I A L R V P S A L A E L
PvuI ChArsRLacZR Start arsR
1441 GGCGGGCGACAATGCGATCGCTGTGCTCTTCGGTCATGGGATTCCAATGGTTCAAGTTAG
CCGCCCGCTGTTACGCTAGCGACACGAGAAGCCAGTACCCTAAGGTTACCAAGTTCAATC
R A V I R D S H E E T M
ArsR-RT Fw
ChArsRF
1501 ATTGAATTATTGACAAATGTTTTTGGCTCGTCAAGACTACCGGCAATTTCACCCAAGGAG
TAACTTAATAACTGTTTACAAAAACCGAGCAGTTCTGATGGCCGTTAAAGTGGGTTCCTC
Start arsB
ChArsBLacZR
1561 TACAGGATGCTTGCCCTGGCCATCTTCGTTGTGACCCTGATCCTCGTCATCTGGCAGCCC
ATGTCCTACGAACGGGACCGGTAGAAGCAACACTGGGACTAGGAGCAGTAGACCGTCGGG
M L A L A I F V V T L I L V I W Q P
1621 AAGGGACTCCAGATCGGTTGGAGTGCCATGGGCGGTGCGGTCGTCGCCCTGGCCACCGGG
TTCCCTGAGGTCTAGCCAACCTCACGGTACCCGCCACGCCAGCAGCGGGACCGGTGGCCC
K G L Q I G W S A M G G A V V A L A T G
1681 GTCATCACCTGGAGCGATATTCCTGTGGTCTGGCATATCGTCTGGGATGCTACCTTTACC
CAGTAGTGGACCTCGCTATAAGGACACCAGACCGTATAGCAGACCCTACGATGGAAATGG
V I T W S D I P V V W H I V W D A T F T
1741 TTCGTCGCCCTCATCATCATCTCCCTGATTCTGGACGAAGCCGGTTTCTTCCATTGGGCG
AAGCAGCGGGAGTAGTAGTAGAGGGACTAAGACCTGCTTCGGCCAAAGAAGGTAACCCGC
F V A L I I I S L I L D E A G F F H W A
1801 GCTCTGCACGTGGCGCGCTGGGGCGGTGGGCGTGGGCGTCGGCTCTTCCCCCTCATCGTC
CGAGACGTGCACCGCGCGACCCCGCCACCCGCACCCGCAGCCGAGAAGGGGGAGTAGCAG
A L H V A R W G G G R G R R L F P L I V
1861 CTGCTGGGGGCGATGATTTCTGCCGTATTCGCCAACGACGGGGCGGCGCTGCTGCTCACC
GACGACCCCCGCTACTAAAGACGGCATAAGCGGTTGCTGCCCCGCCGCGACGACGAGTGG
L L G A M I S A V F A N D G A A L L L T
ChArsRLacZF
1921 CCCATCGTCATGGCCATCCTGTTGCGGCTCGACTTCTCACCCAAGGCCACCTTCGCCTTT
GGGTAGCAGTACCGGTAGGACAACGCCGAGCTGAAGAGTGGGTTCCGGTGGAAGCGGAAA
P I V M A I L L R L D F S P K A T F A F
1981 GTCATCGCCACCGGTTTTGTGGCCGATACCACCAGTCTGCCGCTGATGATCTCTAACCTC
CAGTAGCGGTGGCCAAAACACCGGCTATGGTGGTCAGACGGCGACTACTAGAGATTGGAG
V I A T G F V A D T T S L P L M I S N L

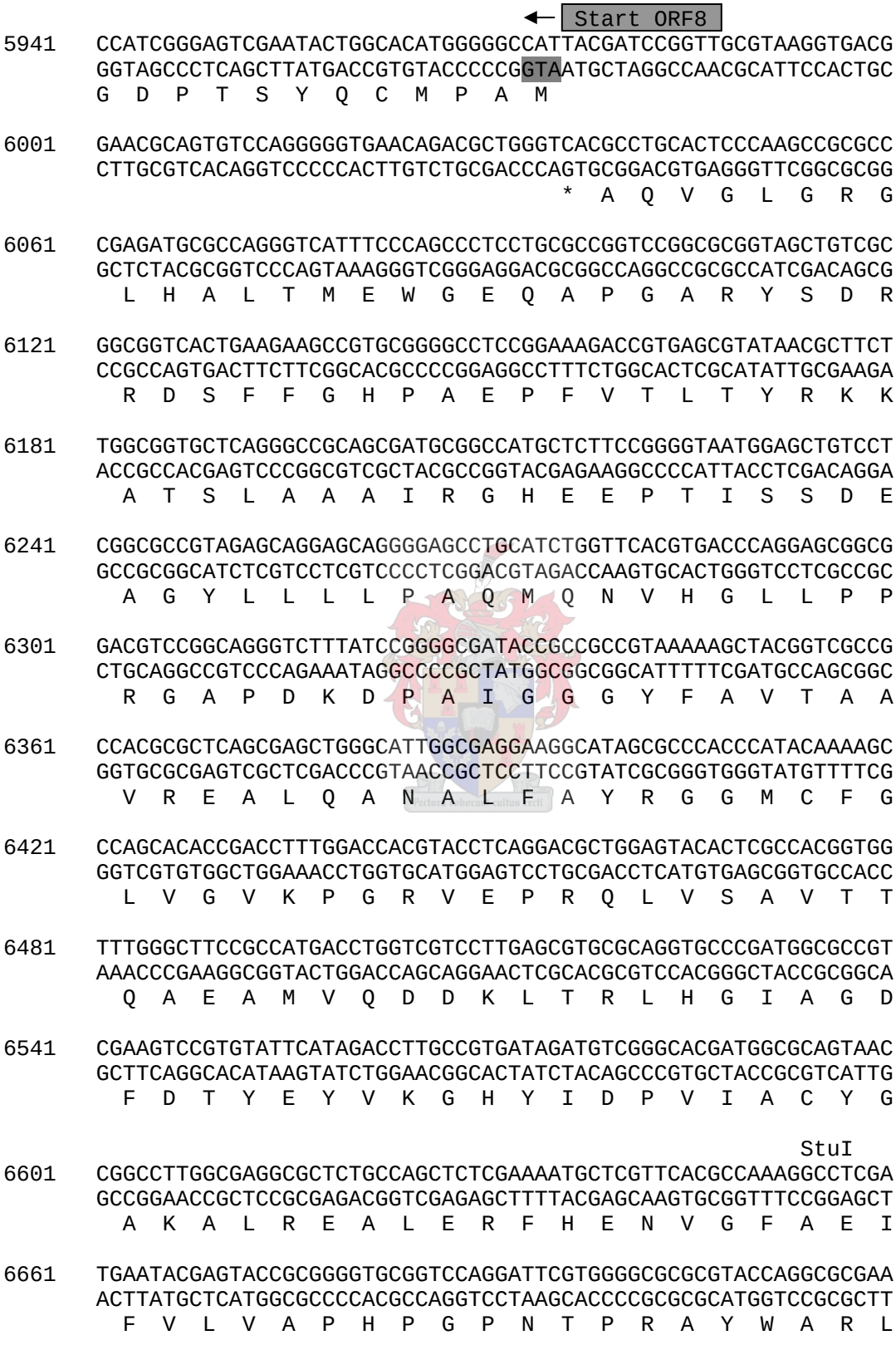
```
2041   GTCAACATCGTCAGCGCCAATTATTTCAATATCACGTTTGATCGCTATGCCCTGGTGATG
       CAGTTGTAGCAGTCGCGGTTAATAAAGTTATAGTGCAAACTAGCGATACGGGACCACTAC
        V  N  I  V  S  A  N  Y  F  N  I  T  F  D  R  Y  A  L  V  M

2101   GTACCGGTGGATCTGGTGGCACTGGCGGCCACCCTGGGGGTGCTCTGGCTTTACTTTCGT
       CATGGCCACCTAGACCACCGTGACCGCCGGTGGGACCCCCACGAGACCGAAATGAAAGCA
        V  P  V  D  L  V  A  L  A  A  T  L  G  V  L  W  L  Y  F  R

                                                                 BamHI
2161   CGCTCCGTGCCCGCGTCCTATCCTACCGCCCACCTCGACGCTCCGCGCACGGCCATCCAG
       GCGAGGCACGGGCGCAGGATAGGATGGCGGGTGGAGCTGCGAGGCGCGTGCCGGTAGGTC
        R  S  V  P  A  S  Y  P  T  A  H  L  D  A  P  R  T  A  I  Q

2221   GATCCTGTGGTCTTCCGTGCCGCCTTTCCGGTGCTGGTTTTGCTACTAATCGCCTACTTC
       CTAGGACACCAGAAGGCACGGCGGAAAGGCCACGACCAAAACGATGATTAGCGGATGAAG
        D  P  V  V  F  R  A  A  F  P  V  L  V  L  L  L  I  A  Y  F

2281   GTTACCGCGCAGTGGCAGGTACCGGTATCGGTGGTGACGGGGGCTGGGGCGCTGATCCTC
       CAATGGCGCGTCACCGTCCATGGCCATAGCCACCACTGCCCCCGACCCCGCGACTAGGAG
        V  T  A  Q  W  Q  V  P  V  S  V  V  T  G  A  G  A  L  I  L

2341   CTCGCTTTGGCCGGACGTTGGTTGCAGGGCGGGCGCGGCGCCCGCATTCCCGTACGCAAG
       GAGCGAAACCGGCCTGCAACCAACGTCCCGCCCGCGCCGCGGGCGTAAGGGCATGCGTTC
        L  A  L  A  G  R  W  L  Q  G  G  R  G  A  R  I  P  V  R  K

2401   GTCTTGCACGAGGCCCCCTGGCAGATCGTGATTTTCAGCCTCGGGATGTATCTGGTGGTC
       CAGAACGTGCTCCGGGGGACCGTCTAGCACTAAAAGTCGGAGCCCTACATAGACCACCAG
        V  L  H  E  A  P  W  Q  I  V  I  F  S  L  G  M  Y  L  V  V

2461   TATGGGCTGCGCAATGCCGGTTTGACCAACTATGTGGCCATGGCCCTGGCGTATCTGGCG
       ATACCCGACGCGTTACGGCCAAACTGGTTGATACACCGGTACCGGGACCGCATAGACCGC
        Y  G  L  R  N  A  G  L  T  N  Y  V  A  M  A  L  A  Y  L  A

                 ArsB-RT Fw
2521   GCACAAGGTACCGTTCTGGCGGCGATTGGGTCAGGCTTCCTGGCCGCGGCGCTGTCCTCG
       CGTGTTCCATGGCAAGACCGCCGCTAACCCAGTCCGAAGGACCGGCGCCGCGACAGGAGC
        A  Q  G  T  V  L  A  A  I  G  S  G  F  L  A  A  A  L  S  S

2581   GTGATGAACAACATGCCGGCGGTTTTGGTGGGCGCCTTGGCCATCCACCAACTGCCGGTG
       CACTACTTGTTGTACGGCCGCCAAAACCACCCGCGGAACCGGTAGGTGGTTGACGGCCAC
        V  M  N  N  M  P  A  V  L  V  G  A  L  A  I  H  Q  L  P  V

            BamHI
2641   GCCACGGATCCGTTGATACGGGAAATCTTCGTCTACGCCAACATCATCGGTTGCGATCTT
       CGGTGCCTAGGCAACTATGCCCTTTAGAAGCAGATGCGGTTGTAGTAGCCAACGCTAGAA
        A  T  D  P  L  I  R  E  I  F  V  Y  A  N  I  I  G  C  D  L

2701   GGCCCCAAGTTCACGCCCATCGGCAGTCTGGCCACCCTGCTCTGGCTGCATGTGCTGGGC
       CCGGGGTTCAAGTGCGGGTAGCCGTCAGACCGGTGGGACGAGACCGACGTACACGACCCG
        G  P  K  F  T  P  I  G  S  L  A  T  L  L  W  L  H  V  L  G

2761   CGCAAAGGCATGACCGTGACCTGGGGCCAATACATGCGCACCGGACTCCTCATCACCCCG
       GCGTTTCCGTACTGGCACTGGACCCCGGTTATGTACGCGTGGCCTGAGGAGTAGTGGGGC
        R  K  G  M  T  V  T  W  G  Q  Y  M  R  T  G  L  L  I  T  P
```


2821 CCGGTACTGTTCCTGACGCTGCTGGCCTTGGCGGCTTGGTTGCCACTACTCTGAAAAAGG
GGCCATGACAAGGACTGCGACGACCGGAACCGCCGAACCAACGGTGATGAGACTTTTTCC
P V L F L T L L A L A A W L P L L *
Start ORF5
2881 CGTTTGAGAAATTTTGTCTTCCCTCGGGTCATGCGCGAACGGTTCGGCCATGGCCGGATG
GCAAACTCTTTAAAACAGAAGGGAGCCCAGTACGCGCTTGCCAAGCCGGTACCGGCCTAC
M R E R F G H G R M
ORF5-RT Rev
2941 TTGCGACCCGGTGAGGTCCGCCACCGCAATCCCTGCCAAACGCGGCACACCTGCGCTTCC
AACGCTGGGCCACTCCAGGCGGTGGCGTTAGGGACGGTTTGCGCCGTGTGGACGCGAAGG
L R P G E V R H R N P C Q T R H T C A S
ORF5-RT Fw
3001 ATGTTGCGGACTGCCGGGGAAGATCCTAGGTGGATTGCCGGCCAGATGGGGTATGAAGAC
TACAACGCCTGACGGCCCCTTCTAGGATCCACCTAACGGCCGGTCTACCCCATACTTCTG
M L R T A G E D P R W I A G Q M G Y E D
3061 GGGGTACGACTCGACTGGGGTAGGCTCATCAGGTGCCCCATGTCAACCCATCTGCCTAGA
CCCCATGCTGAGCTGACCCCATCCGAGTAGTCCACGGGGTACAGTTGGGTAGACGGATCT
G V R L D W G R L I R C P M S T H L P R
BstEII
3121 GCAGGACTCTCGCCCTTTGGTCACCATTGGATCAAAAGAACGGGGCGGCGAAGCATAGGG
CGTCCTGAGAGCGGGAAACCAGTGGTAACCTAGTTTTCTTGCCCCGCCGCTTCGTATCCC
A G L S P F G H H W I K R T G R R S I G
3181 CTGGAAATGGCAATCTCATGGATGGTGCGAAAGGGGAGACTCGAACTCCCACGCCTTGCG
GACCTTTACCGTTAGAGTACCTACCACGCTTTCCCCTCTGAGCTTGAGGGTGCGGAACGC
L E M A I S W M V R K G R L E L P R L A
3241 GCGCTGGAACCTAAATCCAGTGCGTCTACCAATTCCGCCACTCTCGCCTTTACGCTATCT
CGCGACCTTGGATTTAGGTCACGCAGATGGTTAAGGCGGTGAGAGCGGAAATGCGATAGA
A L E P K S S A S T N S A T L A F T L S
BamHI
Start ORF6
3301 TAAGCACTAACCGCTCTTTTCACAACGACGAGGATCCCATGGCCAAGTCATCATCCCAAC
ATTCGTGATTGGCGAGAAAAGTGTTGCTGCTCCTAGGGTACCGGTTCAGTAGTAGGGTTG
* M A K S S S Q
3361 CCTCTCGGTTGCGTCGCCTTCTGGCTGTCGAGTCGGCGCGGATCATGGCGGAGGAGGGCA
GGAGAGCCAACGCAGCGGAAGACCGACAGCTCAGCCGCGCCTAGTACCGCCTCCTCCCGT
P S R L R R L L A V E S A R I M A E E G
ORF6-RT Rev
3421 TTGCGGACTACCGCTTTGCCAAGGAGAAGGCGGCCCGCCGCCTGGGTAGCGGCAGCACCC
AACGCCTGATGGCGAAACGGTTCCTCTTCCGCCGGGCGGCGGACCCATCGCCGTCGTGGG
I A D Y R F A K E K A A R R L G S G S T
PvuII
3481 AGCAAGATTGGCCCAGCAACTCGGAGATCCAGGCCGAACTCAAGGCGCGGTTGCAGCTGT
TCGTTCTAACCGGGTCGTTGAGCCTCTAGGTCCGGCTTGAGTTCCGCGCCAACGTCGACA
Q Q D W P S N S E I Q A E L K A R L Q L

```
3541   TTCACGGCGAATCCCAACCCCTCGAACTGCGCCGCCTTCGGGAGGTTGCGCTGGAGGCCA
       AAGTGCCGCTTAGGGTTGGGGAGCTTGACGCGGCGGAAGCCCTCCAACGCGACCTCCGGT
       F  H  G  E  S  Q  P  L  E  L  R  R  L  R  E  V  A  L  E  A

3601   TGGGCTGGCTCAAGGATTTTCGTCCCTTGCTCGCAGGTGCGGTACTCAACGGCACGGCCA
       ACCCGACCGAGTTCCTAAAAGCAGGGAACGAGCGTCCACGCCATGAGTTGCCGTGCCGGT
       M  G  W  L  K  D  F  R  P  L  L  A  G  A  V  L  N  G  T  A

3661   CGCGCCATAGCGCCATCGTTCTGCATCTCTTTGCCGATACCCCGGAGAGCGTCATCTTTT
       GCGCGGTATCGCGGTAGCAAGACGTAGAGAAACGGCTATGGGGCCTCTCGCAGTAGAAAA
       T  R  H  S  A  I  V  L  H  L  F  A  D  T  P  E  S  V  I  F

3721   TTCTCATGGATCAGCAGGTGGCCTATGAAGAGGGCTGGCAGCGGCTGCATTTCGGCGACG
       AAGAGTACCTAGTCGTCCACCGGATACTTCTCCCGACCGTCGCCGACGTAAAGCCGCTGC
       F  L  M  D  Q  Q  V  A  Y  E  E  G  W  Q  R  L  H  F  G  D

3781   AAGCGCCCCAGGAGTATCCCCAGATCCGCATGACCCGTGGGGGGGGCAGAATTGCGTCTGG
       TTCGCGGGGTCCTCATAGGGGTCTAGGCGTACTGGGCACCCCCCCGTCTTAACGCAGACC
       E  A  P  Q  E  Y  P  Q  I  R  M  T  R  G  G  A  E  L  R  L

3841   TCATCTTTGCTCCCGACGAGGATCGGCGCCGTCCGGCATCGGCGGTGGACGGTAAGCCTC
       AGTAGAAACGAGGGCTGCTCCTAGCCGCGGCAGGCCGTAGCCGCCACCTGCCATTCGGAG
       V  I  F  A  P  D  E  D  R  R  R  P  A  S  A  V  D  G  K  P

3901   TACAGCGGGTCAACGCCCAGCAGTTGCAACGTTTGCTGGATTCCGAGGCATCACTGCCCG
       ATGTCGCCCAGTTGCGGGTCGTCAACGTTGCAAACGACCTAAGGCTCCGTAGTGACGGGC
       L  Q  R  V  N  A  Q  Q  L  Q  R  L  L  D  S  E  A  S  L  P

3961   CTTATTCCTGAGGTGTCGGTAAAACCTCTGGACTGGGCGTTGGACGAAGGTGCTGCCAAC
       GAATAAGGACTCCACAGCCATTTTGGAGACCTGACCCGCAACCTGCTTCCACGACGGTTG
       A  Y  S  *
          *  E  Q  P  T  P  L  V  E  P  S  P  T  P  R  L  H  Q  W  S

4021   TTACCAGACTGAGGGCGATGAAGATGGCGATCATGCCCAGGCTTTCCCAGCCGCCCGGTA
       AATGGTCTGACTCCCGCTACTTCTACCGCTAGTACGGGTCCGAAAGGGTCGGCGGGCCAT
         V  L  S  L  A  I  F  I  A  I  M  G  L  S  E  W  G  G  P  V

                PvuII
4081   CCTCGTCCAGCTGCCACCAGGCCGCGAGGACGCCGATGAGCGGCGCCAGCAGCATGGCCA
       GGAGCAGGTCGACGGTGGTCCGGCGCTCCTGCGGCTACTCGCCGCGGTCGTCGTACCGGT
         E  D  L  Q  W  W  A  A  L  V  G  I  L  P  A  L  L  M  A  M

4141   TGCCGGCTACCCCCGACGGTAGATGCTTGAGGGCATAGGCCCAAAGCAGGTAGGCAATGG
       ACGGCCGATGGGGGCTGCCATCTACGAACTCCCGTATCCGGGTTTCGTCCATCCGTTACC
         G  A  V  G  S  P  L  H  K  L  A  Y  A  W  L  L  Y  A  I  A

4201   CCGTACCCGGTATGACATTGTAGACTAGGGCGGCGGTGAAGAGCGGGCTCCACTGGATAC
       GGCATGGGCCATACTGTAACATCTGATCCCGCCGCCACTTCTCGCCCGAGGTGACCTATG
         T  G  P  I  V  N  Y  V  L  A  A  T  F  L  P  S  W  Q  I  R

4261   GCAGGGGGTCTATCCAAAGGCTTACGACAATGAGACCGACGGCGCCCAAGGCCGTCTGCC
       CGTCCCCCAGATAGGTTTCCGAATGCTGTTACTCTGGCTGCCGCGGGTTCCGGCAGACGG
         L  P  D  I  W  L  S  V  V  I  L  G  V  A  G  L  A  T  Q  W

4321   AGAAAGTCGCGTTGAGGAGATCCTGGCCCGGTGGCATGTGGCGTTTTTGCCAGATGGCGG
```

```
        TCTTTCAGCGCAACTCCTCTAGGACCGGGCCACCGTACACCGCAAAAACGGTCTACCGCC
          F  T  A  N  L  L  D  Q  G  P  P  M  H  R  K  Q  W  I  A  A

4381    CCCCCGCCCACAGTAGGCCCGATAACAGCGCAAAGGCCGGTCCTTGCCAGGCGCCATGAA
        GGGGGCGGGTGTCATCCGGGCTATTGTCGCGTTTCCGGCCAGGAACGGTCCGCGGTACTT
          G  A  W  L  L  G  S  L  L  A  F  A  P  G  Q  W  A  G  H  F

4441    AATCCCACGGGCTCAATATGCACAAGAGCCCCAGTAGCGCCAGAAGGAGCGCTGGCCAAT
        TTAGGGTGCCCGAGTTATACGTGTTCTCGGGGTCATCGCGGTCTTCCTCGCGACCGGTTA
          D  W  P  S  L  I  C  L  L  G  L  L  A  L  L  L  A  P  W  H

4501    GCAGTCCGTGCAGACGCTCCCCCAAAAGTGGCCAAGCCAGCAGCATGAGCCAGATGGGCA
        CGTCAGGCACGTCTGCGAGGGGGTTTTCACCGGTTCGGTCGTCGTACTCGGTCTACCCGT
          L  G  H  L  R  E  G  L  L  P  W  A  L  L  M  L  W  I  P  M

4561    TCATGTACACGAGGATGGCCGTCTTTCCCGCGCCGCCATATTCCAGGGCCCAGAGGGTGA
        AGTACATGTGCTCCTACCGGCAGAAAGGGCGCGGCGGTATAAGGTCCCGGGTCTCCCACT
          M  Y  V  L  I  A  T  K  G  A  G  G  Y  E  L  A  W  L  T  F

4621    AACCCACAAAGCCGGTGCTCTGCAGCAGGCCAAAGGGCAGGACGTAGCGTAGGGGAGGTC
        TTGGGTGTTTCGGCCACGAGACGTCGTCCGGTTTCCCGTCCTGCATCGCATCCCCTCCAG
          G  V  F  G  T  S  Q  L  L  G  F  P  L  V  Y  R  L  P  P  R

4681    GCCGCAGGGGACGTTTCAGGAGCAGTAGCAAGGGGAGGAGCACGACCGCCCCCAGGGCCA
        CGGCGTCCCCTGCAAAGTCCTCGTCATCGTTCCCCTCCTCGTGCTGGCGGGGGTCCCGGT
          R  L  P  R  K  L  L  L  L  L  P  L  L  V  V  A  G  L  A  V

4741    CGCGTAGTGCCGCAAAGATCAGGGGTGGGCAGTCCTCTATGGCTACCTTCATGACCACCC
        GCGCATCACGGCGTTTCTAGTCCCCACCCGTCAGGAGATACCGATGGAAGTACTGGTGGG
          R  L  A  A  F  I  L  P  P  C  D  E  I  A  V  K  M  V  V  W

4801    AGTTGTAGCCCCAGATCAGGGCAATGGTGACGATGGCGGCAAGGGGAAGCCAGGGAGCAG
        TCAACATCGGGGTCTAGTCCCGTTACCACTGCTACCGCCGTTCCCCTTCGGTCCCTCGTC
          N  Y  G  W  I  L  A  I  T  V  I  A  A  L  P  L  W  P  A  P

                              ← Start ORF7
4861    GGGTCGGGGAGTGAGTGCGCATGAAAGTCCGCGGGCGGCCAAAAAAGATCGGGAAAGACC
        CCCAGCCCCTCACTCACGCGTACTTTCAGGCGCCCGCCGGTTTTTTCTAGCCCTTTCTGG
          T  P  S  H  T  R  M

4921    ACCGCGCCGCTACGGTGGCCTCTCCCGTATCTTAGAATAGCCTAATGATGCACTTTGGCG
        TGGCGCGGCGATGCCACCGGAGAGGGCATAGAATCTTATCGGATTACTACGTGAAACCGC
                                                  *  H  H  V  K  A

           PvuI                    SalI
4981    CGATCGTAGGCAGCCGGCCACGGCCAGTCGACGCCCAACCGTGCGGCCGCCTGCAAGGGC
        GCTAGCATCCGTCGGCCGGTGCCGGTCAGCTGCGGGTTGGCACGCCGGCGGACGTTCCCG
        R  D  Y  A  A  P  W  P  W  D  V  G  L  R  A  A  A  Q  L  P

5041    CAGTATGGATTACGCAGCAGTTCCCGCGCCAGAGCCACGGCGTCCGCCTGTTCCGTGACG
        GTCATACCTAATGCGTCGTCAAGGGCGCGGTCTCGGTGCCGCAGGCGGACAAGGCACTGC
        W  Y  P  N  R  L  L  E  R  A  L  A  V  A  D  A  Q  E  T  V

5101    AGAATGTGCTCTGCCTGCATGGCCTCAGTGATGAGGCCGACGGCGACGGTAGGAATGCCG
        TCTTACACGAGACGGACGTACCGGAGTCACTACTCCGGCTGCCGCTGCCATCCTTACGGC
        L  I  H  E  A  Q  M  A  E  T  I  L  G  V  A  V  T  P  I  G
```

```
5161   ACGGCCCGACGGATGGCTGCGGCGAAGGGCGTCTGGAAGCCGGGGCCAGCGGGGATTTTC
       TGCCGGGCTGCCTACCGACGCCGCTTCCCGCAGACCTTCGGCCCCGGTCGCCCCTAAAAG
       V  A  R  R  I  A  A  A  F  P  T  Q  F  G  P  G  A  P  I  K

                                                              StuI
5221   GCGTCTGGAGTCAGGCCGCCGCTACTGCAGTGTACGACATCCACGCCGCAGGCCTGGAGT
       CGCAGACCTCAGTCCGGCGGCGATGACGTCACATGCTGTAGGTGCGGCGTCCGGACCTCA
       A  D  P  T  L  G  G  S  S  C  H  V  V  D  V  G  C  A  Q  L

5281   TCCCGACTGAAGGCGATGGACTGCTCCAGATCCCAGCCTCCCTCCACCCAATCCGTGGCC
       AGGGCTGACTTCCGCTACCTGACGAGGTCTAGGGTCGGAGGGAGGTGGGTTAGGCACCGG
       E  R  S  F  A  I  S  Q  E  L  D  W  D  D  E  V  W  D  T  A

5341   GATATGCGCACCCAGAGGGGGAGCTCCTCGGGCCAGACGGCGCGTACCGCGGTCGCCACC
       CTATACGCGTGGGTCTCCCCCTCGAGGAGCCCGGTCTGCCGCGCATGGCGCCAGCGGTGG
       S  I  R  V  W  L  P  L  E  E  P  W  V  A  R  V  A  T  A  V

5401   TCCAGCGGAAAGCGCACACGTCCGGAAAAATCGCCGCCGTAAGCGTCGTCCCGTCGGTTG
       AGGTCGCCTTTCGCGTGTGCAGGCCTTTTTAGCGGCGGCATTCGCAGCAGGGCAGCCAAC
       E  L  P  F  R  V  R  G  S  F  D  G  G  Y  A  D  D  R  R  N

5461   CTGATGGGGGAGAGGAAACTGTGCAGGAGATAGCCATGGGCCATGTGCAGTTCCAGCACC
       GACTACCCCCTCTCCTTTGACACGTCCTCTATCGGTACCCGGTACACGTCAAGGTCGTGG
       S  I  P  S  L  F  S  H  L  L  Y  G  H  A  M  H  L  E  L  V

5521   TGAAAGCCGGCCCGGTGCGCACGCCGGGCGGCCGCCACGAAGTCCGCCTTCACCCGCTCC
       ACTTTCGGCCGGGCCACGCGTGCGGCCCGCCGGCGGTGCTTCAGGCGGAAGTGGGCGAGG
       Q  F  G  A  R  H  A  R  R  A  A  A  V  F  D  A  K  V  R  E

5581   ATATCCTCTGTGCTGAGGGGACGGGGGATGGTGTGGGAAGGGCTGAAGGCCAATGCACTG
       TATAGGAGACACGACTCCCCTGCCCCCTACCACACCCTTCCCGACTTCCGGTTACGTGAC
       M  D  E  T  S  L  P  R  P  I  T  H  S  P  S  F  A  L  A  S

5641   GGGGCTACCGGCTGCCAGCCCAGTTCCGCCTCGGGTGGGATGGGTTGACCTCCATGGAAG
       CCCCGATGGCCGACGGTCGGGTCAAGGCGGAGCCCACCCTACCCAACTGGAGGTACCTTC
       P  A  V  P  Q  W  G  L  E  A  E  P  P  I  P  Q  G  G  H  F

                    StuI
5701   GGGGCTGCCGTCGAGGCCTTGCGGCCCGCATGGGCGATCTGTATGCCGGCAACGGCACCT
       CCCCGACGGCAGCTCCGGAACGCCGGGCGTACCCGCTAGACATACGGCCGTTGCCGTGGA
       P  A  A  T  S  A  K  R  G  A  H  A  I  Q  I  G  A  V  A  G

5761   TGCTCCTGCAAAAAGCGAACGATGGGCGCCAGGGCGTCGGCCTGGACATCGTTCCAGATG
       ACGAGGACGTTTTTCGCTTGCTACCCGCGGTCCCGCAGCCGGACCTGTAGCAAGGTCTAC
       Q  E  Q  L  F  R  V  I  P  A  L  A  D  A  Q  V  D  N  W  I

5821   CCCAGATCTTGTGGCGATATGCGACCTTCCGGACTTACCGCCGATGCCTCCACCAAAATG
       GGGTCTAGAACACCGCTATACGCTGGAAGGCCTGAATGGCGGCTACGGAGGTGGTTTTAC
       G  L  D  Q  P  S  I  R  G  E  P  S  V  A  S  A  E  V  L  I

         StuI
5881   AGGCCTGCTCCGCCCACGGCACGACTGCCCAGGTGCACCCGGTGCCAGTCTCCGGCGACG
       TCCGGACGAGGCGGGTGCCGTGCTGACGGGTCCACGTGGGCCACGGTCAGAGGCCGCTGC
       L  G  A  G  G  V  A  R  S  G  L  H  V  R  H  W  D  G  A  V
```


Start ORF8
5941
CCATCGGGAGTCGAATACTGGCACATGGGGGCCATTACGATCCGGTTGCGTAAGGTGACG
GGTAGCCCTCAGCTTATGACCGTGTACCCCCGGTAATGCTAGGCCAACGCATTCCACTGC
G D P T S Y Q C M P A M
6001
GAACGCAGTGTCCAGGGGGTGAACAGACGCTGGGTCACGCCTGCACTCCCAAGCCGCGCC
CTTGCGTCACAGGTCCCCCACTTGTCTGCGACCCAGTGCGGACGTGAGGGTTCGGCGCGG
* A Q V G L G R G
6061
CGAGATGCGCCAGGGTCATTTCCCAGCCCTCCTGCGCCGGTCCGGCGCGGTAGCTGTCGC
GCTCTACGCGGTCCCAGTAAAGGGTCGGGAGGACGCGGCCAGGCCGCGCCATCGACAGCG
L H A L T M E W G E Q A P G A R Y S D R
6121
GGCGGTCACTGAAGAAGCCGTGCGGGGCCTCCGGAAAGACCGTGAGCGTATAACGCTTCT
CCGCCAGTGACTTCTTCGGCACGCCCCGGAGGCCTTTCTGGCACTCGCATATTGCGAAGA
R D S F F G H P A E P F V T L T Y R K K
6181
TGGCGGTGCTCAGGGCCGCAGCGATGCGGCCATGCTCTTCCGGGGTAATGGAGCTGTCCT
ACCGCCACGAGTCCCGGCGTCGCTACGCCGGTACGAGAAGGCCCCATTACCTCGACAGGA
A T S L A A A I R G H E E P T I S S D E
6241
CGGCGCCGTAGAGCAGGAGCAGGGGAGCCTGCATCTGGTTCACGTGACCCAGGAGCGGCG
GCCGCGGCATCTCGTCCTCGTCCCCTCGGACGTAGACCAAGTGCACTGGGTCCTCGCCGC
A G Y L L L L P A Q M Q N V H G L L P P
6301
GACGTCCGGCAGGGTCTTTATCCGGGGCGATACCGCCGCCGTAAAAAGCTACGGTCGCCG
CTGCAGGCCGTCCCAGAAATAGGCCCCGCTATGGCGGCGGCATTTTTCGATGCCAGCGGC
R G A P D K D P A I G G G Y F A V T A A
6361
CCACGCGCTCAGCGAGCTGGGCATTGGCGAGGAAGGCATAGCGCCCACCCATACAAAAGC
GGTGCGCGAGTCGCTCGACCCGTAACCGCTCCTTCCGTATCGCGGGTGGGTATGTTTTCG
V R E A L Q A N A L F A Y R G G M C F G
6421
CCAGCACACCGACCTTTGGACCACGTACCTCAGGACGCTGGAGTACACTCGCCACGGTGG
GGTCGTGTGGCTGGAAACCTGGTGCATGGAGTCCTGCGACCTCATGTGAGCGGTGCCACC
L V G V K P G R V E P R Q L V S A V T T
6481
TTTGGGCTTCCGCCATGACCTGGTCGTCCTTGAGCGTGCGCAGGTGCCCGATGGCGCCGT
AAACCCGAAGGCGGTACTGGACCAGCAGGAACTCGCACGCGTCCACGGGCTACCGCGGCA
Q A E A M V Q D D K L T R L H G I A G D
6541
CGAAGTCCGTGTATTCATAGACCTTGCCGTGATAGATGTCGGGCACGATGGCGCAGTAAC
GCTTCAGGCACATAAGTATCTGGAACGGCACTATCTACAGCCCGTGCTACCGCGTCATTG
F D T Y E Y V K G H Y I D P V I A C Y G
StuI
6601
CGGCCTTGGCGAGGCGCTCTGCCAGCTCTCGAAAATGCTCGTTCACGCCAAAGGCCTCGA
GCCGGAACCGCTCCGCGAGACGGTCGAGAGCTTTTACGAGCAAGTGCGGTTTCCGGAGCT
A K A L R E A L E R F H E N V G F A E I
6661
TGAATACGAGTACCGCGGGGTGCGGTCCAGGATTCGTGGGGCGCGCGTACCAGGCGCGAA
ACTTATGCTCATGGCGCCCCACGCCAGGTCCTAAGCACCCCGCGCGCATGGTCCGCGCTT
F V L V A P H P G P N T P R A Y W A R L


Start ORF9
6721
GCGAAGGGTTCAGTTCCAGCCACTCGGATTGTATTGCAGACATGGAATCTCCTGTTGCTA
CGCTTCCCAAGTCAAGGTCGGTGAGCCTAACATAACGTCTGTACCTTAGAGGACAACGAT
S P N L E L W E S Q I A S M
PvuI
6781
TTCACGGGGTGGGCGACACCATTATCCTACCGACCGATCGGTACTTTCGTAAAGTCTCTG
AAGTGCCCCACCCGCTGTGGTAATAGGATGGCTGGCTAGCCATGAAAGCATTTCAGAGAC
6841
GAGAGGGCTCGAATCCAGGACTCAGGGACGCAAGGAGCCCAGGTACAGGCGGAGCTGCCC
CTCTCCCGAGCTTAGGTCCTGAGTCCCTGCGTTCCTCGGGTCCATGTCCGCCTCGACGGG
* P R L S G L Y L R L Q G
6901
AAGGCAGCGCCGGTACGCATCCACCGACTGGGCATTTTTGCCAAGACCGACGCAACCCTC
TTCCGTCGCGGCCATGCGTAGGTGGCTGACCCGTAAAAACGGTTCTGGCTGCGTTGGGAG
L C R R Y A D V S Q A N K G L G V C G E
6961
GACTCCGGCAACGATGAACAGGGCCGCCTCAGCGGCGACGACATCGCCACGCACGTGACC
CTGAGGCCGTTGCTACTTGTCCCGGCGGAGTCGCCGCTGCTGTAGCGGTGCGTGCACTGG
V G A V I F L A A E A A V V D G R V H G
7021
CGCCTCCTGACCGCGTGCCAGGGCTTGGGCAAGCACCTCCGTCCAGCGCTCGAGGATCAC
GCGGAGGACTGGCGCACGGTCCCGAACCCGTTCGTGGAGGCAGGTCGCGAGCTCCTAGTG
A E Q G R A L A Q A L V E T W R E L I V
7081
CTGCAGGGCTTGGCGAAATTGCTCGTCCAGCGGGCTCATTTCTTGAATCAGGTTGTTGAG
GACGTCCCGAACCGCTTTAACGAGCAGGTCGCCCGAGTAAAGAACTTAGTCCAACAACTC
Q L A Q R F Q E D L P S M E Q I L N N L
7141
GGGGCAGCCGTAGCGCAAAACCCGTTCGTCCGCGCTGGCGATCTTGTCGCAGAGGATTTT
CCCCGTCGGCATCGCGTTTTGGGCAAGCAGGCGCGACCGCTAGAACAGCGTCTCCTAAAA
P C G Y R L V R E D A S A I K D C L I K
7201
GTCCAGTTGCTCCAGAGGGTCATCGCCCTCGCTGAGGGGCGCAAACCAGCGTCTTGCCAG
CAGGTCAACGAGGTCTCCCAGTAGCGGGAGCGACTCCCCGCGTTTGGTCGCAGAACGGTC
D L Q E L P D D G E S L P A F W R R A L
7261
ATCGGGAGCGATTACGTCGTGCACCACCGCAAGGCCAAGTTCCTGCTTGGTGCCGAAGTG
TAGCCCTCGCTAATGCAGCACGTGGTGGCGTTCCGGTTCAAGGACGAACCACGGCTTCAC
D P A I V D H V V A L G L E Q K T G F H
7321
GTGGTACAGCGCGCCCTTGCTCACAGAGGTCTTGGCCAGGATGCGCTCCAGACTTGCCGC
CACCATGTCGCGCGGGAACGAGTGTCTCCAGAACCGGTCCTACGCGAGGTCTGAACGGCG
H Y L A G K S V S T K A L I R E L S A A
7381
CTGAAAGCCGCATTCCAGTATCTCAGCGTAGGCTGCTGCGAGCAGGTCGTTGCGCGTCTG
GACTTTCGGCGTAAGGTCATAGAGTCGCATCCGACGACGCTCGTCCAGCAACGCGCAGAC
Q F G C E L I E A Y A A A L L D N R T Q
7441
CTCCGGATTACGAGATTTTCTCAGTGTGTCCACCATGGTCTGATCAAAATACCGACCAAT
GAGGCCTAATGCTCTAAAAGAGTCACACAGGTGGTACCAGACTAGTTTTATGGCTGGTTA
E P N R S K R L T D V M T Q D F Y R G I
7501
CGGTATTTCGTCAAGACCTTGTATGCGTCCACCACCTTTGGCACTGGGGCTGATGTTGCA
GCCATAAAGCAGTTCTGGAACATACGCAGGTGGTGGAAACCGTGACCCCGACTACAACGT
P I E D L G Q I R G G G K A S P S I N C

```
7561   GGAGGAAGGATAGCGGTGGTTGTTGGCCGAGTCGATGGTGTGGTCGTTCGGCGTTGTAGA
       CCTCCTTCCTATCGCCACCAACAACCGGCTCAGCTACCACACCAGCAAGCCGCAACATCT
        S  S  P  Y  R  H  N  N  A  S  D  I  T  H  D  N  P  T  T  S

7621   AGACGAGCCACTCGGCGAGTTTTTGGTTGAAGCGGGCGAGGTCGGTGAAGAGCAGGTCTT
       TCTGCTCGGTGAGCCGCTCAAAAACCAACTTCGCCCGCTCCAGCCACTTCTCGTCCAGAA
        S  S  G  S  P  S  N  K  T  S  A  P  S  T  P  S  S  C  T  K

              SalI                                    Start ORF10 ←
7681   CGTGGTAGTCGACGAAGGATTCCTGGATGGTGCGATTGAAGCGTTCGGCATGGGCATTCA
       GCACCATCAGCTGCTTCCTAAGGACCTACCACGCTAACTTCGCAAGCCGTACCCGTAAGT
        T  T  T  S  S  P  N  R  S  P  A  I  S  A  N  P  M  P  M

7741   TTTTTGGGGTCTTGGGGTAGGTATACCAACGGGCGATCTGCTGTTCCGTGAGGGTTTGGG
       AAAAACCCCAGAACCCCATCCATATGGTTGCCCGCTAGACGACAAGGCACTCCCAAACCC

7801   ACGGCGGCGAGGGCCGCCTGGGTGTGGCGTGTGGCTTTGCTGGGTAAAGCGACGGCCAAG
       TGCCGCCGCTCCCGGCGGACCCACACCGCACACCGAAACGACCCATTTCGCTGCCGGTTC

7861   GCAAAGGCGCTGGCTGGATCGATGAAGGTGACGAGGTAGCGTTTCAGCCCGTCGCGAACC
       CGTTTCCGCGACCGACCTAGCTACTTCCACTGCTCCATCGCAAAGTCGGGCAGCGCTTGG

7921   TGCTCGATAGTGTCCACCGCCAGGCACTGCAAGGCAGTGGTGGCCACCCCCTTGGGCTTG
       ACGAGCTATCACAGGTGGCGGTCCGTGACGTTCCGTCACCACCGGTGGGGGAACCCGAAC

7981   CGCGGCTTGCGAGAACGCCGCACAGGCTTGGCGCGACCACGGGCATCGAGGCGCGCGGGG
       GCGCCGAACGCTCTTGCGGCGTGTCCGAACCGCGCTGGTGCCCGTAGCTCCGCGCGCCCC

8041   CTGTGGCGCATTTTATCCGGTGCACGGGCGATGATTCTGCCGAGGGTGGACACCGATGGC
       GACACCGCGTAAAATAGGCCACGTGCCCGCTACTAAGACGGCTCCCACCTGTGGCTACCG

8101   AGCGCGATACCCTTCTCCGCACACCAGGGGCGCAGCAGGACGTGCAGCTTGTCCTTGCCG
       TCGCGCTATGGGAAGAGGCGTGTGGTCCCCGCGTCGTCCTGCACGTCGAACAGGAACGGC

8161   AGATTGGGGTGAAGGGTACGTAGCCGCCGAATTTCCGCTACCAGCCTGGGGTCGGTCTTG
       TCTAACCCCACTTCCCATGCATCGGCGGCTTAAAGGCGATGGTCGGACCCCAGCCAGAAC

8221   GGGCTGCGTCGGCGTTTGGGCGCGCTGGACTGGCTGGCCAGGGCGGCGGGATTGCCGCCT
       CCCGACGCAGCCGCAAACCCGCGCGACCTGACCGACCGGTCCCGCCGCCCTAACGGCGGA

8281   TCGGCCTTCAAGGCCGCTCGCCACCGATAGAGCGTCCTTCGGGACACCCCAAAGGCATCG
       AGCCGGAAGTTCCGGCGAGCGGTGGCTATCTCGCAGGAAGCCCTGTGGGGTTTCCGTAGC

8341   CAGGCGGCCACCAAACCGTGTTTGTCCCAGAACTGTACAATTCTGAGTCGCTCTTGCGCA
       GTCCGCCGGTGGTTTGGCACAAACAGGGTCTTGACATGTTAAGACTCAGCGAGAACGCGT

                                                              BamHI
8401   TCTTTCGGAGTCATCTCCCAACGATGCCCCAAGCCCGCTACCCGGTAAAACCCTCGGATC
       AGAAAGCCTCAGTAGAGGGTTGCTACGGGGTTCGGGCGATGGGCCATTTTGGGAGCCTAG

8461   CAGTTCTTAATAAGAGGTCATTGTTAATACGCAAAAAACGGGGGTTGGGATGCGTTGAAT
       GTCAAGAATTATTCTCCAGTAACAATTATGCGTTTTTTGCCCCCAACCCTACGCAACTTA

8521   AGGGTAGCCTTGGCGACGATCAGCGCAGCCACCACAGCAATAACAGCAATAGCGCGCTGC
       TCCCATCGGAACCGCTGCTAGTCGCGTCGGTGGTGTCGTTATTGTCGTTATCGCGCGACG
```

```
                                                EcoRI
8581    CGATGGTGACGGCACCGCTGATCAGGGCGACGCCGAGAATTCTGCGTGCCTGCACGTACT
        GCTACCACTGCCGTGGCGACTAGTCCCGCTGCGGCTCTTAAGACGCACGGACGTGCATGA

8641    CGCTCTGGGGTGATTGCTCTTTATCGCTCACGGTGTCGGCTCCCGCGCCTCTTCAAAGGC
        GCGAGACCCCACTAACGAGAAATAGCGAGTGCCACAGCCGAGGGCGCGGAGAAGTTTCCG

                                                      EcoRV
8701    ACAGACCCAAATGTCACGCCCGAGTTCCCGCGTCGGTAGCCGATATCCCTCCTCCCGGTC
        TGTCTGGGTTTACAGTGCGGGCTCAAGGGCGCAGCCATCGGCTATAGGGAGGAGGGCCAG

8761    CAGACTGTTGGGAGAACGAAGGCTCAGGGCGTAAGCACCACTCCGACGGAAATCGTCCAC
        GTCTGACAACCCTCTTGCTTCCGAGTCCCGCATTCGTGGTGAGGCTGCCTTTAGCAGGTG

8821    GTATCCGGATGCTCGCCTCTGCGAACAGCGGAGCGGTAGGAACAGATCTCCGCGCAGGCG
        CATAGGCCTACGAGCGGAGACGCTTGTCGCCTCGCCATCCTTGTCTAGAGGCGCGTCCGC

8881    TGCATAAGTTCACAAACCTTTGGCACTTCTTTGGCGAGGGAGGAGTGCCATGCAGGTTCA
        ACGTATTCAAGTGTTTGGAAACCGTGAAGAAACCGCTCCCTCCTCACGGTACGTCCAAGT

                    BamHI
8941    ACATGTCGGGTACCGGATCCAGGCGTCTTCCAACGCTTGCAAAAAATGCGTGTGGCCACC
        TGTACAGCCCATGGCCTAGGTCCGCAGAAGGTTGCGAACGTTTTTTACGCACACCGGTGG

9001    GTCCATAGTCGCTGGAGTGTGTTCAGGAGGTGAGCTGTGGCTGCGGAACGCGGTGGCGGT
        CAGGTATCAGCGACCTCACACAAGTCCTCCACTCGACACCGACGCCTTGCGCCACCGCCA

9061    GTGCGTTTGCTGCGATACAGGGCTAGCGTGTTGTGGTTGGCGCTGCTCTTGTTGGCCTTT
        CACGCAAACGACGCTATGTCCCGATCGCACAACACCAACCGCGACGAGAACAACCGGAAA

9121    TCGGCGCAGGCTCTGGGGGTAGACGAGGGGAATCTGCAGGTTGAGGTACAGCGGGTTGCC
        AGCCGCGTCCGAGACCCCCATCTGCTCCCCTTAGACGTCCAACTCCATGTCGCCCAACGG

9181    GATGGCTTCGCGGTATATGCGGTATTCCCGGTAGCGGTGACGCCAGCTCGAGCCTTTGCG
        CTACCGAAGCGCCATATACGCCATAAGGGCCATCGCCACTGCGGTCGAGCTCGGAAACGC

9241    GTGATGACGGATTACGGGCATATGGCGCAGTTCATCCCGCAGATGCACAAAAGCCAGGTG
        CACTACTGCCTAATGCCCGTATACCGCGTCAAGTAGGGCGTCTACGTGTTTTCGGTCCAC

9301    CTGTGGCGTGACGGAGCGCACGAGTCGGTACTCCAGGAAGGGTCCATCAAACTCCTGTGG
        GACACCGCACTGCCTCGCGTGCTCAGCCATGAGGTCCTTCCCAGGTAGTTTGAGGACACC

9361    TTACGTATCCCCACCTTTGTGGTGATGTCCGTGCATCGGCTGTCGGATCGAGCGGTGCGC
        AATGCATAGGGGTGGAAACACCACTACAGGCACGTAGCCGACAGCCTAGCTCGCCACGCG

9421    TTTCACAGTACCGGAGGGAGCATGGCGATCCGCGGGCAGGCCGAGGTCCGCGCGTCGGCC
        AAAGTGTCATGGCCTCCCTCGTACCGCTAGGCGCCCGTCCGGCTCCAGGCGCGCAGCCGG

9481    GACGGCAGCGTGGTGGAGTATCGCGCCACACTGGAACCCCAGACCCTGGTGCCCATGGGT
        CTGCCGTCGCACCACCTCATAGCGCGGTGTGACCTTGGGGTCTGGGACCACGGGTACCCA

                                                        StuI
9541    CTTGCCCAGAATCTCGTGGCTCGCTACATCCGCGGGCAGATGGAGGCCTTACGCGCGGAA
        GAACGGGTCTTAGAGCACCGAGCGATGTAGGCGCCCGTCTACCTCCGGAATGCGCGCCTT
```

```
9601    ATGCTGCGGTTTGGGTGACGGGCGCCTACATCCAAAAGTCTGCCGCTGTCGTAGTGTGGG
        TACGACGCCAAACCCACTGCCCGCGGATGTAGGTTTTCAGACGGCGACAGCATCACACCC

9661    TGGGCAGCATGCGATTGCCCACGTTTCCTTCCTTCCTGATC
        ACCCGTCGTACGCTAACGGGTGCAAAGGAAGGAAGGACTAG
```

Reference List

1. **Altshul S.F., Madden T.L., Schaffer A.A., Zhang Z., Miller W., and Lipman D.J.** (1997). Gapped BLAST abd PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* **25**:3308-3402.

2. **Arredondo R., Garcia A., and Jerez C.A.** (1994). Partial removal of lipopolysaccharide from *Thiobacillus ferrooxidans* affects its adhesion to solids. *Appl. Envir. Microbiol.* **60**:2846-2851.

3. **Bairoch A.** (1993). A possible mechanism for metal-ion induced DNA-protein dissociation in a family of prokaryotic transcriptional regulators. *Nucl. Acids Res.* **21**:2515.

4. **Bennett M.S., Guan Z., Laurberg M., and Su X.D.** (2001). *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases. *PNAS* **98**:13577-13582.

5. **Bernat B.A., Laughlin L.T., and Armstrong R.N.** (1997). Fosfomycin resistance protein (FosA) is a manganese metalloglutathione transferase related to glyoxalase I and the extradiol dioxygenases. *Biochemistry* **36**:3050-3055.

6. **Bhattacharjee H., Li J., Ksenzenko M., and Rosen B.** (1995). Role of cysteinyl residues in metalloactivation of the oxyanion- translocating ArsA ATPase. *J. Biol. Chem.* **270**:11245-11250.

7. **Bhattacharjee H., Zhou T., Li J., Gatti D.L., Walmsley A.R., and Rosen B.P.** (2000). Structure-function relationships in an anion-translocating ATPase. *Biochem. Soc. Trans.* **28**:520-523.

8. **Bhattacharjee H. and Rosen B.P.** (1996). Spatial Proximity of Cys113, Cys172, and Cys422 in the Metalloactivation Domain of the ArsA ATPase. *J. Biol. Chem.* **271**:24465-24470.

9. **Boyer H.W. and Roulland-Dussoix D.** (1969). A complementation analysis of the restriction and modification DNA in *Escherichia coli*. *J. Mol. Biol.* **41**:459-472.

10. **Busenlehner L.S., Pennella M.A., and Giedroc D.P.** (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. *FEMS Microbiology Rev.* **27**:131-143.

11. **Bushweller J.H., Aslund F., Wuthruch K., and Holmgren A.** (1992). Structural and functional characterization of the mutant *Escherichia coli* glutaredoxin (C14 - S) and its mixed disulfide with glutathione. *Biochemistry* **31**:9288-9293.

12. **Butcher B.G., Deane S.M., and Rawlings D.E.** (2000). The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *Appl. Environ. Microbiol.* **66**:1826-1833.

13. **Butcher B.G. and Rawlings D.E.** (2002). The divergent chromosomal *ars* operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. *Microbiology* **148**:3983-3992.

14. **Cai J., Salmon K., and DuBow M.S.** (1998). A chromosomal *ars* operon homolog of *Pseudomonas aeruginosa* confers increased resistance to arsenic and antimony in *Escherichia coli*. *Microbiology* **144**:2705-2713.

15. **Carlin A., Shi W., Dey S., and Rosen B.P.** (1995). The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. *J. Bacteriol.* **177**:981-986.

16. **Casadaban M.J., Martinez-Arias A., Shapira S.K., and Chou J.** (1983). β-Galactosidase gene fusions for analysing gene expression in *Escherichia coli* and yeast. *Methods Enzymol.* **100**:293-308.

17. **Cervantes C., ji G., Ramirez J., and Silver S.** (1994). Resistance to arsenic compounds in microorganisms. *FEMS Microbiology Reviews* **15**:355-367.

18. **Chen C.M., Misra T.K., Silver S., and Rosen B.P.** (1986). Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *J. Biol. Chem.***261**:15030-15038.

19. **Chen Y., Dey S., and Rosen B.P.** (1996). Soft metal thiol chemistry is not involved in the transport of arsenite by the Ars pump. *J. Bacteriol.* **178**:911-913.

20. **Chen Y. and Rosen B.P.** (1997). Metalloregulatory Properties of the ArsD Repressor. *J. Biol. Chem.***272**:14257-14262.

21. **Ching M., Kaur P., Karkaria C., Steiner R., and Rosen B.** (1991). Substrate-induced dimerization of the ArsA protein, the catalytic component of an anion-translocating ATPase. *J. Biol. Chem.* **266**:2327-2332.

22. **Coddington K.** (1986). A review of Arsenicals in Biology. *Toxicol. Environ. Chem.* **11**:281-290.

23. **Cook W.J., Kar S.R., Taylor K.B., and Hall L.M.** (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J. Mol. Biol.* **275**:337-346.

24. **De Groot P., Deane S.M., and Rawlings D.E.** (2003). A transposon-located arsenic resistance mechanism from a strain of *Acidithiobacillus caldus* isolated from commercial, arsenopyrite biooxidation tanks. *Hydrometallurgy* **71**:115-123.

25. **DeMel S., Shi J., Martin P., Rosen B.P., and Edwards B.F.P.** (2004). Arginine 60 in the ArsC arsenate reductase of *E. coli* plasmid R773 determines the chemical nature of the bound As(III) product. *Protein Sci.* **13**:2330-2340.

26. **Dew D.W., Lawson E.N., and Broadhurst J.L.** (1997). The BIOX® process for bioxidation of gold-bearing ores or concentrates. In *Biomining: Theory, Microbes and Industrial Processes.* p45-80. Edited by D.E. Rawlings. Berlin:Springer.

27. **Dey S., Dou D., and Rosen B.P.** (1994). ATP-dependent arsenite transport in everted membrane vesicles of *Escherichia coli*. *J. Biol. Chem.* **269**:25442-25446.

28. **Dey S. and Rosen B.P.** (1995). Dual mode of energy coupling by the oxyanion-t ranslocating ArsB protein. *J. Bacteriol.***177**:385-389.

29. **Diorio C., Cai J., Marmor J., Shinder R., and DuBow M.S.** (1995). An *Escherichia coli* chromosomal *ars* operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. *J. Bacteriol.* **177**:2050-2056.

30. **Dopson M., Lindstrom E.B., and Hallberg K.B.** (2001). Chromosomally encoded arsenical resistance of the moderately thermophilic acidophile *Acidithiobacillus caldus*. *Extremophiles* **5**:247-255.

31. **Dopson M. and Lindstrom E.B.** (1999). Potential Role of *Thiobacillus caldus* in Arsenopyrite Bioleaching. *Appl. Environ. Microbiol.* **65**:36-40.

32. **Dou D., Owolabi J.B., Dey S., and Rosen B.P.** (1992). Construction of a chimeric ArsA-ArsB protein for overexpression of the oxyanion-translocating ATPase. *J. Biol. Chem.* **267**:25768-25775.

33. **Gihring T.M., Bond P.L., Peters S.C., and Banfield J.F.** (2003). Arsenic resistance in the archaeon "*Ferroplasma acidarmanus*": new insights into the structure and evolution of the *ars* genes. *Extremophiles* **7**:123-130.

34. **Gladysheva T., Liu J., and Rosen B.P.** (1996). His-8 lowers the pKa of the essential Cys-12 residue of the ArsC arsenate reductase of plasmid R773. *J. Biol. Chem.* **271**:33256-33260.

35. **Goebel B.M. and Stackebrandt E.** (1994). Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl. Environ. Microbiol.* **60**:1614-1621.

36. **Guillamet E., Creus A., Ponti J., Sabbioni E., Fortaner S., and Marcos R.** (2004). In vitro DNA damage by arsenic compounds in a human lymphoblastoid cell line (TK6) assessed by the alkaline Comet assay. *Mutagenesis* **19**:129-135.

37. **Guo X., Li Y., Peng K., Hu Y., Li C., Xia B., and Jin C.** (2005). Solution Structures and Backbone Dynamics of Arsenate Reductase from *Bacillus subtilis*: Reversible conformational switch associated with arsenate reduction. *J. Biol. Chem.* **280**:39601-39608.

38. **Hallberg K.B. and Lindstrom E.B.** (1994). Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. *Microbiology* **140**:3451-3456.

39. **Hallberg K.B., Dopson M., and Lindstrom E.B.** (1996a). Reduced sulfur compound oxidation by *Thiobacillus caldus*. *J. Bacteriol.* **178**:6-11.

40. **Hallberg K.B. and Lindstrom E.B.** (1996). Multiple serotypes of the moderate thermophile *Thiobacillus caldus*, a limitation of immunological assays for biomining microorganisms. *Appl. Environ. Microbiol.* **62**:4243-4246.

41. **Hallberg K.B., Dopson M., and Lindstrom E.B.** (1996b). Arsenic toxicity is not due to a direct effect on the oxidation of reduced inorganic sulfur compounds by *Thiobacillus caldus*. *FEMS Microbiol. Lett.* **145**:409-414.

42. **Hotter G.S., Wilson T., and Collins D.M.** (2001). Identification of a cadmium-induced gene in *Mycobacterium bovis* and *Mycobacterium tuberculosis*. *FEMS Microbiol. Lett.* **200**:151-155.

43. **Hsu C.M. and Rosen B.P.** (1989). Characterization of the catalytic subunit of an anion pump. *J. Biol. Chem.* **264**:17349-17354.

44. **Jackson C.R. and Dugas S.L.** (2003). Phylogenetic analysis of bacterial and archaeal *arsC* gene sequences suggests an ancient, common origin for arsenate reductase. *BMC Evol. Biol.* **3:**18.

45. **Ji G., Garber E., Armes L., Chen C., Fuchs J., and Silver S.** (1994). Arsenate reductase of *Staphylococcus aureus* plasmid pI258. *Biochemistry* **33**:7294-7299.

46. **Ji G. and Silver S.** (1992). Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. *PNAS* **89**:9474-9478.

47. **Jia H. and Kaur P.** (2001). Role of the Linker Region of the Anion-stimulated ATPase ArsA. Effect of deletion and point mutations in the linker region *J. Biol. Chem.* **276**:29582-29587.

48. **Jia H. and Kaur P.** (2003). Biochemical evidence for interaction between the two nucleotide binding domains of ArsA. Insights from mutants and ATP analogs. *J. Biol. Chem.* **278**:6603-6609.

49. **Jiang Y., Bhattacharjee H., Zhou T., Rosen B.P., Ambudkar S.V., and Sauna Z.E.** (2005). Nonequivalence of the nucleotide binding domains of the ArsA ATPase. *J. Biol. Chem.* **280**:9921-9926.

50. **Jin S.M., Yan W.M., and Wang Z.N.** (1992). Transfer of IncP plasmids to extremely acidophilic *Thiobacillus thiooxidans*. *Appl. Environ. Microbiol.* **58**:429-430.

51. **Karkaria C.E., Chen C.M., and Rosen B.P.** (1990). Mutagenesis of a nucleotide-binding site of an anion-translocating ATPase. *J. Biol. Chem.* **265**:7832-7836.

52. **Kaur P. and Rosen B.P.** (1992). Mutagenesis of the C-terminal nucleotide-binding site of an anion- translocating ATPase. *J. Biol. Chem.* **267**:19272-19277.

53. **Kaur P.** (1999). The anion-stimulated ATPase ArsA shows unisite and multisite catalytic activity. *J. Biol. Chem.* **274**:25849-25854.

54. **Kelly D.P. and Wood A.P.** (2000). Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov.,

Halothiobacillus gen. nov. and *Thermithiobacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**:511-516.

55. **Kenney L.J. and Kaplan J.H.** (1988). Arsenate substitutes for phosphate in the human red cell sodium pump and anion exchanger. *J. Biol. Chem.* **263**:7954-7960.

56. **Kondratyeva T.F., Muntyan L.N. and Karavaiko G.I.** (1995). Zinc- and arsenic-resistant strains of *Thiobacillus ferrooxidans* have increased copy numbers of chromosomal resistence genes. *Microbiol*. **141**:1157-1162.

57. **Kostal J., Yang R., Wu C.H., Mulchandani A., and Chen W.** (2004). Enhanced arsenic accumulation in engineered bacterial cells expressing ArsR. *Appl. Environ. Microbiol.* **70**:4582-4587.

58. **Kuroda M., Dey S., Sanders O.I., and Rosen B.P.** (1997). Alternate energy coupling of ArsB, the membrane subunit of the *ars* anion-translocating ATPase. *J. Biol. Chem.* **272**:326-331.

59. **Kusano T., Sugawara K., Inoue C., Takeshima T., Numata M., and Shiratori T.** (1992). Electrotransformation of *Thiobacillus ferrooxidans* with plasmids containing a *mer* determinant. *J. Bacteriol.* **174**:6617-6623.

60. **Lah N., Lah J., Zegers I., Wyns L., and Messens J.** (2003). Specific potassium binding stabilizes pI258 arsenate reductase from *Staphylococcus aureus*. *J. Biol. Chem.* **278**:24673-24679.

61. **Li J. and Rosen B.P.** (2000). The linker peptide of the ArsA ATPase. *Mol. Microbiol.* **35**:361-367.

62. **Li J., Liu S., and Rosen B.P.** (1996). Interaction of ATP binding sites in the ArsA ATPase, the catalytic subunit of the *ars* pump. *J. Biol. Chem.* **271**:25247-25252.

63. **Li J. and Rosen B.P.** (1998). Steric limitations in the interaction of the ATP binding domains of the ArsA ATPase. *J. Biol. Chem.* **273**:6796-6800.

64. **Li R., Haile J.D., and Kennelly P.J.** (2003). An arsenate reductase from *Synechocystis* sp. strain PCC 6803 exhibits a novel combination of catalytic characteristics. *J. Bacteriol.* **185**:6780-6789.

65. **Li S., Chen Y., and Rosen B.P.** (2001). Role of vicinal cysteine pairs in metalloid sensing by the ArsD As(III)-responsive repressor. *Mol. Microbiol.* **41**:687-696.

66. **Li S., Rosen B.P., Borges-Walmsley M.I., and Walmsley A.R.** (2002). Evidence for cooperativity between the four binding sites of dimeric ArsD, an As(III)-responsive transcriptional regulator. *J. Biol. Chem.* **277**:25992-26002.

67. **Lim C.J., Gleason F.K., and Fuchs J.A.** (1986). Cloning, expression, and characterization of the *Anabaena* thioredoxin gene in *Escherichia coli*. *J. Bacteriol.* **168**:1258-1264.

68. **Liu J., Gladysheva T.B., Lee L., and Rosen B.P.** (1995). Identification of an essential cysteinyl residue in the ArsC arsenate reductase of plasmid R773. *Biochemistry* **34**:13472-13476.

69. **Liu J. and Rosen B.P.** (1997). Ligand interactions of the ArsC arsenate reductase. *J. Biol. Chem.* **272**:21084-21089.

70. **Liu Z., Borne F., Ratouchniak J., and Bonnefoy V.** (1999). Genetic tranfer of IncP, IncQ, IncW plasmids to four *Thiobacillus ferrooxidans* strains by conjugation. In: *Biohydrometallurgy and the environment toward the 21st century, PartB, Elsevier, Amsterdam.* p39-50. Edited by R. Amils and A. Ballester.

71. **Lopez-Maury L., Florencio F.J., and Reyes J.C.** (2003). Arsenic sensing and resistance system in the *Cyanobacterium synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **185**:5363-5371.

72. **Martin A.** (1990). Keeping a neutral cytoplasm: the bioenergetics of obligate acidophiles. *FEMS Microbiol. Rev.* **75**:307-318.

73. **Martin P., DeMel S., Shi J., Gladysheva T., Gatti D.L., Rosen B.P., and Edwards B.F.P.** (2001). Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure* **9**:1071-1081.

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

75. **Messens J., Hayburn G., Desmyter A., Laus G., and Wyns L.** (1999). The essential catalytic redox couple in arsenate reductase from *Staphylococcus aureus*. *Biochemistry* **38**:16857-16865.

76. **Messens J., Martins J.C., Van Belle K., Brosens E., Desmyter A., De Gieter M., Wieruszeski J.M., Willem R., Wyns L., and Zegers I.** (2002). All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. *PNAS* **99**:8506-8511.

77. **Messens J., Molle I.V., Vanhaesebrouck P., Limbourg M., Belle K.V., Wahni K., Martins J.C., Loris R., and Wyns L.** (2004). How thioredoxin can reduce a buried disulphide bond. *J. Mol. Biol.* **339**:527-537.

78. **Miller J.H.** (1972). *Experiments in molecular genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

79. **Moore C.M., Gaballa A., Hui M., Ye R.W., and Helmann J.D.** (2005). Genetic and physiological responses *of Bacillus subtilis* to metal ion stress. *Mol. Microbiol.* **57**:27-40.

80. **Mukhopadhyay R., Rosen B.P., Phung L.T., and Silver S.** (2002). Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol. Rev.* **26**:311-325.

81. **Mukhopadhyay R. and Rosen B.P.** (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environ. Health Perspec.* **110**:745-748.

82. **Neyt C., Iriarte M., Thi V.H., and Cornelis G.R.** (1997). Virulence and arsenic resistance in *Yersiniae*. *J. Bacteriol.* **179**:612-619.

83. **Oden K.L., Gladysheva T.B., and Rosen B.P.** (1994). Arsenate reduction mediated by the plasmid-encoded ArsC protein is coupled to glutathione. *Mol. Microbiol.* **12**:301-306.

84. **Okibe N., Gericke M., Hallberg K.B., and Johnson D.B.** (2003). Enumeration and characterization of acidophilic microorganisms isolated from a pilot plant stirred-tank bioleaching operation. *Appl. Environ. Microbiol.* **69**:1936-1943.

85. **Okibe N. and Johnson B.** (2004). Biooxidation of pyrite by defined mixed cultures of moderately thermophilic acidophiles in pH-controlled bioreactors: Significance of microbial interactions. *Biotechnol. Bioeng.* **87**:574-583.

86. **Oremland R.S. and Stolz J.F.** (2003). The ecology of arsenic. *Science* **300**:939-944.

87. **Pai E.F., Krengel U., Petsko G.A., Goody R.S., Kabsch W., and Wittinghofer A.** (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**:2351-2359.

88. **Peng J.B., Yan W.M., and Bao X.Z.** (1994). Expression of heterogenous arsenic resistance genes in the obligately autotrophic biomining bacterium *Thiobacillus ferrooxidans*. *Appl. Environ. Microbiol.* **60**:2653-2656.

89. **Powles R.E., Deane S.M., and Rawlings D.E.** (1995). Molecular genetic analysis of a thioredoxin gene from *Thiobacillus ferrooxidans*. *Microbiology* **141**:2175-2181.

90. **Rawlings D.E. and Woods D.R.** (1995). Development of improved biomining bacteria. In *Bioextraction and Biodeterioration of Metals*. p64-84. Edited by Gaylarde, C and Videla, H. Cambridge University Press, UK.

91. **Rawlings D.E. and Kusano T.** (1994). Molecular genetics of T*hiobacillus ferrooxidans*. *Microbiol. Rev.* **58**:39-55.

92. **Rawlings D.E., Tributsch H., and Hansford G.S.** (1999a). Reasons why '*Leptospirillum*'-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. *Microbiology* **145**:5-13.

93. **Rawlings D.E.** (1995). Restriction enzyme analysis of 16S rRNA genes for the rapid identification of *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans* and *Leptospirillum ferrooxidans* strains in leaching environments. In: *Biohydrometallurgical Processing*. p9-17. Edited by Jerez, C.A., Vargas, T., Toledo, H and Wiertz, J.V. University of Chile.

94. **Rawlings D.E.** (2002). Heavy metal mining using microbes. *Annu. Rev. Microbiol.* **56**:65-91.

95. **Rawlings D.E.** (2005). Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microb. Cell Fact.* **4**:13.

96. **Rawlings D.E., Coram N.J., Gardner M.N. and Deane S.M.** (1999b). *Thiobacillus caldus* and *Leptospirillum ferrooxidans* are widely distributed in continuous flow biooxidation tanks used to treat a variety of metal containing ores and concentrates. In: *Biohydrometallurgy and the environment: toward the mining of the 21st century. Part A. Elsevier, Amsterdam*. p777-786. Edited by R. Amils and A. Ballester. .

97. **Rawlings D.E., Dorrington R.A., Rohrer J., and Clennel A.M.** (1993). A molecular analysis of a broad-host-range plasmid isolated from *Thiobacillus ferrooxidans*. *FEMS Microbiol. Rev.* **11**:3-7.

98. **Rensing C., Ghosh M., and Rosen B.P.** (1999). Families of soft-metal-ion-transporting ATPases. *J. Bacteriol.* **181**:5891-5897.

99. **Rohwerder T., Gehrke T., Kinzler K., and Sand W.** (2003). Bioleaching review part A. *Appl. Microbiol. and Biotech.* **63**:239-248.

100. **Rosen B.P.** (2002a). Biochemistry of arsenic detoxification. *FEBS Lett.* **529**:86-92.

101. **Rosen B.P., Bhattacharjee H., Zhou T., and Walmsley A.R.** (1999). Mechanism of the ArsA ATPase. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1461**:207-215.

102. **Rosen B.P.** (1999). Families of arsenic transporters. *Trends Microbiol.* **7**:207-212.

103. **Rosen B.P.** (2002b). Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **133**:689-693.

104. **Ryan D. and Colleran E.** (2002). Arsenical resistance in the IncHI2 plasmids. *Plasmid* **47**:234-240.

105. **Saltikov C.W. and Olson B.H.** (2002). Homology of *Escherichia coli* R773 *arsA*, *arsB*, and *arsC* genes in arsenic-resistant bacteria isolated from raw sewage and arsenic-enriched creek waters. *Appl. Environ. Microbiol.* **68**:280-288.

106. **Sambrook J., Fritsch E.F., and Maniatis T.** (1989). *Molecular cloning: a Laboratory Manual, 2nd edn.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

107. **San Francisco M.J.D., Tisa L.S., and Rosen B.P.** (1989). Identification of the membrane component of the anion pump encoded by the arsenical resistance operon of R-factor R773. *Mol. Microbiol.* **3**:15-21.

108. **Sanders O.I., Rensing C., Kuroda M., Mitra B., and Rosen B.P.** (1997). Antimonite is accumulated by the glycerol facilitator GlpF in *Escherichia coli*. *J. Bacteriol.* **179**:3365-3367.

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

110. **Shi J., Vlamis-Gardikas A., Aslund F., Holmgren A., and Rosen B.P.** (1999). Reactivity of glutaredoxins 1, 2, and 3 from *Escherichia coli* shows that glutaredoxin 2 is the primary hydrogen donor to ArsC-catalyzed arsenate reduction. *J. Biol. Chem.* **274**:36039-36042.

111. **Shi J., Mukhopadhyay R., and Rosen B.P.** (2003). Identification of a triad of arginine residues in the active site of the ArsC arsenate reductase of plasmid R773. *FEMS Microbiol. Lett.* **227**:295-301.

112. **Shi W., Wu J., and Rosen B.P.** (1994). Identification of a putative metal binding site in a new family of metalloregulatory proteins. *J. Biol. Chem.* **269**:19826-19829.

113. **Shi W., Dong J., Scott R.A., Ksenzenko M.Y., and Rosen B.P.** (1996). The role of arsenic-thiol interactions in metalloregulation of the *ars* operon. *J. Biol. Chem.* **271**:9291-9297.

114. **Silver S., Budd K., Leahy K.M., Shaw W.V., Hammond D., Novick R.P., Willsky G.R., Malamy M.H., and Rosenberg H.** (1981). Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in *Escherichia coli* and *Staphylococcus aureus*. *J. Bacteriol.* **146**:983-996.

115. **Silver S.** (1996). Bacterial resistances to toxic metal ions - a review. *Gene* **179**:9-19.

116. **Silver S. and Phung L.T.** (1996). Bacterial heavy metal resistance: New Surprises. *Annu. Rev. Microbiol.* **50**:753-789.

117. **Suzuki I.** (2001). Microbial leaching of metals from sulfide minerals. *Biotechnol. Adv.* **19** :119-132.

118. **Suzuki K., Wakao N., Kimura T., Sakka K., and Ohmiya K.** (1998). Expression and regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU 301 plasmid pKW301 in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**:411-418.

119. **Tait R.C., Lundquist R.C., and Kado C.I.** (1982). Genetic map of the crown gall supressive IncW plasmid pSa. *Mol. Gen. Genet.* **186**:10-15.

120. **Thornalley P.J.** (1996). Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification. A role in pathogenesis and antiproliferative chemotherapy. *Gen. Pharmacol.* **27**:565-573.

121. **Tisa L.S. and Rosen B.P.** (1990). Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein. *J. Biol. Chem.* **265** :190-194.

122. **Trinidade M., Abratt V.R., and Reid S.J.** (2003). Induction of sucrose utilisation genes from *Bifidobacterium lactis* by sucrose and raffinose. *Appl. Environ. Microbiol.* **69**:24-32.

123. **Tsai K.-J., Hsu C.-M., and Rosen B.P.** (1997). Efflux mechanisms of resistance to cadmium, arsenic and antimony in prokaryotes and eukaryotes. *Zoological Studies* **31**:1-16.

124. **Tuffin I.M., De Groot P., Deane S.M., and Rawlings D.E.** (2005). An unusual Tn21-like transposon containing an ars operon is present in highly arsenic-resistant strains of the biomining bacterium *Acidithiobacillus caldus*. *Microbiology* **151**:3027-3039.

125. **Tuffin I.M., Hector S.B., Deane S.M., and Rawlings D.E.** (2006). Resistance determinants of a highly arsenic-resistant strain of *Leptospirillum ferriphilum*

isolated from a commercial biooxidation tank. *Appl. Environ. Microbiol.* **72**:2247-2253

126. **Tuffin M., De Groot P., Deane S.M., and Rawlings D.E.** (2004). Multiple sets of arsenic resistance genes are present within highly arsenic-resistant industrial strains of the biomining bacterium, *Acidithiobacillus caldus*. *International Congress Series* **1275** :165-172.

127. **Van Aswegen P.C., Godfrey M.W., Miller D.M., and Haines A.K.** (1991). Developments and innovations in bacterial oxidation of refractory ores. *Miner Metallurg Processing* **8**:188-192.

128. **Vickers T.J., Greig N., and Fairlamb A.H.** (2004). A trypanothione-dependent glyoxalase I with a prokaryotic ancestry in *Leishmania major*. *PNAS* **101**:13186-13191.

129. **Walmsley A.R., Zhou T., Borges-Walmsley M.I., and Rosen B.P.** (1999). The ATPase mechanism of ArsA, the catalytic subunit of the arsenite pump. *J. Biol. Chem.* **274**:16153-16161.

130. **Walmsley A.R., Zhou T., Borges-Walmsley M.I., and Rosen B.P.** (2001). A kinetic model for the action of a resistance efflux pump. *J. Biol. Chem.* **276**:6378-6391.

131. **Wang G., Kennedy S.P., Fasiludeen S., Rensing C., and DasSarma S.** (2004). Arsenic resistance in *Halobacterium* sp. Strain NRC-1 examined by using an improved gene knockout system. *J. Bacteriol.* **186**:3187-3194.

132. **Woese C.R.** (2004). A new biology for a new century. *Microbiol. Mol. Biol. Rev*. **68**:173-186.

133. **Wu J. and Rosen B.P.** (1991). The ArsR protein is a *trans*-acting regulatory protein. *Mol. Microbiol.* **5**:1331-1336.

134. **Wu J. and Rosen B.P.** (1993b). The *arsD* gene encodes a second trans-acting regulatory protein of the plasmid-encoded arsenical resistance operon. *Mol. Microbiol.* **8**:615-623.

135. **Wu J., Tisa L.S., and Rosen B.P.** (1992). Membrane topology of the ArsB protein, the membrane subunit of an anion- translocating ATPase. *J. Biol. Chem.* **267**:12570-12576.

136. **Wu J. and Rosen B.P.** (1993a). Metalloregulated expression of the *ars* operon. *J. Biol. Chem.* **268**:52-58.

137. **Xu C., Shi W., and Rosen B.P.** (1996). The chromosomal *arsR* gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. *J. Biol. Chem.* **271**:2427-2432.

138. **Xu C. and Rosen B.P.** (1997). Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of *Escherichia coli*. *J. Biol. Chem.* **272**:15734-15738.

139. **Yang H.C., Cheng J., Finan T.M., Rosen B.P., and Bhattacharjee H.** (2005). Novel pathway for arsenic detoxification in the legume symbiont *Sinorhizobium meliloti*. *J. Bacteriol.* **187**:6991-6997.

140. **Zabeau M. and Stanley K.K.** (1982). Enhanced expression of the cro-β-galactosidase fusion proteins under the control of the P_R promoter of the bacteriophage lambda. *EMBO J.* **1**:1217-1224.

141. **Zegers I., Martins J.C., Willem R., Wyns L., and Messens J.** (2001). Arsenate reductase from *S. aureus* plasmid pI258 is a phosphatase drafted for redox duty. *Nat. Struct. Biol.* **8**:843-847.

142. **Zhou T., Liu S., and Rosen B.P.** (1995). Interaction of substrate and effector binding sites in the ArsA ATPase. *Biochemistry* **34**:13622-13626.

143. **Zhou T. and Rosen B.P.** (1997). Tryptophan fluorescence reports nucleotide-induced conformational changes in a domain of the ArsA ATPase. *J. Biol. Chem.* **272**:19731-19737.

144. **Zhou T. and Rosen B.P.** (1999). Asp45 is a Mg2+ ligand in the ArsA ATPase. *J. Biol. Chem.* **274**:13854-13858.

145. **Zhou T., Radaev S., Rosen B.P., and Gatti D.L.** (2000). Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *EMBO J.* **19**:4838-4845.

146. **Zhou T., Radaev S., Rosen B.P., and Gatti D.L.** (2001). Conformational changes in four regions of the *Escherichia coli* ArsA ATPase link ATP hydrolysis to ion translocation. *J. Biol. Chem.* **276**:30414-30422.

147. **Zhou T., Shen J., Liu Y., and Rosen B.P.** (2002). Unisite and multisite catalysis in the ArsA ATPase. *J. Biol. Chem.* **277**:23815-23820.