Molecular Genetics of Arsenic Resistance of the Biomining Bacterium *Acidithiobacillus ferrooxidans*.

Bronwyn Gwyneth Butcher

Dissertation presented for the degree of Doctor of Philosophy at the University of Stellenbosch.

Supervisor: Professor Douglas E. Rawlings

December 2003

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously, in its entirety or part, submitted it to any university for a degree.

(Bronwyn G. Butcher)

Abstract

The acidophilic, chemolithoautotrophic bacterium, *Acidothiobacillus ferrooxidans* is one of a consortium of bacteria involved in biomining, including the recovery of gold from arsenopyrite ores. The genes conferring arsenic resistance to *At. ferrooxidans* were cloned and sequenced and shown to be chromosomally located. Homologues to the *arsB* (membrane located arsenite efflux pump), *arsC* (arsenate reductase) and *arsH* (unknown function) genes from known arsenic resistance (*ars*) operons were identified. A fourth gene was found to have weak homology to the ArsR-family of regulators. The arsenic resistance genes of *At. ferrooxidans* are arranged in an unusual manner, with the *arsRC* and *arsBH* genes divergently transcribed. This divergent arrangement was found to be conserved in all four of the *At. ferrooxidans* strains we tested.

All of the *At. ferrooxidans ars* genes were expressed in *Escherichia coli* and the *arsB* and *arsC* genes conferred arsenite (and antimonite) and arsenate resistance, respectively, to an *E. coli ars* mutant (AW3110). Analysis of the putative amino acid sequences of these *ars* genes revealed that the ArsB from *At. ferrooxidans* is closely related to the ArsB proteins from other Gram-negative bacteria. However, the ArsC protein is more closely related to the ArsC proteins from Gram-positive bacteria. Furthermore, a functional thioredoxin (*trxA*) gene was required for ArsC-mediated arsenate resistance in *E. coli*. This suggests that reduction of arsenate by *At. ferrooxidans* has a similar reaction mechanism as that by Gram-positive ArsC proteins. While *arsH* was expressed in an *E. coli*-derived in *vitro* transcription-translation system, the presence of this gene was not required for, nor enhanced, arsenite or arsenate resistance in *E. coli*. We predict that the function provided by this gene is not required in *E. coli*.

While the putative ArsR from *At. ferrooxidans* does contain a potential DNA-binding helix-turn-helix (HTH) domain, it does not contain the arsenite binding motif (ELCVCDL), required for response to the presence of inducer. Instead, the ArsR-like protein from *At. ferrooxidans* is related to a group of unstudied ArsR-like proteins that have been associated with other *ars*-like genes identified during genome sequencing projects. Using *arsB-lacZ*, *arsC-lacZ*, and *arsR-lacZ* fusions, it has been shown that this

atypical ArsR protein from *At. ferrooxidans* did repress expression from the *arsBH* and *arsRC* promoters and that this repression was relieved by the presence of either arsenite or arsenate. Deletion of 19 amino acids from the C-terminus of the ArsR protein did not affect regulation, while deletion of a further 28 amino acids inactivated ArsR. Northern blot hybridization confirmed that expression of the *arsRC* and *arsBH* transcripts is increased in the presence of either arsenite or arsenate.

This study is the first to show that the *ars* genes from the acidophilic bioming bacterium *At. ferrooxidans* are able to be studied in the neutrophilic bacterium, *E. coli*. We have also shown that the atypical ArsR found in this *ars* operon is able to regulate expression of these genes in response to arsenic, despite not containing the arsenite binding domain, suggesting that this protein senses arsenic by a different mechanism to that used by the ArsR family members already studied.

Opsomming

Acidothiobacillus ferrooxidans, 'n asidofiliese, chemolitotrofiese bakterium, is een van 'n konsortium bakterieë betrokke by biologiese ontgunnig ("biomining") asook by die herwinning van goud uit arsenopiriet erts. Die gene wat aan *At. ferrooxidans* weerstandbiedendheid teen arseen verleen, is gekloneer. Die DNA-volgorde van hierdie gene is bepaal en daar is bewys dat die gene op die chromosoom geleë is. Homoloë van die *arsB* (membraan geleë pomp wat arseniet uitpomp), *arsC* (arsenaat reduktase) en die *arsH* (funksie onbekend) gene is in bekende arseenweerstanbiedheidsoperons (*ars*-operons) geïdentifiseer. Verder is daar 'n vierde geen geïdentifiseer wat lae homologie met die ArsR-familie van reguleerders toon. *At. ferrooxidans* se *ars* gene is op 'n ongewone manier gerangskik met twee van die gene, *arsRC* en *arsBH* wat in teenoorgestelde rigtings getranskribeer word. Hierdie rangskikking van gene is waargeneem in al vier die *At. ferrooxidans* rasse wat getoets is.

Al die *At. ferrooxidans ars* gene is in *Escherichia coli* uitgedruk. Die *arsB* en *arsC* gene het aan 'n *E.coli ars* mutant (AW3110) weerstandbiedendheid teen aseniet, antimoniet en arseen verleen. Analiese van die afgeleide aminosuurvolgorde van die *ars* proteïene het getoon dat die *At. ferrooxidans* ArsB naby verwant aan die ArsB-proteïene van ander Gram negatiewe bakterieë is. In teenstelling hiermee, is gevind dat die ArsC-proteïene nader verwant aan die ArsC-proteïene van Gram positiewe bakterieë is. Daar is ook gevind dat 'n funksionele tioredoksien (*trxA*) geen vir ArsC-bemiddelde arsenaat weerstandbiedendheid in *E.coli* benodig word. Dit dui daarop dat die meganisme van arsenaatreduksie deur *At. ferrooxidans* soortgelyk is aan die ArsC-proteïen-meganisme van Gram positiewe bakteriee. *In vitro* studies met behulp van 'n *E. coli* gebaseerde transkripsie-translasie sisteem het getoon dat *arsH* nie nodig is vir arsenaat of aseniet weerstanbiedendheid in sensitiewe *E.coli* rasse nie en ook nie help om weerstand in hierdie rasse te verhoog nie. Daarom kan daar aangeneem word dat die funskie van die *arsH* geen nie deur *E. coli* benodig word nie.

Die vermeende ArsR van *At. ferrooxidans* bevat 'n potensiële DNA-binding heliks-draai-heliks motief, maar nie die arsiniet binding motief (ELCVCDL) wat nodig is vir reaksie in die teenwoordigheid van 'n induseerder nie. Die ArsA-proteïen van *At. ferrooxidans* is soortgelyk aan 'n groep ArsA-proteïene wat tydens genoom DNA-

volgordebepalingsprojekte geïdentifiseer is. Hierdie groep gene is egter nog nie verder bestudeer nie. Deur gebruik te maak van 'n stel fusie gene, *arsB-lacZ*, *arsC-lacZ* en *arsR-lacZ* kon daar bewys word dat die ongewone ArsH-proteïen van *At. ferrooxidans* uitdrukking van *arsBH* en *arsRC* onderdruk en dat die onderdrukking deur arseniet of arsenaat opgehef kan word. Delesie van die eerste 19 aminosure vanaf die C-terminus van die ArsA-proteïen het geen uitwerking op die regulering van die proteïen nie, maar delesie van 'n vedere 28 aminosure het ArsR geïnaktiveer. Verhoogde vlakke van transkripsie van *arsRC* en *arsBH* in die teenwoordigheid van arseniet en arsenaat is met behulp van Noordelike kladanalise bewys.

Hierdie is die eerste studie waarin daar bewys word dat die *ars* gene van die asidofiliese bakterium *At.ferrooxidans* in die neutrofiliese bacterium *E. coli* bestudeer kan word. Daar is ook bewys dat ten spyte daarvan dat die ArsR in die *ars* operon nie 'n arseniet bindingsdomein het nie, dit die uitdrukking van die gene in hierdie operon reguleer in reaksie op arseen. Dit dui dus daarop dat hierdie proteïen op arseen in die omgewing reageer met behulp van 'n meganisme wat verskil van die ArsR-proteïene wat tot dusver bestudeer is.

Acknowledgements

This work was funded by grants from the National Research Foundation (Pretoria, South Africa) and BHP Billiton Minerals Technology (Randburg, South Africa).

I received financial support in the form of Harry Crossley and Stellenbosch 2000 scholarships from the University of Stellenbosch.

I would like to thank Barry Rosen for providing the *E. coli* W3110 and AW3110 strains, Michael S. DuBow for phage $P1_{vir}$ and Aresa Toukdarien for plasmid pGL10.

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

My supervisor, Doug Rawlings, who encouraged me to join this lab in the first place and gave me an exciting project to work on and who always had time to give advice and support even in the final stages of this thesis, when I moved so far away.

Shelly Deane, who started this project off and watched over my shoulder as I learned the ropes in the lab. Thank you for the friendship and encouragement and for always being willing to lend an ear and share ideas. Thank you also for proof-reading the thesis and putting up with the many glitches while sending manuscripts electronically around the world.

Murray Gardner, for your friendship and for providing many a fun and amusing moment in the lab.

The rest of the TRU/BRG lab members and the members of the Microbiology departments of both UCT and Sellenbosch University. I would also like to thank Danie le Grange for translating my abstract into Afrikaans.

Lois Banta (Williams College), who has been extremely supportive and has allowed me time to complete my thesis while working for her.

My family and especially my parents, who have always encouraged me and have supported me in so many ways, and who asked every time they phoned me how the writing of the thesis was going.

Finally, my husband, Dirk, who spent many evenings cooking and cleaning, etc. so that I would have no excuses not to work on the writing after work. Thank you for all the love and encouragement and for being proud of me.

TABLE OF CONTENTS

Chapter One: General introduction

TABLE OF CONTENTS

1.1. Acidithiobacillus ferrooxidans

Acidithiobacillus ferrooxidans was the first bacterium discovered that was capable of oxidizing minerals. This bacterium was previously known as *Thiobacillus ferrooxidans*, but because the genus *Thiobacillus* was found to contain members that fell into the α-, β- and γ-subclasses of *Proteobacteria*, a subclass of this genus was reassigned to a new genus, *Acidithiobacillus*. This genus includes the species *At. ferrooxidans*, *At. thiooxidans*, *At. caldus* and *At. alberticus* (Kelly and Wood, 2000). All four of these species are acidophilic and are members of the γ-subclass of *Proteobacteria*. *At. ferrooxidans* is obligately chemolithoautotrophic and is able to use ferrous iron or reduced inorganic sulfur sources as electron donors. This property allows this bacteria to convert insoluble metal sulphides into soluble metal sulphates, which can then be leached from the surroundings. However, whether this is a direct or indirect mechanism is a subject of debate (Rawlings, 2002). It is important to note that within the species *At. ferrooxidans* there is considerable genetic variation among the strains (Kelly and Wood, 2000), which may in future result in some strains being reassigned to new genera.

The (above) ability of these bacteria to convert insoluble sulphides to soluble sulphates has been made use of to extract various minerals from ore, and is especially useful in the extraction of minerals from recalcitrant or lower grade ores. There are two main types of mining processes which make use of bacteria: irrigation-type (which probably takes place over a range of temperatures reaching as high as 65°C if there is a lot of sulphide present in the heaps) and stirred tank-type processes (commercial processes take place at either 40°C or 50°C). Irrigation processes involve stacking the crushed ore in heaps or dumps and then irrigating these with recycled leach liquor. The bacteria present on the ore then solubilize the mineral allowing it to be recovered from the leach liquor by subsequent extraction processes. These processes are used mainly for the recovery of copper. The second type involves treating the ore in highly aerated stirred tank bioreactors. These are usually arranged in series and the feed (consisting of the mineral concentrate to be treated, inorganic nutrients and water) is added to the first tank and then overflows to the tanks downstream until complete treatment of the ore has occurred. This process has been

used mainly to recover gold, but a process to recover cobalt is in use and a process to recover nickel has been demonstrated at pilot scale. In the gold recovery processes, the ore is solubilized by the bacteria allowing greater access of cyanide to the gold, whereas with cobalt and nickel ores, the metal is solubilized as a consequence of microbially mediated mineral decomposition. High temperature (75°C) stirred-tank processes for recovery of copper from chalcopyrite are being developed.

At. ferrooxidans was originally thought to be the dominant bacterium in the biooxidation processes that took place at 40°C or less, because it was readily isolated from these environments by batch-culture methods. With recent developments in molecular techniques such as polymerase chain reaction (PCR) and fluorescence *in situ* hybridization it has been shown that, in fact, other bacteria such as *At. thiooxidans, At. caldus* and *Leptospirillum ferrooxidans* are more abundant and more important in most of these processes. PCR of 16S and 23S rDNA intergenic regions has yielded no product corresponding to that from *At. ferrooxidans* in copper heap-leaching environments. Instead other species such as *At. thiooxidans* and *Leptospirillum ferrooxidans* dominated the process (Rawlings *et al.*, 1999). However, if ferrous iron was added to the environment, a PCR product corresponding to *At. ferrooxidans* dominated the population. In continuous-flow reactors *L. ferrooxidans* and *At. caldus* were the dominant species, and while *At. ferrooxidans* appears to be present, it is in the minority (Rawlings *et al.*, 1999).

It is believed that *L. ferrooxidans* has a higher affinity for ferrous iron and is inhibited less by ferric iron than *At. ferrooxidans*, explaining why *L. ferrooxidans* dominates these processes (Rawlings *et al.*, 1999). This may also explain why adding more ferrous iron to the copper-leaching heaps allowed *At. ferrooxidans* to dominate. For this reason *At. ferrooxidans* may still dominate in dump or heap-leaching environments if the ferrous iron concentration in solution is high (Rawlings, 2002).

Of interest to us was an application of the above stirred-tank process for extracting gold from recalcitrant and lower-grade arsenopyrite ores (Figure 1.1). In the case of the recalcitrant ores the gold is encased in a matrix of pyrite and arsenopyrite and is

unable to be solubilized as usual by cyanide. Using vat or tank leaching, bacteria are able to "decompose" the ore as described by the following equation:

$$\underset{\text{(arsenopyrite)}}{2FeAsS} + 7O_2 + 2H_2O \rightarrow 2FeAsO_4 + 2H_2SO_4$$

This gold is now accessible to cyanide allowing recovery of the previously unattainable gold from the ore by conventional methods.

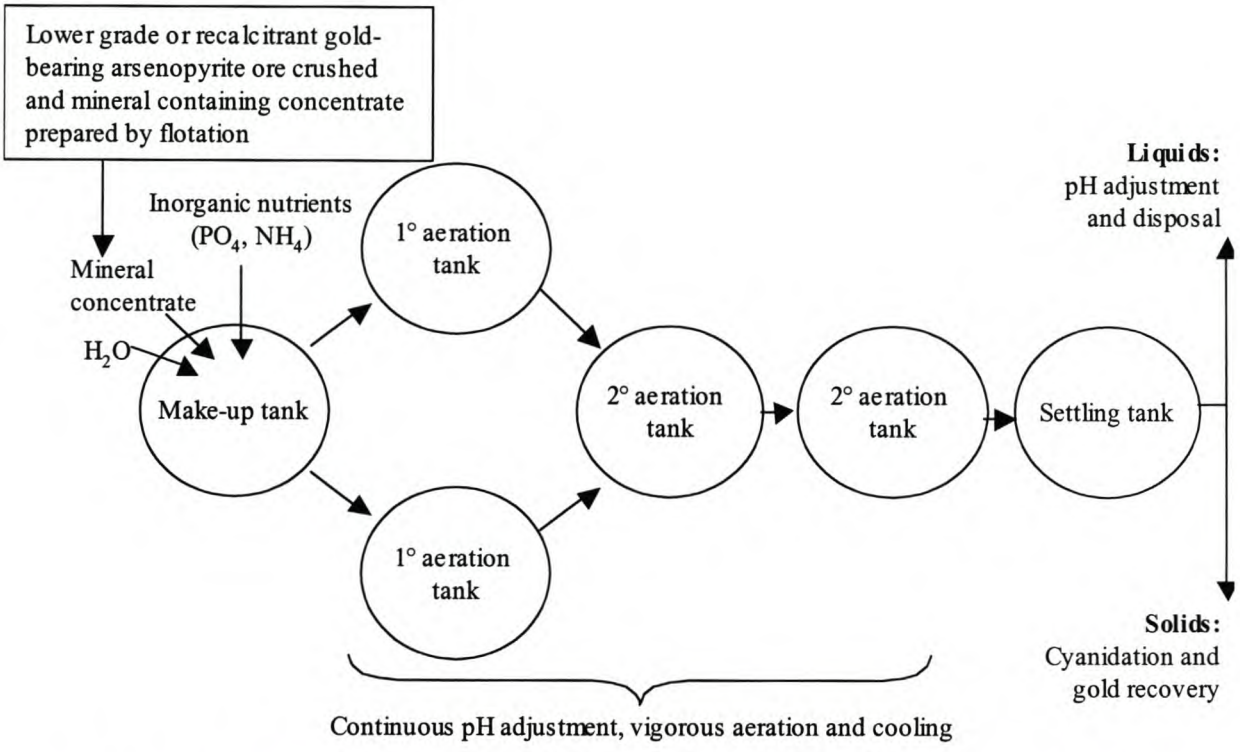


Figure 1.1: Flow diagram showing a typical commercial process for the extraction of gold from arsenopyrite ore. Due to the low concentrations of gold in the ore, the ore is first crushed to a fine powder and a concentrate of this prepared. In the make-up tank nutrients and water are added to this concentrate. These nutrients are necessary for the efficient growth of the bacteria. By continuous-flow mode the feed overflows from one tank to another until sufficient decomposition of the ore has occurred. The parallel primary aeration tanks are necessary to increase retention time thereby allowing sufficient bacterial growth and preventing wash-out. The tanks are aerated allowing for bacterial growth and agitated to keep the solids in suspension. They are also cooled, because the biooxidation process is exothermic. Finally, the solids which still contain the gold are recovered in a settling tank and sent to the cyanide gold recovery process. (Adapted from Rawlings, 2002)

However, during this process arsenic compounds are leached from the arsenopyrite ore and these are toxic to the bacteria. When this process was developed at the Fairview mine at Barberton, South Africa, using the naturally occurring bacterial population, it was found that the process required long retention times to oxidize the

ore sufficiently. One reason for this was that these bacteria were inhibited by these toxic arsenic compounds. It was found that the bacteria were sensitive to less that 1 g/l of total soluble arsenic. After selection using a bacterial chemostat, bacterial populations that were resistant to 13 g/l soluble arsenic were isolated and these reduced the retention time in the tanks from 12 to 7 days (Rawlings and Woods, 1995). It was obvious that through mutation and selection the bacterial population had developed an increased level of arsenic resistance. An interest in the nature of this resistance led to the work discussed in this dissertation.

In a preliminary study carried out in 1996, *At. ferrooxidans* ATCC33020 was found to contain chromosomally encoded genes homologous to arsenic resistance genes from other known arsenic resistance systems. As mentioned earlier, it has been discovered that while *At. ferrooxidans* is present in the biooxidation tanks, it is not the dominant bacterium. However, when this work was started the molecular tools for the study of the other biomining bacteria were not available. Much work had been done on the genetics of *At. ferrooxidans* and our research group had already constructed plasmid and cosmid libraries of the genome of this bacterium. It had also been shown that many other genes from *At. ferrooxidans* were expressed and functional in *E. coli* despite *At. ferrooxidans'* unusual physiology (e.g. genes for glutamine synthetase (Barros *et al.*, 1985) and pyruvate dehydrogenase (Powles and Rawlings, 1997)). This made *At. ferrooxidans* a good subject for an investigation into arsenic resistance in these acidophilic bacteria. This was also not a strain of *At. ferrooxidans* isolated from a mining or an arsenic-rich environment, yet it still contained genes encoding a membrane located arsenite efflux pump (*arsB*), an arsenate reductase (*arsC*) and a third gene of unknown function (*arsH*). These genes were also arranged in an unusual manner with the *arsB* and *arsC* genes divergently transcribed. These findings are described in this dissertation.

1.2. Properties of arsenic

Arsenic is the 33rd element in the periodic table and is often found in sulphide ores (Shriver *et al.*, 1994). It exists in the 3+ (arsenite) and 5+ (arsenate) oxidation states in water and more readily forms oxides with the 3+ oxidation state. Arsenic is present naturally in the soil, atmospheric dust and water, but usually at very low

concentrations. It was discovered in 1250 by Albertus Magnus and has subsequently been used for many purposes, primarily making use of its toxic properties (eg. poisons, herbicides and pesticides and antimicrobial agents).

Arsenite (As(III)) is believed to be at least 100 times more toxic than arsenate (As(V)). Nakamuro and Sayato (1981) found that in cultured leukocytes and human skin fibroblasts, the chromosome-breaking activity of trivalent arsenic was significantly higher than that of pentavalent arsenic. The general toxicity of arsenite is due to the fact that arsenite acts as a metal and forms strong metal-thiol bonds with vicinal cysteines, thereby inhibiting enzymes such as pyruvate dehydrogenase. Arsenate oxyanions have the same solubilities as phosphates and, therefore, form toxic analogs for inorganic phosphorylating activities. For example, it has been shown that in human red blood cells arsenate can substitute for phosphate in both the sodium pump and the anion exchanger (Kenney and Kaplan, 1988). In bacteria, arsenate is taken into the cells via the Pit and Pst phosphate transport systems. Mutants defective in the Pit pathway are usually arsenate resistant (Cervantes *et al.*, 1994). The mechanisms by which arsenite enters the cells are still unknown. It has been shown that inactivating the glycerol facilitator, GlpF, by insertion of Tn*phoA* confers antimonite resistance on an antimonite sensitive strain (Sanders *et al.*, 1997). However, this insertion did not confer resistance to arsenite.

Due to the fact that arsenic is present in the environment and that it is continually being added to the environment in the form of herbicides and pesticides and during industrial processes and metal smelting, it is not surprising that microorganisms have developed various resistance mechanisms.

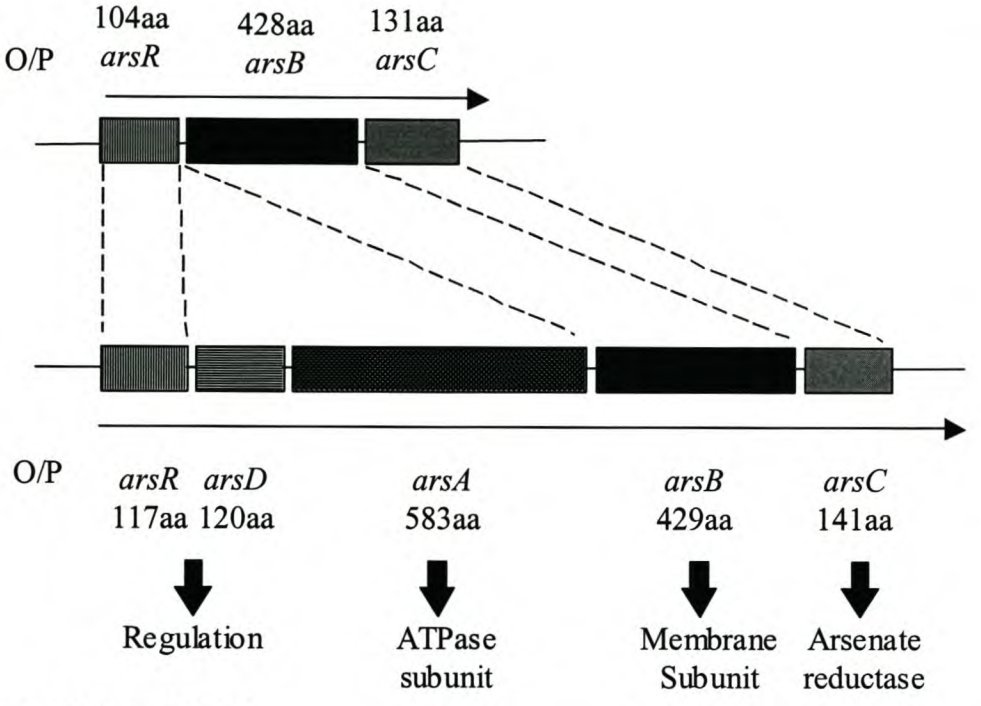
Another element often found in combination with sulfide, and with similar properties to arsenic, is antimony. It is in the same group on the periodic table as arsenic and shares many of the same chemical properties. Many arsenic resistance systems also confer resistance to antimonite.

1.3.Arsenic resistance in bacteria

Microorganisms have been reported to employ a variety of resistance mechanisms, such as oxidation of the more toxic arsenite to arsenate (e.g. *Alcaligenes faecalis*), resistance to arsenate by having phosphate pathways that do not transport arsenate as efficiently (e.g. cyanobacteria), overexpression of intracellular thiols (e.g the protozoan, *Leishmania*) and sequestration of the toxic ion in the vacuole (e.g. fungi) (Cervantes *et al.*, 1994 and Rosen, 1999). However, the best characterized mechanism in bacteria, and the focus of this discussion, is an efflux system where the toxic arsenic ions are pumped from the cell, preventing them from inhibiting enzymes in the cytosol. This system is encoded by the *ars* (arsenic resistance) operon and has been extensively reviewed (Cervantes *et al.*, 1994; Xu *et al.*, 1998; Rosen, 1999, Rensing *et al.*, 1999 and Mukhopadhyay *et al.*, 2002)

There are two common forms of the *ars* operon. One consists of 5 genes (*arsRDABC*) and has to date only been found on plasmids in Gram-negative bacteria. The second, more common form, consists of only 3 of the above genes (*arsRBC*) (Figure 1.2). Both types of operons are regulated by the *arsR* gene, which encodes an arsenite responsive transcriptional repressor. The *arsD* gene, found only in the 5-gene operon, is also a trans-acting repressor, but appears to be inducer-independent and is thought to control the upper level of expression of this operon. The *arsB* gene encodes the membrane spanning arsenite efflux pump. ArsA is an arsenite-stimulated ATPase that interacts with the ArsB membrane protein to form a more efficient ATP-driven pump. The final gene, *arsC*, encodes an arsenate reductase. This protein reduces arsenate to arsenite, which can then be pumped from the cell via the ArsAB or ArsB pumps. These systems have been extensively studied in *E. coli* and *Staphylococcus* species and the individual proteins will be discussed in detail later.

Staphylococcus aureus pI258, *S. xylosus* pSX267 and *Escherichia coli* chromosome



Escherichia coli R773

Figure 1.2: The organisation and structure of the arsenic resistance (*ars*) operons of *E. coli* and *Staphylococcus* species. (adapted from Silver *et al*, 1993). O/P, operator/promoter

Due to the increase in genome sequencing projects and the volume of sequence information being generated, a database search (e.g. at www.ncbi.nih.nlm.gov) will yield many potential *ars* gene homologues, indicating that this method of arsenic resistance appears to be widespread among bacteria. Unfortunately most of these genes or operons have not been characterised. Those operons that have been cloned and shown to confer resistance to arsenic are listed in Table 1.1. Other studies where researchers have used probes of known arsenic resistance genes (e.g. from *E. coli*) in Southern hybridisation experiments, or PCR to look for the presence of the genes in other bacteria, have shown that many different bacteria from many different locations contain homologues to these arsenic resistance genes. For example, Saltikov and Olson (2002) identified arsenic resistance genes in the bacterial population isolated from raw sewage and arsenic-enriched creek waters using colony hybridisation and PCR, and Dopson *et al.* (2001) identified homologues to the *E. coli arsB* gene in *At. caldus* and other gram-negative acidophiles using Southern hybridisation.

Table 1.1: Bacterial arsenic resistance operons cloned and characterised to date.

Bacterium	**Location**	**Arrangement**	**Reference#**
Staphylococcus aureus	plasmid, pI258	*arsRBC*	Ji and Silver,1992a
Staphyloccocus xylosus	plasmid, pSX267	*arsRBC*	Rosenstein *et al.*, 1992
Escherichia coli	plasmid, R773	*arsRDABC*	Chen *et al.*, 1985
Escherichia coli	plasmid, R46	*arsRDABC*	Bruhn *et al.*, 1996
Escherichia coli	chromosome	*arsRBC*	Diorio *et al.*, 1995; Carlin *et al*, 1995
Yersiniae enterocolitica	plasmid, pYV	*arsRBC* and divergent *arsH*	Neyt *et al.*, 1997
Acidiphilium multivorum	plasmid, pKW301	*arsRDABC*	Suzuki *et al.*, 1998
Bacillus subtilis	skin element	*arsR ORF2 arsBC*	Sato and Kobayashi *et al.*, 1998
Pseudomonas aeruginosa	chromosome	*arsRBC*	Cai *et al.*, 1998
Prototype IncH12 plasmid first isolated from *Serratia marcesens*	plasmid, R478	*arsRBC* and divergent *arsH*	Ryan and Colleran, 2002

Reference where the genes were cloned and sequenced

1.4.The proteins of the *ars* operon in detail

1.4.1. The efflux pump

The most extensively studied efflux pump is that encoded by the *arsAB* genes of the operon found on plasmid R773. In this system the ArsB protein is the membrane spanning subunit and the carrier for the export of the arsenite anion, while the ArsA protein is an ATPase providing the energy for the efflux of the arsenite (Figure 1.3).

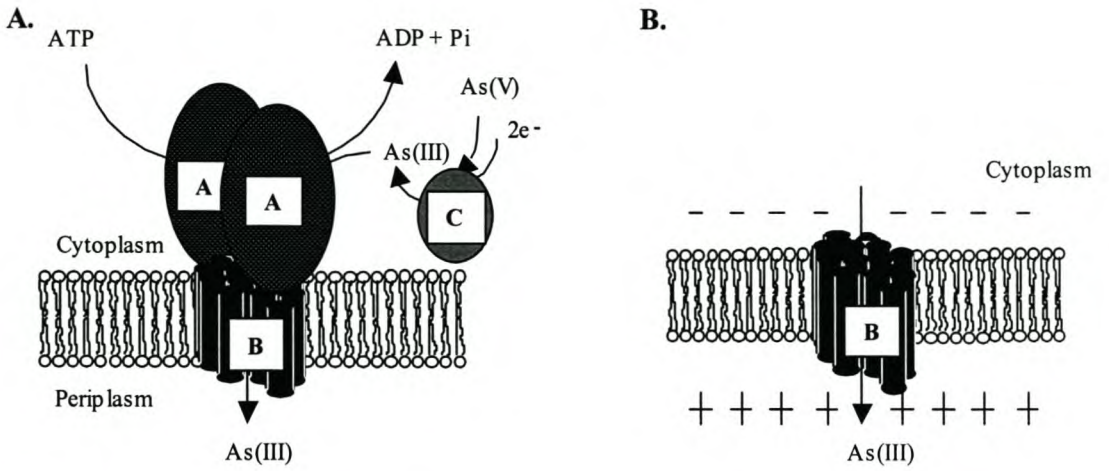


Figure 1.3: A. The function of the proteins encoded by the *ars* operon of *E. coli* plasmid R773 (see text for details). **B**. In the absence of the ArsA ATPase subunit (e.g. in the case of the chromosomal *E. coli* or the *Stapphylococcal ars* operons) the efflux of arsenite from the cell through ArsB is coupled to the proton motive force. (adapted from Dey and Rosen, 1995)

Rosen and colleagues have studied the ArsA subunit of this pump for many years. The 63 kDa protein is a peripheral membrane protein when bound to ArsB, but is soluble when expressed in the absence of *arsB*. This agreed with the amino acid composition and the hydropathy profile of the protein, which showed that the protein should be soluble (Chen *et al.*, 1986). The protein consists of two related domains A1 (N-terminal half) and A2 (C-terminal half), which share 23 % identity with each other (Figure 1.4). These halves also showed homology with several ATP-binding proteins. Each domain contains a consensus nucleotide binding site (NBS) and it was shown that both NBSs were required for enzymatic activity. Li *et al.* (1996) showed, using second site suppressor analysis, that the mutation of a glycine residue in the glycine–rich NBS of A1, which caused a partial loss in arsenic resistance, could be rescued by a second mutation in A2. This indicated that the two domains did in fact interact to form an enzymatically active protein.

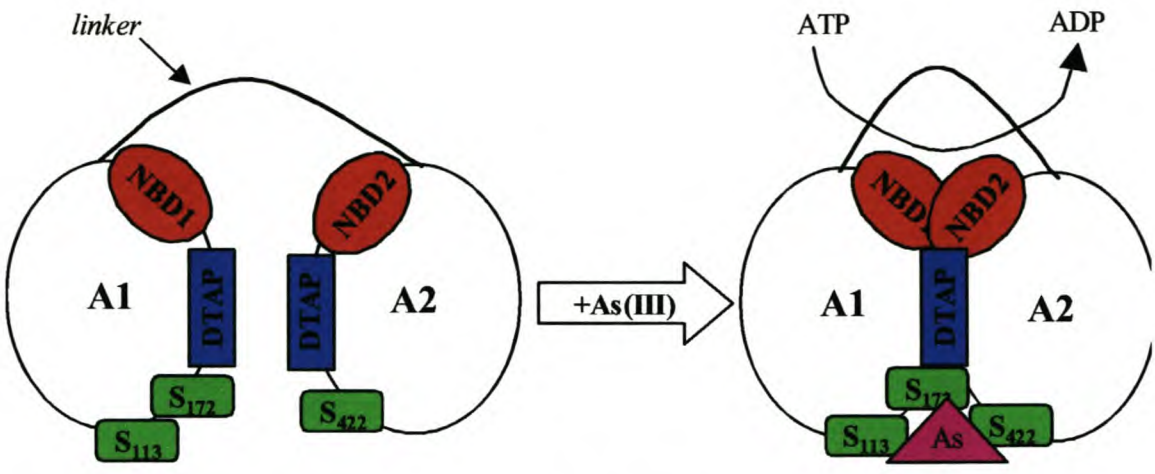


Figure 1.4: Model of the structure of ArsA from *E. coli* plasmid R773. The protein consists of two homologous domains (A1 and A2) separated by a flexible linker region. Each domain contains a nucleotide binding domain (NBD) and a signal transduction domain (DTAP). The cysteine residues shown have been found to comprise the allosteric arsenic binding domain. When arsenite is present it binds to these cysteine residues bringing the A1 and A2 domains together, thereby allowing hydrolysis of ATP. (adapted from Li and Rosen, 2000).

These two domains are separated by a flexible linker region. Li and Rosen (2000) investigated the function of this linker, by lengthening it by 5 glycine residues or shortening it by the deletion of 5, 10, or 15 residues. They found that the addition of residues to the linker did not affect the level of resistance, while cells expressing the *arsA* genes with deleted linkers had increasing levels of arsenite sensitivity as well as a decrease in affinity for ATP and antimonite. Together with the observation that the sequence of the linker is not conserved among the ArsA homologues, these results led the researchers to propose that the length of the linker had evolved to the shortest length that allowed the two halves of the protein to interact efficiently and that it was the length and not the sequence of the linker that was important. However, Jia and Kaur (2001) performed the following complementation experiments and mutational analysis and believe that in fact the residues in the linker do appear to be important for the correct conformation of the nucleotide binding domains and for catalytic activity of the protein. They started by expressing the N-terminal A1 domain without a linker on one clone and the C-terminal A2 domain with or without a linker on another. In other words the two domains were not covalently linked. Since each domain was expressed as a separate polypeptide, changes to the linker should affect the activity of ArsA only if the linker is involved in the function of the protein. They found that the

C-terminal half could only complement the N-terminal clone when the linker was present. It was also shown that mutations of certain residues in the linker region resulted in a loss of arsenite resistance and antimonite-stimulated ATPase activity. These proteins also appear to have conformational changes as shown by their trypsin cleavage patterns. The authors therefore propose that the residues in the linker region of ArsA do play an important role in the function of the protein and that there is interaction between the linker region and the nucleotide-binding domains.

An interesting observation was that while the ArsA protein has vicinal cysteine residues, it is not inhibited by the presence of arsenic as other enzymes are, instead it is allosterically activated by the presence of arsenite or antimonite. Mutagenesis of the cysteine residues showed that three of these residues are involved in the allosteric activation of the ATPase activity (Bhattacharjee *et al.*, 1995). In the primary sequence of the protein these residues are far apart, but it was proposed that in the tertiary structure these residues must be close enough to interact with the metal. Using dibromobimane (dBBr), which has two bromomethyl groups that can cross-link a thiol pair located within 3-6Å of each other forming a fluorescent adduct, it was shown that these three cysteine residues were close enough together in the tertiary structure to interact with arsenite or antimonite (Bhattacharjee and Rosen, 1996). It was hypothesised that binding of arsenite or antimonite pulls the domains of the ArsA closer together thereby triggering the ATP hydrolysis.

Another conserved motif (DTAPTGHTIRLL) was identified between the ArsA homologs and named the DTAP motif (it is located in each of the A1 and A2 domains of ArsA). ArsA proteins containing only single tryptophan residues on either side of the DTAP motif were engineered. Then using intrinsic tryptophan fluorescence (where the fluorescence of the tryptophan residue changes depending on the environment in which it is found) it was shown that during ATP hydrolysis the C-terminal end of this motif moves to a less polar region, while the N-terminal end moves to a more hydrophilic region (Zhou and Rosen, 1997). In other words this DTAP motif appears to experience a "flipping" movement during catalysis. Modelling of ArsA based on its distant homology to NifH (iron protein subunit of the nitrogenase complex) led Xu *et al.* (1998) to propose that the DTAP domain may be a

transduction domain involved in the transmission of energy from ATP hydrolysis to the other functions of the pump.

Recently the crystal structure of ArsA has been determined. Crystals of ArsA bound to MgADP confirmed many of the above observations and highlighted interesting new discoveries (Zhou *et al.*, 2001). Firstly, they showed that the two NBS are situated at the interface between the A1 and A2 domains, and were formed from residues from both halves of the proteins. However, the surface of one NBS is formed mostly of residues from one domain (eg. A1) and there is a much smaller contribution from residues from the other domain (eg. A2). This is, therefore, still named the A1 NBS. They also showed that the metal binding site is situated at the opposite end of the protein with respect to the NBS. This site, also at the interface of A1 and A2 was shown to bind three arsenite or antimonite ions through interactions with three cysteines, two histidines and a serine residue. Each ion was bound by one residue from the A1 domain and one from the A2 domain, confirming previous suggestions that binding of the metal ion brought the domains together. A signal transduction pathway was also proposed where the aspartic residues from the DTAP motifs are in close proximity to the phosphate chain of the adenine nucleotide, while the histidine residues from the other end of the DTAP motif are involved in the interactions with arsenite or antimonite.

Zhou *et al.* (2000) proposed that, based on the structure of the ArsA protein, it may be possible that the metal ions might be injected into ArsB as a result of conformational changes during ATP hydrolysis. This was contrary to previous ideas that the ions that allosterically activate ArsA are not transported by ArsB. This was investigated in further work, where ArsA was crystallized in complex with ATP, the non-hydrolysable ATP analogue AMP-PNP and the transition state analogue of ATP hydrolysis, ADP-AlF_3.

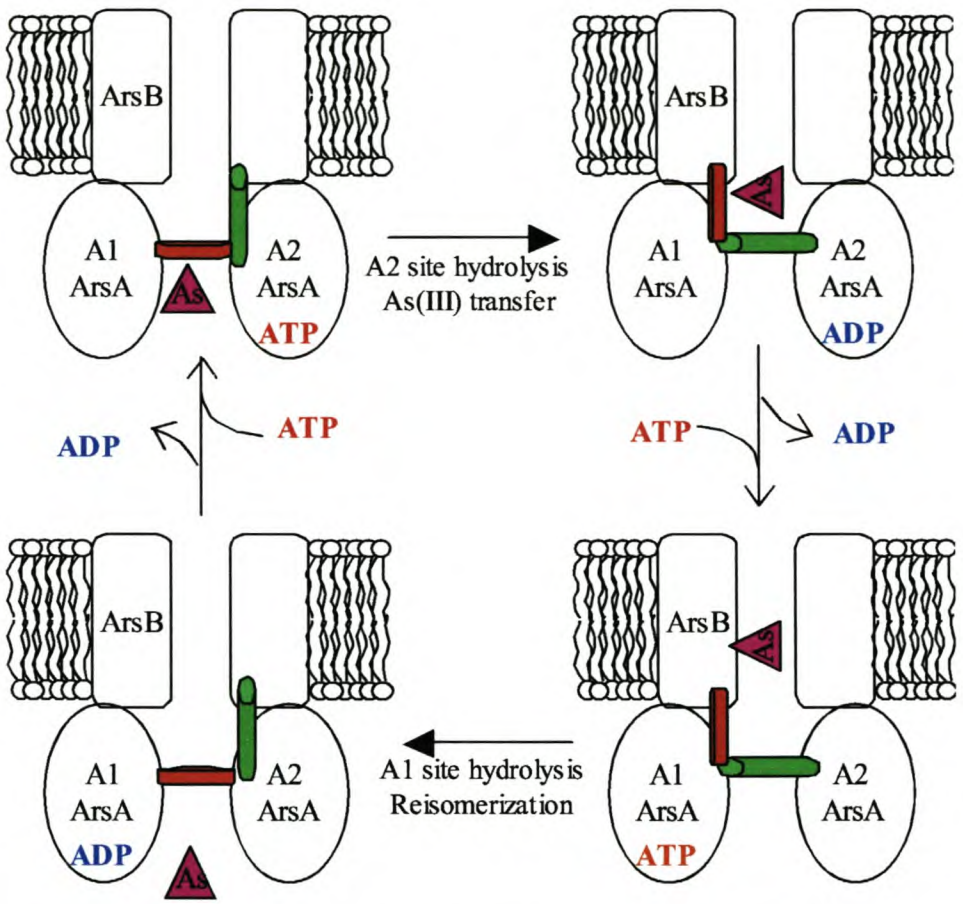


Figure 1.5: Schematic model of the proposed ArsA catalytic cycle based on the crystal structure of ArsA (taken from Zhou *et al.*, 2001). Helices from A1 (red) and A2 (green) are proposed to alternate between the "open" and "closed" positions. An arsenite ion is depicted as a triangle. See text for details.

The authors found that when the crystals were formed, either in the presence of or when incubated with ATP, ATP was found at the A2 NBD, while ADP was found at the A1 NBD (Zhou *et al.*, 2001). They also found that when crystals were incubated with the non-hydrolysable analogue, AMP-PNP was found only at the A2 NBD. It was not possible to form crystals in the presence of AMP-PNP, suggesting that crystals can only be formed if the A1 NBD contains ADP. It is speculated that this is due to the fact that there are conformational changes required for the binding and hydolysis of ATP at the A1 NBD that are not allowed in the crystal. Using steady-state kinetics it had already been shown that the release of ADP from A2 NBS is associated with the release of the As(III) or Sb(III) ion and that binding of ATP favors the uptake of these ions. The authors had also observed changes in the positions of

helices in the A1 and A2 domains that corresponded to hydrolysis of ATP. All of the above led the authors to propose the model shown in Figure 1.5. Here helices (H9-H10) from A1 and A2 alternate positions at the interface with ArsB. When As(III) first interacts with ArsA the helices from A1 form the ceiling of the cavity where the ion binds. During hydrolysis of ATP at the A2 site there is a conformational change where arsenite moves to the protected pocket at the interface with ArsB. Release of ADP from A2 NBD would then result in release of the arsenite ion inside this pocket. The conformation of ArsA would then be returned to ground state by hydrolysis of a second ATP at A1 NBD.

This model is contrary to all previous ideas that the arsenite ion activating ArsA is not transported and that transport through the pump is of arsenite oxyanions. Many of the elements of this model remain to be proven. The position of the H9-H10 helices in the A2 domain have always been disordered in crystals obtained to date and the determination of how many ATP molecules are hydrolyzed per cycle of ion translocation has not yet been possible.

The *arsB* gene of the R773 *ars* operon encodes a hydrophobic protein of about 46 kDa. Hydopathy plots of the predicted amino acid sequence showed that there were at least 10 regions of 19 or more residues with a high hydropathy value. These were considered to be potential membrane-spanning α-helices (Chen *et al.*, 1986). Researchers had been unable to visualize the ArsB protein and, therefore, constructed an ArsB-β-galactosidase hybrid protein, that was found to still give partial resistance to arsenite. When cell extracts from cells expressing this fusion were analyzed using antibodies to the β-galactosidase protein or by measuring β-galactosidase activity of the different extracts, it appeared that the fusion protein was localized to the inner membrane (San Francisco *et al.*, 1989). Membrane proteins are often difficult to identify due to low levels of expression and San Francisco and colleagues propose that in the case of ArsB the level of expression is controlled at a translational level. They identified two regions of potential mRNA secondary structure, one immediately upstream of the predicted ribosome binding site and one at the third codon. These may interfere with the ribosome during translation causing it to pause, thereby limiting the amounts of ArsB produced (San Francisco *et al.*, 1989). When the R773

ars genes were expressed under the control of the T7 RNA polymerase promoter the authors were able to identify the ArsB protein. A protein with an apparent molecular weight of 36 kDa on SDS-PAGE gels was identified as ArsB and this was found only in the membrane fraction (San Fransisco *et al.*, 1989). As mentioned the predicted size of the ArsB protein is about 46 kDa. This difference could be due to the binding of increased amounts of SDS by the basic protein, causing it the migrate faster through the gel. The same was later observed for the ArsB protein of other bacterial *ars* operons (Ji and Silver, 1992 and Rosenstein *et al.*, 1992).

As mentioned the structure of ArsB had only been predicted from hydropathy plots. However, using a method in which regions of the *arsB* gene were fused with a selection of reporter genes, a more accurate prediction of the structure of the ArsB across the membrane could be determined. The *pho*A gene, encoding for alkaline phosphatase, shows increased activity when in a periplasmic position. The *lacZ* gene, encoding β-galactosidase, shows increased activity when in a cytosolic position. Lastly, the *blaM* gene (β-lactamase) gives resistance to high concentrations of ampicillin when transported into the periplasmic space. Based on results from fusions of the above genes with the *arsB* gene a model consisting of 12 membrane-spanning α-helices joined by 5 cytoplasmic loops (with a net positive charge) and 6 periplasmic loops (with a net negative charge) was proposed (Wu *et al.*, 1992).

The ArsAB complex is similar in both structure and function to the ABC (ATP-binding cassette) transporters such as the human multidrug resistance P-glycoprotein. ArsA and ArsB together form an obligatory ATP-driven pump. The surprising observation is that there are many forms of the *ars* operon where there is no *arsA* homolog. This led researchers to ask whether in the absence of ArsA, energy was supplied by another ATPase found elsewhere in the genome, or whether another form of energy was driving this pump. They noticed that the ArsB protein has the same membrane topology as secondary carrier proteins (12 membrane-spanning domains) suggesting that it may act as a secondary arsenite transporter when no ArsA is present. They also investigated whether the R773 ArsB was able to utilize at least two forms of energy. A first experiment was carried out in an *unc E. coli* strain. This strain lacks the H^+-translocating ATPase and can, therefore, not interconvert ATP and the

electrochemical gradient. An *arsA* deletion of the R773 *ars* operon was created (pBC101) as well as a clone containing only the *arsA* gene (pArsA) and these were transformed into the *unc* deletion strain. These cells were grown with glucose (generates ATP through substrate-level phosphorylation) or succinate (little generation of ATP) as the sole energy source and the level of arsenite uptake was measured as an indication of the activity of the membrane efflux pump (Dey and Rosen, 1995). It was found that cells expressing both the *arsA* (from pArsA) and *arsB* (from pBC101) only actively excluded arsenite when the cells were grown in glucose. This exclusion was inhibited by the addition of fluoride (glycolysis inhibitor, which prevents ATP synthesis) and not cyanide (prevents respiration without affecting ATP levels). This indicated that the exclusion of arsenite from cells expressing both *arsA* and *arsB* was dependent on ATP levels. However, cells expressing only the *arsB* gene (from pBC101) were able to exclude arsenite when grown in the presence of either glucose or succinate and this exclusion was inhibited by cyanide, showing that the presence of ATP was not sufficient for arsenite exclusion. When an uncoupler, CCCP (carbonyl cyanide *m*-chlorophenylhydrozone), which destroys the pH and ion gradients of the cell without directly affecting ATP levels, was added the cells immediately accumulated arsenite. This indicated that efflux of arsenite from the cell in the presence of *arsB* only was dependent on the proton motive force. This was confirmed by the observation that uptake of labeled arsenite by everted membrane vesicles expressing only *arsB* was coupled to the electrochemical gradient (Kuroda *et al.*, 1997) (Figure 1.3).

These results led researchers to ask why cells have ArsA, if ArsB is sufficient for arsenic resistance? Rosen and colleagues proposed the following hypothesis (Dey and Rosen, 1995 and Rosen, 1999): The addition of ArsA confers more effective resistance to the cells as ATP levels drop more slowly under conditions of stress than membrane potential. The cells would, therefore, be able to respond to the presence of arsenite better when the ATPase driven pump was present.

A question remaining is whether the arsenite pumped through ArsB is transported through soft-metal interactions with thiol groups or whether it is transported as an oxyanion? There is only one cysteine residue in ArsB and it is predicted to be located in the 11th membrane-spanning region. When this single cysteine residue in ArsB was

mutated to either a serine or alanine residue there was no change in the levels of arsenite or antimonite resistance or the rate of arsenite exclusion from the cells (Chen *et al.*, 1996). This indicated that, unlike the activation of ArsA (discussed above), or the sensing of arsenite by ArsR (discussed later), which both make use of strong metal-thiol bonds, the ArsB protein appears to make use of much weaker nonmetal interactions to transport arsenite anions. These weaker interactions would allow for the release of the anion outside of the cell (Rosen, 1999).

1.4.2. Arsenate reductase

It has been proposed that resistances to metals evolved earlier than antibiotic resistances (Rensing *et al.*, 1999). In the early primordial anaerobic conditions arsenite would have been the major form of arsenic in the environment. This may be the reason that resistance mechanisms are based on systems that transport this trivalent form of arsenic. However, as the environment became more oxidizing, arsenate (the pentavalent form of arsenic) would become more prevalent, requiring the addition of arsenate resistance to the picture. This is the hypothesis put forward by Rosen (1999) to explain the presence of ArsC, an arsenate reductase, that reduces the less toxic arsenate into the more toxic arsenite so that it can be transported out of the cell (Figure 1.3A).

The 16 kDa ArsC protein from R773 was located in the cytosolic fraction of cells, but showed no homology to other proteins in the database when it was originally identified (Chen *et al.*, 1986). It was known to be required for resistance to arsenate, but not arsenite so it was thought to modify the efflux pump in some way allowing recognition of arsenate by the pump. However, in 1992 Ji and Silver showed that in fact the ArsC from the *Staphyloccocus aureus* plasmid pI258 was an arsenate reductase. They showed that purified ArsC protein coupled *in vitro* with thioredoxin and dihiothreitol, but not β-mercaptoethanol or glutathione, was able to reduce arsenate to arsenite.

Although the ArsC protein from *S. aureus* shows less than 20% homology to the ArsC protein from *E. coli* plasmid R773, the R773 ArsC was also shown to be an arsenate reductase (Oden *et al.*, 1994). However, *E. coli* strains containing mutations in two

genes required for glutathione synthesis, *gshA* (γ-glutamylcysteinyl synthetase) and *gshB* (glutathione synthetase) containing the R773 *ars* operon showed wild-type levels of resistance to arsenite, but reduced levels of arsenate resistance. When strains defective in the gene for glutathione reductase (*gor*) were examined these strains were also more sensitive to arsenate than wild-type cells. The addition of exogenous glutathione to the *gshA* mutant strains restored their resistance to arsenate. Strains defective in the genes for thioredoxin reductase (*trxB*) and thioredoxin (*trxA*) showed no difference in either arsenite or arsenate resistance (Oden *et al.*, 1994). These results indicated that unlike the ArsC of *S. aureus*, reduction of arsenite by the ArsC of R773 requires reduction of glutathione and not thioredoxin. Shi *et al.* (1999) showed that of the three glutaredoxin genes found in *E. coli*, *grx2* is the most effective hydrogen donor for arsenate reduction.

When comparing all the ArsC proteins from known *ars* operons it appeared that the ArsC proteins from Gram-negative bacteria grouped together and those from Gram-positive bacteria grouped together (see Chapter 2, Figure 2.6). However, both families of ArsC proteins (along with the non-homologous arsenate reductase from *Saccharomyces cerevisiae*, Acr2p) did appear to share a conserved Cys-X_n-Arg sequence, which is also the conserved catalytic motif (the anion-binding P-loop) found in phosphotyrosine phosphatases (PTPases) (Zegers *et al.*, 2001).

Recently, the crystal structures of representatives of both the gram-negative (*E. coli* R773) and the gram-positive (*Baciliius subtilis* and *S. aureus* pI258) families of ArsC proteins have been determined.

The crystal structure of R773 ArsC on its own and complexed with arsenate and arsenite (Martin *et al.*, 2001) has been determined. Sulphate and sulphite bind as analogues of arsenate and arsenite at the active site of the native protein. The overall structure of the protein consists of α-helices and β-sheets. A search of databases did not show any overall similarity to any known protein, however it does contain a lower βαβ substructure that has some homology to crambin and glutaredoxin. The structure did not confirm previous studies that suggested that the active site may resemble that of low molecular weight tyrosine phosphatases (LMW PTPase) as their topology

contains a four-stranded parallel β-sheet, while the R773 ArsC consists of a mixed β sheet. The catalytic activity of R773 ArsC was previously shown to require the presence of a single cysteine residue (Cys-12). However the pK_a of cysteine residues was known to be higher than that of the optimal pH of the ArsC catalysed reaction. It was thought that a histidine residue (His-8) was responsible for ion-pairing with Cys-12 thereby lowering the pK_a of this residue, because cells bearing each of four different mutations to the His-8 residue were all arsenate sensitive (Gladysheva *et al.*, 1996). However, this model is not supported by the crystal structure. Cys-12 appears to be activated by hydrogen bonds from Arg-94 and Arg-197, while His-8 is more than 7Å away and stabilizes the active loop by forming a side chain hydrogen bond with Ser-15. Based on the crystal structure and previous research, Martin and colleagues proposed the following reaction mechanism: (Figure 1.6).

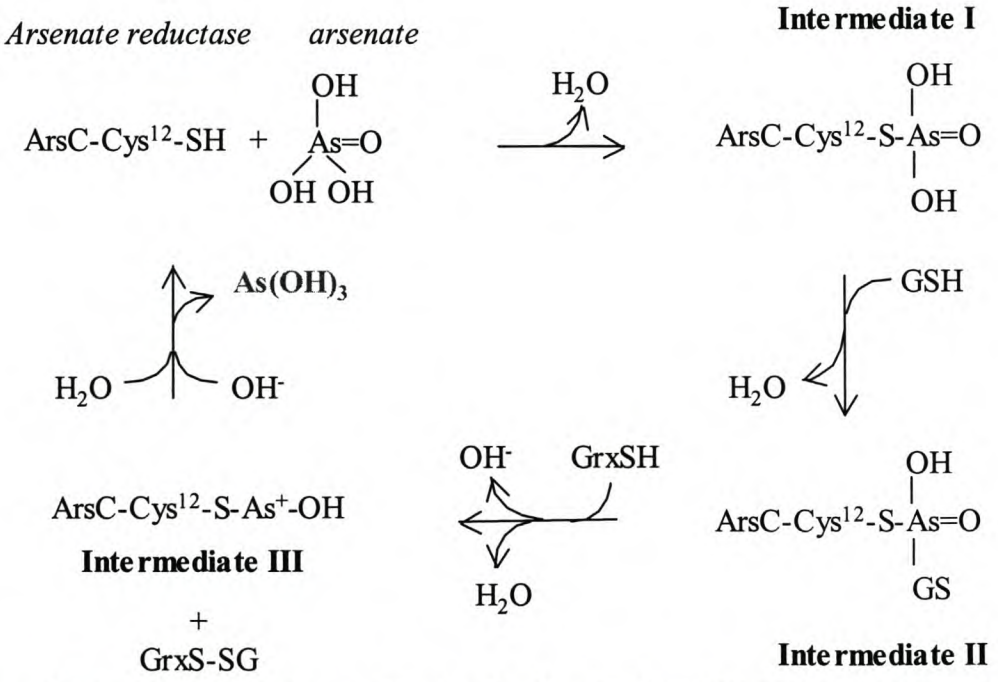


Figure 1.6: Proposed reaction mechanism for *E. coli* plasmid R773 ArsC (taken from Martin *et al*, 2001). See text for details.

Step 1: <u>The formation of a thioarsenate binary adduct (intermediate I):</u> Intermediate I was demonstrated by a difference in electron density. The native structure observed makes use of the Cys-12 thiol group and the three arginine residues to trap the arsenate. Once bound the overall surface charge changes from

positive to negative, facilitating the binding of the other cofactors and enzymes.

Step 2: A tertiary complex of {ArsC Cys-12}S-arsenate-S{glutathione} (intermediate II) is formed: Intermediate II was inferred from biochemical studies that showed that glutathione reacts only after arsenate binds and that a free thiol is required on ArsC and glutathione for the reaction to take place. However, the researchers were unable to image this complex by X-ray crystallography.

Step 3: Arsenate it reduced to arsenite: A short-lived quaternary complex involving glutaredoxin is formed that dissociates into thioarsahydroxy adduct of ArsC (intermediate III) and a mixed disulphide complex of glutathione and glutaredoxin. Although this quaternary complex appears to be too short-lived to be captured by crystallographic techniques, it has been shown previously that glutaredoxin binds to ArsC only if glutathione and arsenate are present (Liu and Rosen, 1997). The thioarsahydroxy adduct of ArsC has an unusual structure with only two bonds. The positively charged arsenite atom is also unusual, but the researchers believe that it may be stabilized by resonance. They argue that this causes the adduct to be less stable than other cysteine-arsenite complexes ensuring that arsenite does not inhibit this protein as it does other enzymes.

Step 4: Hydrolysis of the ArsC-arsenite bond releasing arsenite and returning ArsC to the original configuration to reduce another arsenate ion.

Determination of the crystal structures of the ArsC proteins from *B. subtilis* (Guan *et al.,* 2001 and Bennett *et al.*, 2001) and the *S. aureus* plasmid pI258 (Zegers *et al.*, 2001) indicated that these proteins, unlike the R773 ArsC, have structures highly similar to that of low molecular weight protein tyrosine phophatases (LMW PTPases). The proteins consist of a single α/β domain consisting of a central four-stranded, parallel open-twisted β-sheet, with α-helices on either side (Bennett *et al.*, 2001). Previous research had shown that three of the four cysteine residues found in this group of ArsC proteins are required for the function of the *S. aureus* ArsC enzyme (Cys-10, Cys-82 and Cys-89). The two latter cysteine residues are not found in the gram-negative group of ArsC proteins and are believed to form a disulphide bridge upon oxidation (Messens *et al.*, 1999). Due to the similarity between this ArsC and

the LMW PTPases, comparisons were drawn between the active sites of these two enzymes. The Gram-positive ArsC proteins as well as the LMW PTPases have a conserved CX_5R motif. The *B. subtilis* ArsC was crystallised in native form only and the crystal structures show that this region forms an oxyanion binding loop called the arsenate binding loop (AB loop) resembling the PTP loop that is the catalytic site of the PTPases (Bennett *et al.*, 2001). The AB loop is slightly larger than the PTP loop to enable the binding of the larger arsenate ion. The conserved arginine residue (Arg-16) is nearby to the Cys-10 and Cys-82 residues. This residue has an overall positive charge and its placement makes it an ideal candidate to lower the pK_a values of the cysteine residues, thereby stabilizing the thiolate ions. The third cysteine residue mentioned earlier (Cys-89) is situated at the other end of a flexible region from Cys-82. It is believed that this region can move during the reaction cycle, bringing Cys-89 close to the active site, allowing it to form a disulphide bridge with Cys-82. Finally there is an aspartate residue (Asp-105) that is conserved in all Gram-positive arsenate reductases. This residue is situated quite close to the active site and is believed to play a role as a general acid/base catalyst (as it does in the PTPases). Based on the above observations and comparisons with the LMW PTPases the following model for the reaction mechanism of the *B. subtilis* ArsC has been put forward (Bennett *et al.*, 2001): (Figure 1.7)

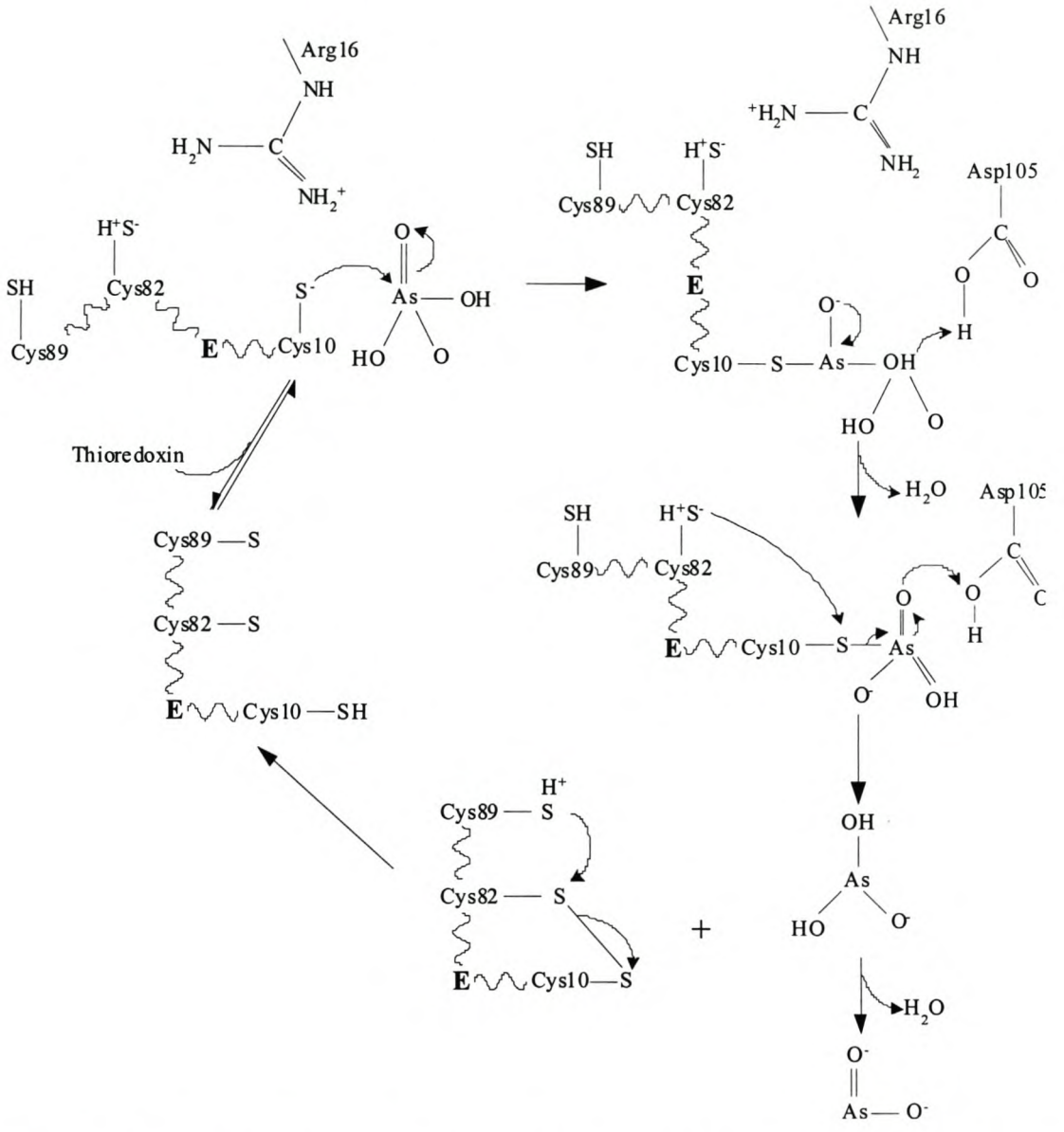


Figure 1.7: Reaction scheme for *B. subtilis* ArsC (taken from Bennett *et al*, 2001). **E** represents the ArsC enzyme. See text for details.

Step1: Nucleophilic attack on arsenate: Cys-10 of ArsC acts as an attacking nucleophile forming an arsenylated enzyme substrate. The active site of ArsC is surrounded by polar residues allowing the entry of arsenate ions into the active site. Here the Asp-105 residue would help this reaction as a general acid.

Step 2: Reduction of arsenate: This step requires a triple cysteine redox relay involving the three cysteine residues mentioned above, ultimately resulting in the reduction of arsenate. First the {ArsC Cys-10}S-arsenate bond is attacked

by the adjacent Cys-82. The arsenate ion is reduced to arsenite by obtaining electrons from Cys-10 and Cys-82. A disulphide bridge is formed between Cys-10 and Cys-82. Cys-89, which is in a flexible region, can then come close to Cys-82 and be activated by the positive charge on Arg-16. This would allow oxidation of Cys-82 resulting in a disulphide bond between Cys-82 and Cys-89. Cys-10 is now free to interact with another arsenate ion. Arg-16 has three roles in this process. It stabilizes the AB loop and the binding of the arsenate ion and as mentioned earlier, its positive charge lowers the pK_a of the cysteine residues thereby activating them for the reactions.

Step 3: Regeneration of ArsC: Thioredoxin reduces the disulphide bond between Cys-82 and Cys-89 and regenerates the whole system.

The ArsC protein from *S. aureus* pI258 was crystallised in both the oxidised and reduced forms (Zegers *et al.*, 2001). In the reduced form Cys-82 and Cys-89 form a disulphide bridge. This is associated with a conformational change allowing the Cys-89 to move closer to Cys-82 in order to form this disulphide bond. Zegers *et al.* (2001) found that ArsC was also able to catalyse a dephosphorylation reaction and propose that the ArsC protein evolved from a PTPase ancestor. The catalytic mechanism put forward by Zegers *et al.* is the same as that shown for the ArsC from *B. subtilis*. The reaction mechanism proposed was further confirmed by creating various mutations of the important residues and thereby visualising the intermediates of the reaction pathway (Messens *et al.*, 2002). These authors were able to show that the removal of the disulphide bond formed between Cys-10 and Cys-82 restored the protein to the original conformation. This indicated that it is the formation of the Cys-10-Cys-82 disulphide bond that triggers the change in conformation that allows the Cys-89 to move close enough to attack Cys-82.

All the above results indicate that the two different types of ArsC proteins found in bacteria have evolved separately to perform the same function of reducing arsenate to arsenite. However, both reaction mechanisms make use of sequential nucleophilic attacks by three different thiolates. In the case of the *E. coli* plasmid R773 these thiolates are located intermolecularly (on ArsC, glutathione and glutaredoxin), while the reduction of arsenate to arsenite in *B. subtilis* and *S. aureus* makes use of three cysteine residues all located on the ArsC protein itself.

1.4.3. Regulation of the arsenic resistance genes

1.4.3.1.ArsR

The *arsR* gene encodes a dimeric trans-acting repressor. The ArsR protein from the *E.coli* plasmid R773 *ars* operon has been studied in great detail. Wu and Rosen (1991) showed that the expression of ArsR from *E. coli* was autoregulated and that this expression was repressed by ArsR in the absence of inducers, and induced in their presence. Further studies using gel retardation and footprinting analysis showed that the ArsR binds to a DNA region containing an imperfectly symmetrical dyad sequence just upstream of the –35 site of the *ars* promoter (Wu and Rosen, 1993). Higher resolution analysis showed that only two small regions (4 bp each) within this area were actually protected. These were separated by 10 bp indicating that the repressor binds only to one side of the DNA. Wu and Rosen (1993) also showed that the ArsR protein acts as a dimer which is stable in the cytosol.

It was found that the ArsR proteins belong to a new class of DNA binding proteins which also includes the cadmium regulatory proteins and the SmtB repressor (activated by zinc) which regulates the *smtA* gene encoding for metallothionien. These proteins have a helix-turn-helix region containing two conserved cysteine residues (Cys-32 and Cys-34) and either one or two histidine residues (Bairoch, 1993). Alignments of the regulators belonging to the ArsR-family showed that a conserved sequence, ELCVCDL, was present in all members. Mutants with tyrosine or phenylalanine instead of the cysteine residues were found to bind to the DNA operator (shown with gel mobility shift assays and DNase I footprinting analysis), but responded poorly to the inducers. This indicated that the cysteine residues are involved in inducer binding, but not DNA binding (Shi *et al.*, 1994). This also suggests that the cysteine residues are located outside of the DNA binding helix-turn-helix motif. Mutation of the histidine residue located in the second helix of this motif resulted in decreased affinity for the operator. The observation that these cysteine residues are involved in the inducer binding suggests that the arsenite and antimonite react as soft metals with the thiolates of the cysteine residues. These results were confirmed by substituting the cysteine residues with glycine instead of tyrosine and phenylalanine. This confirmed that the previous results were not due to the addition of residues with bulky aromatic side chains (Shi *et al.*, 1996).

It is known that As(III) can form tricoordinate complexes with either sulphur (eg. arsenite-glutathione complexes, $As(GS)_3$, or with oxygen (arsenite-dithiothreitol complex, where the arsenite forms soft metal bonds with two sulphur thiolates and one hydroxyl oxygen). For this reason the involvement of another residue in the binding of arsenite to ArsR was investigated (Shi *et al.*, 1996). Of the three other cysteine residues found in R773 ArsR, two were known not to be required for ArsR function (Wu and Rosen, 1993). The third (Cys-37) was mutated to alanine. While this mutation did not affect the regulation of ArsR by inducers or the ability of ArsR to bind to DNA, it was shown using affinity chromatography that binding of As(III) to Cys-37 does occur. X-ray absorption spectroscopy studies showed that three sulphur ligands are involved in the interactions with As(III) and no serine or threonine ligands are involved (Shi *et al.*, 1996). A model whereby As(III) is bound in a cage formed by the three cysteine thiolates was proposed. Although arsenite binds to all three thiolates, binding at Cys-32 and Cys-34 produces a conformational change in the DNA-binding domain that results in the dissociation of ArsR from the promoter, allowing for transcription of the arsenic resistance genes (Shi *et al.*, 1996). ArsR has a high affinity for the inducer, which implies that it must sense and respond to low concentrations of the inducer (Shi *et al.*, 1996). The thiol groups of the cysteine residues in ArsR recognise the metallic As(III) and provide this function.

Studies on the chromosomal ArsR protein from *E. coli* confirmed the above results even though the protein was only 75% identical to the plasmid-borne ArsR. It was also found that the repressor from R773 and the chromosomal repressor were interchangeable. Although the overall sequence of the protected sequences in the promoter of the chromosomal and plasmid genes differed, it was found that the chromosomal version also contained the same 4 bp sequences separated by 10 bp. This is not conserved in promoter regions of other *ars* operons (Xu *et al.*, 1996).

It was proposed that all members of the ArsR family contain at least three domains: the DNA binding domain, a metal binding domain and a dimerization domain. Xu and Rosen (1997) attempted to isolate the dimerization region. Using a yeast two-hybrid system they were able to show that residues 1-8 and 90-117 of *E. coli* ArsR are not required for dimerization. They also showed that without dimerization ArsR was

unable to regulate expression from the *ars* promoter. It was, therefore, proposed that a core region of about 80 residues contains all the regulatory properties (dimerization, DNA-binding and metal binding domains) of the ArsR repressor.

The crystal structure of SmtB, which shows homology with the ArsR proteins has been determined (Cook *et al.*, 1998). The authors predict that the ArsR proteins will have a similar structure, which includes a winged HTH DNA-binding motif.

1.4.3.2.ArsD

The *arsD* gene from the R773 operon has been identified as a second *ars* regulator (Wu and Rosen, 1993). It was previously thought that the product of this open reading frame was unnecessary for the function of the *ars* operon. When a frameshift mutation was introduced into the *arsD* gene, it was found that the cells containing this mutation were sensitive to arsenic even though the *arsABC* genes were intact. It is thought that this was due to the overexpression of the ArsB protein which is deleterious to the cell. The resistance could be restored by the addition of an intact *arsD* in *trans*. When translational *blaM* fusions were constructed with the *arsA*, *arsB* and *arsC* genes, it was found that the presence of the mutated *arsD* gene resulted in overexpression of the fusion proteins. Once again this expression could be reduced to normal levels with the addition of an intact *arsD* in *trans* (Wu and Rosen, 1993).

It was shown that ArsD binds to the *ars* promoter at the same position as the ArsR dimer, but that ArsD has a lower affinity for the promoter (Chen and Rosen, 1997). It was previously thought that regulation by the ArsD protein was inducer-independent, but it was subsequently shown that repression by ArsD could be relieved, although this required much higher levels of arsenite than for ArsR (Chen and Rosen, 1997).

Of the eight cysteine residues found in ArsD, six are arranged in pairs (Cys-12-Cys-13; Cys-112-Cys-113 and Cys-119-Cys-120). A recent study has shown that two of the vicinal cysteine pairs, Cys-12-Cys-13 and Cys-112-Cys-113, are involved in the inducer response (Li *et al.*, 2001). Mutations of these cysteine residues resulted an inability to respond to inducer, but did not affect DNA binding. The authors propose that the ArsD protein contains two arsenite binding sites. Since arsenite is three-coordinate, it is predicted that it would bind a third ligand, yet to be identified.

Binding to both of the sites is believed to be required to produce a conformational change resulting in the release of the promoter. However, ArsD also binds to the promoter region as a homodimer, meaning that there are potentially four arsenite binding sites. Further research has shown that there is positive cooperativity between adjacent sites on the homodimer (Li *et al.*, 2002). Based on their results, these authors have proposed a model whereby the arsenite binding sites are formed by pairs of equivalent cysteine residues from each of the subunits of the ArsD dimer. The arsenite ions bind to these four arsenite-binding sites sequentially, starting by binding to Cys-113 residues and then being transferred to the Cys-112, Cys-13 and finally the Cys-12 residues. When the concentration of arsenite is high enough, all sites of the ArsD dimer will be filled and dissociation from the promoter occurs.

Based on the above observations the following model for the ArsR-ArsD regulatory circuit was proposed (Figure 1.8): Small amounts of all proteins will be synthesised at a basal level. Due to the fact that the ArsR has a higher affinity for the *ars* promoter, it would preferentially bind to the operator and repress *ars* expression. ArsR also has a higher affinity for the inducer than ArsD. As a result, a low concentration of inducer will cause the ArsR to dissociate from the promoter and the *ars* genes would be expressed (including *arsD)*. When the concentration of ArsD reaches a critical level it will bind to the operator and shut down the expression of the genes. However if there is a high concentration of the inducer, ArsD will dissociate from the operator allowing further expression of the genes. This regulatory circuit only allows the expression of the *ars* genes under a narrow range of conditions (Chen and Rosen, 1997).

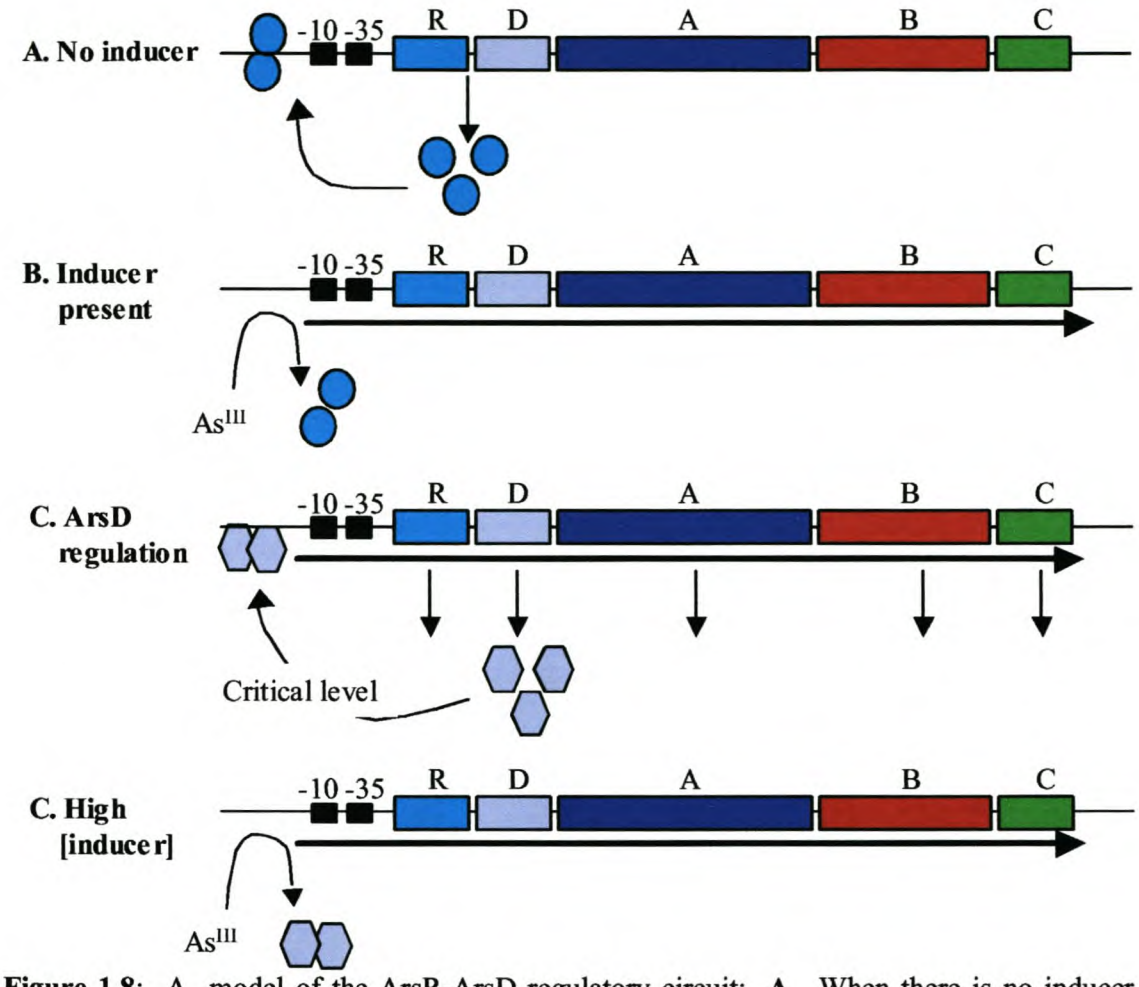


Figure 1.8: A model of the ArsR-ArsD regulatory circuit: **A**. When there is no inducer present, there is a basal level of transcription allowing production of ArsR which prevents further expression of the operon. **B**. When the inducer is present it binds to ArsR causing it to dissociate from the promoter, allowing expression of the *ars* genes. **C**. When ArsD accumulates to a certain critical level it binds to the promoter and prevents further expression of the operon. **D**. When there is a high concentration of inducer, this binds to ArsD causing it to dissociate from the promoter, thereby allowing further expression of the operon. (adapted from Chen and Rosen, 1997)

1.4.4. Other genes within ars operons

As more *ars* operons have been discovered most have either the 5 gene or 3 gene arrangement (see Table 1.1). However there are two other genes that have been found associated with *ars* operons. The first is found within the *ars* operon on the skin element of *Bacillus subtilis*. This operon has homologues to the *arsRBC* genes, but contains a fourth gene (*ORF2*) located between the *arsR* and *arsB* genes (Sato and Kobayashi, 1998). This gene did not show homology to any genes with known function, but it did show homology to an open reading frame (ORF) of unknown

function situated before a predicted *arsRBC* operon in *Mycobacterium tuberculosis*. The experiments conducted on the *ars* genes of the *B. subtilis* operon did not elucidate the role of *ORF2* in arsenic resistance. A BLAST search (www.ncbi.nlm.nih.gov/BLAST) reveals that homologues to this gene are also found in *Streptomyces coelicolor* and *Nostoc* sp. along with homologues to other *ars* genes.

The second gene was first identified associated with the *ars* genes from *Yersinia enterocolitica* plasmid pYV. The *arsRBC* operon found on this plasmid also has a fourth gene (*arsH*) which is divergently transcribed upstream of the *arsR* gene (Neyt *et al.*, 1997). Although the function of this gene is unknown, it was required for resistance to arsenic. A *Y. enterocolitica* strain containing the *arsRBC* genes, but not *arsH* was constructed. This strain was sensitive to arsenic and resistance could be restored by the addition in *trans* of the *arsH* gene under the control of the *plac* promoter. This is surprising as the *arsRBC* genes are sufficient to confer resistance in other bacteria. The authors suggest that *arsH* may function as some sort of regulator, similar to *arsD*. However, no further work has been done to confirm this hypothesis.

Recently, another *arsH* homologue was identified in the *ars* operon found on the IncH12 plasmid R478 (Ryan and Colleran, 2002). This plasmid was found to contain an *ars* operon similar to that found on pYV from *Yersinia enterocolitica*. The *arsH* gene is also divergently transcribed from the *arsRBC* genes. Amplification by means of the PCR, using primers to the *arsH* gene, showed that this gene is found in many other arsenic resistant IncH12 plasmids. A clone constructed from R478, which was missing approximately 16 % of the C-terminal of ArsH was more sensitive to arsenite than R478. Complete removal of *arsH*, resulted in a total loss of arsenite resistance. However, the authors have not added the *arsH* gene *in trans* in order to show that this lack of arsenite resistance was not due to a disruption of the promoter of the *arsRBC* operon.

1.5. Aim of Thesis

As mentioned earlier, the aim of this study was to characterize the arsenic resistance genes identified during a preliminary study of *At. ferrooxidans* ATCC33020. This preliminary study had involved screening a plasmid library for the ability to confer

arsenite resistance to the *E. coli* arsenic mutant, AW3110. The plasmid isolated was able to confer resistance to both arsenite and arsenate. Analysis of the partial sequence revealed homologues to both the *arsB* and *arsC* genes from other *ars* operons (figure 1.9).

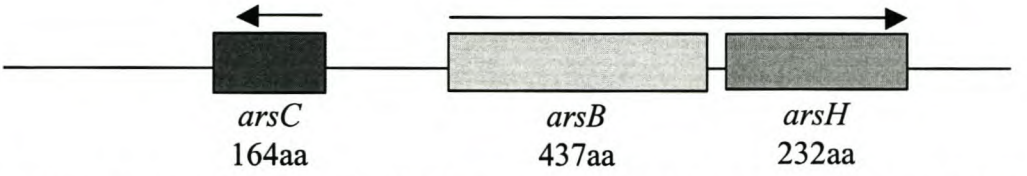


Figure 1.9: Arrangement of the *ars* operon from *At. ferrooxidans* (based on preliminary results). This operon has an unusual divergent arrangement of the *arsB* and *arsC* genes. It also contain a homologue of the *arsH* gene of unknown function. No regulator of the operon has been identified at this stage.

However, these genes were divergently transcribed, unlike any other arsenic resistance system studied to date. Furthermore, no *arsR*-like gene was identified. Another interesting discovery was the presence of an *arsH* homologue downstream of the *arsB* gene. This interesting divergent arrangement and the presence of the *arsH* gene raised many questions, for example: Is *arsH* required for resistance to arsenic in this system and if so can the function of the *arsH* gene in this system be determined? How are these divergently transcribed genes regulated? Is the *arsC* gene expressed only in the presence of arsenate? An attempt to address these and other questions has led to the work presented in the dissertation to follow.

Chapter Two: The *Acidithiobacillus ferrooxidans* chromosomal arsenic resistance genes have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *

TABLE OF CONTENTS

* These results have been published (Butcher *et al*, 2000), however the sequence analyses, figures and text in this chapter have been updated to include new data in the NCBI database and results published since the publication of our results .

2.1.INTRODUCTION

Acidithiobacillus ferrooxidans is an acidophilic (optimum pH 1.8 to 2.5), obligately chemolithotrophic bacterium that obtains its energy through the oxidation of ferrous iron to ferric iron or reduced inorganic sulphur compounds to sulphuric acid. It is a member of a consortium of bacteria (which includes *Acidithiobacillus caldus* and *Leptospirillum ferrooxidans*) that is used in commercial biooxidation processes for the recovery of gold from arsenopyrite ores (Rawlings and Silver, 1995). Although recent analysis of microbial populations in continuous-flow biooxidation tanks has indicated that *At. ferrooxidans* may not be as dominant as was once thought, it is nevertheless usually present in such tanks. Levels in excess of 13 g l^{-1} total arsenic may be present in these arsenopyrite biooxidation tanks and the microorganisms present therefore require a mechanism of resistance to arsenic (Dew *et al.*, 1997).

Plasmid-associated arsenic efflux resistance mechanisms have been known for many years and have been extensively reviewed (Silver *et al.*, 1993; Cervantes *et al.*, 1994; Silver and Phung, 1996; Rosen, 1999; Rensing *et al.*, 1999 and Mukhopadhyay *et al.*, 2002). Although the number of components of these systems varies, in the case of *Escherichia coli* plasmids R773 and R46 as well as the *Acidiphilium multivorum* plasmid pKW301 (Suzuki *et al.*, 1998) as many as five genes (*arsRDABC*) are present. In the case of R773, the genes are transcribed in a single operon. The *arsR* and *arsD* genes encode repressors that control the basal and upper level, respectively, of *ars* operon expression while the *arsABC* genes encode the structural components of the arsenic resistance mechanism. ArsA is an ATPase that forms a complex with ArsB, the trans-membrane arsenite efflux pump. ArsC is a small, cytoplasmic-located, arsenate reductase that reduces arsenate to arsenite which can then be pumped out of the cell. The ArsB protein is capable of exporting arsenite even in the absence of ArsA (Dey and Rosen, 1995).

The arsenic resistance genes of the *Staphylococcus* plasmids pSX267 and pI258 as well as the chromosomally located arsenic resistance genes of *E. coli* (Carlin *et al.*, 1995) and *Pseudomonas aeruginosa* (Cai *et al.*, 1998) consist of only three genes, *arsRBC*. Nevertheless, these *ars* operons are capable of exporting arsenate, arsenite and antimony oxyanions in the absence of *arsA* using membrane potential rather than ATP as an energy

source (Dey and Rosen, 1995 and Kuroda *et al.*, 1997). Recently an arsenic resistance operon that consists of *arsRBC* and a fourth ORF of unknown function was discovered in the *skin* element of *Bacillus subtilis* (Sato and Kobayashi, 1998). An arsenic resistance mechanism has been discovered in Tn*2502* located on plasmid pYV of *Yersinia enterocolitica*, that consists of *arsRBC* as well as a divergently transcribed gene, *arsH* (Neyt *et al.*, 1997). A second operon with the same function was also identified on the IncH12 plasmid, R478, isolated from *Serratia marcescens* (Ryan and Colleran, 2002). The function of *arsH* is unknown, but its presence either in *cis* or in *trans* was essential for arsenic resistance in *Y. enterocolitica.*

Here we report the isolation and analysis of the evolutionary relationship of arsenic resistance genes from *At. ferrooxidans.* We further show that these genes are functional in *E. coli* and have an unusual divergent *arsCRBH* operon structure that appears to be conserved in all *At. ferrooxidans* strains examined.

2.2.METHODS

Bacterial strains, plasmids, primers and media. The strains, plasmids and primers used in this study are shown in Table 2.1. *E. coli* strains were grown on Luria-Bertani liquid (LB) or agar (LA) medium (Sambrook *et al.*, 1989). *At. ferrooxidans* strains were grown in tetrathionate medium or iron sulphate medium (Powles *et al.*, 1995) at 30°C. Ampicillin (100 μg/ml), chloramphenicol (20 μg/ml) and tetracycline (20 μg/ml) were used as required.

Table 2.1: Strains, plasmids and primers used in this chapter.

Strains, plasmids and primers	Description	Reference or source
Strains		
Escherichia coli		
W3110	K-12 F⁻ IN(*rrnD-rrnE*)	B. Rosen, Wayne State University, Michigan
AW3110	W3110 Δ*ars*::*cam*	Carlin *et al.*, 1995
JM109	*endA*1 *recA*1 *gyrA*96 *thi hsdR*17 (r_k^-,m_k^+) *relA*1 *supE*44 Δ(*lac-proAB*) [F′ *trad*36 *proAB lacI*^{q}ZΔM15]	Promega Corp. USA

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 Corp. USA
MC1061	K-12 F^- *araD*139? *galU galK hsdR rpsL* Δ(*araABC-leu*) 7679 D*lacX*74	Britt Persson, University of Umea, Sweden
BH5262	K-12 *araD*139? *galU galK hsdR rpsL argH*1 *trxA*7004 *ilvC*::Tn10 *gshA*	Lim *et al.*, 1986
BH2012	K-12 *araD*139? *galU galK hsdR rpsL metA*46 *argH*1 *trxA*7004 *ilvC*::Tn5	Lim *et al.*, 1986
Acidithiobacillus ferrooxidans		
ATCC 33020	Wild type	Rockville, Md.
ATCC 23270	Wild type	Rockville, Md.
ATCC 19859	Wild type	Rockville, Md.
ATCC 13598	Wild type	Rockville, Md.
Acidithiobacillus caldus		
MNG	Wild type	This laboratory
Acidithiobacillus thiooxidans		
ATCC 19377	Wild type	Rockville, Md.
Leptospirillum ferrooxidans		
DSM 2705	Wild type	Braunschweig, FRG
Plasmids		
pBluescript (SK)	Amp^r *lacZ'*, cloning vector	Stratagene, USA
pUCBM21	Amp^r *lacZ'*, cloning vector	Stratagene, USA
pACYC184	Tc^r Cm^r, cloning vector	Chang and Cohen
pEcoR251	Ap^r, cloning vector	
pTfars1a	Ap^r (from *At.* ferrooxidans plasmid library ~7 kb *Sau*3A1 fragment cloned into pEcoR251 digested with *Bgl*II)	Library made by Ramesar
pTfars1b	Ap^r (6.7kb *Hin*dIII–*Bgl*II fragment of pTfars1a cloned into pBluescript)	This study
pTfarsCBH	Ap^r (5.3kb *Hin*dIII-*Stu*I fragment of pTfars1b cloned into pBluescript cut with *Eco*RV and *Hin*dIII)	This study
pTfarsCB	Ap^r (The *Eco*RV-*Xba*I fragment of pTfars1b, containing the 5′ end of *arsB* and the whole of *arsH*, was replaced with the *Eco*RV-*Xba*I fragment of pBB08, containing only the 5′ end of *arsH*.) see figure 2.1.	This study
pTfarsBH	Ap^r (*Kpn*I deletion of pTfarsCBH)	This study
pTfarsB	Ap^r (*Pst*I deletion of pTfarsBH)	This study
pTfarsH	Ap^r (PCR product using the ARSHF and ARSHR primers, cloned into pBluescript)	This study
pTfarsC	Ap^r (PCR product using the BBARSCF and BBARSCR primers, cloned into pBluescript)	This study
pTfarsR	Ap^r (0.55kb *Sph*I fragment of pTfarsCBH, cloned into pUCBM21)	This study
pBB08	Ap^r (0.8kb *Eco*RV-*Pst*I fragment from pTfars1b, cloned into pBluescript)	This study
pTfarsCBH-Cm	Cm^r (6.7 kb *Hin*dIII-*Bgl*II fragment from pTfars1a, cloned into the *tet* gene of pACYC184)	This study
pTrx6	Ap^r (contains *At. ferrooxidans* ATCC33020 *trxA* gene, cloned into pBluescript)	Powles *et al.*, 1995)

pTTn1	Ap^r (contains ORF8 from pTF-FC2 - with similarity to a glutaredoxin-like protein, cloned into pBluescript)	Clennel *et al.*, 1995)
Primers		
BBARSB	(*Bam*HI) 5′-GCGGATCCAGGGTGACGAGAAATATGGC-3′	This study
BBARSC	(*Bam*HI) 5′-GCGGATCCGGGGTTTTCATCACTGG-3′	This study
ARSHF	(*Eco*RI) 5′-GCGAATTCTGGTGGCTGCCGCTGGCTTG-3′	This study
ARSHR	(*Hin*dIII) 5′-GAAAGCTTGCGTACCCCCAACCTCATGCC3′	This study
BBARSCF	(*Hin*dIII) 5′-GCAAGCTTCGGTGAAACCCGCTCTCCCT-3′	This study
BBARSCR	(*Eco*RI) 5′-GCGAATTCTGTGCGCGCCTTGATGGGTGGC-3′	This study

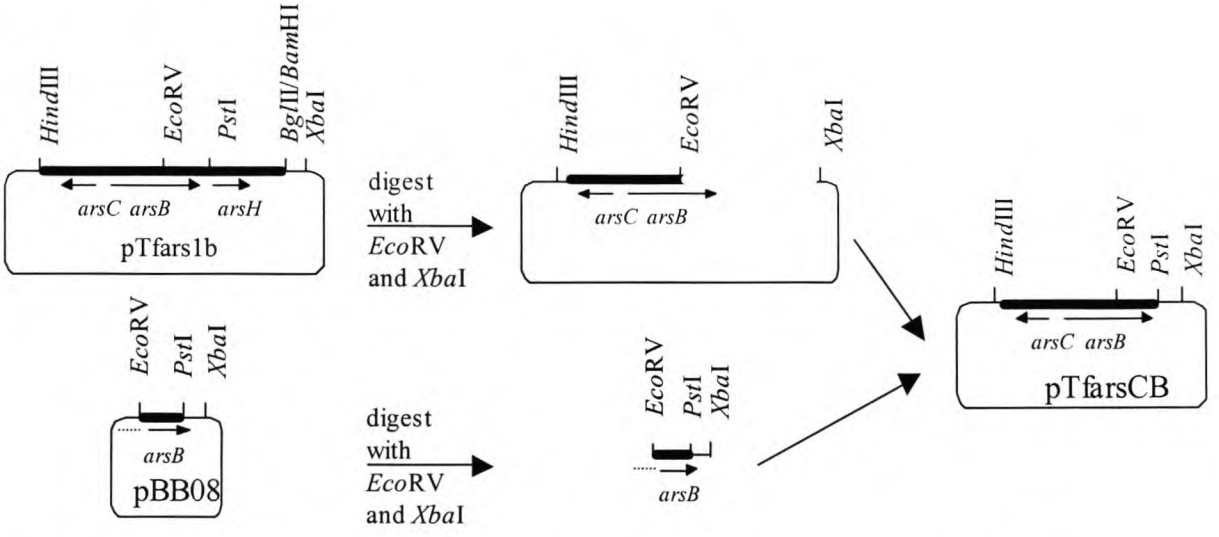


Figure 2.1: Diagram showing the strategy used to construct the *arsH* deletion clone, pTfarsCB. The plasmid, pTfars1b, was digested with *Eco*RV (located within the *arsB* gene) and *Xba*I (located in the multiple cloning site of the vector). This resulted in the deletion of *arsH* and the 3' end of *arsB*. The *Eco*RV-*Xba*I fragment from another clone, pBB08, was used to reconstruct *arsB* without adding *arsH*.

DNA techniques, sequencing and analysis. An *At. ferrooxidans* ATCC33020 gene bank consisting of 4 to 9 kb fragments from a partial *Sau*3A1 digest cloned into the *Bgl*II site of the suicide vector pEcoR251 (Ramesar, 1988) was transformed into the *E. coli ars* deletion mutant AW3110, made competent by the SEM method (Inoue *et al.*, 1990). Plasmid preparation, restriction endonuclease digestions, gel electrophoresis, ligations and Southern hybridisation were carried out using standard methods (Sambrook *et al.*, 1989). Labelling of probes, hybridisation and detection were done using the dioxygenin-dUTP non-radioactive DNA labelling and detection kit (Roche). Sequencing was by the dideoxy chain termination method (Sanger *et al.*, 1977), using the thermosequenase flourescent labelled primer cycle sequencing kit (Amersham Pharmacia Biotech UK Ltd.). Sequencing reactions were run on

an ALFexpress automated sequencer (Pharmacia Biotech, Uppsala, Sweden) at the University of Cape Town Sequencing Service. Additional clones constructed for the sequencing of pTfars1b are listed in Appendix 1. Results were analysed using the VAX based Genetic Computer Group Inc. (GCG) sequence analysis package (version 7.1) and its associated programs and the PC based DNAMAN (version 4.1) from Lynnon BioSoft. Comparison searches were performed using the gapped-BLAST program at the National Center for Biotechnology Information (Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST). Alignments and subsequent phylogenetic and homology trees were constructed using the Multiple Sequence Alignment tool in DNAMAN. Multiple sequence alignments were shaded using the Genedoc Multiple Sequence Alignment Editor and Shading Utility (version 2.5.000) by Karl Nicholas.

Requirement for thioredoxin for arsenate resistance. A 7.1 kb *Hin*dIII - *Bgl*II fragment from pTfars1a, containing the *ars* genes of *At. ferrooxidans*, was cloned into pACYC184 digested with *Hin*dIII and *Bam*HI. The resulting clone, pTfarsCRBH-Cm, allowed the testing of pBluescriptSK-based plasmids *in trans*. *E. coli* strains BH2012, BH5262 and MC1061 were made competent using $CaCl_2$ and were transformed with pTfarsCRBH-Cm. *E. coli* BH5262 was also transformed with pTrx6, pTTn1 and pTT150 (Table 2.1). All of the above strains were streaked on LA plates containing 0, 2, 5, 7, 10 and 15 mM sodium arsenate and incubated at 37°C overnight.

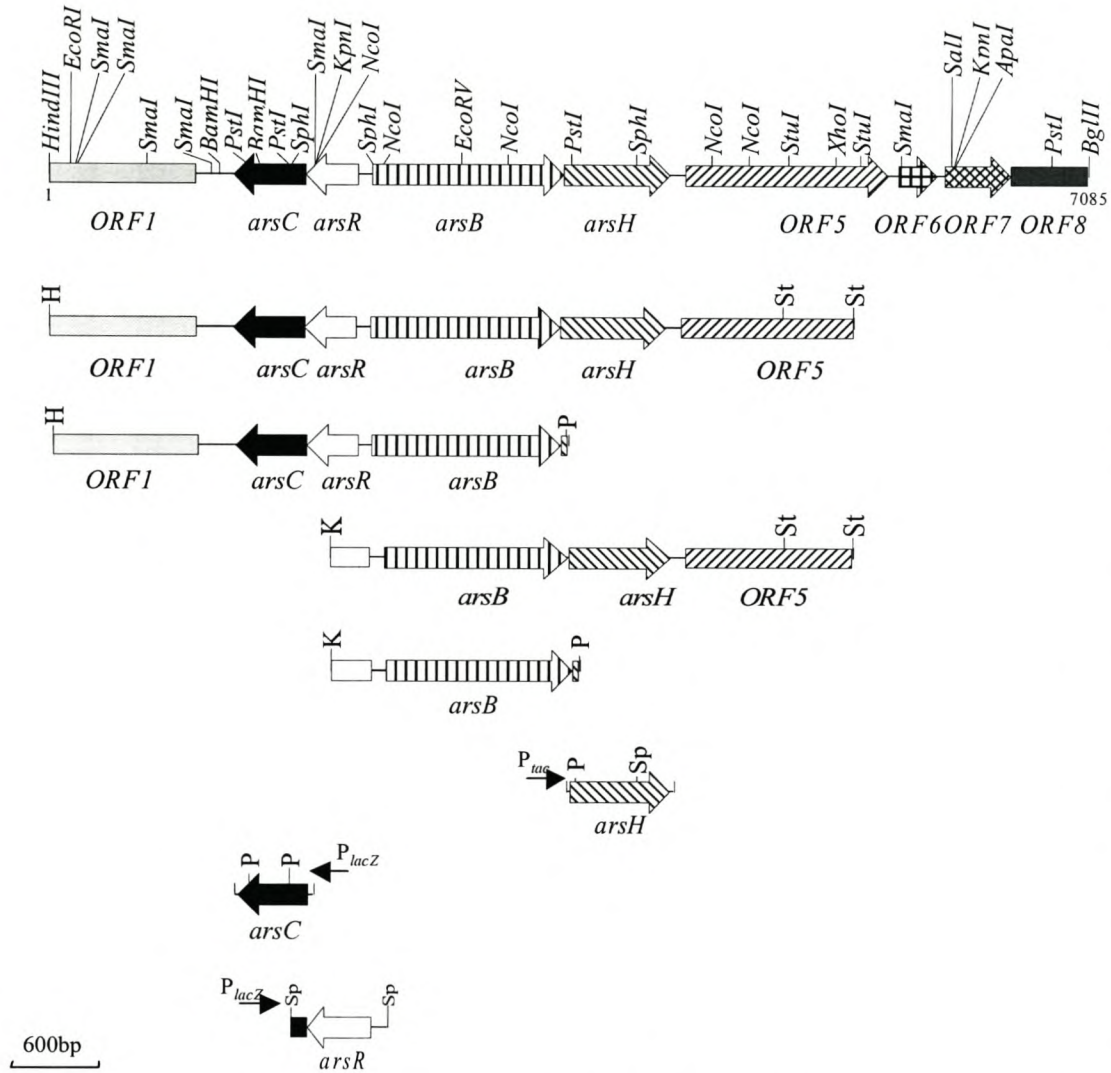
Polymerase chain reaction. The polymerase chain reaction was carried out using the primers described in Table 2.1 and synthesized by the Synthetic DNA Laboratory, Department of Biochemistry (University of Cape Town). The reaction was performed in a Biometra Personal Cycler using Redhot polymerase (Advanced Biotechnologies). After initial denaturation of 60s at 94°C, 25 cycles of 30s at 94°C, 30s at 57°C (for primers BBARSB and BBARSC) or 63°C (for primers ARSHF and ARSHR) and 90s at 72°C were performed. A final extension step of 120s at 72°C before cooling to 25°C completed the reaction.

Arsenic and antimony resistance assays. Constructs were transformed into competent *E. coli* AW3110 cells. Assays of resistance to arsenite and antimonite were carried out in Luria Bertani broth (LB) medium. Assays for resistance to arsenate were done on cells grown 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 increasing concentrations of sodium arsenite, potassium antimonite or sodium arsenate. Cultures were incubated at 37°C for 5 hr in the case of LB medium or 12 hr in the case of low phosphate medium and the absorbance read at 600 nm. These times corresponded to the end of the log phase of growth of a control cultured under the same conditions.

***In vitro* transcription-translation analysis.** A prokaryotic-DNA-directed, *E. coli* S30 extract-based, *in vitro* transcription-translation kit for circular DNA (Promega) was used to analyse the polypeptides synthesised from clones containing various combinations of the arsenic resistance genes. The ^{35}S-methionine-labelled translation products were separated on a SDS-PAGE gel (4 % stacking gel and 12 % separating gel) and detected by autoradiography.

Nucleotide sequence accession number: GenBank AF173880



	$As^{III(*)}$	$As^{V(\#)}$
pTfars1b	+	+
pTfarsCRBH	+	+
pTfarsCRB	+	+
pTfarsBH	+	-
pTfarsB	+	-
pTfarsH	ND	ND
pTfarsC	ND	ND
pTfarsR	ND	ND

Figure 2.2: pTfars1b and deletion clones constructed in this study. This diagram shows the restriction endonuclease maps and whether the clones are resistant to arsenite (As^{III}) or arsenate (As^{V}). * + indicates growth on 0.5 mM sodium arsenite, - indicates no growth ; # + indicates growth on 0.2 mM sodium arsenate; ND = not determined.

2.3.RESULTS

2.3.1. Cloning of the *ars* genes from *At. ferrooxidans*

The *E. coli ars* operon deletion mutant (AW3110) was transformed with an *At. ferrooxidans* plasmid bank and colonies were selected for their ability to complement the mutant on LA plates containing 0.5mM sodium arsenite. One plasmid (pTfars1a), containing a 7.4kb insert, which retransformed *E. coli* AW3110 at high frequency to arsenite resistance was selected for further study. The plasmid was mapped to determine the positions of restriction endonuclease sites (Figure 2.2), and a 7.1 kb *Hin*dIII-*Bgl*II fragment was cloned into pBluescriptSK (creating pTfars1b). The source of the insert DNA was confirmed using Southern hybridisation. An internal *Hind*III-*Stu*I fragment of 5.3 kb was labelled and used to probe the chromosomal DNA from *At. ferrooxidans* ATCC 33020 and pTfars1b digested with *Pst*I (Figure 2.3). The 3.5 kb and 2.0 kb *Pst*I fragments which are internal to the insert of pTfars1b corresponded exactly in size to *Pst*I fragments of the *At. ferrooxidans* ATCC 33020 chromosomal DNA. This indicated that the insert DNA originated from *At. ferrooxidans* ATCC 33020, that it was present in a single copy and that no rearrangements within the region which included the two *Pst*I fragments had occurred during cloning.

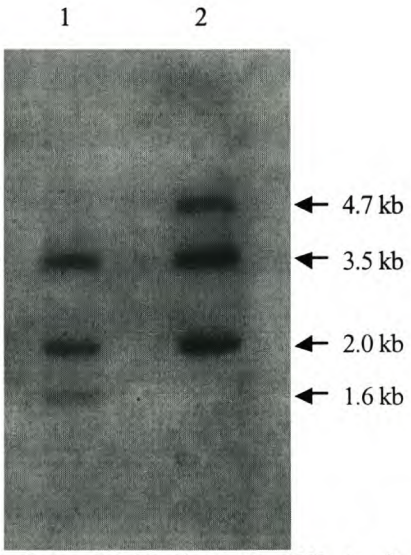


Figure 2.3: Hybridisation of the 5.3kb *Stu*I-*Hind*III *At. ferrooxidans arsCRBH* probe to the *Pst*I digests of *At. ferrooxidans* ATCC33020 chromosomal DNA (lane 1) and pTfars1b (lane 2)

Chromosomal DNA was also digested with two rarely cutting 8-bp recognition sequence restriction enzymes , *Pac*I and *Swa*I, as well as with a combination of these enzymes. Restriction fragments were separated by pulsed-field gel electrophoresis and were hybridised to an *At. ferrooxidans arsBH* probe. Signals of hybridisation to *Pac*I, *Swa*I and *Pac*I-*Swa*I fragments of approximately 450, 620 and 280 kb, respectively, were obtained (Figure 2.4). During chromosomal mapping experiments, DNA fragments, of the same sizes described above, hybridised to an *At. ferrooxidans ntrBC* chromosomal gene probe (unpublished data, S. Deane, pers. comm.). This indicated that the two sets of genes are located within 280 kb of each other and that the *At. ferrooxidans ars* genes are located on the chromosome.

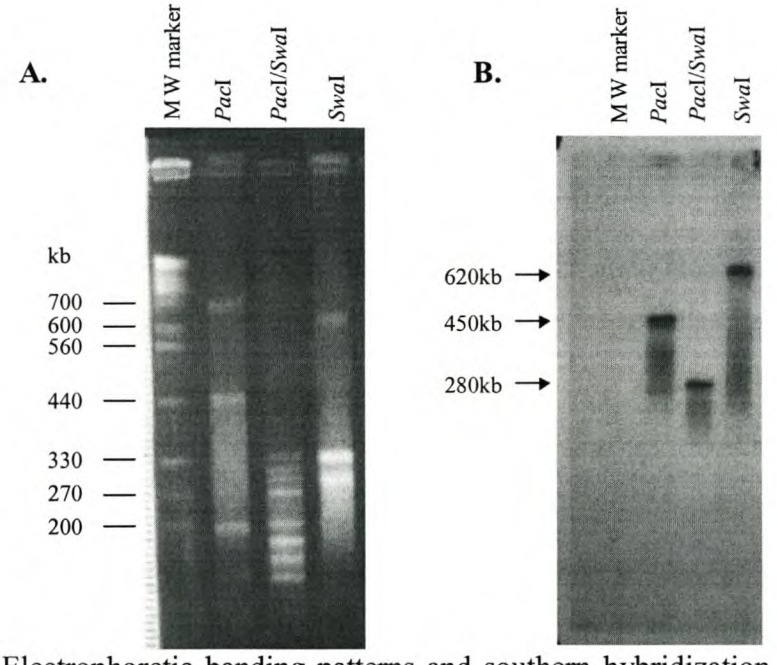


Figure 2.4: Electrophoretic banding patterns and southern hybridization analysis of *At. ferrooxidans* genomic DNA using Transverse Alternating Field Electrophoresis (TAFE) run with a 32 sec pulse for 16h at 150mA. **A.** Ethidium bromide stained gel. **B.** Southern hybridization with DIG-labeled pTfarsBH. (These experiments were performed by Shelly Deane.)

2.3.2. Sequence analysis of pTfars1b.

The entire insert DNA was sequenced on both strands and 9 open reading frames (ORFs) or partial ORFs were identified (Figure 2.2). The complete annotated sequence is shown in appendix 2. The predicted amino acid sequences from these nine open reading frames were analysed for homology to other proteins using the standard protein-protein BLAST tool from NCBI. These results and the characteristics of the predicted products of the 9 ORFs are shown in Table 2.2.

Table 2.2: Location and size of ORFs in pTfars1b

ORF	Position	No. of aa	size (kDa)	Closest relationship to known proteins	product detected [a]
ORF1	1014 - *	338	*	43/60% aa identity/similarity to hypothetical protein from *Azotobacter vinelandii* **ZP_00090840**[b]	yes
arsC	1752 - 1261	163	18.1	49/65% aa identity/similarity to the putative ArsC from *P. putida* **CAC18654**	yes
arsR	1752 - 2108	118	13.1	48/63% aa identity/ similarity to a hypothetical transcription regulator protein from *Sinorhizobium meliloti* **NP_385183**	yes
arsB	2202 - 3512	436	48.4	61/75% aa identity/similarity to hypothetical protein from *P. fluorescens* **ZP_00087013**	yes
arsH	3518 - 4240	240	26.7	79/89% aa identity/similarity to ArsH of Tn2502 from *S. typhimurium* **BAB91594**	yes
ORF5	4348 - 5724	458	44.7	62/79% aa identity/similarity to signal recognition particle protein (fifty-four homologue) *Neisseria meningitidis* **NP_273349**	yes
ORF6	5795 - 6055	86	9.56	63/80% aa identity/similarity to 30S ribosomal protein S16 of *P. aeruginosa* **NP_252434**	nd
ORF7	6111 - 6554	147	12.6	42/62% aa identity/similarity to 16S rRNA processing protein RIMM of *P. aeruginosa* **NP_252433**	nd
ORF8	6554 -*	94	*	56/73% aa identity/similarity to tRNA (guanine-N1)-methyltransferase of *Xylella fastidiosa* **NP_297402**	nd

[a] Protein products detected using a an *E. coli in vitro* transcription-translation system

[b] GenEMBL database accession numbers indicated on boldface.

* these open reading frames are incomplete and the size (kD) has not been calculated, however the no. of amino acids is shown.

nd = not determined.

Homologues to the *arsB* and *arsC* genes of other bacteria were identified, but unlike other systems where the *arsC* gene is downstream of the *arsB* gene, the *At. ferrooxidans arsC* gene is upstream of *arsB* and the genes are divergently transcribed (Figure 2.2). A fourth ORF was identified between the *arsC* and *arsB* genes, also divergent to *arsB*. This ORF has weak but clear homology (30-40% identity) to many transcriptional regulators including some members of the ArsR family. Although the putative ArsR from *At. ferrooxidans* contains two possible helix-turn-helix motifs these do not correspond to the position of the helix-turn-helix DNA binding domain identified in other studied ArsR proteins. The putative ArsR from *At. ferrooxidans* also does not contain the conserved motif including the two cysteine residues (ELCVCDL) which has been shown to be required for binding of the arsenite inducer (Shi *et al.*,1994) (see chapter 3 for a detailed discussion of ArsR from *At. ferrooxidans*).

The BLAST results show that the ArsB from *At. ferrooxidans* has the highest homology with the ArsB from *Y. enterocolitica* pYVe227 and a putative ArsB from *Klebsiella oxytoca* plasmid pMH12 (both with an E-value of e^{-113}). Phylogenetic analysis of the ArsB proteins, including some putative ArsB proteins, shows that the ArsB from *At. ferrooxidans* groups with the ArsB proteins from other Gram-negative bacteria (Figure 2.4), and an alignment shows that these proteins are highly homologous (overall identity of 61.61% between all the proteins in the Gram-positive and Gram-negative groups and 68.17% between the ArsB proteins of the Gram-positive group only). However, the *At. ferrooxidans* ArsB has the longest branch-length within the Gram-positive ArsB group indicating that this protein is the least similar of the group (Figure 2.5). Using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html), a web-based program for predicting trans-membrane spanning domains, we predict that the ArsB protein from *At. ferrooxidans* has 12 trans-membrane domains as is the case for other known ArsB proteins. The hydropathy plot for the ArsB protein from *At. ferrooxidans* is very similar to that for both the ArsB proteins from *E. coli* R773 and *S. aureus* pI258 (Figure 2.6).

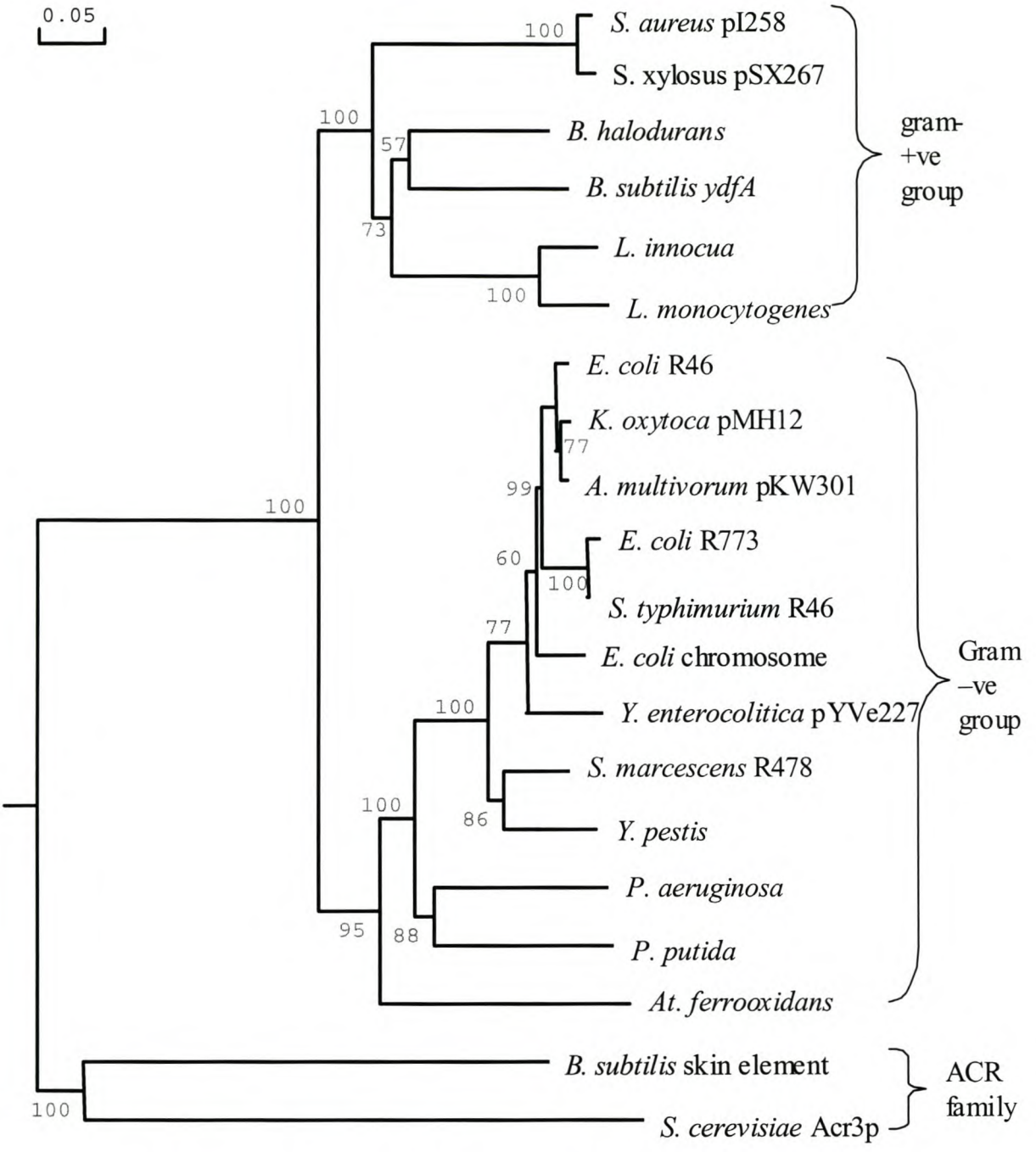
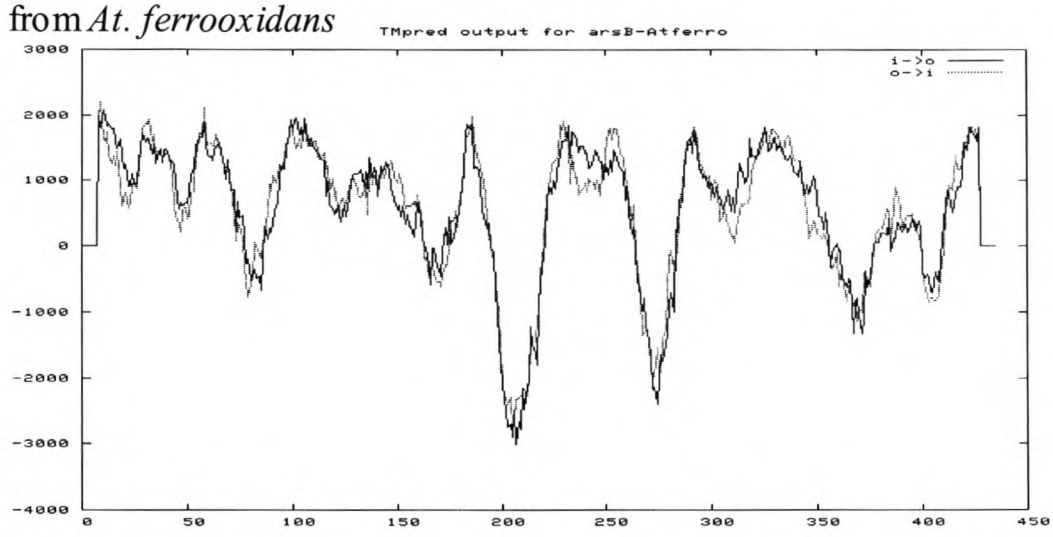
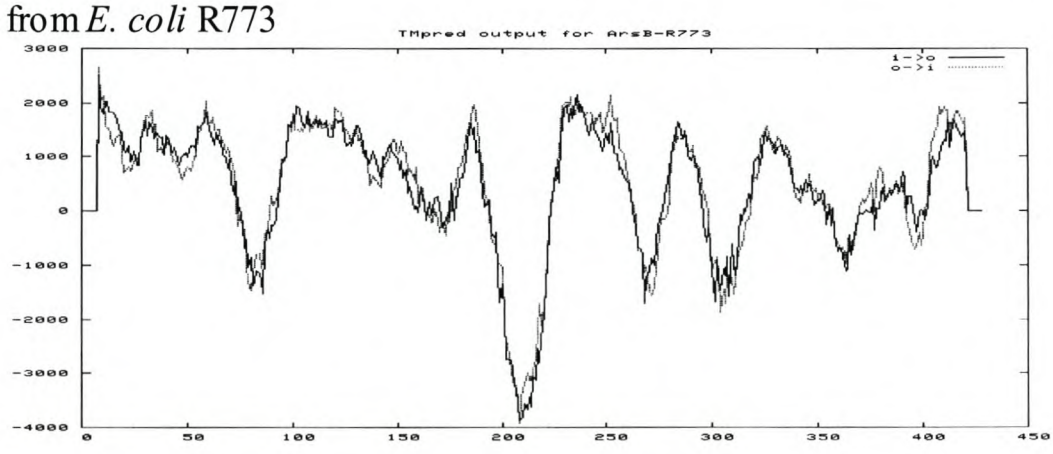


Figure 2.5: Phylogenetic tree of the ArsB proteins from known *ars* operons and putative ArsB homologues. Accession numbers: *S. xylosus* pSX267, Q01255; *S. aureus* pI258, P30329; *B. halodurans*, NP_243865; *B. subtilis ydfA*, NP_388415; *L. monocytogenes*, NP_464825; *L. innocua*, NP_470674; *E. coli* R46, NP_511239; *K. oxytoca* pMH12, AAF89641; *A. multivorum* pKW301, BAA24823; *E. coli* R773, ARB1_ECOLI; *S. typhimurium* R46, BAB91586; *E. coli* chromosome, NP_417959; *Y. entercolitica* pYVe227, NP_052439; *S. marcescens* R478, CAB88405; *Y. pestis*, NP_406810; *P. aeruginosa*, NP_250968; *At. ferrooxidans*, AAF69238; *B. subtilis* skin element, BAA0696; *S. cerevisiae* Acr3p, ACR3_YEAST.

A. ArsB from *At. ferrooxidans*



B. ArsB from *E. coli* R773



C. ArsB from *S. aureus* pI258

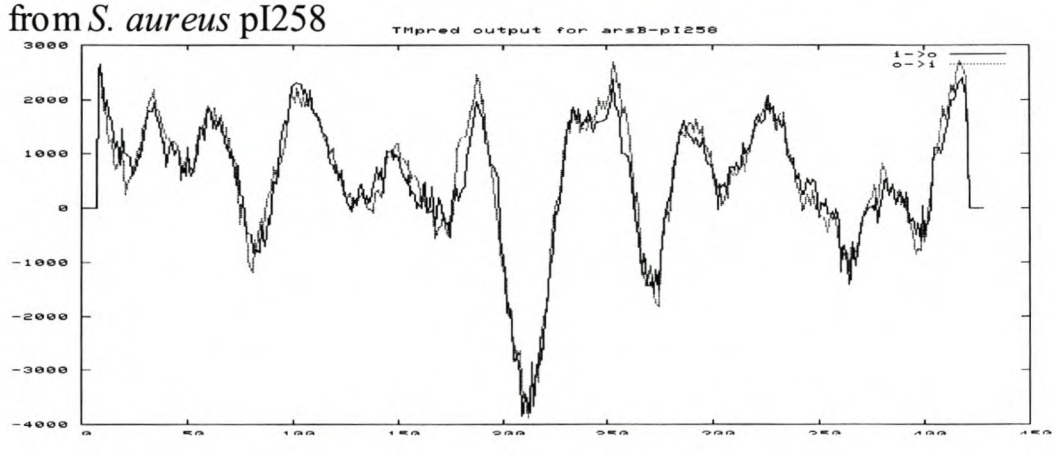


Figure 2.6: Prediction of trans-membrane spanning domains of the ArsB protein from *At. ferrooxidans* by TMpred program (**A**). The predictions under the same conditions for ArsB proteins from *E. coli* R773 (**B**) and *S. aureus* pI258 (**C**) are included for comparison.

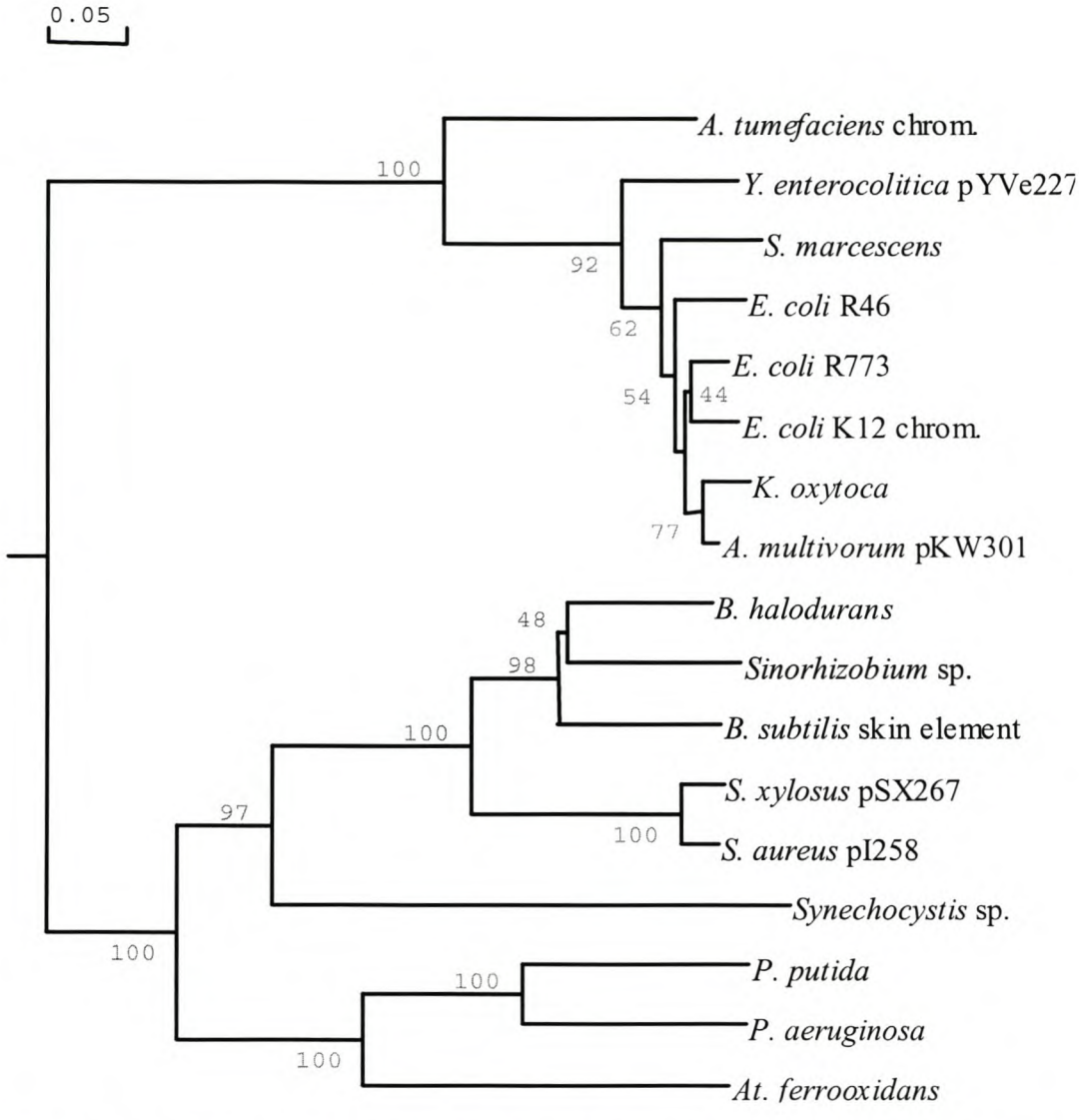
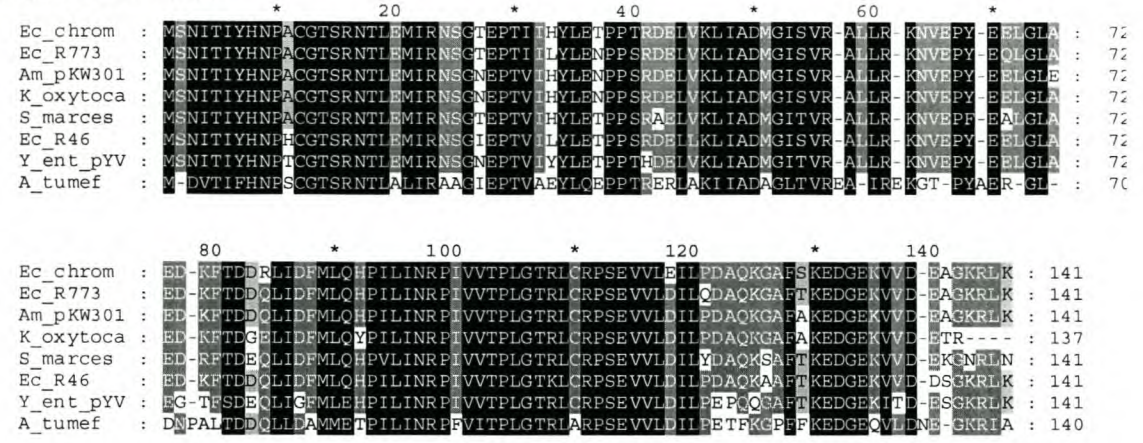


Figure 2.7: Phylogenetic tree of the ArsC protein from *At. ferrooxidans* and other ArsC proteins from known *ars* operons including putative ArsC proteins from the database. Accession numbers: *Agrobacterium tumefaciens* C58, NP_532180; *Yersinia enterocolitica* pYVe227, NP_052438; *Serratia marcescens*, CAB88404; *Escherichia coli* R46, NP_511240; *E. coli* R773, AAA21096; *E. coli* K12 chromosome, NP_417960; *Klebsiella oxytoca*, AF168737_5; *Acidiphilium multivorum* pKW301, BAA24824; *Bacillus halodurans*, NP_243864; *Sinorhizobium* sp., AF178758_3; *Bacillus subtilis* skin element, ARSC_BACSU; *Staphylococcus xylosus* pSX267, ARSC_STAXY; *Staphylococcus aureus* pI258, ARSC_STAAN, *Synechocystis* sp., NP_441727; *Pseudomonas putida*, CAC18654; *Pseudomonas aeruginosa*, NP_250969.

As mentioned, a homologue of *arsC* genes from other bacteria was identified upstream and divergently transcribed from the *arsB* gene. Comparison of the predicted protein from this ORF with other known ArsC proteins and unknown ArsC-homologues in the NCBI database, indicated that the ArsC protein from *At. ferrooxidans* is most closely related to the ArsC proteins from *P. aeruginosa* (e-value of $3e^{-36}$) and *P. putida* (e-value of $5e^{-37}$). All three of these bacteria are unusual in that, while their ArsB proteins group as expected with the ArsB proteins from other Gram-negative bacteria, their ArsC proteins are more closely related to the ArsC proteins from Gram-positive bacteria (Figure 2.7).

The crystal structures of the ArsC proteins from both *S. aureus* and *B. subtilis* have been determined subsequent to the publication of our results (see Chapter 1). It is clear from the crystal structures that the two families of ArsC proteins have different structures and different reaction mechanisms, although they both perform the same function of reducing arsenate to arsenite. As discussed in chapter one the ArsC proteins of the Gram-positive group require glutathione (GSH) and glutaredoxin (Grx) to function, while the proteins of the Gram-negative group require only thioredoxin (Trx). These two groups were named the GSH/Grx clade and the Trx clade in a recent review (Mukhopadhyay, 2002). Alignments of the two groups of ArsC proteins show that the members of the GSH/Grx clade are more closely related to each other than those of the Trx clade (overall sequence identities of 86.64 % and 51.3 %, respectively, between the members within each of the groups shown in Figures 2.7 and 2.8). In fact, the ArsC proteins from the Gram-negative bacteria in the group are the least similar of the ArsC proteins in the Trx clade (Figures 2.7 and 2.8). The predicted reaction scheme for the ArsC protein from *B. subtilis* skin element suggests that the important residues in the reaction are the cysteine residues in positions 10, 82 and 89, an arginine in position 16 and an aspartate residue in position 105 (see Figure 1.7, chapter 1). All of these residues are conserved in *At. ferrooxidans*, except that the second and third cysteine residues are separated by 7 amino acids instead of 6. This is the case for the ArsC proteins from *P. putida* and *P. aeruginosa* as well.

A. GSH/Grx clade



B. Trx clade

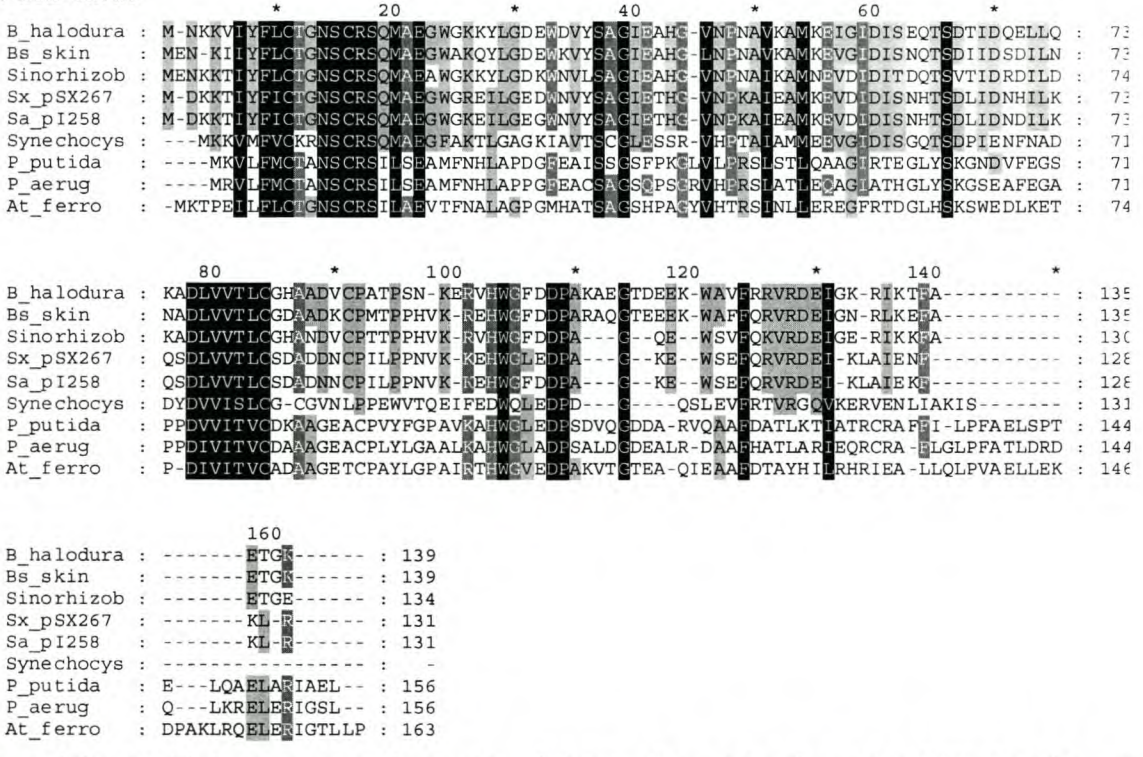


Figure 2.8: Multiple sequence alignments of the ArsC proteins from the GSH/Grx(**A**) and Trx (**B**) clades. The ArsC from *At. ferrooxidans* groups with the Trx clade along with the ArsC proteins from *P. putida* and *P. aeruginosa*. See figure 2.6 for accession numbers of the proteins.

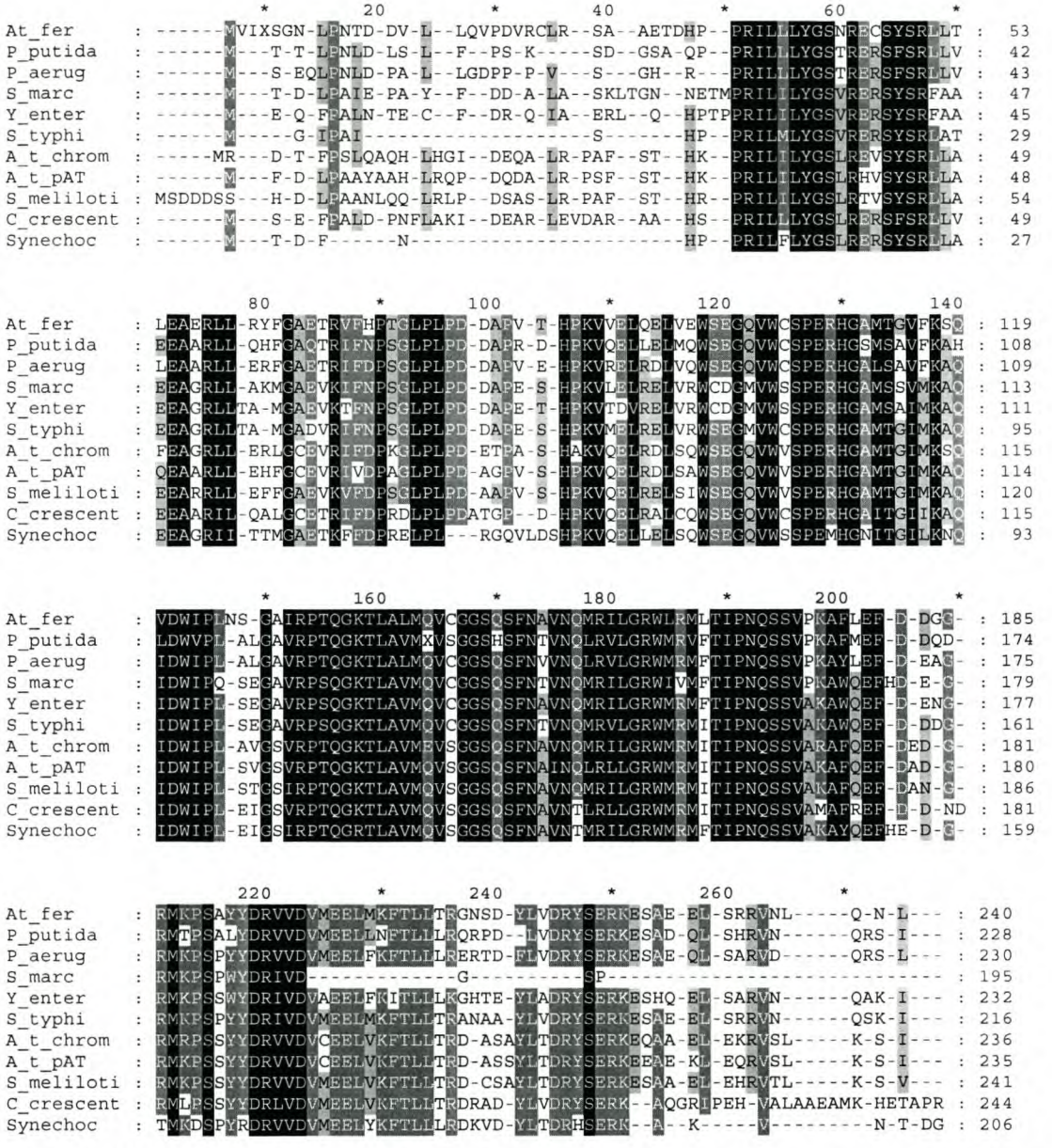
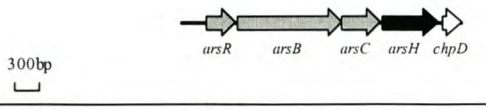
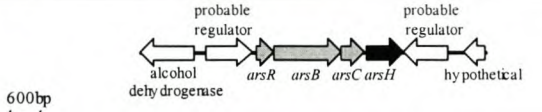
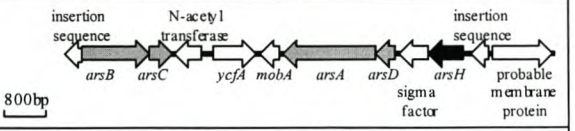
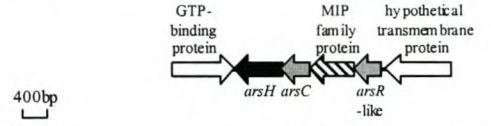
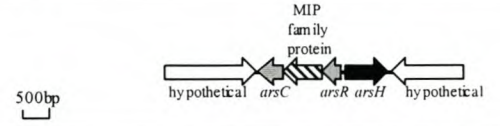
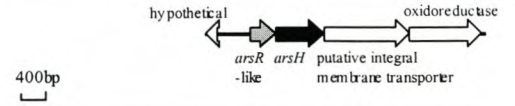
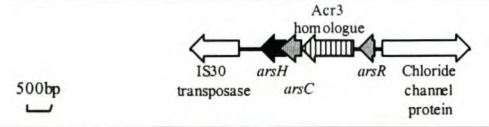
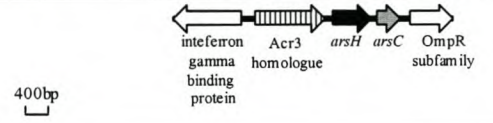
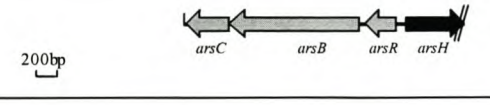
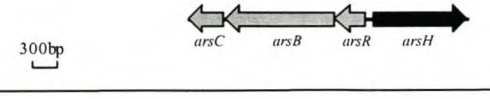


Figure 2.9: Multiple sequence alignment of ArsH proteins and putative ArsH proteins from database (note that the ArsH from *S. marcescens* is incomplete). Accession numbers: *P. putida*, CAC18655; *P. aeruginosa*, NP_250970; *S. marcescens* R748, CAB88407; *Y. enterocolitica* pYVe227, NP_052441, *S. thyphimurium*, BAB91594; *A. tumefaciens* chromosome, NP_532179; *A. tumefaciens* pAT, NP_396255; *S. meliloti*, NP_385180; *C. crescentus*, NP_420317; *Synechocystis* sp., NP_441726.

Downstream of the *arsB* gene is a homologue to a gene previously identified as *arsH*. This gene was first discovered in *Y. enterocolitica* (Neyt *et al.*, 1997) and reported to be essential for arsenic resistance although its function is unknown. More recently a second *arsH* homologue was identified in the *ars* operon of the IncH12 plasmid, R748, from *S. marcescens* (Ryan and Colleran, 2002). In both of these cases the *arsH* gene is divergently transcribed from the *arsRBC* genes and deletion of the divergent *arsH* gene resulted in arsenic sensitive strains. Additional putative proteins from *S. typhimurium*, *P. aeruginosa*, *S. meliloti*, *A. tumefaciens*, *C. cresentus* and *Synechocystis* sp. all showed higher than 67% identity with the predicted *At. ferrooxidans* ArsH protein sequence. An alignment of these proteins shows that there is strong homology between the proteins throughout the whole protein, except for the first 25 to 35 N-terminal amino acids (Figure 2.9).

The arrangement of the *ars* (or putative *ars*) operons in which these *arsH*-like genes have been detected is shown in Table 2.3. All the ArsH homologues identified were found along with other *ars* gene homologues, and all operons coded for either an ArsB homologue or a membrane protein from another family (e.g. major intrinsic protein (MIP) family of proteins). It therefore seems likely that most of these ArsH homologues will have the same function as those from *Y. enterocolitca* and plasmid R748.

Table 2.3: Arrangement of genes surrounding all ArsH homologues identified during a BLAST search with the *At. ferrooxidans* ArsH.

Strain	Accession no. of ArsH homologue	Arrangement of the surrounding genes
Pseudomonas putida	CA18655	
Pseudomonas aeruginosa	NP_250970	
Salmonella typhimurium R64	BAB91594	
Sinorhizobium meliloti	NP_385180	
Caulobacter cresentus	NP_420317	
Agrobacterium tumefaciens pAT	NP_396255	
Agrobacterium tumefaciens chromosome	NP_532179	
Synecocystis sp.	NP_441726	
Serratia marcescens R478	CAB88407	
Yersinia entercolitica pYVe227	NP_052441	

2.3.3. Thioredoxin is required for arsenate reduction by *At. ferrooxidans* ArsC.

As mentioned earlier, a major difference between the ArsC proteins of Gram-positive and Gram-negative bacteria is the source of reducing power for the reduction of arsenate. It has been shown that reduction of arsenate by the Gram-positive ArsC is coupled to thioredoxin (Ji and Silver, 1992a) and reduction by the Gram-negative ArsC is coupled to glutathione (Oden *et al.*, 1994). Since the ArsC of *At. ferrooxidans* was clearly like those of Gram-positive bacteria it should follow that thioredoxin and not glutathione was required for arsenate reduction. To investigate this, *E. coli* strains with mutations in the thioredoxin gene (*trxA*) or both the thioredoxin and γ-glutamylcysteinyl synthetase (*gshA*) gene (responsible for the synthesis of glutathione) were tested for their resistance to arsenate. The *trxA* mutant strain was resistant to arsenate and able to grow on 15mM sodium arsenate, while the double mutant (*trxA* and *gshA*) was sensitive to arsenate and was unable to grown on 2mM sodium arsenate (Table 2.4 and Figure 2.10). This indicated that the glutathione-requiring *E. coli* chromosomal *ars* genes were able to confer resistance to arsenate in the absence of thioredoxin but not the absence of glutathione.

Table 2.4: Arsenate resistance of *E. coli* mutant strains

E. coli strains	Plasmids Transformed			
	none	pTfarsCBH-Cm	pTrx6	pTrx6, pTfarsCBH-Cm
MC1061 (wild type)	+ [a]	+	ND	ND
BH2012(*trxA*)	+	+	ND	ND
BH5262 (*trxA*, *gshA*)	-	-	-	+

[a] + indicates growth on LA + 7mM sodium arsenate , - indicates no growth, ND = not determined.

When the *E. coli* double mutant strain was transformed with the *At. ferrooxidans ars* genes (pTfarsCRBH-Cm) a similar result was obtained because neither the *E. coli* nor the *At. ferrooxidans ars* genes were functional in the double mutant. However, when a plasmid, pTrx6, containing the thioredoxin gene from *At. ferrooxidans*, was added *in trans* together with the *At. ferrooxidans ars* genes, to the *E. coli* (*trxA, gshA*) double mutant, resistance to arsenate was restored (Table 2.4 and Figure 2.10). If the double mutant was transformed with only pTrx6, the cells remained sensitive to arsenate. This result provided genetic evidence that reduction of arsenate by the ArsC of *At. ferrooxidans* is coupled to thioredoxin.

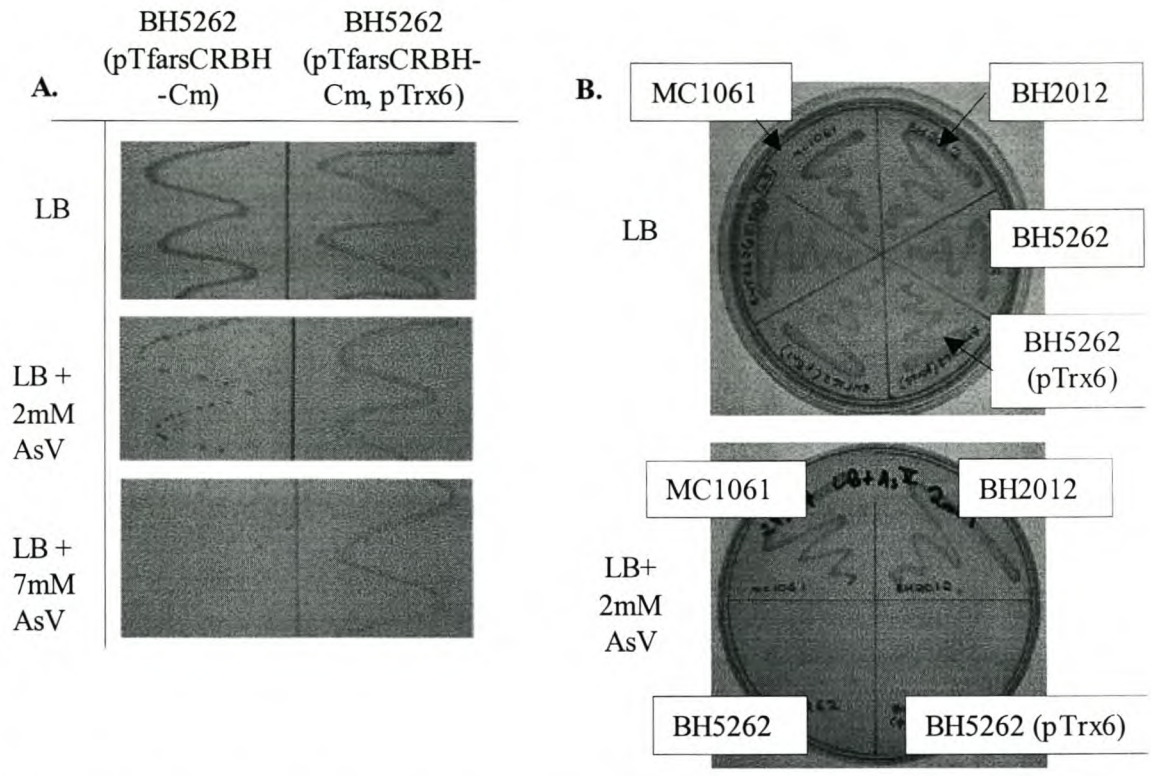


Figure 2.10: Thioredoxin is required for sodium arsenate resistance. **A.** *At. ferrooxidans ars* genes (pTfarsCRBH-Cm) in an *E. coli* thioredoxin and glutathione mutant (BH5262) with or without the thioredoxin gene added in *trans* (pTrx6) on increasing sodium arsenate concentrations. **B.** Control strains on LB and LB with 2mM sodium arsenate, showing that the BH5262 is unable to grow on 2mM sodium arsenate, even with the addition of the *At. ferrooxidans* thioredoxin gene (as the *E. coli ars* genes require glutathione)

2.3.4. Conservation of the unusual *ars* operon structure in other *At. ferrooxidans* strains.

To determine whether the divergent arrangement of the *arsBH* and putative *arsR* and *arsC* genes, found in *At. ferrooxidans* ATCC 33020 was unique to this strain, primers BBARSB and BBARSC (Table 2.1) were designed within the 5′-ends of the *arsB* and *arsC* genes, which allowed for the amplification of a 450 bp region containing the putative *arsR* and any promoter regions (Figure 2.11A). If other strains of *At. ferrooxidans* also have divergent *arsBH* and putative *arsR* and *arsC* genes, the primers would be orientated towards each other and the amplification of a 450 bp fragment would occur. If the genes were not divergently arranged, the primers would face in the same direction and no amplification would occur. A PCR product of the

predicted size was obtained from chromosomal preparations of *At. ferrooxidans* ATCC 33020, ATCC 23270, ATCC 19859 and ATCC 13598, but not from other acidophilic bacteria such as *Leptospirillum ferrooxidans* DSM 2705, *At. thiooxidans* ATCC 19377 and *At. caldus* MNG or from *E. coli* DH5α (Figure 2.11B).

PCR experiments using primers to amplify the *arsH* gene (ARSHF and ARSHR, Table 2.1) gave a product of the predicted size for all *At. ferrooxidans* strains tested, but not for the other bacteria (data not shown). This indicated that other strains of *At. ferrooxidans* also possessed the *arsH* gene.

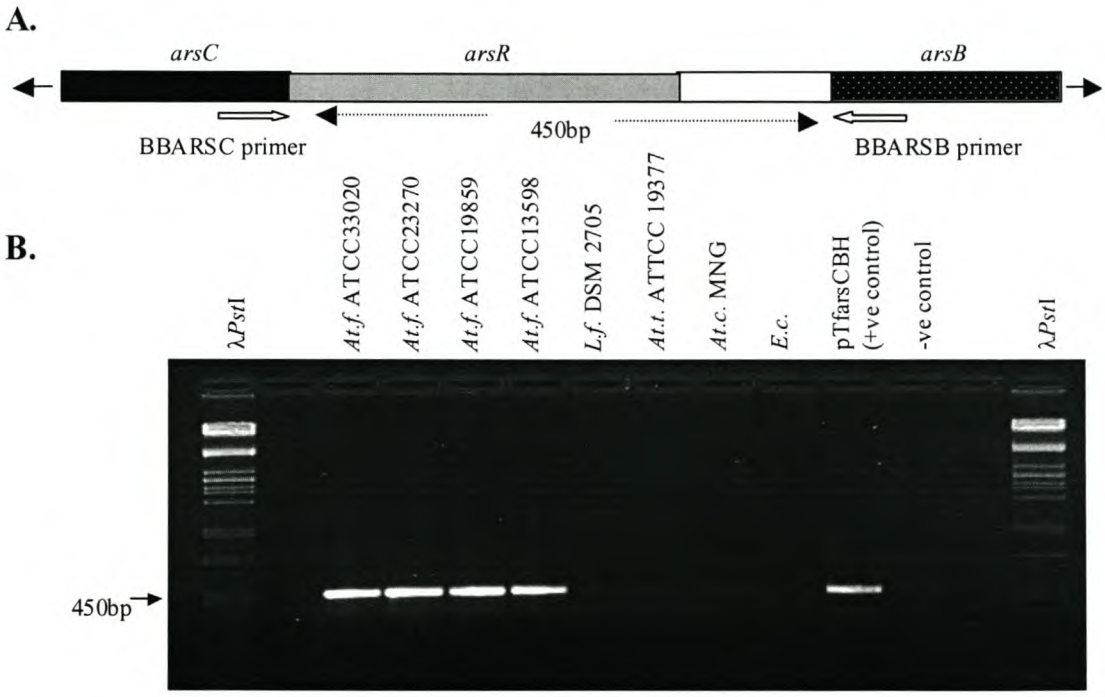


Figure 2.11: **A**. locations of primers used to determine the divergent arrangement of the putative *arsRC* genes and the *arsBH* genes in different strains of *At. ferrooxidans* and other biomining bacteria. Primers located within the *arsB* and *arsC* genes were used to amplify the 450 bp region between these two genes. **B**. Ethidium bromide-stained agarose gel of the PCR amplification products prepared using chromosomal DNA from different biomining bacteria. Abbreviations: *At.f.*, *At. ferrooxidans*; *L.f., L. ferrooxidans; At.c., At. caldus* and *At.t., At. thiooxidans*; *E.c., E. coli.* +ve, positive; -ve, negative.

2.3.5. Expression of *At. ferrooxidans ars* gene products using an *E. coli in vitro* transcription translation system.

Before investigating which of the predicted ORFs described in Table 2 contributed to arsenic resistance in *E. coli*, we examined which polypetides were expressed in an *E. coli in vitro* transcription translation system. Compared with the vector (Figure 2.12 lanes 5 and 10), the pTfars1b clone gave additional polypeptides of approximately 45, 41, 35, 27, 25, 18, 14 and 12 kDa (Figure 2.12 lane 6). The 18 kDa protein was clearly identified as ArsC. This size was in agreement with 18.2 kDa predicted from the sequence, was present only when *arsC* was included in a test construct and was the only polypeptide produced by pTfarsC (not shown). The 27 kDa protein was identified as ArsH as it is close to the predicted size of 26.7 kDa, was synthesised only when an *arsH* gene was present and was the only polypeptide synthesised by pTfarsH (Figure 2.12B lane 9). A broad band at about 35 kDa was the only polypeptide produced by pTfarsB and was always synthesized when *arsB* was present. This protein appears to be smaller than the 48.5 kDa predicted, but membrane-located proteins often run anomalously (San Fransico *et al.*, 1989; Ji and Silver, 1992b; Rosenstein *et al.*, 1992; Cai and DuBow, 1996) and ArsB must be synthesized since all cells containing *arsB* constructs were resistant to arsenite. The 14 kDa band corresponds to the putative ArsR protein as it is the only additional band, compared to the vector, produced by pTfarsR (Figure 2.12 lane 7). Based on the polypeptides produced by pTfars1b and predicted sizes of the ORFs, the 41 and 45 kDa polypeptides are consistent with being the products of ORFs 1 and 5, respectively. The 25 and 12 kDa proteins present in pTfars1b but not pTfarsCRBH are presumed to be synthesised from genes downstream of ORF5.

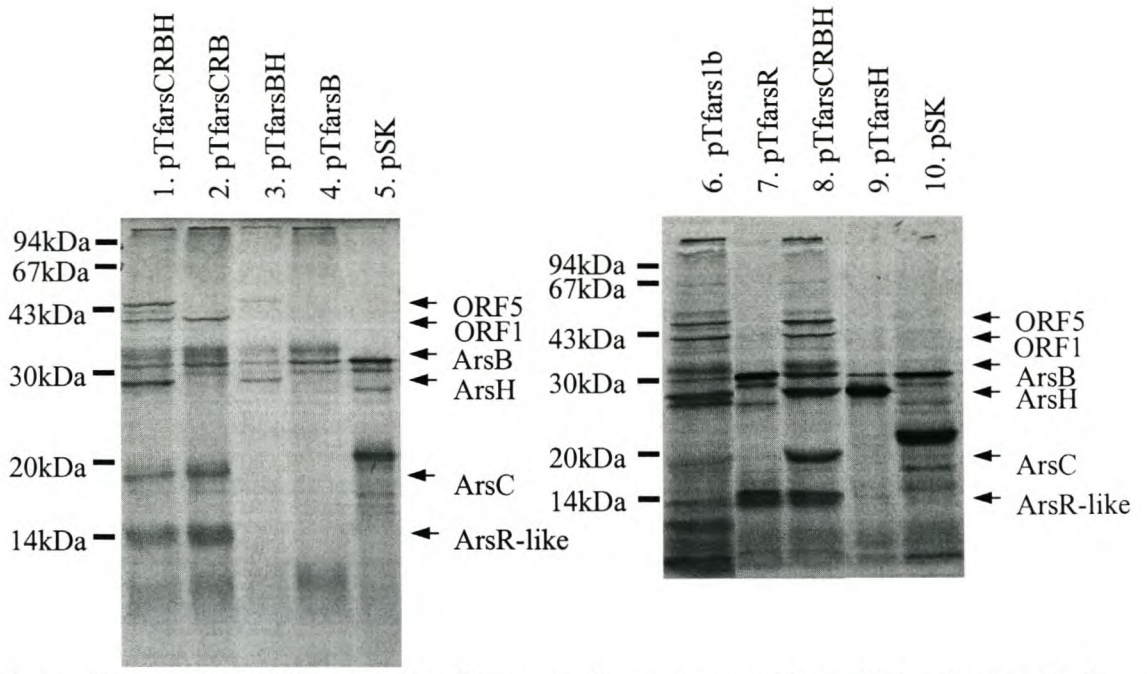


Figure 2.12: SDS-PAGE analysis of the proteins expressed from pTfars1b and subclones using an *E. coli*-derived *in vitro* transcription-translation system (Promega).

2.3.6. Ability of the cloned *At. ferrooxidans ars* gene products to confer increased resistance to arsenic compounds and antimonite in *E. coli* AW3110.

Constructs pTfars1a, pTfars1b and pTfarsCRBH conferred equal levels of resistance to arsenite (As^{III}) and arsenate (As^{V}) in *E. coli* AW3110 (Figure 2.13).

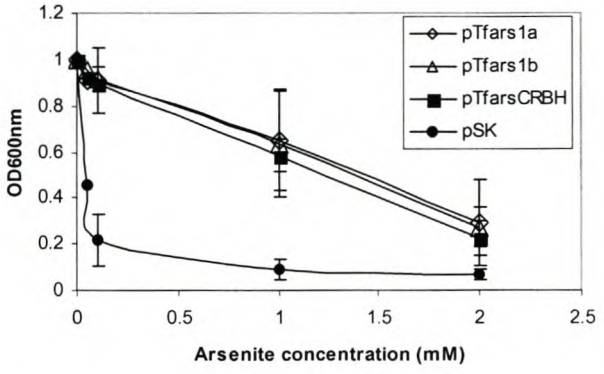


Figure 2.13: Comparison of level of arsenite resistance conferred by pTfars1a (◇), pTfars1b (△) and pTfarsCRBH (■). The values were normalised to OD600 of 1 at 0 mM Arsenite. These results represent the average of 3 experiments each performed in triplicate.

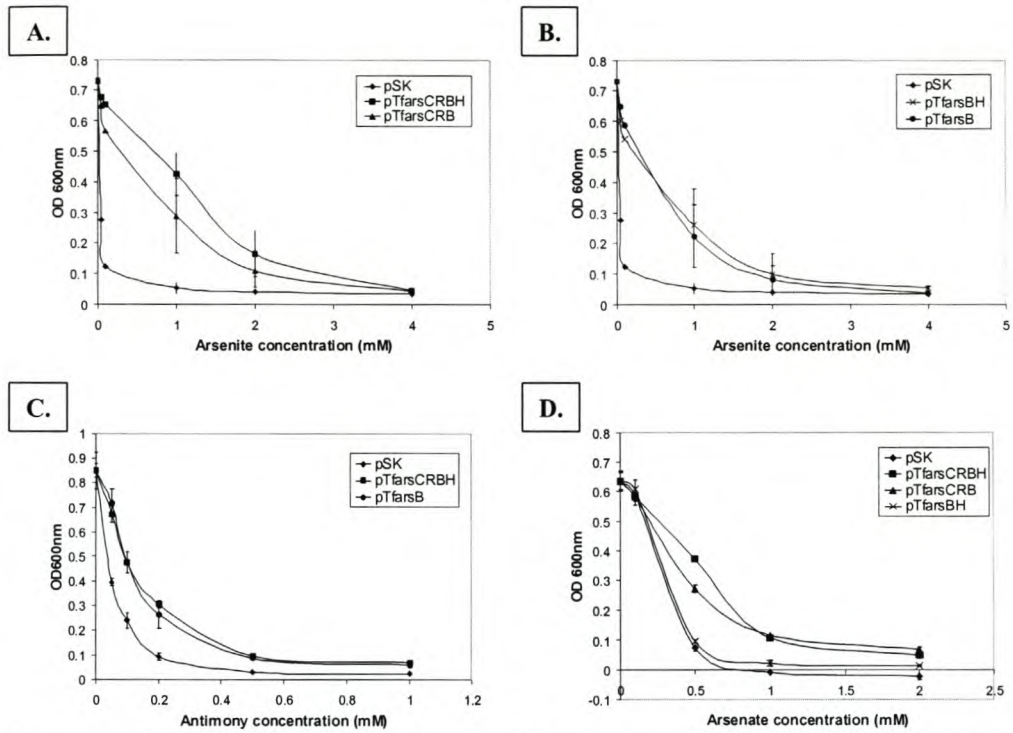


Figure 2.14: Growth of *E. coli* AW3110 containing pSK (♦), pTfarsCRBH(■), pTfarsCRB(▲), pTfarsBH (×) and pTfarsB (•) in arsenite (**A** and **B**), antimony (**C**) and arsenate (**D**).

The ability of pTfarsCRBH and subclones to confer resistance to arsenite in *E. coli* AW3110 was tested further as shown in Figures 2.14A and 2.14B. A construct containing only the *arsB* gene (pTfarsB) conferred similar levels of resistance to arsenite as pTfarsCRBH, pTfarsCRB or pTFarsBH. This experiment was repeated four times and although the results varied between experiments no clear evidence that arsenite resistance in *E. coli* was improved by the presence of *arsC* or *arsH* was obtained. It was clear that only *arsB* was required for resistance to arsenite in *E. coli*. The cloned *At. ferrooxidans arsB* was required to also enhance the resistance of *E. coli* to antimony and this resistance was not further enhanced by the presence of *arsC* or *arsH* (Figure 2.14C). As expected, both the *arsC* and *arsB* genes were essential for increased resistance of *E. coli* to arsenate and this resistance was not increased by the presence of *arsH* (Figure 2.14D).

Out of interest we compared the level of arsenite resistance conferred to the *E. coli ars* mutant by the *ars* genes from *At. ferrooxidans* on the high copy plasmid (pTfarsCRBH) with that conferred by the *E. coli* chromosomal and R773 *ars* genes on

high copy plasmids (Figure 2.15). Our results agree with those reported by Carlin *et al.* (1995). *E. coli* W3110, the wild-type strain, shows a low level of arsenite resistance, while AW3110 (the arsenic sensitive mutant created in the above study) is sensitive to arsenic, even at 0.1 mM sodium arsenite. When the *E. coli* chromosomal arsenic resistance genes cloned in the vector pUC19 (pWPS1) are added to the arsenic mutant AW3110, a moderate level of arsenite resistance is achieved. However, the addition of the *arsRDBC* genes from R773 in pBR322, give a high level of arsenic resistance (pBC101). The addition of the *At. ferrooxidans ars* genes on pBluescript gave similar levels of arsenite resistance to those observed with pWSP1 (the *E. coli* chromosomal genes on a vector with the same ColE1 origin of replication). This indicates that at least in *E. coli* the *At. ferrooxidans ars* genes are as effective as those from the *E. coli* chromosome in conferring arsenite resistance to *E. coli*.

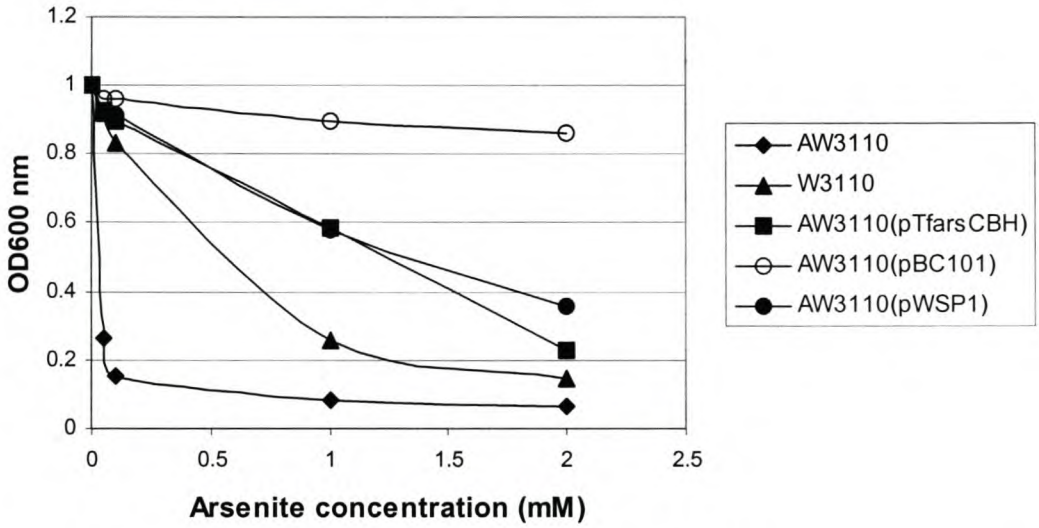


Figure 2.15: Comparison of the cloned *ars* genes from *At. ferrooxidans* (pTfarsCRBH, ■) with those from the *E. coli* chromosome (pWSP1, ●) and resistance plasmid, R773 (pBC101, ○). Also included are the wild-type *E.coli*, W3110 (◆) and the arsenic sensitive mutant AW3110 (▲). The results are normalised to an OD600 of 1 at 0 mM arsenite. These results are the average of one experiment performed in triplicate.

2.4.DISCUSSION

During the biooxidation of arsenopyrite ores and concentrates, large quantities of arsenic are released into the surrounding solution. As a member of the consortium of

bacteria involved in arsenopyrite biooxidation it was expected that *At. ferrooxidans* should possess arsenic resistance genes. What was not expected was the unusual divergent arrangement of the *arsB* and *arsC* genes within the *At. ferrooxidans ars* operon. No *arsD* gene nor an *arsA*-like gene (ATPase subunit) were found in the immediate vicinity of the arsenic resistance genes and only *arsC, arsB* and *arsH*-like genes were identified based on initial sequence comparisons. A tblastn search of the incomplete genome of *At. ferrooxidans* strain ATTC23270 (available at http://www.tigr.org) also did not reveal putative ArsA or ArsD homologues indicating that it is most probable that these genes do not exist elsewhere on the genome. More careful analysis resulted in the identification of a putative regulator between the *arsB* and *arsC* genes. However, the predicted protein has only relatively weak homology to the ArsR protein of known *ars* operons. The protein also lacks the conserved metal-binding box to which the arsenite inducer binds causing a conformational change in the helix-turn-helix domain that results in dissociation of the repressor from the DNA (Shi *et al.*, 1994). Further investigation into the function of this putative regulator was carried out and is discussed in chapter three.

The orientation of the *arsRC* and *arsBH* genes indicated that the genes must be divergently transcribed. While the divergent arrangement of an *arsH* gene relative to the other *ars* genes, has been described previously, there has been no report of divergently transcribed *arsB* and *arsC* genes. When the incomplete sequence of *At. ferrooxidans* ATCC23270 was made available it was confirmed that this strain had the same unusual arrangement of the *ars* genes (contig 10040:a_ferrooxidans shows an overall identity of 99% at the DNA level). Furthermore, this divergent gene arrangement was conserved for other *At. ferrooxidans* strains that originated from Canada, the USA and Japan but was not found in *L. ferrooxidans*, *At. thiooxidans* or *At. caldus*. In such a divergent arrangement, the potential exists for the transcription of *arsC* and *arsB* genes to be regulated separately. This possibility has been addressed in chapter three.

An open reading frame with homology to the *arsH* gene was located immediately downstream of the *arsB* gene. The *arsH* gene was first identified in *Y. enterocolitica*, where it is divergently transcribed from the *arsRBC* genes (Neyt *et al.*, 1997), and a second homologue located in the same orientation was recently described in plasmid

R748 from *S. marcescens* (Ryan and Colleran, 2002). In *Y. enterocolitca* the *arsH* gene was shown to be necessary for arsenic resistance, as a deletion of this gene resulted in a strain that was sensitive to arsenic and the reintroduction of the *arsH* gene in *trans* restored the level of resistance to wild-type levels. While similar experiments were not performed in the case of R748, the authors found that a clone containing *arsRBC* and a truncated *arsH* gene showed lower levels of resistance to arsenic and the deletion of the *arsH* gene did result in a loss of arsenic resistance to the host bacterium (*Serratia marcesens*). The actual function of the gene is unknown and both sets of authors hypothesized that it might act as some type of regulator. The finding that the *At. ferrooxidans arsH*–like gene was expressed in an *E. coli in-vitro* transcription translation system but was not required for resistance to arsenite, arsenate or antimony in *E. coli* is in apparent contrast to the arsenic resistance genes present on the *Y. enterocolitica* pYV virulence plasmid. The high level of homology between the ArsH from *At. ferrooxidans* and those from *Y. enterocolitica* and plasmid R748 would suggest that the function of these genes might be the same. However, the effect of the pYV *arsH* was studied in *Y. enterocolitica* and it appears that the R748 arsenic resistance experiments were also performed in their host bacterium *Serratia marcesens*, while the effect of the *At. ferrooxidans arsH* was studied in a heterologous *E. coli* host. It is possible that *arsH* has an affect on arsenic resistance in *At. ferrooxidans* that is not observed when cloned in *E. coli.* This might be due to a physiological characteristic of the strain, or it could be that the function is performed by another gene in the *E. coli* genome. (Neither of the previous *arsH* studies mentions if they have studied the requirement for *arsH* in *E. coli.*) The presence of an *arsH*-like gene appears to be a feature of *At. ferrooxidans* as it was detected in all four strains examined. As mentioned, a number of unstudied genes also showed homology to *arsH*. When the genes surrounding these putative-*arsH* genes were analysed it was found that all the *arsH* homologues were in the vicinity of other *ars* genes (Table 2.3). None of these bacteria contained *ars* genes arranged in the manner of *At. ferrooxidans*, and there is only one case where the *arsH* gene is downstream of a putative *arsB* (*Synechocystis*). However, this *arsB* homologue groups with the ACR3 family of arsenite transporters. There is, therefore, no clear ancestor of the *At. ferrooxidans ars* gene arrangement at present. Analysis of these unstudied operons shows that all the *arsH* homologues identified so far are found in bacteria of the gamma and alpha groups of proteobacteria, except for one example from

cyanobacteria (*Synechocystis* sp.). This could be related to the evolution of this gene in the *ars* operon or possibly have to do with the particular function of ArsH.

As an obligately chemolithotrophic, acidophilic bacterium, *At. ferrooxidans* comes from an environment which could be predicted to be genetically more isolated than most other bacteria in which *ars* genes have been studied. It was, therefore, interesting to compare the predicted amino acid sequences of the *At. ferrooxidans* ArsB and ArsC proteins with those of other bacteria. The *At. ferrooxidans* ArsB is clearly grouped with the ArsB proteins of other Gram-negative bacteria, however it is also the most divergent, being separated from the others by the longest branch length (Figure 2.5). This protein did confer arsenite resistance to the *E. coli* mutant (Figure 2.14A) and these levels were similar to the levels of arsenite resistance conferred by the cloned *E. coli ars* genes (Figure 2.15). In contrast to the ArsB proteins, the ArsC proteins of *At. ferrooxidans* and both *P. aeruginosa* and *P. putida* are more similar to ArsC proteins from Gram-positive bacteria. Genetic evidence that ArsC of *At. ferrooxidans* is biochemically like the thioredoxin-requiring 'Gram-positive type' ArsC proteins supports this grouping (Table 2.4). A glutaredoxin-like gene from *At. ferrooxidans* had previously been identified, but addition of a plasmid containing this gene (pTTn1) was not able to restore arsenate resistance (results not shown) to the *E. coli* thioredoxin and glutathione mutant containing the *At. ferrooxidans ars* genes. The predicted amino acid sequence of the *At. ferrooxidans* ArsC was most closely related to the ArsC of *P. aeruginosa*. However, the *P. aeruginosa ars* operon did not confer increased arsenate resistance on an *E. coli* wild-type strain (Cai *et al.*, 1998), and the authors suggest that this may be due to suboptimal functioning of the Gram-positive-type ArsC in the Gram-negative host. The cloned *arsC* of *At. ferrooxidans* was expressed and functional in the heterologous host, but levels of arsenate resistance conferred to *E. coli* by the *At. ferrooxidans ars* genes are lower than those conferred by *E. coli* genes themselves (The cloned chromosomal *ars* genes from *E. coli* conferred resistance to at least 2 mM sodium arsenate (Carlin *et al.*, 1995), while the *At. ferrooxidans ars* genes only conferred resistance to about 0.5 mM sodium arsenate). While growth curves in sodium arsenate were not generated for the growth of *E. coli* BH5262(pTfarsCRBH-Cm, pTrx6) there is an indication that when the *At. ferrooxidans* thioredoxin gene was provided the levels of arsenate resistance conferred on the *E. coli* strain were higher (up to 15 mM on LB plates and up to 2

mM on low-phosphate plates). This would suggest that the *At. ferrooxidans* ArsC is not able to make efficient use of the *E. coli* thioredoxin. It is clear that at least two groups of ArsC proteins exist, a four-conserved-cysteine thioredoxin-requiring ArsC and a two-conserved-cysteine glutathione/glutaredoxin-requiring ArsC (Silver and Phung, 1996). Since it is clear that there are now at least three Gram-negative bacteria that have a four-conserved-cysteine-thioredoxin-requiring ArsC, the correlation of this type of ArsC with Gram-positive bacteria, which appeared to be a valid assumption from the initial several arsenic resistance genes examined, appears to be an oversimplification.

Like other workers we found ArsB protein difficult to detect unequivocally in an *E. coli in vitro* system (Suzuki *et al.*, 1998). Nevertheless, the *arsB* and *arsC* genes from *At. ferrooxidans* were clearly expressed and functional in *E. coli.* As found with other *ars* operons, the *At. ferrooxidans arsB* gene on its own was able to confer resistance to arsenite and antimony, but *arsC* was required in addition to *arsB* for arsenate resistance (Fig. 2.14). The ability of the *At. ferrooxidans ars* system to function in *E. coli* is noteworthy when it is considered how arsenic resistance systems may be energised. In arsenic resistance systems that lack the ArsA (ATPase) it is believed that membrane potential rather than ATP hydrolysis serves as the energy source (Ji and Silver, 1992b, Dey and Rosen, 1995, Tsai *et al.*, 1997). The obligately acidophilic *At. ferrooxidans* has an internal pH of about 6.5 and when growing on Fe^{2+} at pH 2.0, the cells maintain a ΔpH of 4.5 units (Cox *et al.*, 1979). Unlike other bacteria, acidophilic bacteria with a large ΔpH have a positive-inside rather than negative-inside membrane potential that subtracts from the H^+ gradient instead of augmenting it (Matin, 1990). Nevertheless, in spite of originating from a bacterium having a positive-inside membrane potential, the *At. ferrooxidans* ArsB arsenite efflux pump membrane protein was functional in *E. coli*.

This initial investigation of the *At. ferrooxidans* arsenic resistance system has been carried out in *E. coli.* Future work will have to include studies of arsenic resistance and gene expression in *At. ferrooxidans* which, because of the rudimentary genetic system available, lack of mutants and difficulties of readily obtaining large quantities of cells, presents a considerable challenge.

Chapter Three: The divergent chromosomal *ars* operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein.*

TABLE OF CONTENTS

* These results have been published (Butcher and Rawlings, 2002).

3.1.INTRODUCTION

Acidithiobacillus ferrooxidans has been shown to contain genes for arsenic resistance on its chromosome (Butcher *et al.*, 2000). Arsenic resistance has been found on both the chromosomes and plasmids of various bacteria and has been extensively reviewed (Mukhopadhyay *et al*, 2002; Rosen, 1999; Xu *et al.*, 1998; Tsai *et al.*, 1997; Silver and Phung, 1996 and Cervantes *et al.*, 1994). These arsenic resistance genes are usually arranged in an operon, the two most common forms of which contain either three (*arsRBC*) or five (*arsRDABC)* genes. The ArsR and ArsD proteins are *trans*-acting regulators. ArsR is a repressor controlling the basal level of expression in response to the presence of arsenite, while ArsD controls the upper level of expression of the operon. ArsA is an ATPase that forms a complex with the ArsB transmembrane protein to pump arsenite out of the cell. Another gene, *arsH*, was discovered in *Yersinia enterocolitica* (Neyt *et al.*, 1997) and more recently in plasmid R478 isolated from *Serratia marcesens* (Ryan and Colleran, 2002). The function of this gene is unknown, but it was shown to be required for resistance to arsenite and arsenate in *Y. enterocolitica*.

The arsenic resistance operon from *At. ferrooxidans* was found to consist of two divergent elements, *arsRC* and *arsBH*. These genes were shown to confer arsenic resistance to an *E. coli ars* mutant AW3110 (Butcher *et al.*, 2000). The *arsH* gene was however not required for arsenic resistance in *Escherichia coli* and it is not known whether it is required in *At. ferrooxidans*. This unusual divergent arrangement raised questions about the regulation of this operon. The finding that the *arsB* and *arsC* genes are located on divergent elements, may mean that they are independently regulated. The putative ArsR from *At. ferrooxidans* had only weak homology to ArsR proteins from well-studied *ars* operons and did not contain the conserved metal-binding box, ELCVCDL, to which the arsenite inducer binds (Shi *et al.,* 1994 and Shi *et al.,* 1996). It was, therefore, uncertain whether the putative ArsR which is present within the *At. ferrooxidans* operon was able to regulate these genes in response to arsenite.

Here we describe an investigation into the regulation of the chromosomal *ars* operon of *At. ferrooxidans* and into whether the ArsR-like protein is a regulator of the operon in response to arsenite, arsenate and antimonite.

3.2.METHODS

Bacterial strains, plasmids, primers and media. The strains, plasmids and primers used in this study are shown in Table 3.1. *E. coli* strains were grown on Luria-Bertani medium (Sambrook *et al.*, 1989) at 30°C. *At. ferrooxidans* strains were grown in tetrathionate medium (Powles *et al.*, 1995) at 30°C. Ampicillin (100μg/ml), chloramphenicol (20μg/ml), tetracycline (20μg/ml) and kanamycin (30μg/ml) were used when indicated.

DNA techniques and analysis. Plasmid preparation, restriction endonuclease digestion, gel electrophoresis and cloning techniques were carried out by standard methods (Sambrook *et al.*, 1989). Searches for sequences related to ArsR were performed using the gapped-BLAST program of the National Center of Biotechnology (NCBI) at http://www.ncbi.nih.nlm.gov (Altshul *et al.*, 1997). The PC-based DNAMAN software (version 4.1) from Lynnon BioSoft was used to create the sequence alignments and homology trees. Prediction of protein secondary structure was performed using the web-based PSIPRED v2.4 (http://bioinf.cs.ucl.ac.uk/psipred/).

Construction of the arsenic sensitive *lac*⁻ *E. coli* strain. The generalised transduction method using phage $P1_{vir}$ (Miller, 1972) was used to construct this strain. Phage $P1_{vir}$ was isolated from *E. coli* AW3110 (Carlin *et al.*, 1995) and used to transduce *E. coli* CSH50I^{q}. The transduced cells were selected on chloramphenicol and X-gal plates. They were then checked for sensitivity to arsenic on 0.5 mM sodium arsenite plates. This indicated that the Δ*ars*::*cam* from *E. coli* AW3110 had replaced the *ars* genes from *E. coli* CSH50I^{q}, resulting in a new strain, *E. coli* ACSH50I^{q}.

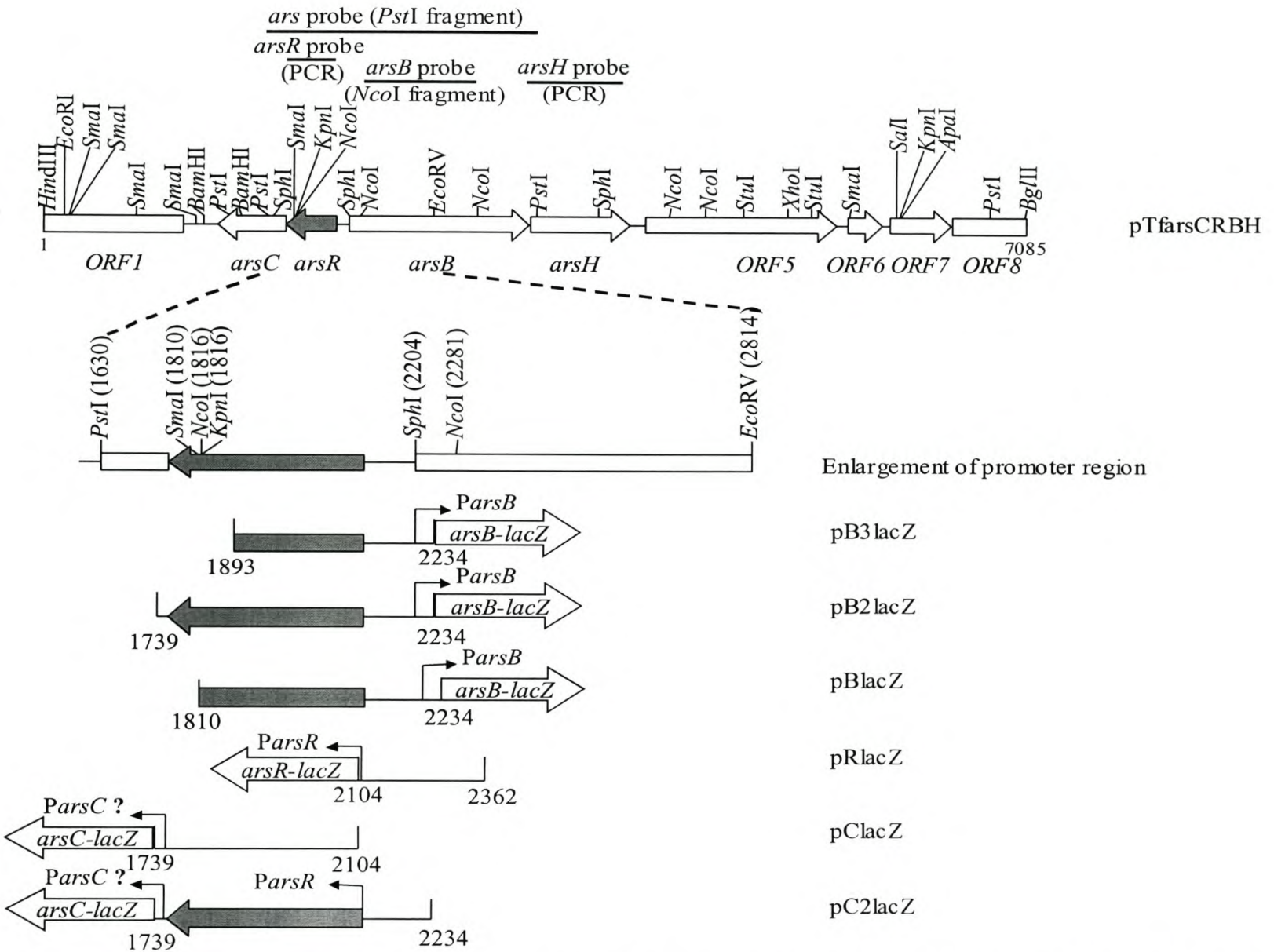


Figure 3.1. Arrangement of the *At. ferrooxidans* arsenic resistance genes (accession number AF173880) showing the restriction endonuclease map, open reading frames and translational fusions of putative *ars* promoter regions (P_{arsB}, P_{arsR}, P_{arsC}) to the *lacZ* reporter gene. The *lacZ* gene is shown by the open arrow, the *arsR* gene is shaded grey, and the direction of the promoters is indicated by small arrows. The positions of the probes used in the northern blot experiments are shown above with the method used to isolate the probes shown in brackets.

Construction of the promoter-*lacZ* reporter constructs. The putative promoter regions were amplified by the PCR using the following primers (listed in Table 3.1): Promoter for *arsB* (BBARSB and BLACZE) and *arsR* promoter (RLACZB and RLACZE). These PCR products were digested with *Bam*HI and *Eco*RI (these sites had been included in the primers) and ligated to the promoterless *lacZ* gene of pMC1403 digested with the same enzymes, resulting in pB3lacZ and pRlacZ, which are translational fusions of *arsB* and *arsR* with the *lacZ* gene respectively (Figure 3.1). Primers BBARSB and BBARSC were used to amplify a region containing the *arsB* promoter and the intact *arsR* gene (Figure 3.2). This PCR product was digested with *Bam*HI and cloned into pMC1403 digested with *Bam*HI. The product could be cloned in either direction. Cloning in one direction resulted in pB2lacZ in which the *arsB* promoter was fused to the *lacZ* gene of pMC1403 and the intact *arsR* gene is transcribed in the opposite direction (Figure 3.1). With the insert in the other direction the *lacZ* was fused in frame at the start of the *arsC* gene, resulting in pC2lacZ (Figure 3.1). In both constructs, the *arsR* promoter and the intact *arsR* were also included. The above PCR product was also digested with *Bam*HI and *Sma*I and cloned into pMC1403 digested with the same enzymes. The resultant construct, pBlacZ, contains a translational fusion of *arsB* with the *lacZ* gene and a partly truncated *arsR* gene in *cis* (Figure 3.1).

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

Strain, plasmid, primers and probes	Description	Reference or source[a]
Strains		
Escherichia coli		
W3110	K-12 F⁻ IN(*rrnD-rrnE*)	Rosen
AW3110	W3110 Δ*ars*::*cam*	Carlin *et al.* (1995)
CSH50I^q	*rspL* Δ(*lac-pro*) [F′ *traD36 proAB lacI^q* ΔM15	Smith and Rawlings (1998)
ACSH50I^q	CSH50I^q Δ*ars*::*cam*	This work
Acidithiobacillus ferrooxidans		
ATCC 33020	Wild type	ATCC
Plasmids		
pBluescript(SK)	Amp^R *lacZ′*ColE1 replicon	Stratagene
pMC1403	Amp^R promoterless *lacZYA* operon ColE1 replicon	Casadaban *et al.* (1983)
pKK223-3	Amp^R *tac* promoter ColE1 replicon	Pharmacia Biotech
pGL10	Kan^R RK2/RP4 replicon	Toukdarian
pACYC184	Tet^R Cm^R p15A replicon	Chang *et al.* (1978)
pTfarsCRBH-ACYC	Tet^R (5.3 kb *Hin*dIII-*Stu*I fragment containing *ars* genes cloned into pBluescript cut with *Eco*RV-*Hin*dIII, this was	This study[b]

	then cut with *Eco*RI and cloned into pACYC cut with *Eco*RI)	
pTfarsCRB-ACYC	Tet^R (The *Eco*RV-*Xba*I fragment of pTfars1b, containing the 5′ end of *arsB* and the whole of *arsH*, was replaced with the *Eco*RV-*Xba*I fragment of pBB08, containing only the 5′ end of *arsH*. This was then cloned into pACYC)	This study[b]
pTfarsBH-ACYC	Tet^R (*Kpn*I deletion of pTfarsCBH, cut with *Kpn*I, blunted and cut with *Eco*RI. This was then cloned into pACYC cut with *Sca*I and *Eco*RI)	This study[b]
pTfarsB-ACYC	Tet^R (*Pst*I deletion of pTfarsBH. This was then cut with *Kpn*I, blunted, cut with *Bam*HI and cloned into pACYC cut with *Sca*I and *Bcl*I)	This study[b]
pTfarsC-ACYC	Tet^R (PCR product containing *arsC* cloned into pBluescript. This was then cut with *Hin*dIII, blunted, cut with *Eco*RI and the fragment containing *arsC* cloned into pACYC cut with *Sca*I and *Eco*RI))	This study[b]
pTacArsR-ACYC	Tet^R (PCR product from arsR-eco and arsR-hind primers cloned into pKK223-3 cut with *Eco*RI and *Hin*dIII. P_{tac}-*arsR* was then excised with *Bam*HI and *Sca*I and cloned into pACYC cut with *Bcl*I and *Sca*I)	This study
pTacArsH-ACYC	Tet^R(PCR product from ARSHF and ARSHR primers cut with *Eco*RI and *Hin*dIII cloned into pKK223-3 cut with the same restriction enzymes. P_{tac}-*arsH* was then excised with *Bam*HI and *Sca*I and cloned into pACYC cut with *Bcl*I and *Sca*I)	This study
pBlacZ	Amp^R (arsB-lacZ fusion, 100 aa of ArsR, see Fig. 1 and text)	This study
pB2lacZ	Amp^R (*arsB-lacZ* fusion, complete ArsR *in cis*, see Fig. 1 and text)	This study
pB2lacZ-GL10	Kan^R (as above but in low copy number pGL10 vector)	This study
pB3lacZ	Amp^R (*arsB-lacZ* fusion, 71 aa of ArsR, see Fig. 1 and text)	This study
pB3lacZ-GL10	Kan^R (as above but in low copy number pGL10 vector)	This study
pRlacZ	Amp^R (*arsR-lacZ* fusion, see Fig. 1 and text)	This study
pRlacC-GL10	Kan^R (as above but in low copy number pGL10 vector)	This study
pClacZ	Amp^R (*arsC-lacZ* fusion, *arsR in cis* but no *arsR* promoter)	This study
pClacZ-GL10	Kan^R (as above but in low copy number pGL10 vector)	This study
pC2lacZ	Amp^R (*arsC-lacZ* fusion, *arsR in cis* with *arsR* promoter)	This study
pC2lacZ-GL10	Kan^R (as above but in low copy number pGL10 vector)	This study
Primers		
BBARSB	(*Bam*HI) 5′-GCGGATCCAGGGTGACGAGAAATATGGC-3′	Butcher *et al.* (2000)
BBARSC	(*Bam*HI) 5′-GCGGATCCGGGGTTTTCATCACTGG-3′	Butcher *et al.* (2000)
BLACZE	(*Eco*RI) 5′-GGGAATTCCTTCGCGCTGCACGGTGA-3′	This study
RLACZB	(*Bam*HI) 5′-CGGGATCCATCGTATTATCCACG-3′	This study
RLACZE	(*Eco*RI) 5′-CGGAATTCCCAGACAATGTGCCA-3′	This study
arsR-hind	(*Hin*d III) 5′-ACGAAGCTTGGGGTTTTCATCACTGG-3′	This study
arsR-eco	(*Eco*RI) 5′-GCGAATTCATGGAACCACTACAAGACCC-3′	This study
ARSHF	(*Eco*RI) 5′-GCGAATTCTGGTGGCTGCCGCTGGCTTG-3′	Butcher *et al.* (2000)
ARSHR	(*Hin*dIII) 5′-GAAAGCTTGCGTACCCCCAACCTCATGCC3′	Butcher *et al.* (2000)
Probes for Northern blots		
ars	1.9kb *Pst*I fragment from pTfarsCRB	This study
arsB	0.85kb *Nco*I fragment from pTfarsCRBH	This study
arsR	PCR product using ars-eco and ars-hind primers	This study
arsH	PCR product using the ARSHF and ARSHR primers	This study

[a] Rosen, B. Rosen, Wayne State University, Detroit, Mich.; Toukdarian, A. Toukdarian, University of California, San Diego, Ca.

[b] the first cloning step was carried out in Butcher *et al.*, 2000 (see Chapter 2).

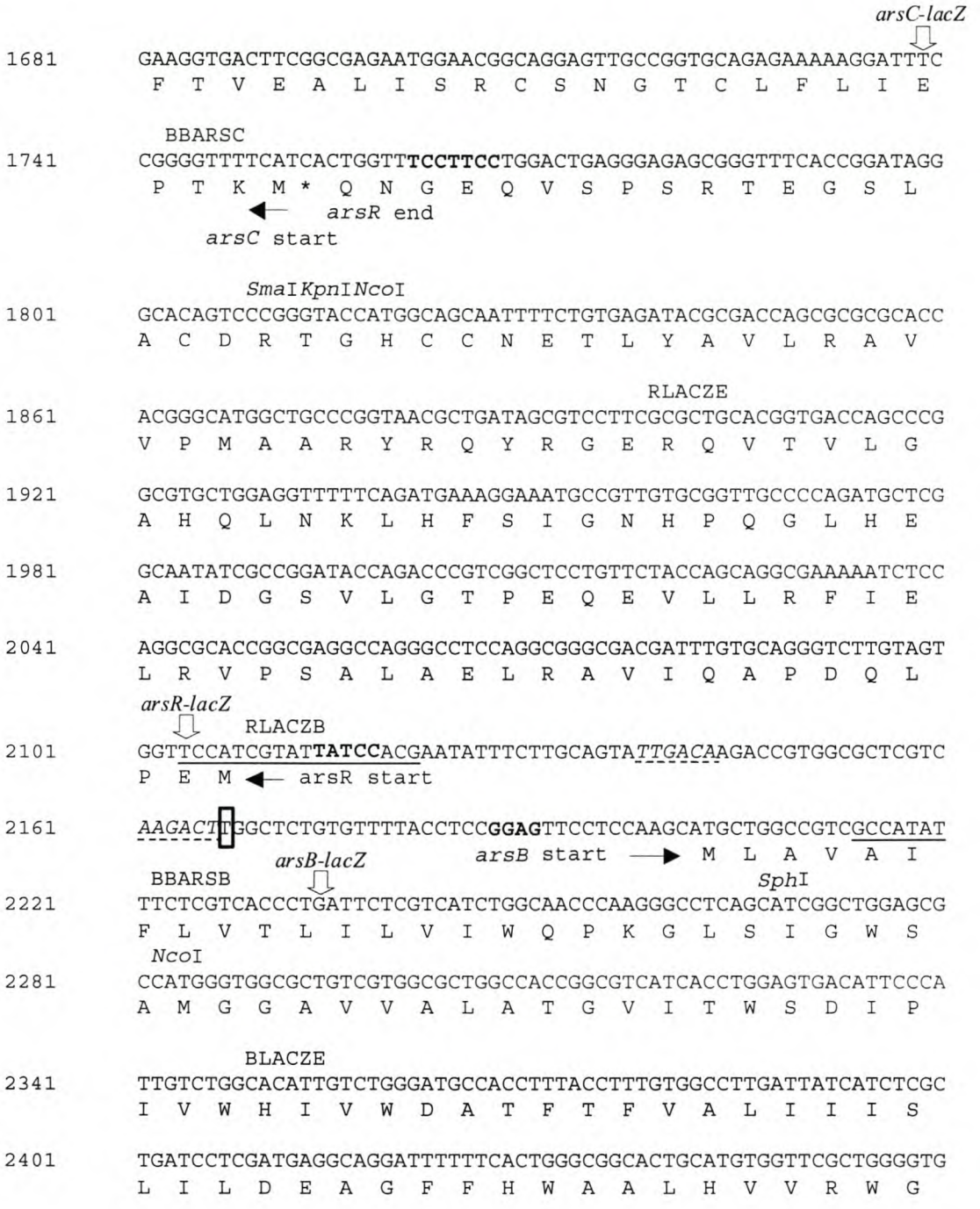


Figure 3.2. Sequence of the *arsR* gene and the 93 bp intergenic region, showing the location of primers used to create the *ars-lacZ* fusions (underlined) and the positions of the translational-fusions (open arrows). Putative ribosome binding sites are marked in bold, while a broken line underlines a putative –10 and –35 sequence for *arsB* that has a good consensus for a sigma 70 promoter. The T residue (blocked) was shown in a primer extension experiment to be the start of the *arsBH* transcript. The numbering is according to the *At. ferrooxidans ars* sequence, Genbank AF173880.

The translational fusions from pB3lacZ, pRlacZ, pB2lacZ and pC2lacZ were cloned into the low copy number, RK2 replicon-based vector, pGL10. The fusions were obtained by digesting the construct with *Sal*I, blunting this site with T4 DNA polymerase (Roche Molecular Biochemicals), and then digesting with *Eco*RI. The fragment containing the fusion was then cloned into pGL10 digested with *Sma*I and *Eco*RI. The constructs pB3lacZ-GL10, pRlacZ-GL10, pB2lacZ-GL10 and pC2lacZ-GL10 were constructed in this way. The *ars* genes were also cloned into the relatively low copy number, ColE1 replicon compatible vector, pACYC184, to enable expression in *trans* with the promoter-fusion constructs (Table 3.1). Correct construction of all *lacZ* fusions was confirmed by DNA sequencing.

β-Galactosidase assays: Overnight cultures were diluted 1:100 into fresh medium containing the appropriate antibiotics. These were grown at 30°C until an OD_{600} of between 0.3 and 0.4 was obtained. Either 25 μM sodium arsenite, sodium arsenate, or potassium antimonite was added as an inducer and the cultures were grown for a further hour after which the β-galactosidase activity of the cultures was measured using the method of Miller, (1972).

Isolation of total RNA. RNA was isolated from *E. coli* ACSH50I^q strains carrying the pTfarsCRBH-ACYC plasmid, which contains the complete *ars* operon from *At. ferrooxidans*, using the yeast total RNA isolation protocol of Liang and Pretorius, 1992. The cells were grown overnight in LB and diluted 1:100 into fresh media containing 25 μM sodium arsenite or 50 μM sodium arsenate. These were then grown at 30°C for a further 4hrs, before RNA was isolated from the cultures. RNA was also isolated from *At. ferrooxidans* ATCC 33020. *At. ferrooxidans* was first grown without arsenic in 800 ml tetrathionate medium. These cells were resuspended in 1.5 ml acid water and used to inoculate cultures containing 25 μM sodium arsenite, 25 μM sodium arsenate or no arsenic. These were then incubated at 30°C overnight and used to extract RNA.

Northern blot analysis. 10 μg RNA from each sample was separated on a 1% denaturing agarose/formaldehyde gel. The gel was soaked in 20 X SSC for 1h and

then transferred to Hybond-N nylon membrane under capillary blotting in the presence of 20 X SSC. The required probes (Figure 3.1 and Table 3.1) were labelled with ^{32}P-dATP using the Random Primed DNA labelling kit from Roche Molecular Biochemicals and hybridised to the RNA overnight at 60°C in hybridisation buffer (7% SDS, 1% BSA, 1 mM EDTA and 0.25 M Na_2HPO_4). The membrane was washed in a 1 X SSC and 0.1% SDS solution and then in a 0.1 X SSC and 0.1% SDS solution, before exposing the membrane to X-ray film to detect the bound probe.

3.3.RESULTS

3.3.1. Comparison of ArsR sequences.

The predicted amino acid sequence of the putative ArsR from *At. ferrooxidans* was compared with sequences in the NCBI database using the BLAST search tool and there were no proteins with known functions that showed strong matches. The best matches were to unstudied genes from chromosomal sequencing projects, with the closest match (E-value of $9e^{-18}$) being to a hypothetical transcriptional regulator from *Sinorhizobium meliloti* (NP_385183). Alignments of *At. ferrooxidans* ArsR with the ten most similar matches from the BLAST search and with other known ArsR proteins were performed (Figure 3.3). Of the ten closest matches from the database, only one protein (NP_486806; *Nostoc* sp. PCC7120) contained the conserved cysteine residues in the metal-binding domain (ELCVCDL) (Shi *et al.*, 1996). There are, however, a number of residues that are conserved throughout all the proteins including the known ArsR proteins. These are grouped in two regions, one before and one after the metal-binding motif previously identified in the known ArsR proteins. Besides a totally conserved arginine, serine and leucine, other conserved residues are mainly hydrophobic.

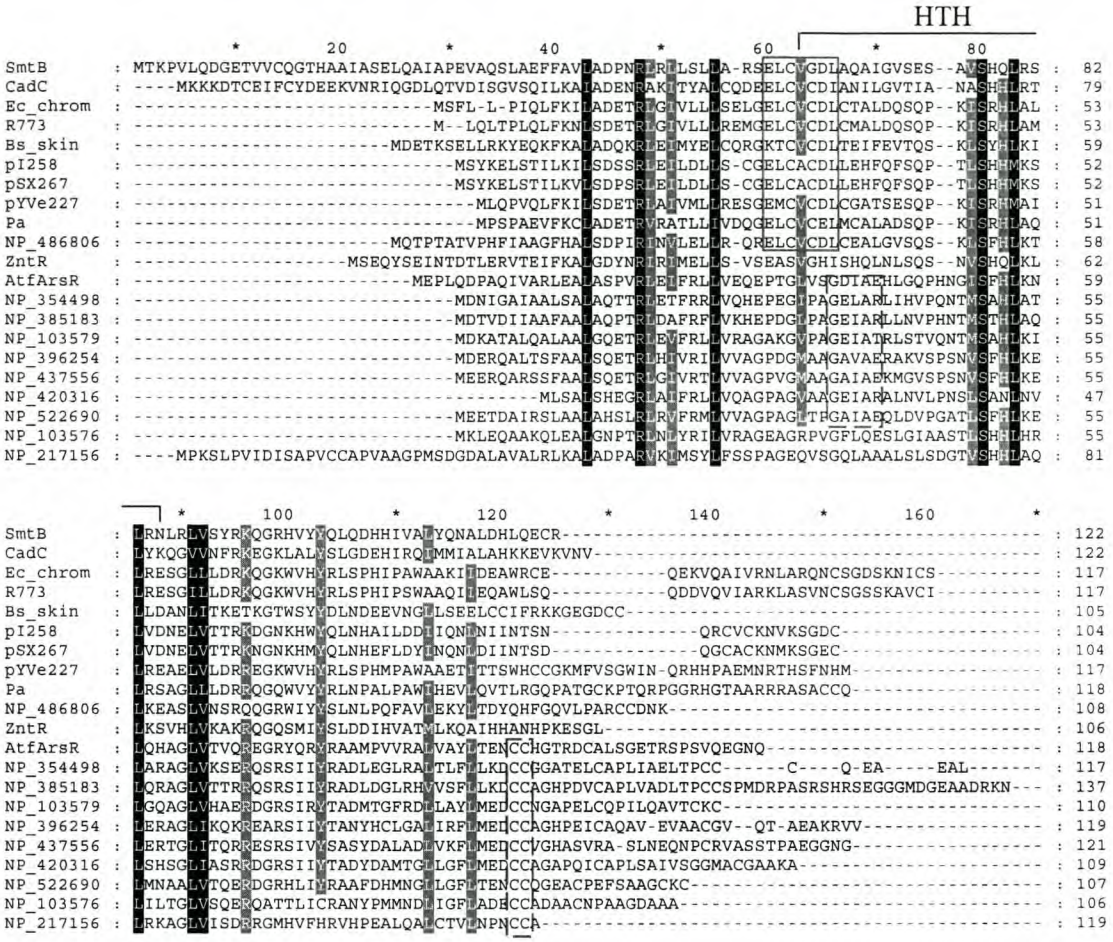


Figure 3.3. Multiple sequence alignment of a selection of functionally characterised ArsR proteins compared with the ArsR from *At. ferrooxidans* and the nine closest matches from the BLAST search. Also included is SmtB from *Synechococcus* PCC7942 (accession number S31197), ZntR from *Staphylococcus aureus* (AAC32484) and CadC from *S. aureus* pI258 (B32561). Known ArsR proteins are labelled as follows: pI258, *S. auerus* pI258 (accession number AAA25636); pSX267, *S. xylosus* (AAA27587); Ec chrom, *Escherichia coli* chromosome (AAC76526); R773, *E. coli* R773 (CAA34168); pYVe267, *Yersinia enterocolitica* pYVe267 (AAD16860); Bs skin, *Bacillus subtilis* skin element (BAA06967); Pa, *Pseudomonas aeruginosa* chromosome (AAC69642). The ten closest matches are labelled by their accession numbers: NP_486806, *Nostoc* sp. PCC7120; NP_358183, *Sinorhizobium meliloti*; NP_522690, *Ralstonia solanacearum*; NP_103579, *Mesorhizobium loti*; NP_354498, *Agrobacterium tumefaciens* chromosome; NP_396254, *Agrobacterium tumefaciens* pAT; NP_437556, *Sinorhizobium meliloti*; NP_420316, *Caulobacter crescentus*; NP_103576, *M. loti*; NP_486806, *Nostoc* sp. PCC7120. Black shading highlights amino acids with 100% conservation and grey shading highlights those with greater than 70% conservation across all proteins. For the comparison, amino acids assumed to be functionally equivalent were I/L/V; R/K and G/A. The metal-binding motif found in the characterised ArsR proteins has been boxed using solid red lines, while the regions conserved between the proteins of the new ArsR subgroup are boxed with broken blue lines. The helix-turn-helix (HTH) motif marked corresponds to that in SmtB and the known ArsR proteins. The alignments were performed using the Clustal W-based optimal alignment tool of DNAMAN.

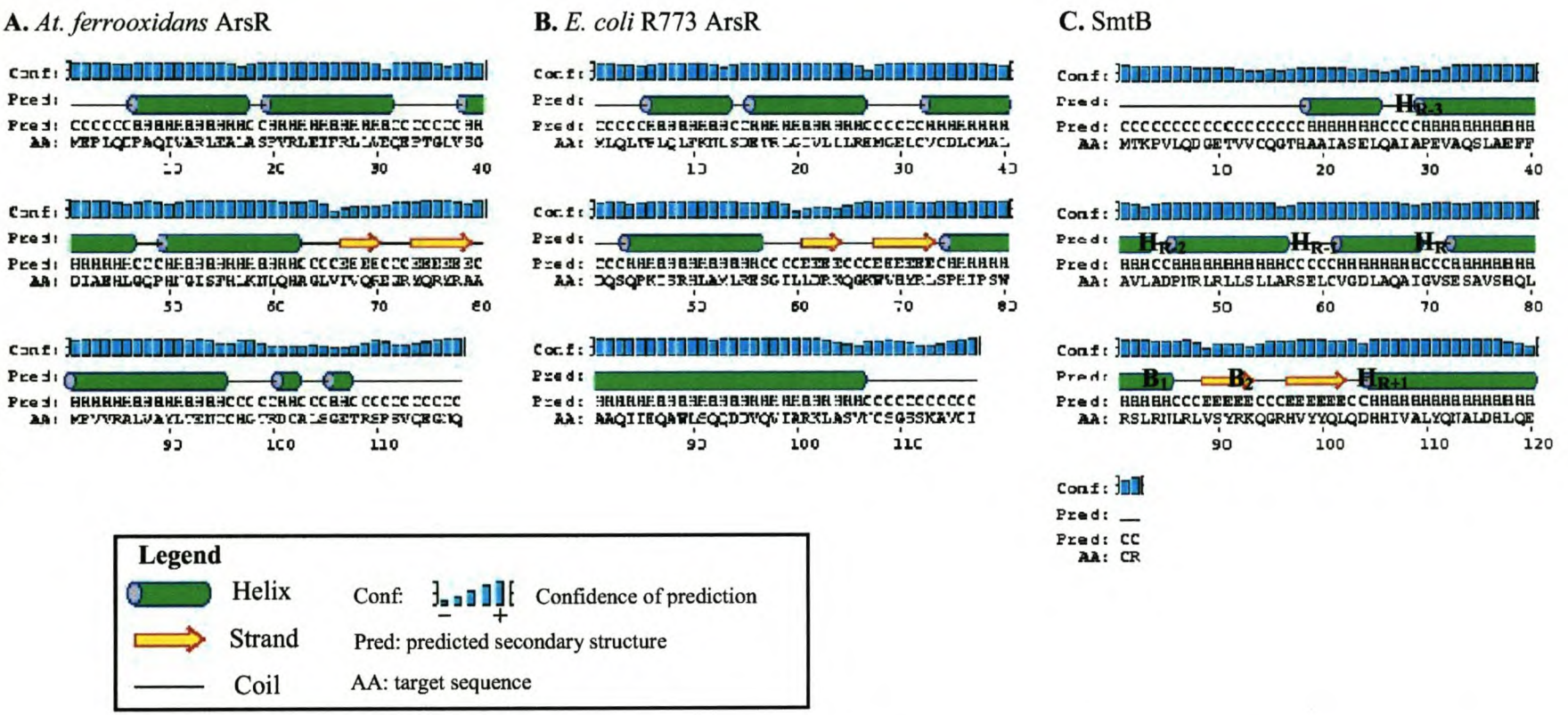


Figure 3.4: Secondary structure prediction of the *At. ferrooxidans* ArsR performed by the web-based program PSIPRED v2.4. The *E. coli* R773 ArsR and SmtB were included for comparison. The structure predicted for SmtB compares well with that previously reported and the helices and β-strands have been labeled as in Cook *et al* (1998) relative to the putative DNA recognition helix (H_R).

Several observations can be made from the multiple sequence alignment. Firstly, while ArsR from *At. ferrooxidans* and its closest matches do not contain the putative metal-binding sequence (ELCVCDL), they do contain a conserved sequence among themselves (GX(L/I)A) situated immediately downstream of the region corresponding to the metal-binding motif of the already studied ArsR proteins. Secondly, all proteins that lack the putative metal-binding sequence appear to have two extra residues before the second helix of the DNA-binding HTH motif. Lastly, all the atypical ArsR-like proteins contain a cysteine doublet towards their C-terminals. When the secondary structure of the *At. ferrooxidans* ArsR was predicted and this prediction compared with a proven ArsR regulator (from *E. coli* R773 *ars* operon) and with SmtB (a member of the ArsR family of regulators for which the crystal structure has been determined (Cook *et al.*, 1998)) it can been seen that all three of these proteins have a similar secondary structure consisting of a least 5 helices and two β-strands.

The *At. ferrooxidans* ArsR protein and nine of the ten closest matches cluster as a second subgroup of ArsR regulators (Figure 3.5). Analysis of the open reading frames up- and downstream of these putative atypical ArsR-like regulators indicated that all, except the proteins NP_217156 (*M. tuberculosis*) and NP_103576 (*M. loti*), located adjacent to a putative membrane binding protein and often an ArsC and ArsH homologue (Table 3.2).

Although all of the members of this new ArsR subgroup do share conserved regions among themselves and with SmtB and other known ArsR regulators, in the absence of a metal-binding domain it was important to discover whether the atypical *At. ferrooxidans* ArsR protein was able to regulate the arsenic resistance genes in response to arsenic.

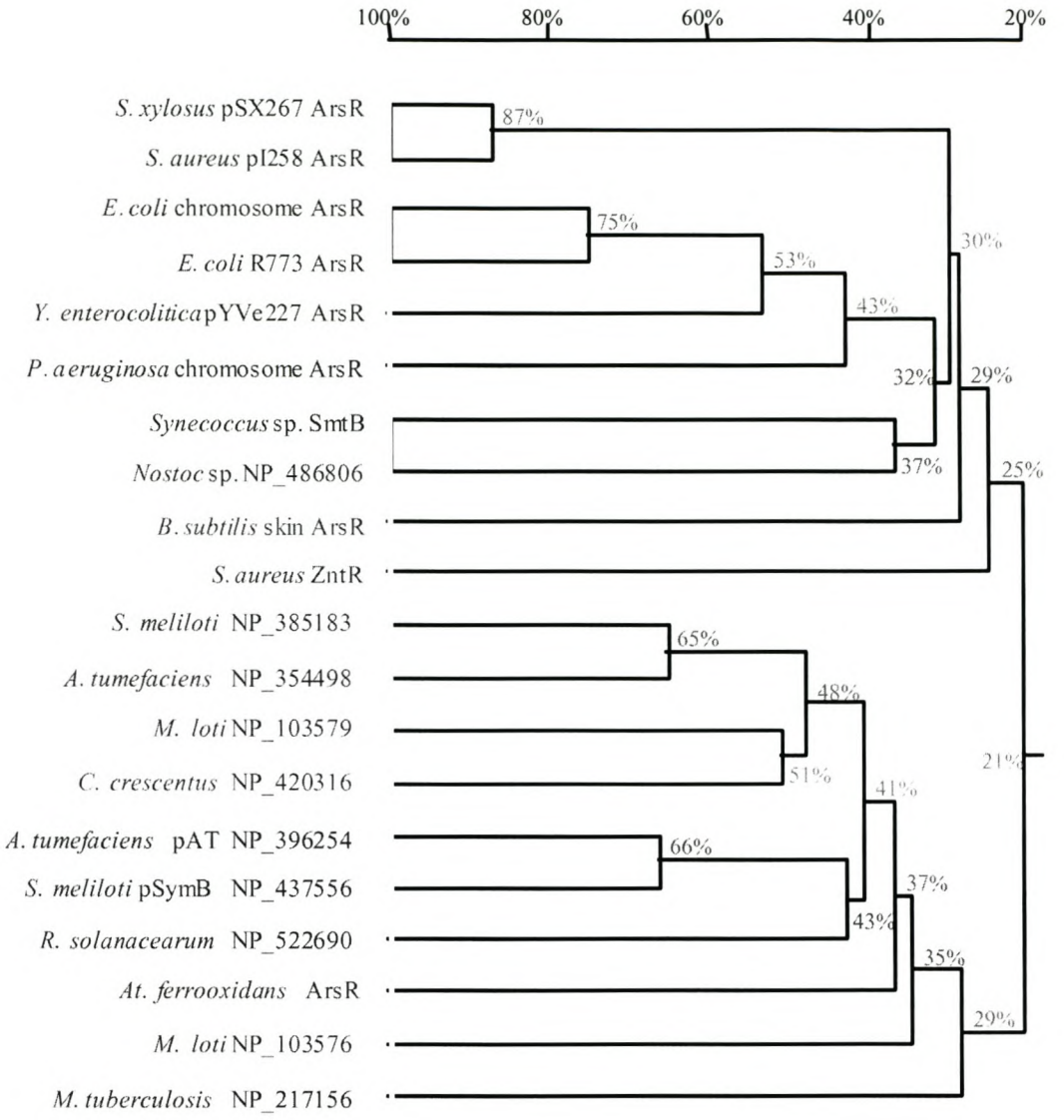
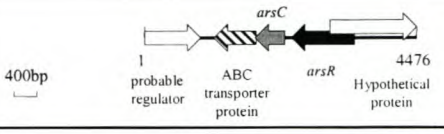
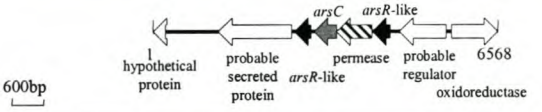
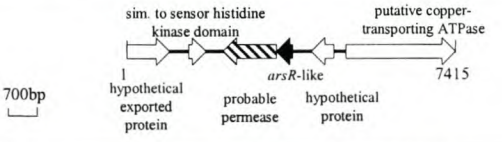
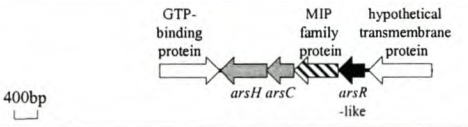
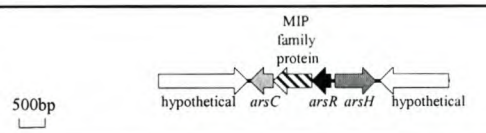
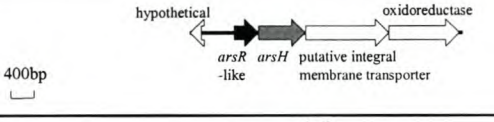
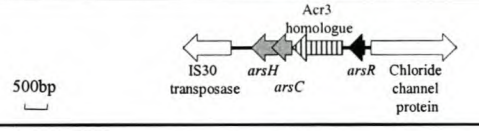
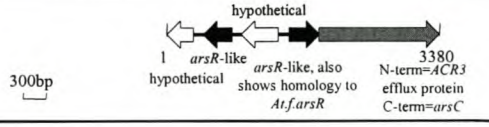


Figure 3.5: Dendogram of the same proteins as aligned in Figure 3.3 showing the possible grouping of *At. ferrooxidans* ArsR in a separate subclass of ArsR regulators. Values represent the percentage amino acid identity between proteins. (see Figure 3.3 for accession numbers).

Table 3.2 :Arrangement of ORFs surrounding the subgroup of *At. ferrooxidans* ArsR-like proteins.

Strain	Accession no. of ArsR homologue	Arrangement of the surrounding genes
Ralstonia solanacearum	NP_522690	
Mezorhizobium loti	NP_103579 and NP_103576	
Sinorhizobium meliloti megaplasmid pSymB	NP_437556	
Sinorhizobium meliloti	NP_385183	
Caulobacter cresentus	NP_420316	
Agrobacterium tumefaciens pAT	NP_396234	
Agrobacterium tumefaciens chromosome	NP_354498	
Mycobacterium tuberculosis	NP_217156	

3.3.2. The *arsBH* and *arsRC* promoter studies.

The arsenic sensitive mutant *E. coli* strain AW3110 was wild-type for *lac* and we therefore constructed *E. coli* strain ACSH50I^q that is both arsenic sensitive and *lac*$^-$. This would ensure that there was no host-cell background β-galactosidase activity and that there were no *E. coli* chromosomal *ars* genes present that may interact with the promoter-fusion constructs. This strain was used for all work presented here. This strain also contains the *lacI*^{q} on the F′ plasmid. This would ensure that any genes that

were added in *trans* and controlled by the P_{tac} promoter, were repressed in the absence of IPTG.

Although we had difficulty obtaining primer extension products, one primer extension experiment in the direction of the *arsBH* genes, gave a start signal corresponding to a T, 35 bp upstream of the ATG codon (Figure 3.2). This is only 1 bp downstream of an *E. coli* σ^{70}-like promoter sequence (TTGACA-N17-AAGACT). No transcription start signal could be detected in the direction of the *arsRC* genes and no recognisable promoter was identified from the sequence. To study the expression of the *ars* genes from the putative promoters we designed a number of translational gene fusions. The use of reporter genes to study gene expression should ideally be done in a homologous host, with a single gene copy integrated into the chromosome. However, this is not possible with *At. ferrooxidans.* To avoid many integration experiments into a heterologous *E. coli* host, all the *ars*-reporter gene fusions were transferred into the low copy number vector, pGL10. Translational fusions with the *arsB* or *arsC* genes were constructed where an intact *arsR* gene was included in *cis* (pB2lacZ-GL10 and pC2lacZ-GL10). Additional *arsB-lacZ* fusions in which 100 amino acids (pBlacZ) or 71 amino acids (pB3lacZ and pB3lacZ-GL10) of ArsR are included in *cis* were also constructed (Figure 3.1). These and other fusions are shown in Figure 3.1.

3.3.3. Expression of *arsBH.*

When the *arsB-lacZ* fusion constructs were expressed in *E. coli* ACSH50I^{q} in the absence of arsenic, β-galactosidase activity of approximately 19 and 75 units was obtained when the complete ArsR (pB2lacZ-GL10) and 71 amino acids of ArsR (pB3lacZ-GL10) were present, respectively (Figure 3.6). This indicated that in the absence of arsenic, the presence of an intact ArsR repressed the *arsBH* genes approximately 4-fold. With the addition of either 25 μM sodium arsenite or sodium arsenate this repression was relieved 3 to 3.5-fold provided that the intact ArsR was present.

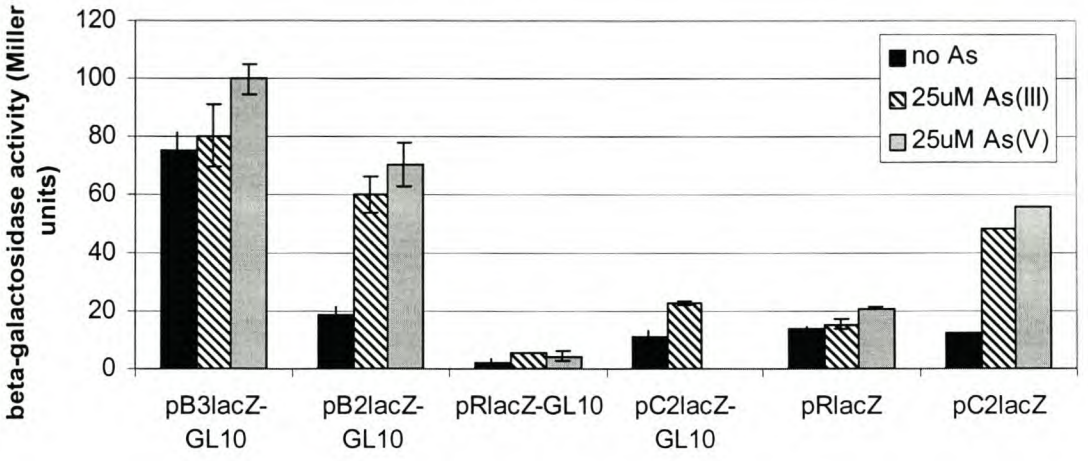


Figure 3.6: Expression of β-galactosidase under the control of the *arsB* (pB3lacZ-GL10) and *arsR* promoters (pRlacZ-GL10) in the presence of 25 μM sodium arsenite and sodium arsenate. Also shown is the *arsB* promoter-fusion with an intact *arsR* gene in *cis* (pB2lacZ-GL10) and the *arsRC-lacZ* fusion (pC2lacZ-GL10). These fusions were cloned into the low copy number vector, pGL10. Fusions to *arsR* (pRlacZ) and *arsRC* (pC2lacZ) are also shown in a higher copy number pMC1403- based vector as expression was very low. (All assays were carried out in triplicate and bars represent standard deviations of two or more experiments).

When similar experiments were carried out with fusions in a higher copy number pMC1403-based vector, the presence of an intact ArsR repressed the *arsB-lacZ* fusion about 7-fold (pB3lacZ vs pB2lacZ) in the absence of inducer (Figure 3.7). Following the addition of 25 μM sodium arsenite or sodium arsenate, repression was lifted about 5-fold, while with addition of 25 μM potassium antimonyl tartrate, derepression was about 3-fold (Figure 3.7). An ArsR construct with 18 amino acids from the C-terminal deleted, (pBlacZ) gave similar levels of repression and derepression in the absence and presence of arsenite, arsenate and antimony (Figure 3.7). These results indicate that the product of the *arsR* gene is able to regulate expression from the *arsB* promoter in response to both arsenic and antimony. We also show that a region between amino acids 71 and 100 is required for this regulation and that the C-terminal amino acids after position 100 can be deleted with no effect on regulation.

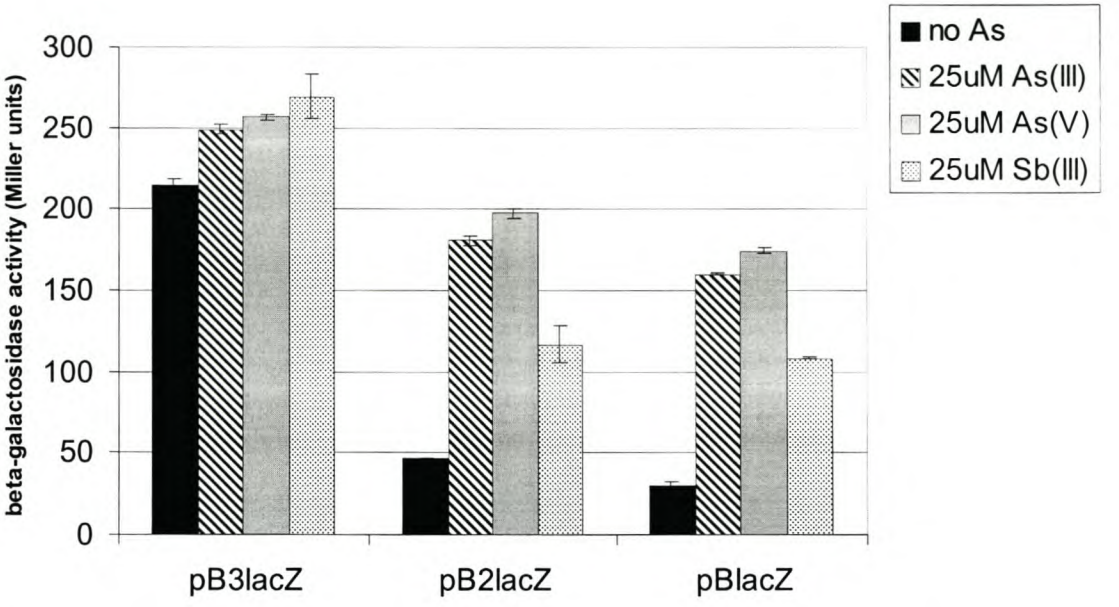


Figure 3.7: *arsB* promoter-*lacZ* fusions containing 71 amino acids (pB3lacZ), 100 amino acids (pBlacZ) and complete (pB2lacZ) ArsR in *cis* induced with 25 μM sodium arsenite, sodium arsenate or antimonyl tartrate.

In *trans* expression of the *arsR* gene on its own under the control of the *tac* promoter gave problems. The *tac-arsR* construct was toxic and host cells grew poorly at 30°C with no growth at 37°C. Furthermore, the addition of IPTG in the absence of the *tac-arsR* construct, reduced expression of the *arsB-lacZ* promoter fusion. We are unsure of the reason for this.

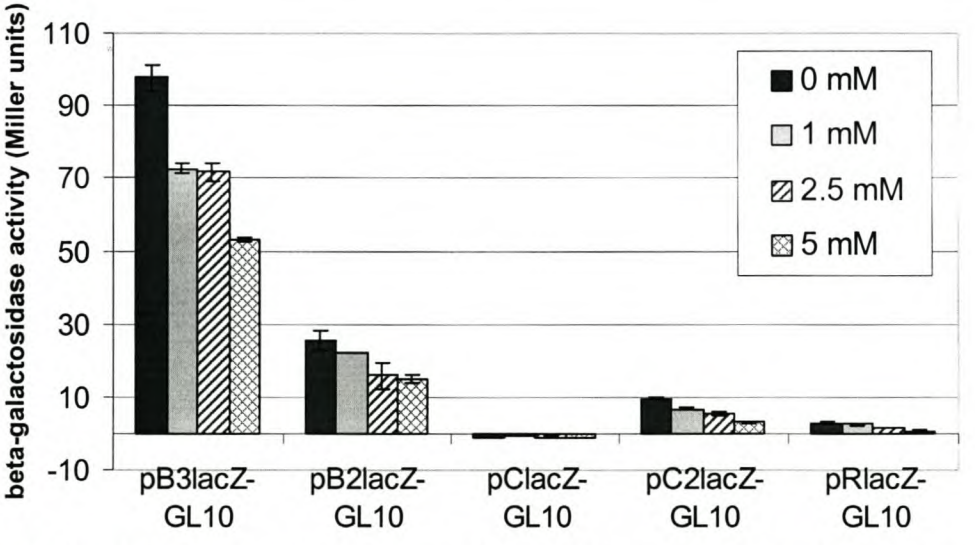


Figure 3.8: The effect of increasing concentrations of IPTG on β-galactosidase expression from the *arsB*, *arsC* and *arsR* promoter fusion constructs. Note that constructs pB2lacZ-GL10 and pC2lacZ-GL10 contain the intact *arsR* gene in *cis*.

This effect of IPTG on the promoter fusion constructs was investigated further (Figures 3.8 and 3.9). Increasing concentrations of IPTG caused increased repression of expression from the promoters (at the highest concentration of 5mM IPTG there was about a 2-fold repression of expression from all the promoter-fusions examined.). This repression occurred even when the *arsR* gene was present in *cis* (pB2lacZ-GL10 and pC2lacZ-GL10).

The repression of the *arsB* promoter caused by ArsR was relieved by the presence of arsenite even in the presence of IPTG (Figure 3.9) However expression was not as high when IPTG was present.

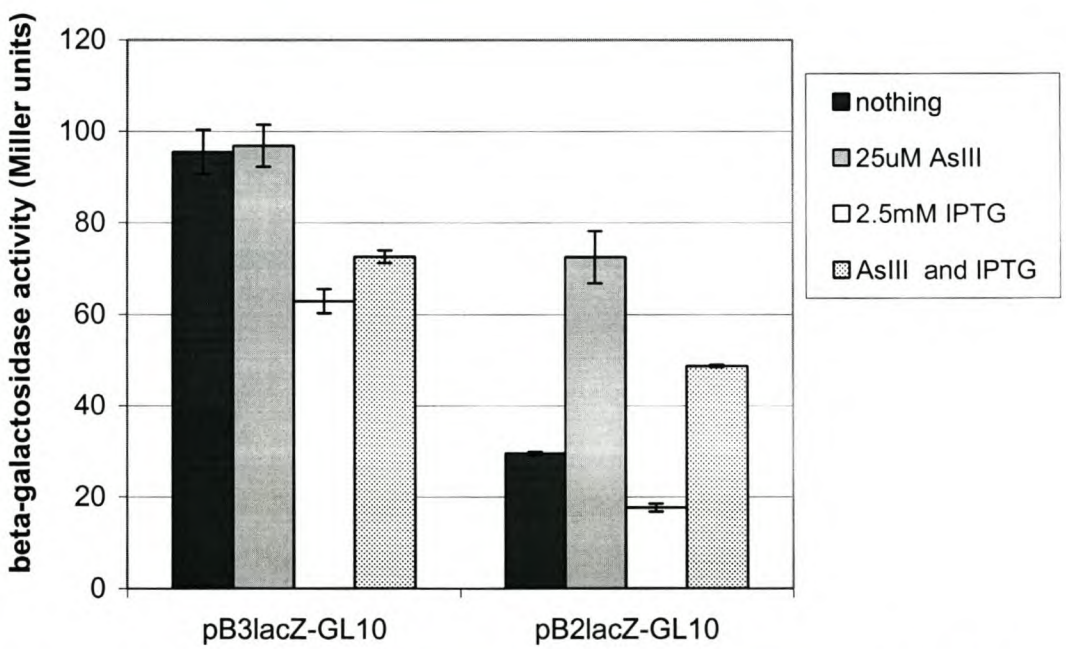


Figure 3.9: The effect of both arsenic and IPTG on expression from the *arsB* promoter.

When constructs containing the *arsR* gene expressed from its own promoter, together with other *ars* genes (pTfarsCRBH-ACYC, pTfarsCRB-ACYC) were added in *trans* with the pB3lacZ-GL10 fusion construct, reporter gene expression was repressed to similar levels as in pB2lacZ-GL10. This repression could be relieved by the addition of both arsenite and arsenate (Figure 3.10) although the level of expression after induction was not as high as the unrepressed β-galactosidase levels. (Note that

pTfarsBH-ACYC and pTfarsB-ACYC contain the same truncated, but functional *arsR* as pBlacZ).

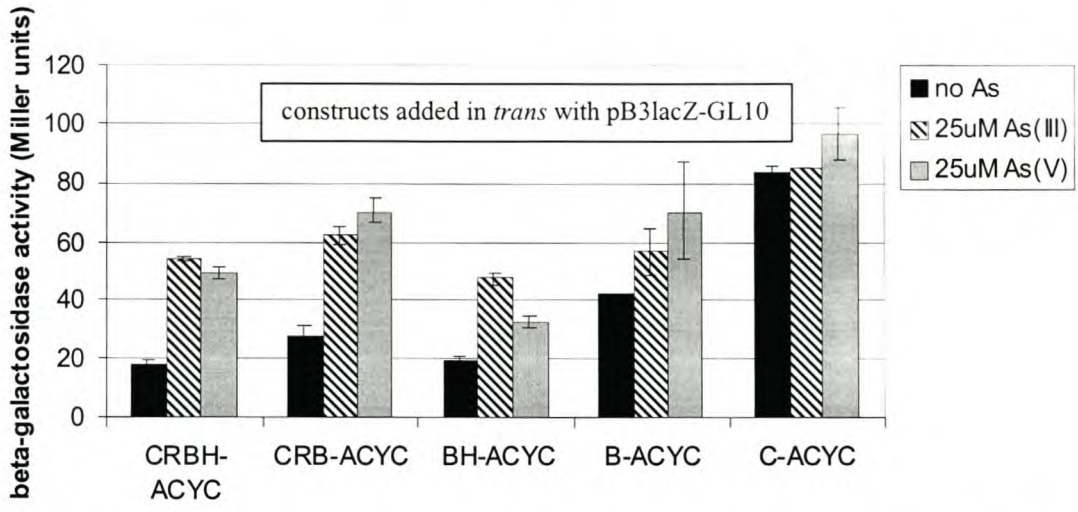


Figure 3.10: Various *ars* constructs added in *trans* with the *arsB* promoter-*lacZ* fusion containing the non-functional 71 aa of ArsR, pB3lacZ-GL10. * pTfarsB-ACYC and pTfarsBH-ACYC contain the same truncated yet functional ArsR as pBlacZ. (All assays were carried out in triplicate and bars represent standard deviations of two or more experiments).

Interestingly, when expression of the pB3lacZ-GL10 fusion in the presence of constructs pTfarsBH-ACYC or pTfarsB-ACYC was compared, it appeared that *arsH* may play a regulatory role in *arsB* expression as expression from the promoter when arsenate was added was higher in the absence of *arsH*. This can also be seen when the presence of pTfarsCRBH-ACYC and pTfarsCRB-ACYC in *trans* with the pB3lacZ-GL10 fusion construct, are compared. However, attempts to investigate reporter gene expression with *arsH* expressed on its own from a *tac* promoter were subject to similar interference when IPTG was added, as found for *tac-arsR* expression described earlier.

3.3.4. Expression of *arsRC*.

When the *arsR-lacZ* fusion constructs were expressed in *E. coli* ACSH50I^q in the absence of arsenic in both a low copy number (pRlacZ-GL10) and a higher copy number vector (pRlacZ), the β-galactosidase activity was 2 and 14 Miller units respectively (Figure 3.6). This is more than 15-fold less than expression from the *arsB-lacZ* fusion in equivalent vectors. The levels of expression from the arsR

promoter in the low copy number vector were so low that further work was carried out only with reporter gene fusions in the higher copy number vector. As might be expected, the addition of arsenic to the *arsR-lacZ* fusion did not result in an increase in reporter gene activity, presumably due to the absence of a functional ArsR (Figure 3.6). Unexpectedly, the addition of combinations of the same *ars* genes shown in Figure 3.10 in *trans* did not affect the expression from pRlacZ with or without the presence of arsenic (Figure 3.11).

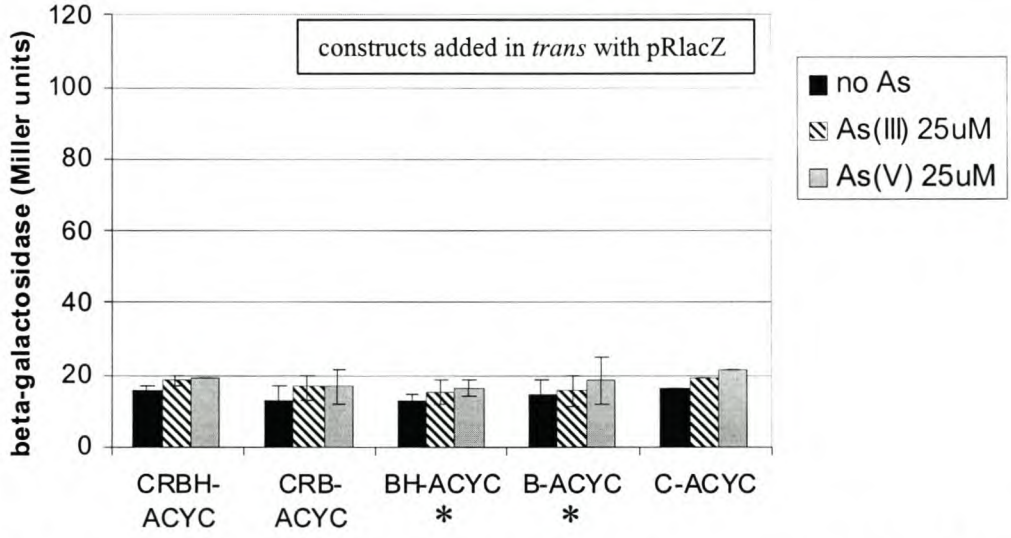


Figure 3.11: Various *ars* constructs added in *trans* with the *arsR* promoter-*lacZ* fusion, pRlacZ. * pTfarsB-ACYC and pTfarsBH-ACYC contain the same truncated yet functional ArsR as pBlacZ. (All assays were carried out in triplicate and bars represent standard deviations of two or more experiments).

An *arsC-lacZ* reporter gene fusion was constructed that included *arsR* and its putative promoter region (pC2lacZ) and a second fusion that excluded the *arsR* promoter region (pClacZ). There was no detectable expression from pClacZ, which suggested that the *arsC* gene does not have a promoter of its own but is transcribed together with *arsR* (Figure 3.8). The *arsC-lacZ* reporter gene fusion, pC2lacZ, which contained an intact *arsR* and its promoter, did respond to added arsenic, and expression was induced 4- and 4.5-fold in the presence of 25 μM arsenite and arsenate, respectively (Figure 3.6). It therefore appears that if ArsR does autoregulate this promoter, it is either required in *cis*, or that there are elements downstream of the promoter required for induction by arsenic that were not included in pRlacZ.

3.3.5. Number and size of transcripts.

We predicted, based on the above promoter-fusion results and on sequence analysis, that the *At. ferrooxidans ars* genes are expressed from two mRNA transcripts. To investigate this prediction, as well as estimate their size and whether they are regulated differently by arsenate and arsenite in *E. coli*, Northern blot analysis was performed. Total RNA was prepared from *E. coli* ACSH50I^q cells containing the pTfarsCRBH-ACYC plasmid. Probes of *arsR, arsB,*and *arsH* as well as a probe covering a region including the *arsB* and *arsR* genes (*ars* probe) (Figure 3.1) were used to probe the Northern blot. About 10 μg of RNA was loaded on the gel shown in Figure 3.12. Attempts to obtain a stronger signal by loading 30 μg of RNA or by preparing RNA from cells expressing the *ars* genes on a higher copy number plasmid (pTfarsCRBH) resulted in an increase in signal from the degraded transcript smear and blot resolution was not improved. When RNA was isolated from *At. ferrooxidans* strains, no transcripts were observed and we believe that the *ars* genes are expressed at low levels.

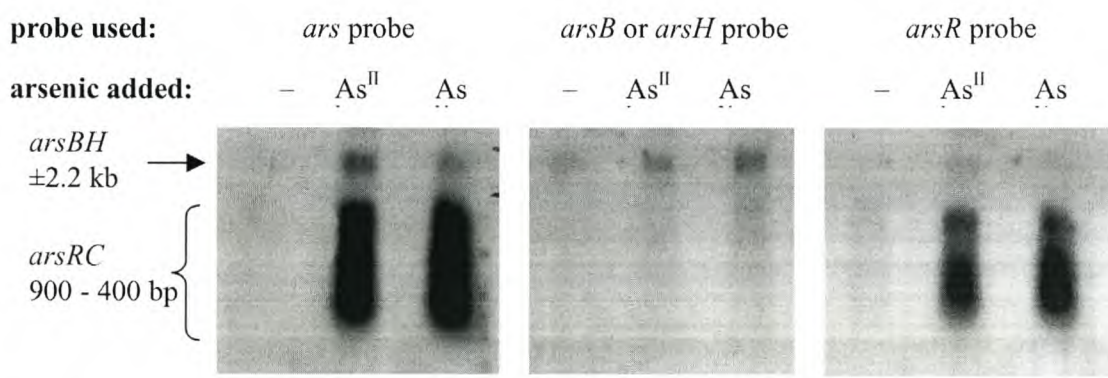


Figure 3.12. Total mRNA prepared from *E. coli* ACSH50I^q (pTfarsCRBH-ACYC) was probed with DNA probes to the *arsB, arsR* and *arsH* genes and a probe to *arsR* and *arsB* including the intergenic region (*ars* probe). RNA was isolated from cells grown in the absence (-) or presence of 25 μM sodium arsenite (AsIII) or 50 μM sodium arsenate (AsV). The positions and sizes of the transcripts are marked with arrows.

Two different patterns of expression were seen depending on whether *arsB* or *arsH* genes or the *arsR* gene were used as probes (Figure 3.12). The size of the transcript detected using either *arsB* or *arsH* as probes was about 2.2 kb, while that detected with the *arsR* probe was a smear of about 400 - 950 bp. When the *ars* probe (containing *arsR* and *arsB*) was used, two transcripts of approximately these sizes

were visible. The calculated sizes of the regions encompassing the *arsBH* and *arsRC* open reading frames are 2,038 and 847 bp respectively, which is consistent with the sizes of the transcripts observed. It was also clear that the quantity of both transcripts, but especially *arsRC*, increased when cells were grown in the presence of either arsenite or arsenate. No clear differences in the levels of arsenite-induced compared to arsenate-induced expression were observed. When a probe from the tetracycline gene of pACYC184 was used, a 1.2 kb signal of approximately equal intensity was observed in all three lanes indicating that approximately equal amounts of RNA had been loaded in each lane (results not shown). The approximately equal levels of induction irrespective of whether arsenite or arsenate was used are in agreement with the reporter gene fusion experiments.

3.4. DISCUSSION

The arsenic resistance genes from *At. ferrooxidans* have an unusual divergent arrangement with the two major genes, *arsB* (encoding for the arsenite membrane pump) and the *arsC* (encoding for the arsenate reductase) transcribed in opposite directions. A gene showing similarity to the regulatory gene, *arsR*, was found in the middle of these genes, transcribed in the *arsC* direction. While the members of the family of ArsR regulatory proteins often show low similarity to each other (e.g. ArsR proteins from the staphylococcal plasmids are only approximately 30% identical to those from the *E. coli* chromosome and R773), those that have been functionally investigated all contain a highly conserved metal-binding domain with an ELCVCDLC consensus sequence. Shi *et al.*, (1994 and 1996) found that the three cysteine residues were able to interact with arsenite. Mutations of the first two cysteine residues resulted in a protein that was able to bind the promoter and repress expression, but that was not able to undergo the conformational change required to release it from the promoter in response to inducer. The finding that most of the closest matching proteins to the *At. ferrooxidans* ArsR are located adjacent to a putative membrane protein, and often an ArsC and ArsH homologue, is circumstantial evidence that they may be *ars* regulators. However, none of the putative ArsR proteins contained the conserved metal-binding motif and the only cysteine residues are found at the C-terminals of the proteins. Nevertheless, we found that mRNA

production from the *At. ferrooxidans ars* genes was increased in response to the presence of arsenite and arsenate. This indicated that regulation of the *ars* genes occurred at the transcriptional level and we showed, using promoter fusions to a β-galactosidase reporter gene, that this regulation was controlled by the atypical ArsR. This suggests that the other related but unstudied *ArsR*-like proteins discovered during genome sequencing projects may do the same. The inducer metalloid binding mechanism of the two families of ArsR regulator proteins may be different. For example, the *ars* operon of *E. coli* plasmid R773 is induced only by arsenite and antimony (Wu and Rosen, 1993). Induction by arsenate takes place after reduction to arsenite, typically by the ArsC reductase. The *At. ferrooxidans* ArsR appeared to respond to arsenate in the *E. coli ars* deletion mutant which contained no ArsC. This suggests that either the ability of arsenate to bind the *At. ferrooxidans* ArsR is different, or some other mechanism exists for reduction of sufficient arsenate to arsenite which then serves as the inducer. In a previous study on ZntR from the chromosome of *S. aureus* (Singh *et al.*, 1999), it was reported that this protein was able to bind to the promoter of the zinc resistance operon and that this binding was inhibited in the presence of zinc even though it did not contain any cysteine residues (Figure 3.3). This protein is related to the ArsR regulatory family (Figure 3.5).

The ArsR regulators are part of the ArsR-family of metalloregulatory proteins, which also includes the CadC (regulator of the Cd^{2+}-efflux ATPase) and SmtB (regulator of the metallothionein SmtA in response to the presence of zinc). We found that although ArsR from *At. ferrooxidans* does not contain the conserved metal binding domain, there are many other strongly conserved areas with SmtB and other known ArsR proteins and the close matches from the BLAST search. The crystal structure of SmtB has recently been determined and has been shown to have a winged helix-turn-helix structure (Cook *et al.*, 1998). The secondary structure of SmtB consists of five α-helices and two β-sheets, with the β-sheets having been shown to form a hairpin structure. Two of the helices (H_{R-1} and H_R, see Figure 3.4) form the standard helix-turn-helix (HTH) motif, while residues from the other three helices interact between monomers to form a hydrophobic core for the dimer, which provides the scaffolding to correctly orient the DNA-binding HTH domain. The conserved Gly-X-X motif (where X represents a large hydrophobic residue) at the end of the DNA-binding

helix-turn-helix (HTH) domain is found in many of the proteins shown in Figure 3.3. Cook *et al.,* (1998) believe that these hydrophobic residues, are important for anchoring the β-sheet that follows. Secondary structure predictions indicate that many of these general features appear to be present in the second family of ArsR-like regulators (Figure 3.4). However, the absence of the established metal-binding motif and the observation that there are extra amino acid residues between the helices, argue against a detailed extrapolation from the secondary structure of SmtB to the ArsR of *At. ferrooxidans*.

We have shown that at least two RNA transcripts, one corresponding to *arsBH* and another to *arsRC* were induced in the presence of arsenite and arsenate. Both Northern hybridisation and *lacZ* reporter gene studies indicated that the level of expression from the *arsB* promoter was low. Although expression was increased following induction, the levels of expression remained low. While this could be due to inefficient expression in *E. coli*, these low levels of expression were also observed when fusions to the *ars* genes from *Bacillus subtilis* skin element were constructed and integrated into the chromosome (Sato and Kobayashi *et al.*, 1998). The reason for such low levels of expression could be due to the toxic nature of the *ars* genes. Over expression of the *E. coli* ArsB from a *tac* promoter has been reported to be toxic to cells (e.g. Cai and DuBow, 1996). Despite the low levels of expression, induction of expression by arsenite, arsenate and antimonite was observed.

The relative levels of expression from the *arsRC* promoter gave apparently conflicting results depending on whether Northern hybridization or reporter gene studies were used. Although the quantity of *arsRC* transcript seen on the Northern blot was higher than for *arsBH*, the level of *arsRC-lacZ* reporter gene expression was much lower than for the *arsB-lacZ* fusion. The observation that *arsRC* mRNA transcript was a smear may indicate that this mRNA has a higher turnover rate than *arsBH* mRNA. If much of the degradation takes place from the 5'end, this may be the reason for the lower level of reporter gene expression. Reporter gene expression from the *arsRC* promoter was so low (less than 10 Miller units) that we worked with a higher copy number vector and obtained results that were consistent with results observed in the low copy number vector. Expression from an *arsRC* promoter fused at the start of

arsR was unresponsive to the presence of arsenic, presumably due to the absence of *arsR*. The addition of an *arsR* gene in *trans* did not induce expression, possibly because sequences downstream of the point of fusion were required. However, when a translational fusion to the *arsC* gene (which included the *arsR* gene and *arsRC* promoter) was investigated, expression was induced in the presence of both arsenite and arsenate. This indicated that *arsR* was required for regulation of this promoter. As an *arsC-lacZ* fusion containing the region upstream of *arsC*, but not including the *arsR* promoter (pClacZ) showed no β-galactosidase activity, it appears that *arsR* and *arsC* are expressed as a single transcript. Therefore, although the two *ars* transcripts of the *At. ferrooxidans ars* operon were transcribed at different levels, both were regulated by arsenite and arsenate.

We found that the last 19 amino acids at the C-terminal of *At. ferrooxidans* ArsR were not required for either repression or induction of the *arsBH* promoter by arsenite, arsenate or antimony. However when a further 28 amino acids of ArsR in *cis* with the promoter fusion was deleted (leaving only 71 amino acids), expression from the promoter was constitutive. Similarly, it was shown that *E. coli* chromosomal ArsR-β-lactamase chimeras were constitutive when fused to regions upstream of the 79th amino acid, but were inducible when β-lactamase was fused downstream of the 92nd amino acid (Xu *et al.*, 1996). These positions align with amino acids 85 and 98 of the *At. ferrooxidans* ArsR. An interesting observation is that the double cysteines conserved in *At. ferrooxidans* ArsR and its closest relatives were retained in the functional truncated ArsR. Further experiments are required to show whether these residues are involved in regulation.

We were unable to find a clear role for ArsH in regulation of the *At. ferrooxidans ars* genes. While addition of constructs containing *arsH* in *trans* with the *arsB* promoter fusions appeared to result in increased repression of the promoter and lower levels of induction especially when arsenate was the inducer, these differences were small (Figure 4B). We were unable to add the *tac*-regulated *arsH* gene back in *trans*, as we found that IPTG on its own decreased expression from the *At. ferrooxidans arsB* promoter. It is possible that the role that ArsH plays is only important in *At.*

ferrooxidans and not in *E. coli* or that this function is provided by a gene already present in *E. coli*.

This is the first functional study of an ArsR regulator that is able to respond to arsenic, but does not contain the conserved metal binding motif. We propose that the ArsR from *At. ferrooxidans* is the first reported member of a subclass of ArsR regulators which may have a different method of binding the inducer.

Chapter Four: General Discussion

TABLE OF CONTENTS

4.1. Introduction

At the start of this study, resistance systems to many heavy metals had been identified in bacteria (Silver and Walderhaug, 1992), including resistance to arsenic metalloids. Researchers had shown that resistance to both arsenic and antimony was encoded by arsenic resistance genes (*ars*) and this system was well characterised in both *E. coli* and *Staphylococcus* species. It was known that there were two types of arsenic resistance operons, one consisting of 5 genes (*arsRDABC*) found on the R773 plasmid in *E. coli* and the second consisting of only 3 genes (*arsRBC*) was found on the plasmids of *S. aureus* (pI258) and *S. xylosus* (pSX267). In 1995, during the sequencing of the *E. coli* chromosome it was found that this bacterium also contained a 3 gene *ars* operon on its chromosome and this operon was found to be functional (Carlin *et al.*, 1995 and Diorio *et al.*, 1995). This was the first time that chromosomally located arsenic resistance genes had been identified. Since then five more *ars* operons were cloned and sequenced (see Table 1.1). Only one was chromosomally located (found in *P. aeruginosa*) and one operon was identified in the skin element of *B. subtilis*. With the advent of chromosomal sequencing projects homologues to the genes of the arsenic resistance operon have been located on many more bacterial chromosomes. Recently, the ArsA protein from *E. coli* R773 (Zhou *et al.*, 2001) and the ArsC proteins from *E. coli* R773 (Martin *et al.*, 2001), *S. aureus* pI258 (Zegers *et al.*, 2001) and *B. subtilis* (Bennett *et al.*, 2001) were crystallised and the structures determined, giving more insight into the function of these proteins in arsenic resistance (see chapter 1 for details). A member of the ArsR-family of regulators (SmtB) has also been crystallised and the structure determined (Cook *et al.*, 1998).

This study was the first to isolate and sequence the arsenic resistance genes from the chromosome of an acidophile. It was interesting that we were able to use the complementation of the *E. coli* arsenic sensitive mutant, which lacks the entire *arsRBC* operon (AW3110), to isolate the resistance genes from a *At. ferrooxidans* gene bank. In particular we were surprised to find that the *arsB* gene, encoding for the membrane-spanning arsenite efflux pump, was highly similar to the *arsB* gene from the neutrophile *E. coli* (63% identity), and that there was no *arsA* homologue

providing energy for the pump. In *E. coli* it has been shown that when no ArsA is present, efflux of arsenite through ArsB is driven by the membrane potential. However, it has been shown that in acidophilic bacteria there is a large ΔpH across the membrane. Furthermore, the bacteria have a positive-inside rather than a negative-inside membrane potential and this subtracts from the H^+ gradient rather than augmenting it (Cox *et al.*, 1979). Despite this difference between the bacteria, we found that a plasmid containing only the *At. ferrooxidans arsB* gene was able to provide the *E. coli ars* mutant with resistance to arsenite (Figure 2.13) and that the cloned *ars* genes from *At. ferrooxidans* conferred similar levels of arsenite resistance to the *E. coli* AW3110 strain as the cloned *E. coli* chromosomal *ars* genes carried on a plasmid with the same origin of replication (Figure 2.14).

The *ars* operon we isolated from the gene bank of *At. ferrooxidans* was shown to be located on the chromosome of this bacteria (Figure 2.3). While various plasmids have been isolated from *At. ferrooxidans* and some have been either partially or completely sequenced (reviewed Rawlings and Kusano, 1994), none of these plasmids has yet been shown to carry homologues to the arsenic resistance genes. The *ars* operon from *At. ferrooxidans* had three unusual features which were further investigated. These features and our main findings are summarised below with a detailed discussion to follow:

i. Unlike all previously studied *ars* operons, the *arsB* and *arsC* genes were divergently transcribed and while the ArsB protein from *At. ferrooxidans* was most closely related to the ArsB homologues from other Gram-negative bacteria, the ArsC protein from *At. ferrooxidans* was found to group with ArsC proteins from Gram-positive bacteria. We showed that reduction of arsenate by the *At. ferrooxidans arsC* was coupled to thioredoxin as is the case for the ArsC proteins from other gram-positive bacteria.
ii. An *arsH* homologue was identified downstream of the *arsB* gene. This was only the second homologue of this gene isolated and the function of the gene is unknown. While other researchers had shown that the *arsH* gene was required for resistance to arsenic in the bacteria from which it was isolated, we found that this gene from *At. ferrooxidans* was not required for arsenic resistance in the heterologous host, *E. coli*.

iii. We identified an open reading frame with weak homology to the ArsR- family of regulators. While this protein did not contain the conserved metal-binding motif found in all previously identified ArsR regulators, we showed that this gene was able to regulate the *At. ferrooxidans ars* genes in response to the presence of arsenic and propose that it is the first member of a subfamily of atypical ArsR regulators.

4.2.The unusual At. ferrooxidans ars operon

4.2.1. The arsB and arsC genes are divergently transcribed

This was and still is the only reported *ars* operon where the *arsB* and *arsC* (encodes arsenate reductase) genes are divergently transcribed. We confirmed that this divergent arrangement was present on the chromosome of *At. ferrooxidans* and was not due to rearrangement during the construction of the gene bank (Figure 2.2). Both these genes were shown to be expressed (Figure 2.11) and functional in an *E. coli ars* mutant (Figure 2.13) and this unusual arrangement was also found in other *At. ferrooxidans* strains (Figure 2.10). While the level of arsenite resistance (requiring only the *arsB* gene), conferred by the *At. ferrooxidans ars* genes on the *E. coli ars* mutant was comparable to that conferred by the *E. coli* chromosomal *ars* genes (Figure 2.14), it appeared that the level of arsenate resistance conferred on the *E. coli* mutant (requiring both the *arsB* and *arsC* genes), was much lower.

We found that phylogenetically the ArsB protein grouped, as expected, with ArsB proteins from other Gram-negative bacteria, but the ArsC protein grouped with other ArsC proteins from Gram-positive bacteria. At this time this was the only such example and until recently the two groups of ArsC proteins were considered to be separated depending on their Gram-positive or Gram-negative origins. However, in 1998, the chromosomal ArsC from *P. aeruginosa* was also found to group with the Gram-positive ArsC proteins (Cai *et al.*, 1998). A comparison of the amino acid sequences indicated that these ArsC proteins did contain the four conserved cysteine residues found in the other gram-positive ArsC proteins. Recently the crystal structures of ArsC proteins from both groups have been determined and the mechanisms of arsenate reduction predicted (Chapter 1, Figures 1.6 and 1.7). The ArsC proteins from the two groups have different structures and reaction mechanisms, but all three proteins crystallised perform the same function of reducing arsenate to arsenite. We aligned the amino acid sequences of the ArsC proteins from *P.*

aeruginosa, *P. putida* and *At. ferrooxidans* with other Gram-positive ArsC proteins and compared these with the structures and predicted reaction mechanism of the *B. subtilis* and *S. aureus* pI258 ArsC proteins. We found that the arsenate binding loop (AB loop), including the CX_5R motif, is conserved among all the ArsC proteins of this group (Figure 2.7). The two cysteine residues at the C-terminal of the protein (corresponding to Cys-82 and Cys-89 of *B. subtilis* ArsC), which have been shown to form a disulphide bridge upon oxidation, are also conserved. The crystal structure shows that Cys-89 is in a flexible region that moves during the reaction cycle to bring this cysteine close to the active site allowing it to form the disulphide bridge. One difference is that the ArsC proteins from Gram-positive bacteria have 6 residues between these two cysteines while the cysteines from ArsC proteins of *At. ferrooxidans* and the *Pseudomonas* species are separated by 7 amino acids. Another important residue involved in the reaction mechanism is an Aspartate residue (Asp-105 in *B. subtilis* ArsC). This residue is conserved in all the ArsC proteins of this group including those from the Gram-negative bacteria. While the ArsC proteins from the Gram-negative bacteria are the least similar of the members of this group, separated by the longest branch-length in the phylogenetic tree (Figure 2.6), they do contain the residues essential for the reaction mechanism predicted for this group of ArsC proteins. We confirmed that like the Gram-positive ArsC proteins, reduction of arsenate by the *At. ferrooxidans* ArsC required the presence of the *At. ferrooxidans* thioredoxin gene (*trxA*). Since there were now two ArsC proteins isolated from Gram-negative bacteria that were more closely related to those from Gram-positive bacteria, we suggested that the separation of these two groups of proteins based on these criteria was an oversimplification (Butcher *et al.*, 2000). A recent review by Mukhopadhyay *et al.* (2002) referred to the groups of ArsC proteins as the GSH/Grx clade and the Trx clade. Although these Gram-negative ArsC proteins are the least similar of the Trx clade, based on the above results and sequencing analysis we predict that they will have the same reaction mechanism as predicted for the *B. subtilis* and *S. aureus* pI258 ArsC proteins.

In all previously studied *ars* operons, these genes confer resistance to a higher concentration of arsenate compared to arsenite. However, the *At. ferrooxidans ars* genes appeared to show a slightly lower level of resistance to arsenate than arsenite (Figure 2.13). The *P. aeruginosa ars* operon conferred increased arsenite resistance

on a wild-type *E. coli* strain, but it did not confer any increase in arsenate resistance (Cai *et al.*, 1998). These results suggest that these ArsC proteins are unable to function efficiently in *E. coli*. This could be as a result of an inefficient interaction with the *E. coli* thioredoxin, or it could be that these ArsC genes are not well expressed in *E. coli*. We did notice that when the *At. ferrooxidans* thioredoxin gene was added in *trans* with the *At. ferrooxidans* genes to the thioredoxin and glutathione double *E. coli* mutant (BH5262) this strain did grow on low-phosphate plates containing higher concentrations of arsenate, however growth curves in liquid culture were not determined for these strains.

4.2.2. The *At. ferrooxidans ars* operon also contains an *arsH* homologue

Downstream of the *arsB* gene we identified a homologue to an *arsH* gene which at the time had only been reported in *Y. enterocolitica* (Neyt *et al.*, 1997). Recently a third *arsH* homologue has been identified in the IncH12 plasmid, R478 from *S. marcescens* (Ryan and Colleran, 2002). In both of these cases the *ars* genes have the same arrangement, with the *arsH* gene divergently transcribed upstream of an *arsRBC* operon, and the authors have shown that the *arsH* gene is required for arsenic resistance in the homologous host. We identified eight other putative *arsH* homologues from the NCBI database which all showed high homology to the already identified *arsH* genes (Figure 2.8). All of these putative *arsH* genes are located adjacent to an *arsB* or ACR3 (encodes for yeast arsenite efflux protein) gene or another putative membrane spanning protein (e.g. MIP, major intrinsic protein). All, except the operon located on the cryptic plasmid (pAT) of *A. tumefaciens*, also contain an *arsC* homologue. The orientation and location of the putative *arsH* genes is not consistent. In the case of *C. crescentus* the putative *arsH* gene is upstream and divergent of the other *ars* genes, however four other putative *ars* operons have the *arsH* gene following the *arsRBC* genes (Table 2.3). None of these putative *arsH*-containing operons has the same orientation as the *At. ferrooxidans ars* operon.

The function of the *arsH* gene has not yet been elucidated, however it has been shown that in *Y. enterocolitica*, deletion of the *arsH* gene upstream of the *arsRBC* operon resulted in a strain that was sensitive to both arsenite and arsenate. The addition of the *arsH* gene in *trans* restored arsenic resistance to the *Y. enterocolitica* strain (Neyt *et al.*, 1997). Similarly, deletion of the *arsH* gene from *S. marcescens* R478 also

resulted in a reduction of arsenic resistance, but the authors have not added the gene in *trans* and shown that arsenic resistance can be restored (Ryan and Colleran, 2002). Conversely, we found that while the *arsH* gene from *At. ferrooxidans* was expressed in *E. coli* (Figure 2.11), it was not required for arsenic resistance in *E. coli* (Figure 2.13). While it appears that our results do not agree with those already published for *arsH*, the effect of the other *arsH* genes was not studied in *E. coli*. Our results may indicate that the function of the *arsH* gene is not required in *E. coli* or that it is fulfilled by another gene on the *E. coli* chromosome.

The NCBI conserved domain database (CDD) indicates that the ArsH protein has homology with a family of predicted flavoproteins (COG0431) with an E-value of $7e^{-32}$ and a family of NAD(P)H-dependent FMN-reductases (Pfam:03358) with an E-value of $7e^{-29}$. These proteins reduce FMN (flavin mononucleotide) to $FMNH_2$ which is required as a cofactor by another associated enzyme. Many of the members of these groups are putative proteins whose function has only been predicted by sequence identity. The following proteins have known functions: MsuE provides the $FMNH_2$ required by MsuD, which catalyses the desulfonation of methanesulfonates (Kertesz *et al.*, 1999); SsuE proteins from *P. putida* (Kahnert *et al.*, 2000) and *E. coli* (van der Ploeg *et al.*, 1999) are involved in the utilization of aromatic and aliphatic sulfonates respectively and provide reduced FMN for the $FMNH_2$-dependent oxygenase, SsuD; and an iron-sulfur flavoprotein (Isf) from *Methanosarcina thermophila*, is predicted to function as an electron carrier in the pathway for the fermentation of acetate (Latimer *et al.*, 1996). On further investigation it was found that alignments of the *At. ferrooxidans* ArsH protein with these members whose functions are known show that these proteins are less than 25% homologous to the ArsH homologues (Figure 4.1). Therefore, it is not possible to induce any function for ArsH from this output of the NCBI CDD search.

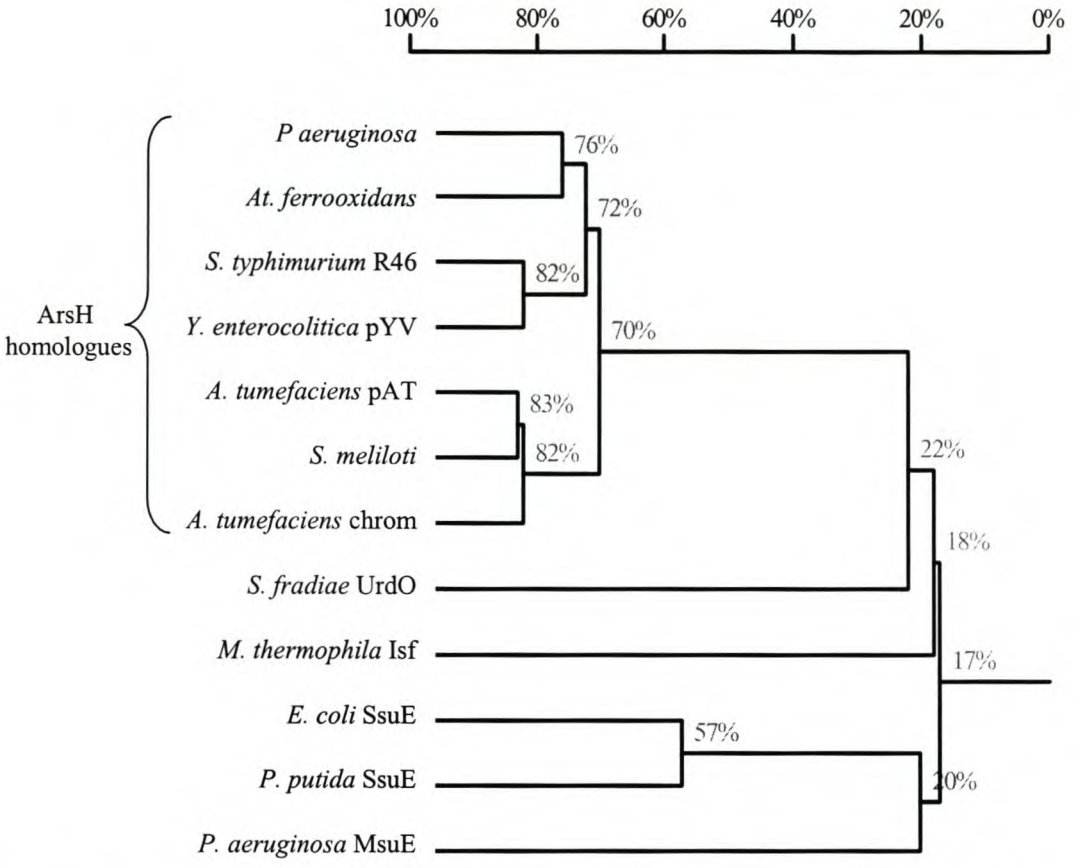


Figure 4.1: Homology tree of the ArsH homologues and the members of the COG0431 and Pfam03358 groups of proteins with known functions. See Table 2.3 for accession numbers of the ArsH homologues. Accession numbers: *S. fardiae* UrdO, AF164961_1; *M. thermophila* Isf, AAC45465; *E. coli* SsuE, P80644; *P. putida* SsuE, SSUE_PSEPU; *P. aeruginosa* MsuE, O31038.

Further investigation into the role of *arsH* is required. Ideally, we would want to find out whether this gene is required for resistance to arsenic in *At. ferrooxidans*, however we do not have sufficient molecular tools to pursue this question. *At. ferrooxidans* has proven difficult to transform and there are not many vectors available for use in this bacterium. The low pH of the growth medium inactivates many antibiotics and the bacterium grows very slowly under laboratory conditions, making it difficult to work with in general (discussed in Rawlings and Kusano, 1994). However, it may be possible to investigate the effect of *At. ferrooxidans arsH* in another host bacterium, such as *Y. enterocolitica* or *S. marsescens* where we already know that their own *arsH* genes are required for arsenic resistance. Since the *arsH* genes show high homology to each other it would be interesting to known if the *At. ferrooxidans arsH* gene would be able to complement the lack of *arsH* in another system, eg. *Y. enterocolitica*.

Another possibility would be to investigate whether bacteria, such as *A. tumefaciens* or *P. aeruginosa* that also contain putative *arsH* homologues, have the same requirement for *arsH*.

Although we found that *arsH* was not required for arsenic resistance in *E. coli*, it appeared that when the *arsH* gene was not present, expression from the *arsB-lacZ* fusion construct was increased, especially in the presence of arsenate (Figure 3.10). We were unable to confirm this effect by adding the *arsH*, driven by the *ptac* promoter in *trans* due to the effect of IPTG on the promoter itself. Both Neyt and co-workers (1997) and Ryan and Colleran (2002) suggested that the *arsH* gene might play a regulatory role. It has been shown that overexpression of the *ars* genes (especially *arsB*) is toxic to the cells. In the 5-gene version of the *ars* operon which has to date only been found on plasmids, the *arsD* gene acts as a regulator of the upper level of expression of the operon, thereby preventing the toxic overexpression of the *ars* operon. When the ArsD protein reaches a critical level it binds to the promoter, thereby shutting down further expression of the *ars* genes. The ArsD and ArsH proteins do not show any homology to each other and they do not contain any known DNA-binding motifs. However, it has been shown that the R773 ArsD does bind to the promoter region of the *arsRDABC* operon, but with a lower affinity than ArsR (Chen and Rosen, 1997). When the *arsD* gene of R773 is inactivated with an insertion leaving the *arsRABC* genes intact, the cells are sensitive to arsenic. Resistance can be restored by adding the intact *arsD* gene in *trans*. It is possible that the *arsH* gene is acting in the same way, as a second regulator of these genes and that it is not required in *E. coli* as the genes are not expressed as well in this heterologous host.

Future experiments to investigate whether ArsH may act as a regulator could include the following:

1. Express the *arsH* gene under the control of a different promoter (one not requiring IPTG for induction), either a constitutive promoter or for example the T7 RNA polymerase promoter. This would enable us to confirm that the effect on *arsB-lacZ* expression in the presence of arsenate (seen in figure 3.10) is due to the presence of *arsH*. By increasing the concentrations of inducer (both arsenite

and arsenate) we could investigate whether this effect becomes more pronounced at higher concentrations of both arsenite and antimony.

2. The DNA-binding activity of the ArsH protein could be investigated. It may be possible to overexpress the *arsH* gene and use cell extracts to investigate the binding activity of ArsH to the promoter region. Alternatively, the protein could be tagged (e.g. HIS-tag) and purified and this purified protein used in the gel mobility shift assays. If the protein were found to bind to the promoter region, the site of binding could be determined using DNaseI footprint analysis.
3. Originally, regulation by the ArsD protein was believed to be inducer-independent, however it has been found that at high concentrations of arsenite or antimonite, binding to the promoter is relieved. It has also been shown that the two vicinal cysteine pairs found on ArsD bind arsenite and are required for the function of the protein (Li *et al.*, 2001). The ArsH protein from *At. ferrooxidans* does contain four cysteine residues (Cys-23, Cys-46, Cys-104 and Cys-144), however only Cys-144 is also conserved in the ArsH proteins from *Y. enterocolitica* and *S. marsescens*. If the ArsH protein does bind the promoter region, it would be important to ask whether this binding could be relieved by the presence of inducer and, if so, if the conserved cysteine residue plays a role.

4.2.3. The *At. ferrooxidans ars* genes are regulated by an atypical ArsR regulator

An open reading frame was identified between the *arsB* and *arsC* genes that showed weak homology to the known ArsR regulators (i.e. 32% identity with *E. coli* R773 ArsR and 15% identity with *S. aureus* pI258 ArsR). More importantly this putative regulator did not contain the conserved metal-binding motif (ELCVCDL) known to be involved in binding arsenite. Arsenite binds to the cysteine residues in this motif thereby causing a conformational change to ArsR that causes the repressor to dissociate from the promoter, ultimately resulting in the expression of the *ars* genes. Despite the lack of the metal-binding motif on the only identified regulator, northern blot analysis indicated that expression of the *At. ferrooxidans ars* gene transcripts in *E. coli* was increased in the presence of arsenic (Figure 3.12). Using fusions of the predicted promoter regions of the *At. ferrooxidans ars* genes with a promoterless reporter gene (*lacZ*) we showed that expression from the *arsB* promoter was repressed in the presence of an intact *arsR* gene and that this repression was relieved when

arsenite, arsenate and, to a lesser extent, antimony was added to the media (Figure 3.6 and 3.10). These results indicated that this atypical ArsR does indeed regulate the expression from the *arsB* promoter. This raises intriguing questions about the mechanism of this regulation.

We identified at least nine other putative proteins that show homology to the atypical ArsR from *At. ferrooxidans* (Table 3.2). Seven of the nine putative ArsR proteins are located along with an *arsC* homologue. None of the putative ArsR proteins are located with an *arsB* homologue, but one is located with the ACR3 homologue and all the rest have a membrane located protein adjacent to them. We propose that the *At. ferrooxidans* ArsR is a member of a new atypical group of ArsR regulators, the first to be shown to function as an *ars* regulator. It is however, unclear whether all of these operons containing the atypical *arsR* genes confer resistance to arsenic or whether the *At. ferrooxidans* operon has acquired a regulator from a different system.

The DNA-binding helix-turn-helix of the atypical ArsRs appears to be conserved compared with the known ArsR regulators as well as the recently crystallized SmtB regulator (Figure 3.4). The second helix (H_R) is believed to be involved in the DNA recognition and many residues in this helix are conserved throughout all the known ArsR regulators and the putative atypical ArsR homologues (Figure 3.3). It has been shown that binding of arsenite to the typical ArsR regulator via interactions with the cysteine residues found in the metal-binding motif, located at the start of the first helix in the helix-turn-helix structure, results in a change of conformation and ultimately dissociation from the promoter, thereby allowing expression of the *ars* genes. Since we know that the *At. ferrooxidans* ArsR is able to regulate the *At. ferrooxidans ars* genes in response to arsenic the next step would be to determine how the regulator responds to arsenic.

It will be important to show that the *At. ferrooxidans* ArsR does bind to the predicted promoter regions. This could be shown using a gel mobility shift assay. Shi *et al.* (1994) performed gel shift assays using crude cell extracts from *E. coli* cells harbouring a plasmid containing the *E. coli* R773 *arsR* gene. We could attempt this procedure or we could try to purify the *At. ferrooxidans* ArsR protein possibly by tagging the protein with a His-tag and then use the purified ArsR protein for the gel

shift assays. However, we have found that overexpression of the *At. ferrooxidans arsR* gene under the control of the *tac* promoter was lethal to *E. coli* cells (clones had to be selected from cells grown at 30°C instead of 37°C and carrying the *lacI*^{q} repressor).

If we were able to perform the gel shift assays there are many questions that could be addressed using this method:

1. Show that the ArsR does not bind in the presence of inducer, confirming that the ArsR repressor responds to the presence of arsenic.
2. As mentioned in chapter three, it appeared that unlike the *E. coli* R773 ArsR which is only induced by arsenite, expression from the *arsB-lacZ* fusion was induced by both arsenite and arsenate even in the absence of *arsC*. This would suggest that either arsenate is reduced to arsenite by processes in the cell other than ArsC, or that the *At. ferrooxidans* ArsR is able to recognise both arsenite and arsenate. It would be possible using a gel shift assay to investigate this further.
3. The different mutant ArsR proteins described below could be assayed for their ability to bind the promoter region and/or respond to the presence of inducer.

While the atypical ArsR proteins did not contain the metal-binding motif, there is a conserved region found among the atypical ArsR proteins corresponding to an area at the beginning of the HTH region (GX(I/L)A), as well as a conserved pair of cysteines at the C-terminus of the proteins (Figure 3.3). It is important to determine how these features are related to the function of these ArsR proteins. We have shown that a region between amino acids 71 and 100 is required for regulation of the *arsB* promoter in response to arsenic (Figure 3.7). This region contains the conserved cysteine doublet. A first step would be to create a fusion construct (similar to pBlacZ) where the *arsR* gene in *cis* with the *arsB-lacZ* fusion was truncated at amino acid 84, i.e. just before the cysteine doublet. If expression from this construct is not repressed in the absence of arsenic, one could conclude that the cysteine residues may be involved in either DNA-binding or that they are required for dimerisation of the ArsR regulator. On the other hand, if expression from the promoter is repressed in the absence of inducer, but is not induced upon addition of either arsenate or arsenite, this would indicate that these cysteines may be required for recognition of the inducer.

This could be investigated further by mutating these cysteine residues (eg. to glycine as done for R773 ArsR by Shi *et al.*, 1996) and the ability of the mutated ArsR to repress or respond to inducer investigated. Mutations of the conserved glycine and alanine residues in the GX(I/L)A motif could also be created and the effect of these mutations on repression and induction investigated. If any of the above mutations result in an inability to repress the promoter, we could use a yeast two-hybrid system to investigate whether they affect dimerisation.

4.3.Regulation of the *arsRC* transcript?

As discussed in chapter three we were unable to detect regulation of an *arsR-lacZ* fusion when the *At. ferrooxidans ars* genes were added in *trans* (Figure 3.11). However, when an *arsC-lacZ* fusion, containing the *arsR* promoter and complete *arsR* gene upstream, was constructed we did observe some induction of expression in the presence of inducer (Figure 3.6). This suggested that either the *arsR* gene is required in *cis* or that there are elements within the *arsR* gene that are required for regulation of expression from the *arsR* promoter. This could be further investigated by making *arsR-lacZ* fusions further downstream in the *arsR* gene.

The typical ArsR proteins from R773 and *E. coli* chromosome have been shown to bind at an inverted repeat (5'-NCNTA-(7bp)-TANGN-3') (Xu *et al.*,1996). While this sequence is not conserved in the promoter regions of other *ars* operons, all of the *ars* operons studied thus far do have an inverted repeat located upstream of the transcription start site. Since the sequence downstream of the *At. ferrooxidans arsR* start may be required for regulation by the *At. ferrooxidans arsR*, this sequence was searched for repeats and inverted repeats. There is one direct repeat and one inverted repeat located within the *arsR* gene, less than 100bp downstream of the *arsR* start (Figure 4.2). Both of these features are included in the pC2lacZ fusion and also all the *arsB-lacZ* fusion constructs, but not the pRlacZ construct. The construction of *arsR-lacZ* fusions at amino acids 25 (including both the direct and inverted repeats), 20 (excluding half of the direct repeat, but including the inverted repeat) and 12 (excluding both the inverted and direct repeats) of the ArsR protein would allow the investigation of the effect of these repeats on regulation from the *arsR* promoter. However, it is possible that these features not only play a role in regulation of expression from the *arsR* promoter, but also the *arsB* promoter. Translational fusions

with the *arsB* gene constructed to include regions upstream with or without the repeats, would allow the study of the effect of these repeats on regulation of the *arsB* promoter. If these constructs do show a difference in expression from these promoters, the involvement of the repeats could be confirmed by mutation of the sequences.

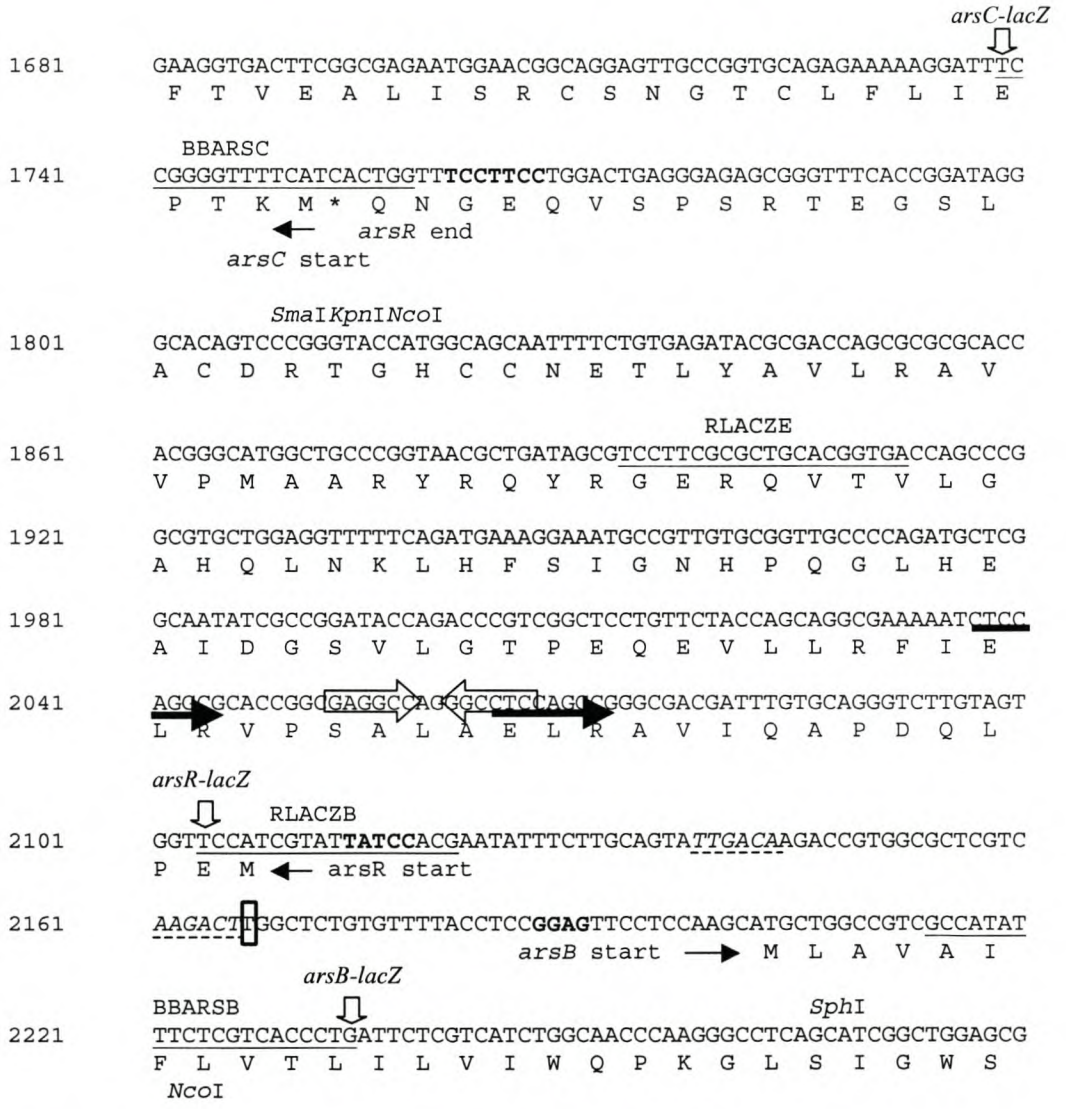


Figure 4.2: Sequence of the *arsR* gene and regions upstream and downstream, showing the position of the inverted repeats (open arrows) and direct repeats (underlined in black arrows). See Figure 3.2 for description of other labels.

4.4.Other open reading frames identified on the pTfars1a clone:

Three complete and two incomplete ORFs with no homology to known *ars* genes were also identified on the chromosomal insert contained in pTfars1a (table 2.2). These are discussed below.

4.4.1. ORF1

This putative protein shows homology to a group of putative membrane proteins, identified through genome sequencing projects, which have homology to the HlyD family of secretion proteins (Pfam accession no: PF00529). These proteins represent one of the components of a secretion mechanism of toxins and drugs which does not make use of the Sec proteins (a commonly used pathway for the secretion of compounds into the medium). Products to be secreted by this mechanism also do not require a cleaved N-terminal signal sequence. Along with the HlyD-like proteins, these systems also require second protein, a member of the ABC transporter family, which is believed to provide the energy for the transport. The various members of this group are shown in Table 2.3 along with the compound they secrete.

Table 2.3: Members of the HlyD family of secretion proteins (taken from Pfam, http://pfam.wustl.edu)

Gene	Species	**Protein transported**
hlyD	*E. coli*	hemolysin
appD	*Actinobacillus pleuropneumoniae*	hemolysin
lcnD	*Lactococcus lactis*	lactococcin A
lktD	*Actinobacillus actinomycetemcomitans* *Pasteurella haemolytica*	Leukotoxin
rtxD	*Actinobacillus pleuropneumoniae*	Toxin-III
cyaD	*Bordetella pertussis*	Calmodulin-sensitive adenylate cyclase-hemolysin (cyclolysin)
cvaA	*E. coli*	Colicin V
prtE	*Erwinia chrysanthemi*	Extracellular proteases B and C
aprE	*Pseudomonas aeruginosa*	Alkaline protease
emrA	*E. coli*	Drugs and toxins
yjcR	*E. coli*	unkown

The best-characterized member of this group is HlyD. This protein is involved in the transport of α-hemolysin (HlyA), which is the best characterized RTX (repeats in toxin) protein that is secreted by a type I secretion system. The synthesis and transport of this toxin is controlled by the *hlyCABD* operon. HlyC is a fatty acid acyltransferase, which activates HlyA via acylation. HlyB and HlyD are inner

membrane proteins specific for the transport of HlyA and they recruit TolC, a multifunctional outer membrane located protein.

Although we did not appear to have the complete ORF on the pTfars1a construct, comparisons of the predicted amino acid sequence with the homologues indicate that the protein present on the construct is almost complete. With this in mind and considering that the homologues of this protein are involved in the secretion of toxins from the cell, it may be worth investigating the effect of the presence of this ORF on arsenic resistance. However, the gene does not appear to be transcribed along with the *arsRC* genes as it is separated from these genes by 738bp. We would, therefore, predict that this ORF is not involved in arsenic resistance.

When we performed a BLAST search on the region downstream of ORF1 on contig 10004a from the *At. ferrooxidans* genome sequencing database at www.tigr.org, we found that the region immediately downstream of ORF1 showed homology to putative ABC-type transporters and ATP-binding proteins. It also contained two domains showing homology to the ABC transporter domain (Pfam0005). It is therefore possible that, similar to HlyD and HylB, these two proteins function together to transport an as yet unidentified protein.

4.4.2. ORF5

The predicted protein from open reading frame five, located 107bp downstream of the *arsBH* genes, shows strong homology to the signal particle protein 54kDa (SRP54) homologue, Ffh, from *E. coli* and *B. subtilis*. This protein forms part of a nucleoprotein, the signal recognition particle or SRP, that is involved in the targeting of integral membrane proteins and secretory proteins to the plasma membrane. The signal recognition particle (SRP) was originally identified in mammalian cells and was found to consist of six polypeptides (named for their apparent molecular weights: SRP6, SRP14, SRP19, SRP54, SRP68 and SRP72) and one RNA molecule (7S SRP RNA). With the sequencing of the genomes of many different organisms, homologues to the signal recognition proteins have been found in many of these organisms and are highly conserved. Bacteria appear to have a minimal SRP, consisting only of a homologue to SRP54 (Ffh) and an RNA molecule (4.5S RNA). An alignment of ORF5 protein from *At. ferrooxidans* with known Ffh proteins from

E. coli and *B. subtilis* show that there is an overall homology of 66.88% between all three proteins (Figure 4.3).

Figure 4.3: Alignment of the ORF5 protein with Signal particle protein 54 homologues (Ffh) from *E. coli* (accession number P07019) and *B. subtilis* (P37105).

A model for the mechanism by which SRP targets proteins to the membranes has been predicted (Figure 4.4, reviewed in Keenan *et al.*, 2001). The SRP targets the protein to the endoplasmic reticulum (in eukaryotes) or the plasma membrane (in prokaryotes) by binding to the hydrophobic signal sequence at the beginning of the

protein just as the nascent polypeptide is emerging from the ribosome. In eukaryotes this binding temporarily stops any further translation by the ribosome. The SRP-nascent polypeptide-ribosome complex then binds to the membrane via a GTP-dependent interaction with a signal recognition particle receptor. In mammalian cells this receptor consists of two subunits, while the receptor from *E. coli* consists of only one polypeptide (FtsY) which shows homology to the α-subunit of the mammalian receptor. In eukaryotes, this interaction with the receptor allows dissociation of the ribosome complex from the SRP and translation continues with the nascent peptide feeding directly through the membrane via the translocon (protein translocation channel). GTP-hydrolysis also releases the SRP from the SRP receptor, freeing the SRP to bind to another signal sequence. This mechanism ensures that proteins do not fold or misfold in the cytoplasm.

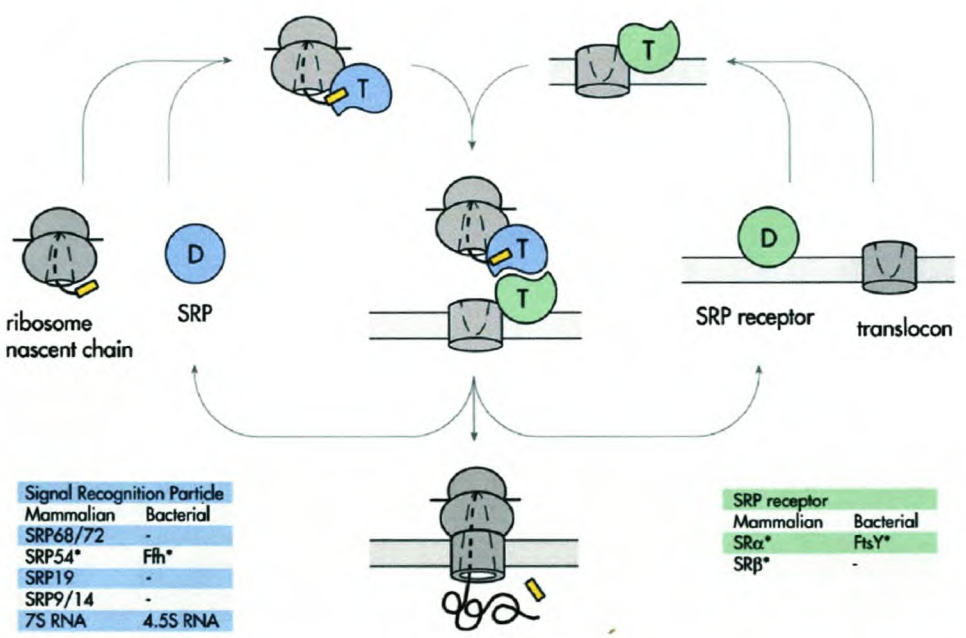


Figure 4.4: The mechanism of protein targeting to the membrane via the signal recognition particle. See text for details. (Taken from Keenen *et al.*, 2001) The SRP and SRP receptor are labeled D when in the GDP-bound state and T when in the GTP-bound state. The membrane represents the eukaryotic endoplasmic reticulum membrane or the prokaryotic plasma membrane. The small rectangle represents the hydrophobic signal sequence of the nascent protein chain.

The Ffh protein consists of three domains: An N-terminal domain of unknown function; the G-domain, which binds GTP; and a methione rich carboxy-terminal or M-domain, which is believed to bind to the hydrophobic signal sequences. All three of these domains are conserved in the *At. ferrooxidans* homologue.

In order to confirm that this open reading frame did not affect resistance to arsenic, we deleted a portion of the 3' end of the *ORF5* gene from pTfarsCRBH to create pTfarsCRBHΔ5. There was no significant difference in the level of arsenite resistance observed between *E. coli* AW3110 cells containing either pTfarsCRBH or pTfarsCRBHΔ5 as determined by growth curves in increasing arsenite concentrations (results not shown).

4.4.3. ORFs 6, 7 and 8:

These three proteins are homologues of three out of the four proteins encoded by the *trmD* operon (Table 2.2). This operon is also located downstream of the *ffh* gene on the *E. coli* genome (Persson *et al.*, 1995). The four proteins encoded by the *trmD* operon are ribosomal protein S16 (encoded by *rpsP*), RimM (encoded by *rimM*, previously called 21K), TrmD (tRNA (m^1G37) methyltransferase encoded by *trmD* gene) and ribosomal protein L19 (encoded by *rplS*).

The first open reading frame of this operon (ORF6) encodes a putative ribosomal protein S16. This protein is believed to play a role in the assembly of the 30S subunit of the ribosome, but does not appear to be actively involved in translation (Held and Nomura, 1975). Recently it has been shown that S16 is a magnesium-dependent DNase (Oberto *et al.*, 1996) that binds preferentially to cruciform DNA structures and nicks the DNA in a sequence specific manner (Bonnefoy, 1997).

The protein encoded by open reading frame seven (ORF7) shows homology to the RimM protein, which is thought to be important for the maturation of 30S ribosomal subunits. This protein is only associated with free 30S subunits and not with those associated with the 70S ribosomal subunit. A deletion of the *rimM* gene results in a seven-fold decrease in growth rate (Lövgren and Wikström, 2001).

The final open reading frame (ORF8) on pTfars1a is incomplete, but shows homology to tRNA (m^1G37) methyltransferases. These proteins modify a guanine at position 37, next to the anticodon, on tRNA molecules reading CCN and CGG from *E. coli* and *S. typhimurium* (Li and Björk, 1995). This modification is believed to be important for maintaining the correct reading frame during translation. A deletion of the *trmD* gene results in a five-fold reduced growth rate.

4.5. Arsenic resistance in *At. ferrooxidans* strains from arsenic rich environments and the possibility of arsenic resistance in other biomining bacteria.

The results discussed above show that arsenic resistance genes are present on the chromosome of *At. ferrooxidans* strain ATCC 33020. However this strain was not isolated from an arsenic rich environment. It has been shown that arsenic resistance in strains isolated from the biomining environment can be increased by selection in a chemostat (Rawlings and Woods, 1995). What is the mechanism of this increased resistance? Do these strains acquire additional genes (e.g. as was proposed for the addition of ArsA) that increase the efficacy of the arsenite efflux pump, do they amplify the arsenic resistance genes present on the chromosome in order to make more copies of the genes, or do the genes present mutate to become more efficient either in expression or function?

Kondratyeva *et al.* (1995) suggested that both arsenic and zinc-resistant *At. ferrooxidans* strains have increased copy numbers of chromosomal genes encoding resistance. The authors compared the *Xba*I restriction endonuclease digestion profile of an *At. ferrooxidans* strain VKM-458 that had been selected for increased arsenic resistance with the parent strain, by pulsed-field gel electrophoresis (PFGE). They found one fragment that showed brighter fluorescence in the arsenic resistant strain compared to the parent strain. The authors proposed that this increase in fluorescence was due to the amplification of the arsenic resistance genes that they propose are located on this fragment. (They observed a similar increase of fluorescence of one fragment when comparing a zinc resistant strain to its parent strain.)

We attempted to select for an *At. ferrooxidans* strain with increased arsenic resistance by subculturing a pure culture of *At ferrooxidans* ATCC33020 in tetrathionate

medium with increasing concentrations of sodium arsenite. We isolated a strain that was able to grow on 38mM sodium arsenite. We investigated whether the observed increase in arsenic resistance was due to the amplification of the *ars* genes by Southern hybridization. The chromosomal DNA from the *At. ferrooxidans* strain with increased arsenite resistance and that from the parent strain was probed with a DNA probe of the *At. ferrooxidans ars* operon. However, we found it difficult to prepare sufficient intact chromosomal DNA from the arsenic resistant strain. In preliminary results (not shown) it appeared that a fragment corresponding to the *arsB* gene, but not the *arsC* gene might be amplified, however we were unable to confirm these results. Nevertheless, the genetic work reported in this thesis will serve as a basis for comparison with the *ars* genes of a highly arsenic resistant *At. ferrooxidans* strain and should help identify the molecular basis for increased arsenic resistance.

As discussed in chapter one, it is now known that other strains such as *At. caldus* and *L. ferrooxidans* are the predominant species in most of the stirred-tank biomining processes. As arsenic is released upon oxidation of the ore (catalyzed by these bacteria), we think that it is likely that these bacteria will also contain arsenic resistance genes. Dopson *et al.* (2001) found that arsenic resistance in *At. caldus* was inducible. If the bacteria were grown overnight in the presence of a nontoxic concentration of arsenite they were then able to grow in previously toxic concentrations of arsenite, arsenate or antimony. They also found that if cultures of uninduced cells were grown for prolonged incubation times in the presence of toxic concentrations of arsenite they did eventually grow. This indicated that over time increased resistance to arsenic did occur within the culture. The authors performed the first biochemical studies on the mechanism of arsenic resistance in acidophiles and found that in *At. caldus* arsenite uptake was energy-independent, while arsenate-uptake appeared to occur via the inorganic phosphate uptake system and was energy-dependent. They also showed that resistance to arsenic was energy-dependent (although they were not able to determine if it was dependent on ATP or membrane potential) and that resistance was due to the efflux of arsenite from the cells (determined by the efflux of 73-arsenite). These results are similar to the results observed for the *ars* systems of neutrophiles, suggesting that *At. caldus* employs a similar arsenic resistance mechanism.

Dopson *et al.* (2001) went on to shown that all the acidophiles they tested (including two species found in the consortium of bacteria involved in biooxidation of ores, *At. thiooxidans* and *At. caldus*) contain an *arsB* homologue, as determined via Southern hybridization with the *arsB* gene from *E. coli*. Under the same conditions they did not observe hybridization to the *arsA* or *arsC* genes from *E. coli*. However, it did appear that the *At. caldus* strain KU reduced arsenate to arsenite. This was determined using an *E. coli arsRD-lacZ* fusion, which is induced by arsenite, to assay the level of arsenite produced from induced and uninduced *At. caldus* KU cultures grown in the presence of arsenate. It may be that the *At. caldus* arsenic resistance operon also contains an *arsC* with homology to the ArsC proteins of the Trx clade. Since this group of proteins has weak homology to the *arsC* genes from *E. coli* and other GSH/Grx family members, the *arsC* gene would not have been identified via Southern hybridization.

Our research has shown that it is possible to use complementation of an *E. coli ars* mutant to isolate the arsenic resistance genes from these acidophiles. If, as the results of Dopson *et al.*, 2001 suggest, the other acidophiles involved in biomining also have the same resistance mechanism, it should be possible to isolate and characterize their *ars* genes by the same means. It would be of particular interest to isolate the *ars* genes from a bacterium isolated from the arsenic-rich biomining environment. Given that this environment is not sterile, these bacteria may have been able to improve their levels of arsenic resistance by horizontal gene transfer. While plasmids have been identified from some of the biomining bacteria, to date none of these have contained homologues to the arsenic resistance genes. It has been proposed that the addition of ArsA to the *E. coli ars* operon, has created a more efficient arsenite efflux pump. It would be interesting to see if these bacteria have acquired this gene or some other gene that would provide an alternative source of energy and whether they have some or all of the unusual features observed in the *At. ferrooxidans ars* operon.

Appendix One: Additional Clones constructed during sequencing of pTfars1b.

Plasmid	Construction	notes
pTfars1a	Ap^r (from *At. ferrooxidans* plasmid library -7 kb *Sau*3A1 fragment cloned into pEcoR251 digested with *Bgl*II)	plasmid bank constructed by Ramesar, 1988.
pBBBH	Ap^r (*Hin*dIII-*Bgl*II fragment from pTfars1a cloned into pBluescript digested with *Bam*HI and *Hin*dIII)	later named pTfars1b
pBBSH	Ap^r (*Hin*dIII-*Stu*I fragment of pTfars1a cloned into pBluescript digested with *Hin*dIII and *Eco*RV)	later named pTfarsCRBH
pBBBEv	Ap^r (*Eco*RV-*Bgl*II fragment of pTfars1a cloned into pBluescript digested with *Eco*RV and *Bam*HI)	
pBBXEv	Ap^r (*Eco*RV-*Xho*I fragment from pTfars1a cloned into pBluescript digested with *Eco*RV and *Xho*I)	
pBBaK	Ap^r (*Kpn*I deletion of pBBSH)	later named pTfarsBH
pBBaB	Ap^r (*Bam*HI deletion of pBBSH)	
pBBa12	Ap^r (1.2 kb *Pst*I-*Eco*RV fragment from pBBBH cloned into pBluescript digested with *Pst*I and *Eco*RV)	
pBB08	Ap^r (0.8 kb *Pst*I-*Eco*RV fragment from pBBBH cloned into pBluescript digested with *Pst*I and *Eco*RV)	
pS18	Ap^r (1.8 kb *Sph*I fragment from pBBaK cloned into pUCBM21 digested with *Sph*I)	
pS06	Ap^r (0.6 kb *Sph*I fragment from pBBaK cloned into pUCBM21 digested with *Sph*I)	later named pTfarsR
pPX180	Ap^r (*Pst*I-*Xho*I fragment from pBBaK cloned into pUCBM21 digested with *Pst*I and *Sal*I)	
pKH175	Ap^r (*Kpn*I-*Hin*dIII fragment from pBBak cloned into pBluescript digested with *Kpn*I and *Hin*dIII)	
pPXdel1	Ap^r (*Sma*I-*Stu*I deletion of pPX180)	
pPXdel2	Ap^r (*Sph*I deletion of pPX180)	
pBHAdel	Ap^r (*Apa*I deletion of pBBBH)	
pBHXdel	Ap^r (*Xho*I deletion of pBBBH)	
pNP-BBBH	Ap^r (2.3 kb *Nco*I-*Pst*I fragment from pBBBH cloned into pUCBM21 digested with *Nco*I and *Pst*I)	
pNP-PX180	Ap^r (1.0 kb *Nco*I-*Pst*I fragment from pPX180 cloned into pUCBM21 digested with *Nco*I and *Pst*I)	
pNPKdel	Ap^r (*Kpn*I deletion of pNP-BBBH)	
pBB06	Ap^r (0.6 kb *Sma*I fragment from pKH175 cloned into pBluescript digested with *Sma*I)	

Appendix Two: Annotated sequence obtained from pTfars1b

Shown below is double stranded sequence of the *At. ferrooxidans* chromosomal fragment contained in pTfars1b determined during the course of this study. This sequence was deposited in the NCBI database: accession number AF173880. Protein translations of the eight open reading frames encoded on this fragment have been shown below the DNA sequence (incomplete open reading frames are indicated with *). Restriction endonuclease sites are labelled and the corresponding sequence italicized. The sequence to which primers were designed is underlined and labelled and putative ribosome binding sites are shown in bold.

*Hin*dIII

1 *AAGCTT*TTGCGGCTGATTGTCTGTCAGCGGAATATGAACGGTGACGGGCATACCCAGGCG
*TTCGAA*AAACGCCGACTAACAGACAGTCGCCTTATACTTGCCACTGCCCGTATGGGTCCGC
L K Q P Q N D T L P I H V T V P M G L R

61 CAGGCGATGCTGCGGATTGCAGGCATAGACCCGCACCCGGTACACCAGTTCCGTACGCAA
GTCCGCTACGACGCCTAACGTCCGTATCTGGGCGTGGGCCATGTGGTCAAGGCATGCGTT
L R H Q P N C A Y V R V R Y V L E T R L

*Eco*RI

121 CTCCGTGGTTTGCACGGTCTTGGGCGT*GAATTC*CGCAGTGGGTGAAATGAAACCCACCCA
GAGGCACCAAACGTGCCAGAACCCGCA*CTTAAG*GCGTCACCCACTTTACTTTGGGTGGGT
E T T Q V T K P T F E A T P S I F G V W

*Sma*I *Sma*I

181 G*CCCGGG*AAAGATTTG*CCCGGG*AAGGAATCACTGCTAATCGTCGCCTTCATGCCCAGCCG
C*GGGCCC*TTTCTAAAC*GGGCCC*TTCCTTAGTGACGATTAGCAGCGGAAGTACGGGTCGGC
G P F S K G P F S D S S I T A K M G L R

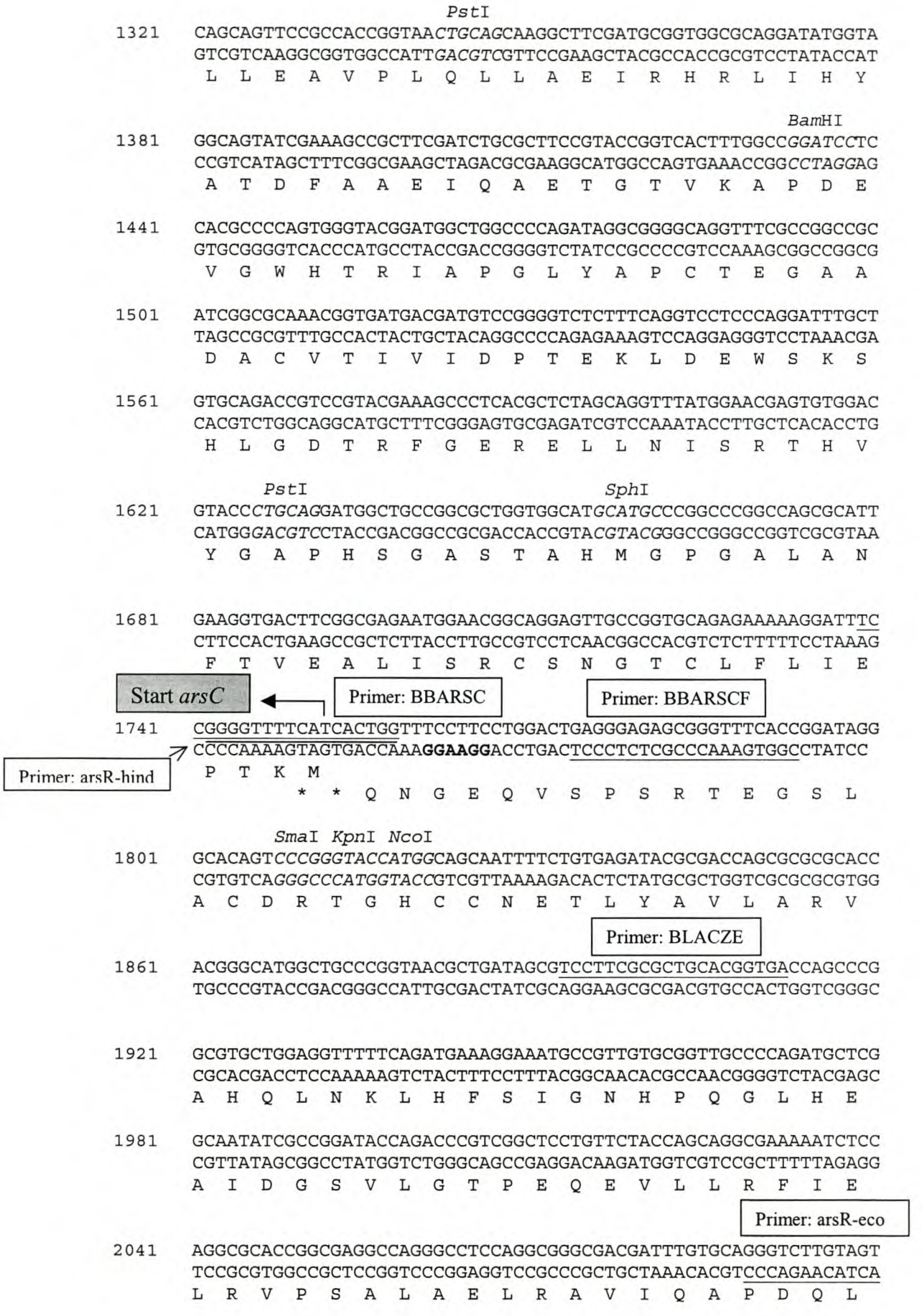
241 CACCTGCCCCAATGCCTTTTCCGGCAGATAGGCGCGCACCCAGACCGGATTGTCCAATGC
GTGGACGGGGTTACGGAAAAGGCCGTCTATCCGCGCGTGGGTCTGGCCTAACAGGTTACG
V Q G L A K E P L Y A R V W V P N D L A

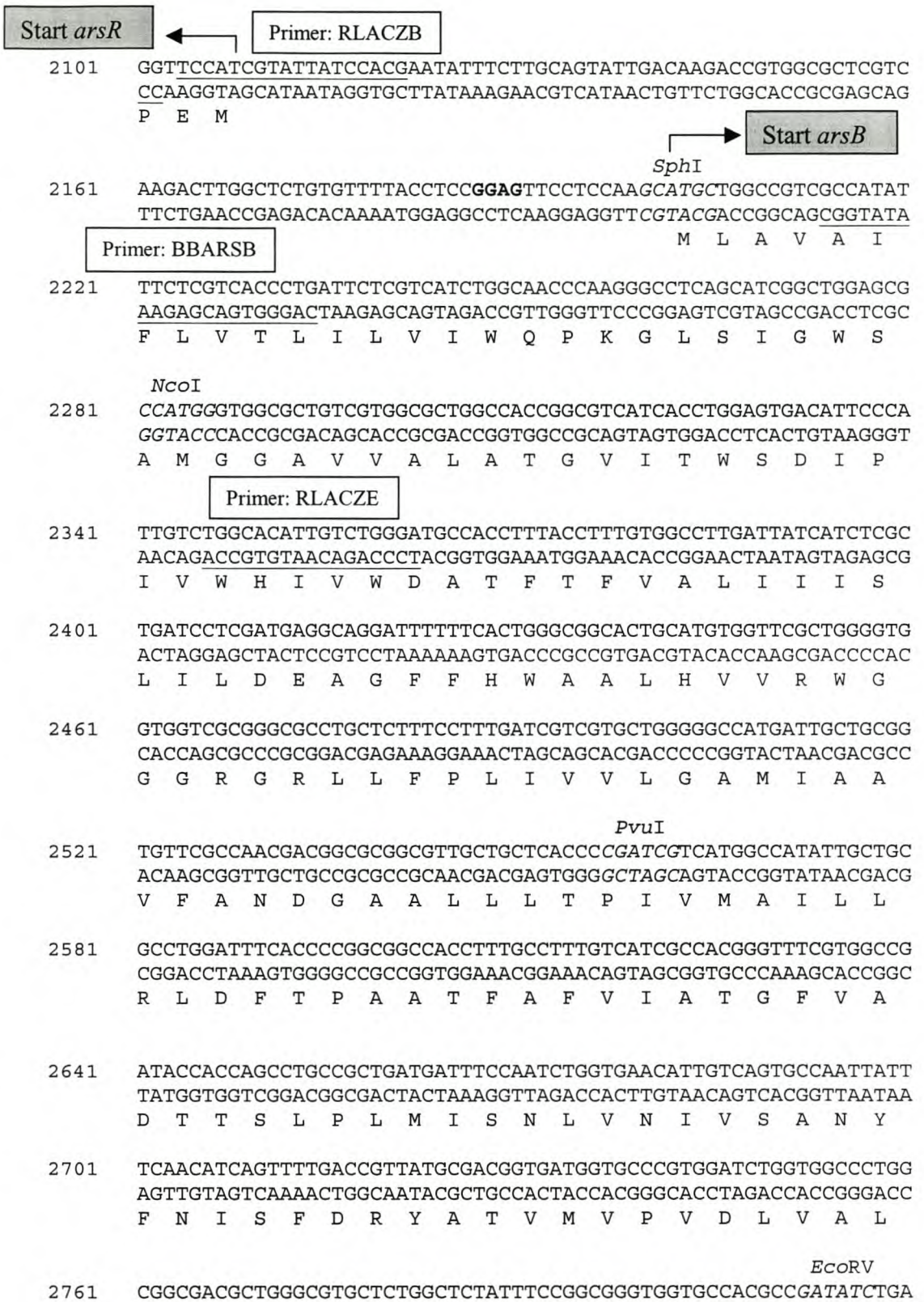
301 CAAAGTGAATACTGGAGTCTGCGGGGAAACCATGTCGCCCGGCTCCAGAATACGGTCCTG
GTTTCACTTATGACCTCAGACGCCCCTTTGGTACAGCGGGCCGAGGTCTTATGCCAGGAC
L T F V P T Q P S V M D G P E L I R D Q

361 AACGACACCGTCTTCTGGCGCATAAAGACGGGTGTCCGTGAGTTCGCGGCGCGCCAGAGA
TTGCTGTGGCAGAAGACCGCGTATTTCTGCCCACAGGCACTCAAGCGCCGCGCGGTCTCT
V V G D E P A Y L R T D T L E R R A L S

421 CAGACCTGCCTTATCGGCTTGTAACTGCTGCCGCGCCGCCGCAATATCTTCTTTGCGCGG
GTCTGGACGGAATAGCCGAACATTGACGACGGCGCGGCGGCGTTATAGAAGAAACGCGCC
L G A K D A Q L Q Q R A A A I D E K R P

```
481   CCCCTTGATGGCCAGGGTCAGTGCCTGCTGGGCGCGGTCCAGATTGGCGCGAGCCGTTTT
      GGGGAACTACCGGTCCCAGTCACGGACGACCCGCGCCAGGTCTAACCGCGCTCGGCAAAA
       G  K  I  A  L  T  L  A  Q  Q  A  R  D  L  N  A  R  A  K  T

541   CAGGGCGGCGGCGGCGTTGTCGAGACTCTGTTTGGGCACATATTGCCGTGCGGCAAGGGC
      GTCCCGCCGCCGCCGCAACAGCTCTGAGACAAACCCGTGTATAACGGCACGCCGTTCCCG
       L  A  A  A  A  N  D  L  S  Q  K  P  V  Y  Q  R  A  A  L  A

601   CTGCTGGCGTTGCCAGGTGATCTCGGCATTGCTCAGGGTCGCCTGAGCGGCAGCGGGCTT
      GACGACCGCAACGGTCCACTAGAGCCGTAACGAGTCCCAGCGGACTCGCCGTCGCCCGAA
       Q  Q  R  Q  W  T  I  E  A  N  S  L  T  A  Q  A  A  A  P  K

          SmaI
661   CGGCCCGGGCCTCGGCGATTTCTTCCGGACGGGAACCCGCCAGCAGGCGCGCCAGCACCT
      GCCGGGCCCGGAGCCGCTAAAGAAGGCCTGCCCTTGGGCGGTCGTCCGCGCGGTCGTGGA
       P  G  P  R  P  S  K  K  R  V  P  V  R  W  C  A  R  W  C  R

721   GCTCTTGCGCGGCCATACTGCGGCATCTTTATCTACGGCGTCCTGAAACCGTACGGGGTC
      CGAGAACGCGCCGGTATGACGCCGTAGAAATAGATGCCGCAGGACTTTGGCATGCCCCAG
       S  K  R  P  W  V  A  A  D  K  D  V  A  D  Q  F  R  V  P  D

781   CAGATCCGCCAGCAATTGTCCCTTTTTTACCCGGTCGCCCTCCTGTACCCGGAGATCGAG
      GTCTAGGCGGTCGTTAACAGGGAAAAAATGGGCCAGCGGGAGGACATGGGCCTCTAGCTC
       L  D  A  L  L  Q  G  K  K  V  R  D  G  E  Q  V  R  L  D  L

841   AAGCCGGCCGTTGTCGTCAAAGGCCGCCTGTACCTGGCGGATGTCGATATTGCCGTAGAT
      TTCGGCCGGCAACAGCAGTTTCCGGCGGACATGGACCGCCTACAGCTATAACGGCATCTA
       L  R  G  N  D  D  F  A  A  Q  V  Q  R  I  D  I  N  G  Y  I

901   GGTGACGGTTTTTTCTGGGGCGTGATGGCGGCTCAGAAAATAATAGGCGACCGCACCAGC
      CCACTGCCAAAAAAGACCCCGCACTACCGCCGAGTCTTTTATTATCCGCTGGCGTGGTCG
       T  V  T  K  E  P  A  H  H  R  S  L  F  Y  Y  A  V  A  G  A

                                          Start ORF1*
961   AACGATAAGCACTACTACCAGTGCCAGAGCACGTTTTTTGGGGGAAATGGCCATATTTAT
      TTGCTATTCGTGATGATGGTCACGGTCTCGTGCAAAAAACCCCCTTTACCGGTATAAATA
       V  I  L  V  V  V  L  A  L  A  R  K  K  P  S  I  A  M

1021  CTCCGATGATTATGCGCGTTATGTTAGAACTTCGCGCCCTATTGACAGGATATTGACGGA
      GAGGCTACTAATACGCGCAATACAATCTTGAAGCGCGGGATAACTGTCCTATAACTGCCT

                                  SmaI
1081  TTCCAGATACGGTCGCGCCGCGCTGCCCGGGAGTTGGGGGGGGGGTCTGGATGTCGGGATC
      AAGGTCTATGCCAGCGCGGCGCGACGGGCCCTCAACCCCCCCCCCAGACCTACAGCCCTAG

                              BamHI
1141  GACTGGGACACAGCCAGCCGGATCCCTCTAGGAGAAACGCATTAAGTGACGGGGATGCTC
      CTGACCCTGTGTCGGTCGGCCTAGGGAGATCCTCTTTGCGTAATTCACTGCCCCTACGAG

                                                  Primer: BBARSCR
1201  ACCAAGCCGTTCGTTTGCAGAATCATCGTACCCCATGCGCTGTGCGCCTTGATGGGTGGC
      TGGTTCGGCAAGCAAACGTCTTAGTAGCATGGGGTACGCGACACGCGGAACTACCCACCG

1261  TTAAGGCAGCAAAGTGCCGATGCGCTCCAGTTCCTGCCGCAGTTTTGCCGGATCTTTCTC
      AATTCCGTCGTTTCACGGCTACGCGAGGTCAAGGACGGCGTCAAAACGGCCTAGAAAGAG
       *  P  L  L  T  G  I  R  E  L  E  Q  R  L  K  A  P  D  K  E
```





```
2881   TCCCGGTTCTGATTCTGTTGTTGGTGGCCTATTTTGTCACTGCGCAATGGCAGGTTCCGG
       AGGGCCAAGACTAAGACAACAACCACCGGATAAAACAGTGACGCGTTACCGTCCAAGGCC
       F  P  V  L  I  L  L  L  V  A  Y  F  V  T  A  Q  W  Q  V  P

2941   TCTCCGTGGTGACCGGAACCGGTGCGTTGATCCTGTTGGCTCTGGCCGGGCGTTGGCTAC
       AGAGGCACCACTGGCCTTGGCCACGCAACTAGGACAACCGAGACCGGCCCGCAACCGATG
       V  S  V  V  T  G  T  G  A  L  I  L  L  A  L  A  G  R  W  L

3001   AGGGCGGGCGCGGCGCGCGGATTCCCGTGCGCAAAGTGCTGCGGGAAGCCCCTTGGGCGA
       TCCCGCCCGCGCCGCGCGCCTAAGGGCACGCGTTTCACGACGCCCTTCGGGGAACCCGCT
       Q  G  G  R  G  A  R  I  P  V  R  K  V  L  R  E  A  P  W  A

       PvuI
3061   TCGTGGTCTTCAGTCTGGGTATGTATCTGGTGGTTTATGGATTACGCAACGCCGGACTCA
       AGCACCAGAAGTCAGACCCATACATAGACCACCAAATACCTAATGCGTTGCGGCCTGAGT
       I  V  V  F  S  L  G  M  Y  L  V  V  Y  G  L  R  N  A  G  L

                 NcoI
3121   CCAGTTATGTGGCCATGGCCCTGCACTGGTTTGCTGGTCACGGGACCGTAGCAGCGGCGC
       GGTCAATACACCGGTACCGGGACGTGACCAAACGACCAGTGCCCTGGCATCGTCGCCGCG
       T  S  Y  V  A  M  A  L  H  W  F  A  G  H  G  T  V  A  A  A

3181   TGGGCACGGGGTTTCTCGCGGCCCTGCTGTCTTCCATCATGAACAATATGCCAGCGGTAC
       ACCCGTGCCCCAAAGAGCGCCGGGACGACAGAAGGTAGTACTTGTTATACGGTCGCCATG
       L  G  T  G  F  L  A  A  L  L  S  S  I  M  N  N  M  P  A  V

3241   TGGTGGGAGCGCTGGCTATCCATCAACTGTCCTCGGTAGCAACAGACCCCTTGCTCCGGG
       ACCACCCTCGCGACCGATAGGTAGTTGACAGGAGCCATCGTTGTCTGGGGAACGAGGCCC
       L  V  G  A  L  A  I  H  Q  L  S  S  V  A  T  D  P  L  L  R

                                                              PvuI
3301   AGATCATGGTCTACGCCAATGTTATCGGCTGTGATCTCGGCCCCAAGTTTACCCCGATCG
       TCTAGTACCAGATGCGGTTACAATAGCCGACACTAGAGCCGGGGTTCAAATGGGGCTAGC
       E  I  M  V  Y  A  N  V  I  G  C  D  L  G  P  K  F  T  P  I

3361   GCAGCCTGGCGACCCTGCTCTGGCTGCATGTGCTGAGTCGCAAAGGTATGACCGTGACCT
       CGTCGGACCGCTGGGACGAGACCGACGTACACGACTCAGCGTTTCCATACTGGCACTGGA
       G  S  L  A  T  L  L  W  L  H  V  L  S  R  K  G  M  T  V  T

3421   GGGGACAGTACATGAGAACGGGTCTGCTCATTACCCCGCCGGTACTGCTGGTCACGCTAC
       CCCCTGTCATGTACTCTTGCCCAGACGAGTAATGGGGCGGCCATGACGACCAGTGCGATG
       W  G  Q  Y  M  R  T  G  L  L  I  T  P  P  V  L  L  V  T  L

                  Primer: ARSHF                  Start arsH
3481   TGGCCCTGGCCTGGTGGCTGCCGCTGGCTTGAGGAGGATGGTCATCATGTCTGGAAATTT
       ACCGGGACCGGACCACCGACGGCGACCGAACTCCTCCTACCAGTAGTACAGACCTTTAAA
                                                  M  V  I  M  S  G  N  L
       L  A  L  A  W  W  L  P  L  A  *

                          PstI
3541   GCCCAATACCGACGATGTGCTGCTGCAGGTGCCGGATGTCCGGTGTTTGCGCAGTGCCGC
       CGGGTTATGGCTGCTACACGACGACGTCCACGGCCTACAGGCCACAAACGCGTCACGGCG
         P  N  T  D  D  V  L  L  Q  V  P  D  V  R  C  L  R  S  A  A

3601   CGAAACGGATCACCCGCCGCGCATCCTGCTGCTGTACGGCTCCAACCGGGAATGTTCCTA
       GCTTTGCCTAGTGGGCGGCGCGTAGGACGACGACATGCCGAGGTTGGCCCTTACAAGGAT
         E  T  D  H  P  P  R  I  L  L  L  Y  G  S  N  R  E  C  S  Y
```

```
3661  CAGTCGCCTGCTGACGCTGGAGGCGGAACGTCTGCTGCGATACTTCGGCGCGGAAACCCG
      GTCAGCGGACGACTGCGACCTCCGCCTTGCAGACGACGCTATGAAGCCGCGCCTTTGGGC
        S  R  L  L  T  L  E  A  E  R  L  L  R  Y  F  G  A  E  T  R

3721  CGTCTTTCACCCGACGGGACTACCCCTGCCCGACGATGCCCCGGTGACACATCCCAAAGT
      GCAGAAAGTGGGCTGCCCTGATGGGGACGGGCTGCTACGGGGCCACTGTGTAGGGTTTCA
        V  F  H  P  T  G  L  P  L  P  D  D  A  P  V  T  H  P  K  V

3781  GGTGGAGTTGCAGGAGTTGGTAGAATGGTCGGAAGGACAGGTCTGGTGTTCTCCGGAGCG
      CCACCTCAACGTCCTCAACCATCTTACCAGCCTTCCTGTCCAGACCACAAGAGGCCTCGC
        V  E  L  Q  E  L  V  E  W  S  E  G  Q  V  W  C  S  P  E  R

3841  TCATGGCGCCATGACGGGGGTGTTCAAATCGCAGGTGGACTGGATTCCCCTCAATTCTGG
      AGTACCGCGGTACTGCCCCCACAAGTTTAGCGTCCACCTGACCTAAGGGGAGTTAAGACC
        H  G  A  M  T  G  V  F  K  S  Q  V  D  W  I  P  L  N  S  G

3901  GGCAATTCGCCCTACCCAGGGCAAAACGCTGGCGTTAATGCAGGTTTGCGGAGGGTCGCA
      CCGTTAAGCGGGATGGGTCCCGTTTTGCGACCGCAATTACGTCCAAACGCCTCCCAGCGT
        A  I  R  P  T  Q  G  K  T  L  A  L  M  Q  V  C  G  G  S  Q

                                                        SphI
3961  ATCGTTCAATGCGGTCAACCAGATGCGCATTCTGGGCCGCTGGCTGCGCATGCTCACCAT
      TAGCAAGTTACGCCAGTTGGTCTACGCGTAAGACCCGGCGACCGACGCGTACGAGTGGTA
        S  F  N  A  V  N  Q  M  R  I  L  G  R  W  L  R  M  L  T  I

4021  CCCCAACCAGTCCTCCGTTCCCAAGGCTTTTCTGGAGTTTGACGATGGTGGCCGGATGAA
      GGGGTTGGTCAGGAGGCAAGGGTTCCGAAAAGACCTCAAACTGCTACCACCGGCCTACTT
        P  N  Q  S  S  V  P  K  A  F  L  E  F  D  D  G  G  R  M  K

4081  GCCTTCTGCCTACTACGACCGGGTCGTGGACGTCATGGAAGAACTCATGAAATTCACCCT
      CGGAAGACGGATGATGCTGGCCCAGCACCTGCAGTACCTTCTTGAGTACTTTAAGTGGGA
        P  S  A  Y  Y  D  R  V  V  D  V  M  E  E  L  M  K  F  T  L

           SacII
4141  GCTTACCCGCGGGAACAGCGATTATCTGGTGGATCGCTATTCCGAGCGCAAAGAATCCGC
      CGAATGGGCGCCCTTGTCGCTAATAGACCACCTAGCGATAAGGCTCGCGTTTCTTAGGCG
        L  T  R  G  N  S  D  Y  L  V  D  R  Y  S  E  R  K  E  S  A

4201  AGAGGAACTTTCCCGGCGCGTCAATCTACAGAACCTATAGCCGCATGGCTGTTGCAAGGC
      TCTCCTTGAAAGGGCCGCGCAGTTAGATGTCTTGGATATCGGCGTACCGACAACGTTCCG
        E  E  L  S  R  R  V  N  L  Q  N  L  *

   Primer: ARSHF
4261  ATGAGGTTGGGGGTACGCAAGAGCCAACTGACGCATTGCGCGACTGCTCCATACAGTCCA
      TACTCCAACCCCCATGCGTTCTCGGTTGACTGCGTAACGCGCTGACGAGGTATGTCAGGT

                                                  Start ORF5
4321  AAATCCACCCCTACACGAGGGGGACCCATGTTTGACAATCTGACCCAGAAACTGACCCAG
      TTTAGGTGGGGATGTGCTCCCCCTGGGTACAAACTGTTAGACTGGGTCTTTGACTGGGTC
                                    M  F  D  N  L  T  Q  K  L  T  Q

4381  ACTTTCAAGAATCTGCGCGGCCAGGGACGGCTGAGCGAGGACAACATCCGTGACGCCCTG
      TGAAAGTTCTTAGACGCGCCGGTCCCTGCCGACTCGCTCCTGTTGTAGGCACTGCGGGAC
        T  F  K  N  L  R  G  Q  G  R  L  S  E  D  N  I  R  D  A  L

4441  CGCGACGTGCGCCTGGCGCTGCTGGAGGCGGATGTTGCCCTGCCGGTGGTCAAGGACTTT
      GCGCTGCACGCGGACCGCGACGACCTCCGCCTACAACGGGACGGCCACCAGTTCCTGAAA
        R  D  V  R  L  A  L  L  E  A  D  V  A  L  P  V  V  K  D  F
```

*Nco*I

4501 ATCAACAACGTGCGCGAATCGG*CCATGG*GCGAAGCCGTGGTCAAAAGCCTCACGCCGGGG
TAGTTGTTGCACGCGCTTAGCC*GGTACC*CGCTTCGGCACCAGTTTTCGGAGTGCGGCCCC
I N N V R E S A M G E A V V K S L T P G

4561 CAAGTCTTCGTCAAGATCGTCCACGACGAACTGGTACAGTTGATGGGCGCGCATAACGAC
GTTCAGAAGCAGTTCTAGCAGGTGCTGCTTGACCATGTCAACTACCCGCGCGTATTGCTG
Q V F V K I V H D E L V Q L M G A H N D

4621 AGTCTCGATCTCAGCGCCCGTCCCCCGGCGGTGATTCTCCTCGCGGGTTTGCAGGGTTCG
TCAGAGCTAGAGTCGCGGGCAGGGGGCCGCCACTAAGAGGAGCGCCCAAACGTCCCAAGC
S L D L S A R P P A V I L L A G L Q G S

4681 GGTAAGACTACCACTAGTGCCAAGCTCGCCCTGTGGCTCAAGGACAAGGAAAAGAAGCGC
CCATTCTGATGGTGATCACGGTTCGAGCGGGACACCGAGTTCCTGTTCCTTTTCTTCGCG
G K T T T S A K L A L W L K D K E K K R

*Nco*I

4741 GTGCTGATGGTCAGTACCGACGTCTATCGTCCGGCCG*CCATGG*AACAGTTGGTGCATCTG
CACGACTACCAGTCATGGCTGCAGATAGCAGGCCGGC*GGTACC*TTGTCAACCACGTAGAC
V L M V S T D V Y R P A A M E Q L V H L

4801 GGTAAGGACATCGAGGTGGATGTATTTCCCTCCAGCCCCGGCCAGCAGCCGGTGCGTATC
CCATTCCTGTAGCTCCACCTACATAAAGGGAGGTCGGGGCCGGTCGTCGGCCACGCATAG
G K D I E V D V F P S S P G Q Q P V R I

4861 GCCCAGGATGCGGTGGAGGCGGCGCAGCGCGGCGTCTATGACGTGCTCATCGTCGATACG
CGGGTCCTACGCCACCTCCGCCGCGTCGCGCCGCAGATACTGCACGAGTAGCAGCTATGC
A Q D A V E A A Q R G V Y D V L I V D T

4921 GCGGGGCGTCTGCATGTGGACGGCGAGATGATGGCGGAGGCACAGGCACTGACGGCGGCG
CGCCCCGCAGACGTACACCTGCCGCTCTACTACCGCCTCCGTGTCCGTGACTGCCGCCGC
A G R L H V D G E M M A E A Q A L T A A

4981 GTCAAGCCGGTGGAACTGCTCTTTGTGGTGGATGCCATGACCGGTCAGGACGCGGTGAAT
CAGTTCGGCCACCTTGACGAGAAACACCACCTACGGTACTGGCCAGTCCTGCGCCACTTA
V K P V E L L F V V D A M T G Q D A V N

*Stu*I

5041 ACCGCCAA*GGCCT*TTAATGACGCTCTGCCGCTGACCGGCGTGATTCTCACCAAGGCGGAC
TGGCGGT*TCCGGA*AATTACTGCGAGACGGCGACTGGCCGCACTAAGAGTGGTTCCGCCTG
T A K A F N D A L P L T G V I L T K A D

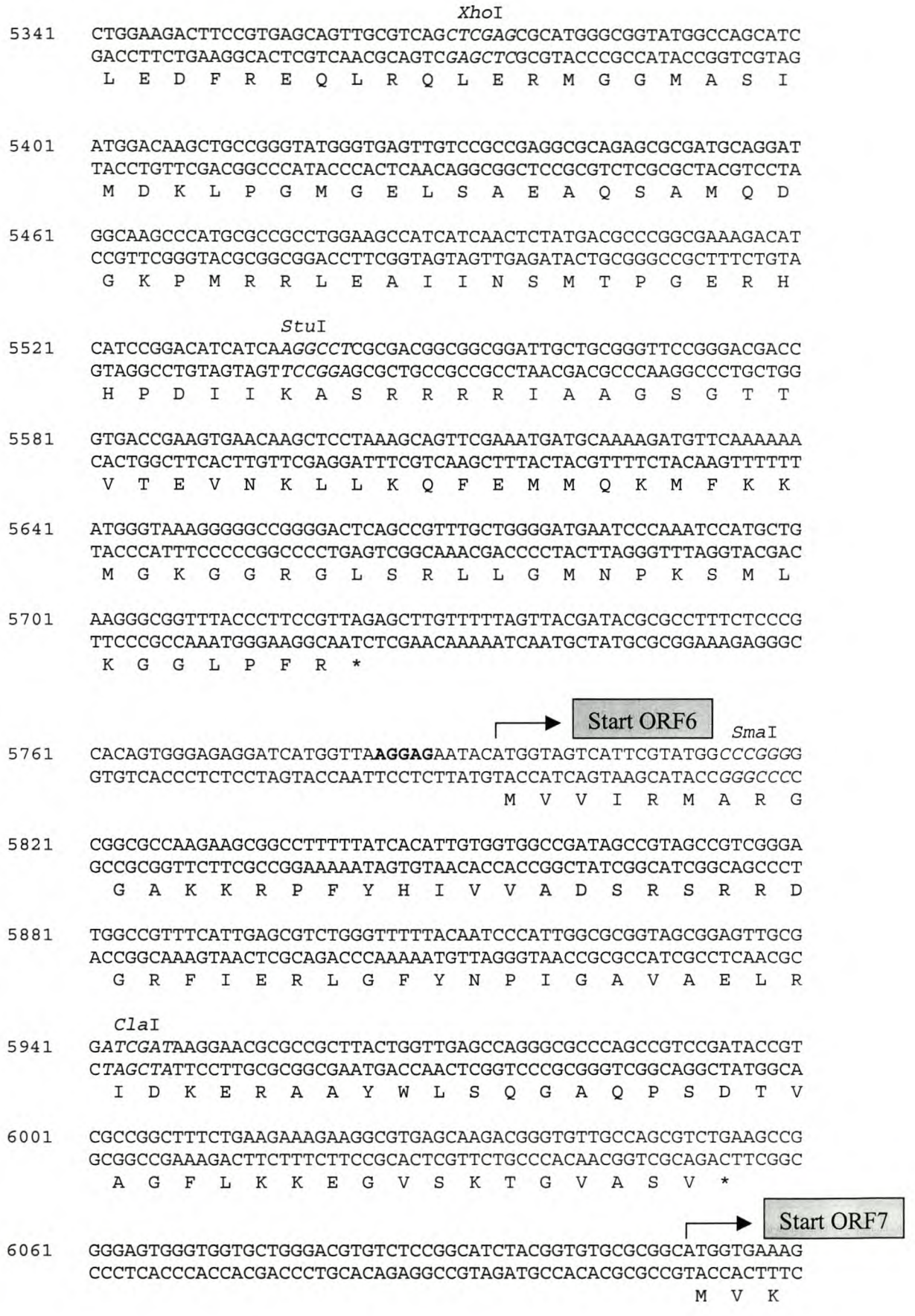
*Sac*II

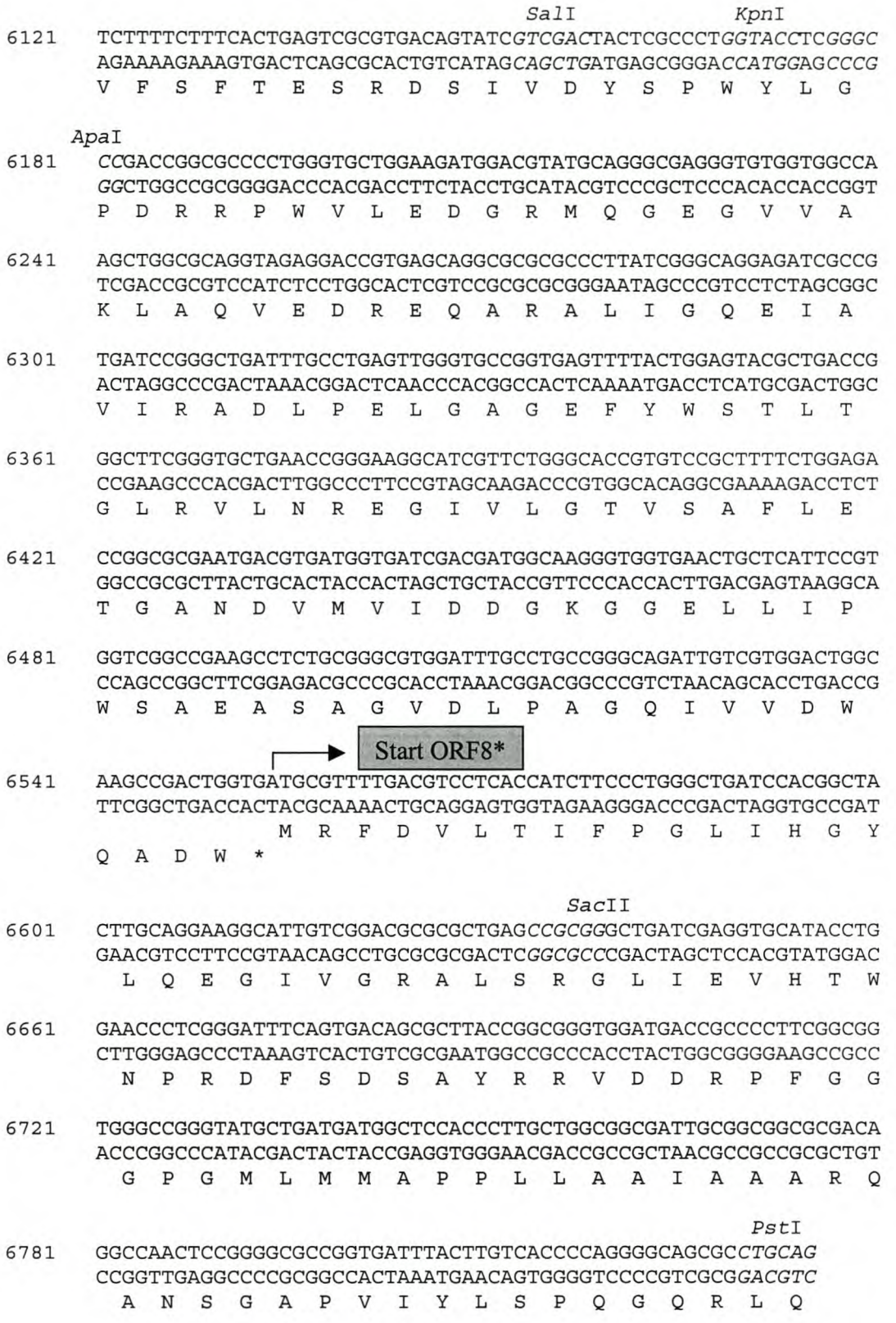
5101 GGCGACGC*CCGCGG*TGGCGCGGCTCTATCGGTGCGAGCCATTACCGGCAAGCCCATCAAG
CCGCTGCG*GGCGCC*ACCGCGCCGAGATAGCCACGCTCGGTAATGGCCGTTCGGGTAGTTC
G D A R G G A A L S V R A I T G K P I K

5161 TTCATCGGTATCGGTGAGAAAATTCGCAAGGGTCTGGAACCGTTTTATCCCGAGCGCATG
AAGTAGCCATAGCCACTCTTTTAAGCGTTCCCAGACCTTGGCAAAATAGGGCTCGCGTAC
F I G I G E K I R K G L E P F Y P E R M

5221 GCCTCGCGCATTCTGGGGATGGGTGACATCGTCAGTTTGGTGGAGCAGGTCCAGCAGGAT
CGGAGCGCGTAAGACCCCTACCCACTGTAGCAGTCAAACCACCTCGTCCAGGTCGTCCTA
A S R I L G M G D I V S L V E Q V Q Q D

5281 GTGGATGAGGATCAGGCCCGGAAAATGACCCAGAAGTTGCGCAGCGGCAAGGGCTTCGAT
CACCTACTCCTAGTCCGGGCCTTTTACTGGGTCTTCAACGCGTCGCCGTTCCCGAAGCTA
V D E D Q A R K M T Q K L R S G K G F D





References

1. **Altshul S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman.** (1997). Gapped BLAST and PSI-BLAST: A new generation of protein data base search programs. *Nucleic Acids Res.* **25**:3308-3402.
2. **Bairoch A.** (1993). A possible mechanism for the metal-ion induced DNA-protein dissociation in a family of prokaryotic transcriptional regulators. *Nucleic Acids Res.* **21**:2515.
3. **Barros M.E., D.E. Rawlings and D.R. Woods.** (1985). Cloning and expression of the *Thiobacillus ferrooxidans* glutamine synthetase gene in *Escherichia coli*. *J. Bacteriol.* **164**:1386-1389.
4. **Bennett M.S.,Z. Guan, M. Laurberg, and X. D. Su.** (2001). *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatase. *Proc. Natl. Acad. Sci USA* **98**:13577-13582.
5. **Bhattacharjee H. and B.P. Rosen.** (1996). Spatial proximity of Cys^{113}, Cys^{172} and Cys^{422} in the metalloactivation domain of the ArsA ATPase. *J. Biol. Chem.* **271**:24465-24470.
6. **Bhattacharjee H. J. Li, M.Y. Ksenzenko and B.P. Rosen.** (1995). Role of cysteinyl residues in metalloactivation of the oxyanion-translocating ArsA ATPase. *J. Biol. Chem.* **270**:11245-11250.
7. **Bhattacharjee H., T. Zhou, J. Li, D.L. Gatti, A.R. Walmsey amd B.P. Rosen.** (2000). Structure-function relationships in an anion-translocating ATPase. *Biochem. Soc. Transactions*. **28**:520-526.
8. **Bonnefoy E.** (1997). The ribosomal S16 protein of *Escherichia coli* displaying a DNA-nicking activity binds to cruciform DNA. *Eur. J. Biochem.* **247**:852-859.
9. **Bruhn D.F., J. Li, S. Silver, F. Roberto and B.P. Rosen.** (1996). The arsenical resistance operon of IncN plasmid R46. *FEMS Microbiol. Lett.* **139**:149-153.

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

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

12. **Cai J. and M.S. DuBow.** (1996). Expression of the *Escherichia coli* chromosomal *ars* operon. *Can. J. Microbiol.* **42,** 662-671.

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

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

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

16. **Cervantes, C., Ji, G., Ramírez, J.L. & Silver, S.** (1994). Resistance to arsenic compounds in microorganisms. *FEMS Microbiol. Rev.* **15,** 355-367.

17. **Chang, A.C.Y. & Cohen, S.N.** (1978). Construction and characterization of amplifyable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J. Bacteriol.* **134,** 1141-1156.

18. **Chen C-M., H. L.T. Mobley and B.P. Rosen.** (1985) Separate resistances to arsenate and arsenite (antimonate) encoded by the arsenical resistance operon of R factor R773. *J. Bacteriol.* **161**:758-763.

19. **Chen C-M., T.K. Misra, S. Silver and B.P. Rosen.** (1986). Nucleotide sequence of the structural genes for the anion pump. *J. Biol. Chem.* **261**:15030-15038.

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

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

22. **Clennel A.M, B. Johnston and D.E. Rawlings.** (1995) Structure and function of Tn5467, a Tn21-like transposon located on the *Thiobacillus ferrooxidans* broad-host range plasmid pTF-FC2. *Appl. Environ. Microbiol.* **61**:4223-4229.

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

24. **Cox, J.C., D.G. Nicholls and W.J. Ingledew.** (1979) Transmembrane electrical potential and transmembrane pH gradient in acidophile *Thiobacillus ferrooxidans*. *Biochem. J.* **178**:195-200.

25. **Dew, D.W., E.N. Lawson and J.L. Broadhurst.** (1997) The Biox® process for biooxidation of gold-bearing ores or concentrates. In *Biomining: Theory, Microbes and Industrial Processes*, pp45-80. Edited by D.E. Rawlings. Berlin:Springer.

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

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

28. **Diorio C. J. Cai, R. Shinder and M.S. DuBow.** (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.

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

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

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

32. **Guan Z., L Hederstedt, J-p. Li and X-D. Su.** (2001) Preparations and crystallization of a *Bacillus subtilis* arsenate reductase. *Acta Cryst. Sect. D.* **57**:1718-1721.

33. **Held W.A. and Nomura M.** (1975). *Escherichia coli* 30 S ribosomal proteins uniquely required for assembly. *J. Biol. Chem.* **250**:3179-3184.

34. **Inoue, H., H. Nojima and H. Okayama.** (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28.

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

36. **Ji, G. and S. Silver.** (1992) Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid, pI258. *Proc. Natl. Acad. Sci. USA* **89**:7974-7978.

37. **Ji, G., E.A. Garber, L.G. Ames, C-M. Chen, J.A. Fuchs and S. Silver.** (1992) Arsenate reductase of *Staphylococcus aureus* plasmid, pI258. *Biochemistry* **33**:7294-7299.

38. **Jia H. and P. Kaur.** (2001). Role of the linker region of the anion-stimulated ATPase ArsA. *J. Biol. Chem.* **276**:29582-29587.

39. **Kahnert A, P. Vermeij, C. Wietek, P. James, T. Leisinger and M.A. Kertesz.** (2000). The ssu locus plays a key role in organosulfur metabolism in *Pseudomonas putida* S-313. *J. Bateriol.* **182**:2869-2878.

40. **Keenan R.J., D.M. Freymann, R.M. Stroud and P. Walter.** (2001) The signal recognition particle. *Annu. Rev. Biochem.* **70**:755-775.

41. **Kelly, D.P. and A.P. Wood.** (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. Evolut. Microbiol.* **50**:511-516.

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

43. **Kertesz M.A., K. Schmidt-Larbig and T. Wuest.** (1999). A novel reduced flavin mononucleotide-dependent methanesulfonate sulfonatase encoded by the sulfur-regulated *msu* operon of *Pseudomonas aeruginosa*. *J. Bacteriol.* **181**:1464-1473.

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

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

46. **Latimer M.T., M.H. Painter and J.G. Ferry.** (1996). Characterization of an iron-sulfur flavoprotein from *Methanosarcina thermophila*. *J. Biol. Chem.* **271**:24023-24028.

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

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

49. **Li J.N. and G.R. Bjork.** (1995). 1-Methylguanosine deficiency of tRNA influences cognate codon interaction and metabolism in *Salmonella typhimurium*. *J. Bacteriol.* **177**:6593-6600.

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

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

52. **Liang E. and I. S. Pretorius.** (1992). Synthesis and secretion of an *Erwinia chrysanthemi* pectate lyase in *Saccharomyces cerevisiae* regulated by different combinations of bacterial and yeast promoter and signal sequences. *Gene* **121,** 35-45.

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

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

55. **Lovgren J.M. and P.M. Wikstrom.** (2001). Hybrid protein between ribosomal protein S16 and RimM of *Escherichia coli* retains the ribosome maturation function of both proteins. *J. Bacteriol.* **183**:5352-5357.

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

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

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

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

60. **Miller J.H.** (1972). *Experiments in Molecular genetics.* Cold Spring Harbor Laboratory. New York: Cold Spring Harbor Press.

61. **Mukhopadhyay R., B.P. Rosen, L.T. Phung and S. Silver.** (2002) Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol. Rev.* **751**:1-15.

62. **Nakamuro K. and Y. Sayato.** (1980). Comparative studies of chromosomal abberation induced by trivalent and pentavalent arsenic. *Mutat. Res.* **88**:73-80.

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

64. **Oberto J, E. Bonnefoy, E. Mouray, O. Pellegrini, P.M. Wikström and J. Rouviere-Yaniv.** (1996). The *Escherichia coli* ribosomal protein S16 is an endonuclease. *Mol. Microbiol.* **19**:1319-1330.

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

66. **Persson B.C., G.O. Bylund, D.E. Berg and P.M. Wikstrom.** (1995). Functional analysis of the *ffh-trmD* region of the *Escherichia coli* chromosome by using reverse genetics. *J. Bacteriol.* **177**:5554-5560.

67. **Powles R.E. and D.E. Rawlings.** (1997) The pyruvate dehydrogenase complex of the chemolithoautotropic bacterium *Thiobacillus ferrooxidans* has an unusual E2-E3 subunit fusion. *Microbiology* **143**:2189-2195.

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

69. **Ramesar R.S.** (1988) *Developmental studies on Thiobacillus ferroxidans*. PhD thesis. University of Cape Town.

70. **Rawlings D.E and D.R. Woods.** (1995) *Development of improved biomining bacteria*. In Bioextraction and biodeterioration of metals, ed. C.C. Gaylarde and H.A. Videla, p 61-84. Cambridge University Press, Cambridge.

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

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

73. **Rawlings D.E. and S. Silver.** (1995) Mining with microbes. *Bio/Technology* **13**:773-778.

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

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

76. **Rosen B.P.** (1999). Families of arsenic transporters. *Trends in Microbiol* **7,** 207-212.

77. **Rosenstein R., A. Peschel, B. Wieland and F. Gotz.** (1992) Expression and regulation of the antimonite, arsenite, and arsenate resistance of *Staphylococcus xylosus* plasmid pSX267. *J. Bacteriol.* **174**:3676-3683.

78. **Ryan D. and E. Colleran.** (2002) Arsenical resistance in the IncH12 plasmids. *Plasmid* **47**:234-240.

79. **Saltikov C.W. and B.H. Olson.** (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. and Environ. Mircobiol.* **68**:280-288.

80. **Sambrook J., E.F. Fritsch and T. Maniatis.** (1989). Molecular Cloning: a Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory. New York:Cold Spring Harbor Press.

81. **San Francisco M.J.D., C.L. Hope, J.B. Owolabi, L.S. Tisa and B.P. Rosen.** (1990). Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon. *Nucleic Acids Res.* **18**:619-624.

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

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

84. **Sanger F., S. Nicklen and A.R. Coulson.** (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.

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

86. **Shi J, A. Vlamis-Gardikas, F. Aslund, A. Holmgren and B.P. Rosen.** (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.

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

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

89. **Shriver D.F., P.W. Atkins and C.H. Langford.** (1994) *Inorganic chemistry*, Second Edition, Oxford University Press, Oxford.

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

91. **Silver S. and M. Walderhaug.** (1992) Gene regulation of plasmid- and chromosome determined inorganic ion transport in bacteria. *Micobiol. Rev.* **56**:195-228.

92. **Silver S., G. Ji, S. Broer, S. Dey, D. Dou and B.P. Rosen.** (1993). Orphan enzyme or patriach of new tribe: the arsenic resistance ATPase of bacterial plasmids. *Mol. Microbiol.* **8:**637-642.

93. **Singh V.K., A. Xiong, T.R. Usgaard, S. Chakrabarti, R. Deora, T.K. Misra, and R.K. Jayaswal.** (1999). ZntR is an autoregulatory protein and negatively regulates the chromosomal zinc resistance operon *znt* of *Staphylococcus aureus*. *Mol Microbiol* **33,** 200-207.

94. **Smith A.S.G. and D.E. Rawlings.** (1998). Efficiency of the pTF-Fc2 pas poison-antidote stability system in *Escherichia coli* is affected by the host strain and antidote degradation requires the Lon protease. *J Bacteriol* **180,** 5458-5462.

95. **Suzuki K., N. Wakao, T. Kimura, K. Sakka and K. Ohmiya.** (1998). Expression and regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU 301 plasmid pK301 in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**:411-418

96. **Tsai K-J., C-M Hsu and B.P. Rosen.** (1997). Efflux mechanisms of resistance to cadmium, arsenic and antimony in prokaryotes and eukaryotes. *Zoological Studies* **36,** 1-16.

97. **van Der Ploeg JR, R. Iwanicka-Nowicka, T Bykowski, M.M. Hryniewicz MM and T. Leisinger.** (1999). The *Escherichia coli ssuEADCB* gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. *J. Biol. Chem.* **274**:29358-29365.

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

99. **Wang H-W., Y-J. Lu, L-J. Li, S. Liu, D-N. Wang and S-f. Sui.** (2000). Trimeric ring-like structure of ArsA ATPase. *FEBS Letters* **469**:105-110.

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

101. **Wu J. and B.P. Rosen.** (1993). Metalloregulated expression of the *ars* operon. *J. Bacteriol.* **268**, 52-58.

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

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

104. **Xu C., T. Zhou, M. Kuroda and B.P. Rosen.** (1998). Metalloid resistance mechanisms in prokaryotes. *J Biochem* **123,** 16-23.

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

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

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

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

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

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