An Investigation into the Arsenic Resistance Genes of *Leptospirillum ferriphilum*

By
Stanton Bevan Ernest Hector

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

Supervisor: Prof. D.E. Rawlings

DECLARATION

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

_____   _____

Stanton Bevan Ernest Hector   **Date**

ACKNOWLEGDEMENTS

Abstract

Leptospirillum ferriphilum is a moderately thermophilic, iron-oxidizing bacterium that was isolated from a continuous-flow biooxidation plant used for the recovery of gold from arsenopyrite ore concentrates. Over many years of continuous selection, *L. ferriphilum* and other bacteria associated with this environment developed resistance to high concentrations of arsenic. We investigated the arsenic resistance genes (*ars*) of *Leptospirillum ferriphilum* strain Fairview and compared these genes to the *ars* genes from other *Leptospirilli*. An arsenic resistance operon (*ars* operon) was isolated from a *L. ferriphilum* Fairview genebank. We discovered that this *ars* operon was situated in between divergently transcribed transposase (*tnpA*) and resolvase (*tnpR*) genes related to the Tn*21* subfamily of transposons. Sequence analysis of this transposon *ars* operon indicated the presence of *arsRCDAB* genes and an additional CBS *orf*, located in between the *arsA* and *arsB* genes. The 8.5 kb *L. ferriphilum* transposon *ars* operon (Tn*LfArs*) was shown to be present only in *L. ferriphilum* strain Fairview and none of the other *Leptospirillum* strains. The Tn*LfArs* conferred resistance to arsenate and arsenite in an *Escherichia coli ars* mutant. We also showed that the Tn*LfArs* is capable of transposition in *Escherichia coli*.
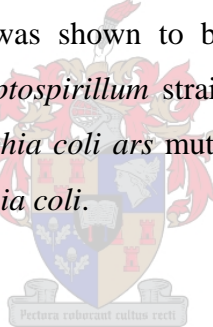
Opsomming

Leptospirillum ferriphilum, 'n matig termofilies, yster-oksiderende bakterium, is een van `n konsortium bakterieë betrokke by die biologiese herwinning van goud uit arsenopiriet erts. Oor vele jare het die selektiewe druk, weens hoë arseen konsentrasies teenwoordig in die erts, veroorsaak dat *L. ferriphilum* en die ander bakteriee geassosieer met die omgewing, verhoogde vlakke van weerstandbiedendheid teen die metaal opgebou het. Die doel van die studie was om die aard van die aanpassing op die molukulere vlak vas te stel deur die gene wat in *L. ferriphilum* (Fairview ras) hiervoor verantwoordelik is te identifiseer en te vergelyk met die van ander *Leptospirilli*. `n Arseen weestandbiedendheids operon (*ars* operon) is met behulp van `n *L. ferriphilum* geen-bank geisoleer. DNA-volgorde bepaling het aangedui dat die operon *arsRCDAB* gene bevat, sowel as `n CBS *orf*, gelee tussen die *arsA* en *arsB* gene. Die hele operon is gelee tussen `n tnpR- (resolvase) en tnpA (transposase) gene wat in teenoorgestelde rigtings getranskribeer word. Hierdie gene behoort aan die Tn*21* familie van transposons. Daar is gevind dat die 8.5 kb *L. ferriphilum* transposon wat die *ars* operon bevat (Tn*LfArs*) slegs teenwoordig is in die Fairview ras van *L. ferriphilum* maar in geen van die ander *Leptospirillum* rasse nie. Die Tn*LfArs* het weerstanbiedendheid verleen, teen beide arsenaat en arseniet, aan `n *Escherichia coli* arseen-sensitiewe mutant. Die vermoë van die transposon (Tn*LfArs*) om transposisie te ondergaan is ook in *E. coli* bevestig.

TABLE OF CONTENTS

Chapter One: General Introduction

TABLE OF CONTENTS

1.1. Biox® Process.

Gencor South Africa Ltd. was a pioneer in the field of biooxidation of refractory gold ore. Gencor developed the Biox® process (which consist of stirred tank, continuous-flow reactors) in the early 1980s, based on innovative research into bacterial oxidation of refractory gold ore prior to cyanidation. A pilot plant was commissioned at the Krugersdorp research laboratories in 1984 in order to treat the flotation concentrate of the Fairview mine and its commercial success led to the decision to construct an industrial plant at the Fairview mine in Barberton in 1986. The Fairview Biox® plant was further extended in 1991, with the intention of treating 40 tonnes of flotation concentrate per day (Dew *et al*., 1997) and its capacity has since been further increased to 55 tonnes (Rawlings *et al*., 2003). The success of the Fairview Biox® plant in Barberton inspired the construction of several similar plants internationally, including Brazil, Australia, Ghana (Dew *et al*., 1997) and more recently Peru and China (Rawlings *et al*., 2003). The Sansu plant in Ghana is the largest fermentation process in the world consisting of 24 tanks of 1,000,000 liters each, processing 1000 tonnes of gold per day (Rawlings, 2002), a testament to the success of the Biox® process. Currently all commercial biooxidation processes use stirred-tanks as reactors with the exception of Youanmi and Beaconsfield in Australia, which use BacTech and Mintek-BioTech technology, respectively. More recently, the Kasese plant in Uganda came in to production using BRGM (Bureau de Recherches Géologiques et Minères) technology, and the Laizhou plant in China, which also uses Mintek-BacTech technology (Rawlings *et al*., 2003). A typical Gencor Biox® flowsheet is shown in Fig. 1.1. It consists of a concentrate feed to the biooxidation plant, a feed make-up tank, a series of mineral aeration tanks, and biooxidation solid/liquid separation tanks. The washed thickener under-flow is sent to cyanidation and gold recovery, while the thickener over-flow is subjected to neutralization, the neutralized product then sent to disposal on a tailings dam.

The operation of a typical biooxidation plant is as follows:

The feed concentrate is finely milled, mixed with water and small amounts of ammonia and phosphate to a density of 18–20 % (Dew *et al*., 1997). This pulp would have a typical residence period of about 4 days in the biooxidation plant. The residence period depends on the oxidation rate achieved, which is dependent on the sulfide content and mineralogical composition of the ore. The biooxidation plant is configured into a primary and secondary stage. Both stages typically consist of equally sized tanks, with three operating in parallel in the primary stage and in series in the secondary stage. This configuration allows a residence period of 2 days in the primary stage and 0.67 days in the secondary stage, depending on sulfide

oxidation rates. A longer period in the primary stage is necessary in order for a stable bacterial population to become established, allowing bacterial attachment to the concentrate. The Biox® culture consists of a mix of mesophilic bacteria that function over a temperature range of 40-50 ºC. Processes operating at 75-80 ºC are still being developed (Rawlings, 2002 and Rawlings *et al.*, 2003). The pulp temperature is controlled at 40-45 ºC to allow maximum rates of sulfide oxidation while minimizing cooling requirements. The biooxidation product is then washed. The washed product contains iron and acid, therefore a second washing step is necessary before gold recovery by cyanidation. Finally, the overflow liquor is neutralized in a two-stage neutralization plant for safe disposal on a tailings dam.
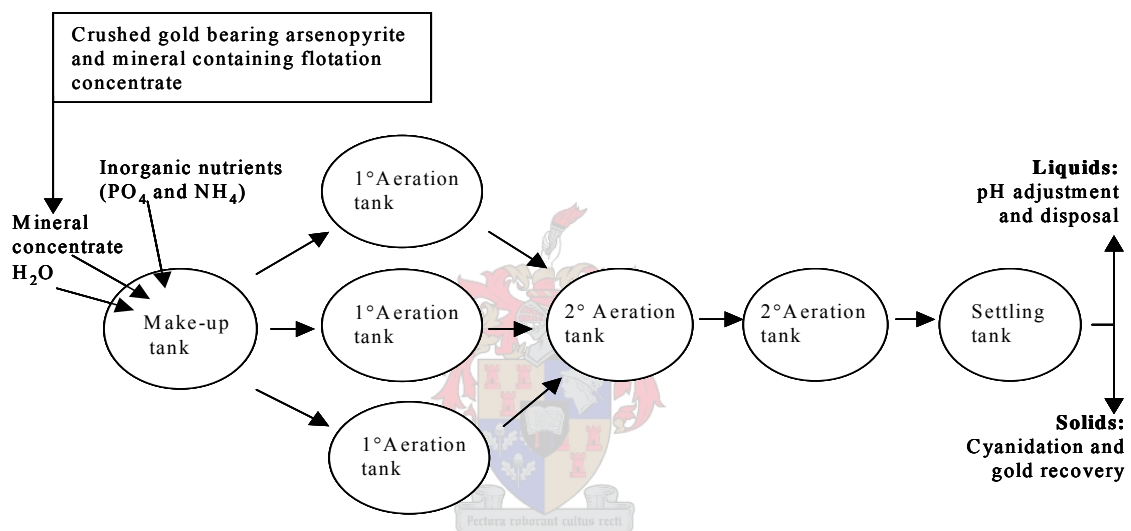
Figure 1.1: Flow diagram depicts a typical commercial biooxidation plant. Gold bearing arsenopyrite ore along with other minerals is crushed to a desired sized powder to prepare a concentrate. Nutrients and water are added to the concentrate in order to insure efficient growth of the bacteria. By continuous-flow the feed from one tank overflows to another until sufficient decomposition of the ore has occurred. In the primary aeration tanks the time of residency is increased allowing for sufficient bacterial growth and attachment. The tanks are continuously aerated for efficient growth of bacteria and to agitate the suspension to stop deposits from forming. Cooling of the tanks is also necessary due to the exothermic biooxidation process. In the final step the solids containing gold collected from the settling tank are sent to cyanidation process for recovery of the gold. (Adapted from Bronwyn Butcher, PhD Thesis)

During the Biox® process the bacteria cause accelerated oxidation of the sulfide minerals releasing gold for recovery by cyanidation. The principle sulfide minerals associated with refractory gold are arsenopyrite, pyrite and pyrrhotite. The bacterial oxidation reactions given for these minerals by Dew *et al.* (1997) are as follows:

$$4FeS_2 \text{ (pyrite)} + 15O_2 + 2H_2O \rightarrow 2Fe_2(SO_4)_3 + 2H_2SO_4 \qquad 1.1$$

$$2FeAsS \text{ (arsenopyrite)} + 7O_2 + H_2SO_4 + 2H_2O \rightarrow 2H_3AsO_4 + Fe_2(SO_4)_3 \qquad 1.2$$

$$4FeS \text{ (pyrrhotite)} + 9O_2 + 2H_2SO_4 \rightarrow 2Fe_2(SO_4)_3 + 2H_2O \qquad 1.3$$
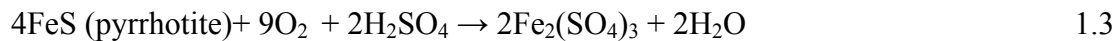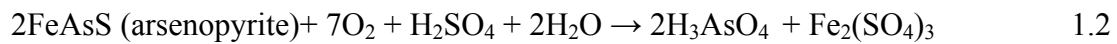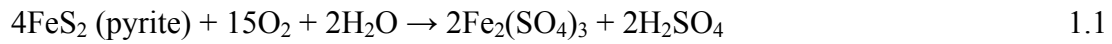
These oxidation reactions indicate the high oxygen requirement of sulfide oxidation. The oxidation reactions are also highly exothermic and occur as follows:

$$FeS + Fe_2(SO_4)_3 \rightarrow 3FeSO_4 + S^{\circ} \qquad 1.4$$

$$FeS_2 + Fe_2(SO_4)_3 \rightarrow 3FeSO_4 + 2S^{\circ} \qquad 1.5$$

$$2FeS + 2H_2SO_4 + O_2 \rightarrow 2FeSO_4 + 2\,S^{\circ} + 2H_2O \qquad 1.6$$

$$4FeSO_4 + 2H_2SO_4 + O_2 \rightarrow 2Fe(SO_4)_3 + 2H_2O \qquad 1.7$$

$$2S^{\circ} + 3O_2 + 2H_2O \rightarrow 2H_2SO_4 \qquad 1.8$$

Secondary reactions resulting from oxidation include precipitation of ferric arsenate ($FeAsO_4$), acid dissolution of carbonates and precipitation of jarosite, which are represented by the following reactions equations:

$$2H_3AsO_4 + Fe(SO_4)_3 \rightarrow FeAsO_4 + 3H_2SO_4 \qquad 1.9$$

$$CaMg(CO_3)_2 + 2H_2SO_4 \rightarrow CaSO_4 + MgSO_4 + 2CO_2 + 2H_2O \qquad 1.10$$

$$3Fe_2(SO_4)_3 + 12H_2O + Mg_2SO_4 \rightarrow 2MFe_3(SO_4)_2(OH)_6 \text{ (jarosite)} + 6\,H_2SO_4 \qquad 1.11$$
where $M^+ = K^+, Na^+, NH_4^+, H_3O^+$
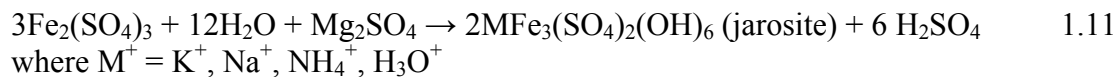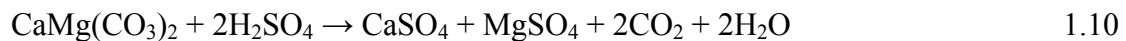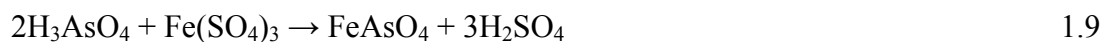
The bacterial oxidation of pyrite/arsenopyrite ore produces arsenic and ferric sulfate.
The conventional method for precipitation of arsenic from solution is by lime neutralization. Biox® liquors are neutralized in a two-stage process employing limestone. In the first stage, As(V) is precipitated as stable ferric arsenate by adjusting the pH to between 4 and 5, after

which the pH is adjusted to an environmentally acceptable level (pH 6-8) in the second stage of neutralization (Dew $et\ al.$, 1997). The overall chemistry of the reactions is represented by the following equations:

Stage I: Neutralization to pH 4-5:

$$Fe_2(SO_4)_3 + H_3AsO_4 + CaCO_3 + 2H_2O \rightarrow Fe(OH)_3(s) + CaSO_4(s)$$
$$+ FeAsO_4(s) + 2H_2SO_4 + CO_2 \qquad\qquad 1.12$$
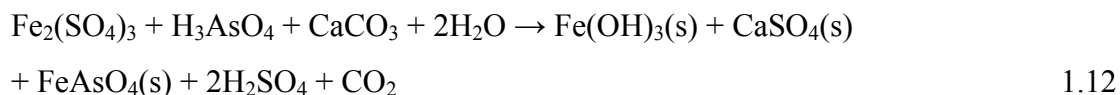
Stage II: Neutralization to pH 6-8:

$$H_2SO_4 + CaCO_3 \rightarrow CaSO_4(s) + CO_2 + H_2O \ / \qquad\qquad 1.13$$
$$H_2SO_4 + Ca(OH)_2 \rightarrow CaSO_4(s) + 2H_2O \qquad\qquad 1.14$$
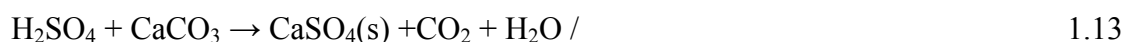
Arsenic is toxic to all life and exists in nature in two forms, As(III) and As(V), with the latter being the less toxic species. The Biox® bacteria break down arsenopyrite and are tolerant to arsenic. Biooxidation at Fairview leads to the production of an arsenate concentration of 12 g/L with little As(III). It is believed that As(III) is oxidized to As(V) by ferric ions, at pyrite surfaces. The Fairview concentrate contains low amounts of pyrrhotite and therefore, biooxidation operates at a high redox potential, promoting oxidation of As(III) to As(V). The presence of high concentrations of As(III) may adversely affect the neutralization process, requiring strong oxidants such as hydrogen peroxide to convert As(III) to As(V) (Dew $et\ al.$, 1997).
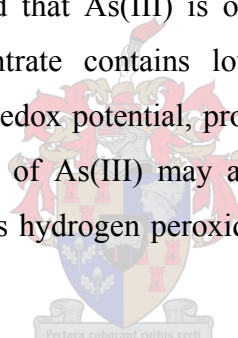
 (Note: Before the paper of Coram and Rawlings (2002) researchers did not distinguish between $Leptospirillum$ $ferrooxidans$ and $Leptospirillum\ ferriphilum$ and therefore species listed as $L.\ ferrooxidans$ could be either $L.$ $ferrooxidans$ or $L.\ ferriphilum$.)

The Biox® bacterial culture is the heart of the process and is reported to consist of a mixed population of $Acidithiobacillus\ ferrooxidans$, $At.\ thiooxidans$ (now known as $At.\ caldus$) and "$Leptospirillum\ ferrooxidans$" (now known as $L.\ ferriphilum$) (Dew $et\ al.$, 1997). They are chemolithotrophic Gram-negative acidophiles and are motile by a single polar flagellum. $At.$ $ferrooxidans$ and "$L.\ ferrooxidans$" are capable of oxidizing iron compounds as electron donors, whereas $At.\ caldus$ obtains energy from the oxidation of reduced sulfur compounds ($At.$ $ferrooxidans$ is also capable of reducing sulfur compounds). These bacteria oxidize arsenopyrite and pyrite by attaching themselves specifically to these metal sulfides. Recent findings support the hypothesis that pyrite is oxidized by means of the indirect mechanism, in

which it is leached chemically by ferric iron produced by the bacteria (Schippers and Sand, 1998). Ferric iron oxidizes pyrite by the following reaction:

$$FeS_{2\,(s)} + 14Fe^{3+}_{(aq)} + 8H_2O_{(l)} \rightarrow 15Fe^{2+}_{(aq)} + 2SO_4^{2-}_{(aq)} + 16H^+ \hspace{2cm} 1.15$$
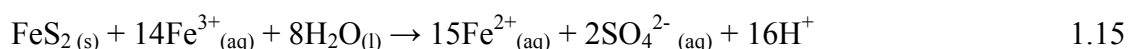
At. ferrooxidans was the first microbe that was isolated from an acid leaching environment that could oxidize minerals. Subsequent research in the field of biomining revolved around this bacterium, believing that it was the primary catalyst in the biomining process (Lundgren and Silver, 1980). This was mainly due to the limited amount of detection techniques and an incomplete understanding of the microorganisms involved in the biomining process. However, advancements in molecular biological techniques such as PCR (Polymerase Chain Reaction) and rRNA gene analysis have improved the ability of microbial detection in bioleaching and biooxidation environments. Goebel and Stackebrandt (1994) isolated and identified *Acidiphilium cryptum*, "*L. ferrooxidans*", *At. thiooxidans* and *At. ferrooxidans* from batch cultures and continuous-flow bioreactors in a study to evaluate the bioleachability of zinc-sulfide ore concentrates (Goebel and Stackebrandt, 1994). However, in a continuous-flow bioreactor at steady-state conditions the authors only managed to identify "*L. ferrooxidans*" and *At. thiooxidans* (now known as *At. caldus*).

Subsequently, a similar study was carried out on the commercial scale, continuous-flow bioreactors at the Fairview mine (Barberton, South Africa), which operated at 40 ºC and pH 1.6. Using a PCR based-technique, the 16S rDNA from known cultures of *At. ferrooxidans*, *At. thiooxidans* and "*Leptospirillum*" was amplified and subjected to restriction enzyme digestion. A distinct restriction pattern for each of species was recorded, which enabled rapid identification by researchers. Comparisons of these patterns to PCR products obtained from total DNA isolated from the biooxidation tanks indicated that a restriction pattern corresponding to *At. ferrooxidans* was undetectable and that the population was dominated by "*Leptospirillum*" and *At. thiooxidans* (Rawlings, 1995). Further studies have indicated through 16S rDNA amplification of crude DNA extracted from the Biox® culture that only "*Leptospirillum*" and *At. caldus* were present in the biooxidation tanks and that *At. thiooxidans* remained undetected (Gardner and Rawlings, 2000). "*Leptospirillum*"-like bacteria isolated from the biooxidation tanks at the Fairview mine was shown to be *L. ferriphilum* (Coram and Rawlings, 2002).

Results obtained from an investigation of the bacteria in commercial biooxidation tanks using a microscopic immuno-fluorescent antibody count detection technique, indicated that *At. ferrooxidans* cells were detected in most samples, although in minority (Schloter *et al*., 1995). Results from analysis of the bacterial population in continuous-flow biooxidation tanks from Sao Bento (Brazil) and Fairview (South Africa) indicated that 48-57 % were "*L. ferrooxidans*" (*L. ferriphilum*), 26-34 % *At. thiooxidans* (*At. caldus*) and 10-17 % *At. ferrooxidans* (Dew *et al*., 1997).

Currently, several reasons exist for the dominance of "*Leptospirillum*" over *At. ferrooxidans* in industrial processes. The first and major reason for this is the high ferric-ferrous iron ratio (Redox potential) in the biooxidation tanks, which is inhibitory to *At. ferrooxidans* (Rawlings *et al*., 1999). Other contributing factors are pH and temperature. The pH optimum for *At. ferrooxidans* is between pH 1.8-2.5, while "*Leptospirillum*"-like bacteria are more acid resistant and grow at pH 1.2 (Norris, 1983). *At. ferrooxidans* is less tolerant to high temperatures than "*L. ferrooxidans*", with an upper limit of 35 ºC. "*Leptospirillum*"-like bacteria have been reported to have an upper limit of 45 ºC (Norris *et al*., 1988), which are more suited to the continuous-flow biooxidation process operating at 40 ºC (Dew *et al*., 1997). Under the above-mentioned conditions, "*L. ferrooxidans*" is the predominant iron-oxidizer and *At. caldus* the predominant sulfur-oxidizer. However, "*Leptospirillum*" spp. seem to be distributed in a wide variety of highly acidic and metal rich environments, including those associated with mines, mine drainage and mine tailings. In a study performed by Bond *et al*. (2000a), the slime streamers found within an extreme acid mine drainage site at Iron Mountain, California, were examined using 16S rDNA PCR. Phylogenetic analyses of the 16S rRNA genes revealed that "*Leptospirillum*" spp. were the most abundant. A previous study on the distribution of "*L. ferrooxidans*" at the same site, using fluorescent *in situ* hybridization (FISH) found that "*L. ferrooxidans*" was the dominant iron-oxidizing bacterium within the mine effluent (Edwards *et al*., 1999). More recently, total genomic DNA extracted from a naturally occurring acidic biofilm from the Richmond Mine at Iron Mountain was shotgun cloned and sequenced (Tyson *et al*., 2004). They reported the discovery of three bacterial, as well as three archeal lineages, of which *Leptospirillum* spp. were the most abundant.

1.2. *Leptospirillum ferriphilum* **and** *Leptospirillum ferrooxidans*

In earlier studies evidence started to accumulate verifying the existence of more than one species of "*L. ferrooxidans*" (Harrison and Norris, 1985; Lane *et al*., 1992; Sand *et al*., 1992; Hallmann *et al*., 1992; Battaglia, 1994; Goebel and Stackebrandt, 1994; Bond *et al*., 2000a and Bond *et al*., 2000b). Previously, only two *Leptospirillum* species were described, *L. ferrooxidans* DSM2705 (Markosyan, 1972) and *L. thermoferrooxidans* (Golovacheva *et al*., 1993). *L. ferrooxidans* DSM2705 was isolated from a copper deposit in Armenia, while *L. thermoferrooxidans* was isolated from hydrothermal hot springs of 45 °C (this culture has subsequently been lost). Harrison and Norris using DNA base composition analysis, six *L. ferrooxidans*-like isolates with relatively low (51-52 %) and relatively high (55-56 %) mol % GC content were grouped followed DNA-DNA hybridization of the same six isolates thus identifying at least two hybridization groups (Harrison and Norris, 1985). Subsequent studies by Lane *et al*. (1992) using 16S rRNA sequence comparisons of *L. ferrooxidans* and two *L. ferrooxidans*-like isolates revealed that these bacteria were related to one another (94 % similar), but were not significantly related to any other bacterium. Similarly, Hallmann *et al*. (1992), found that among six *Leptospirillum* isolates two groups comprising of two strains each were 100 % related and that there was 38-50 % relatedness between these groups and 31-50 % relatedness among all other isolates.

Although sufficient evidence existed to suggest separation into different species, a complete taxonomic study involving a wide variety of *Leptospirillum* strains was necessary. Recently, a study done by Coram and Rawlings (2002) on different *Leptospirillum* strains from around the world, suggested that the *Leptospirilli* used in their study be divided into two species. Using DNA-DNA hybridization experiments and 16S-23S rRNA profiling of the different strains they concluded that *Leptospirilli* belonging to group I have 3 *rrn* gene copies, while those belonging to group II have 2 *rrn* gene copies. The name *L. ferrooxidans* was proposed for strains in group I with DSM2705 as the type strain, while a new species name, *L. ferriphilum*, was proposed for strains belonging to group II, with ATCC49881 as the type strain (Coram and Rawlings, 2002).

A third group of *Leptospirilli* has recently been identified through 16S rDNA PCR amplification from the slime streamers found at an extreme acid mine drainage site at Iron Mountain, California (Bond *et al*., 2000a). Evidence provided by 16S rDNA sequence analysis of the suggested that this group and groups II and I were not significantly related, indicating a

new species within the genus. Even more recently, a 16S rRNA clone library was constructed from DNA extracted directly from a naturally occurring acidophilic biofilm from the same site at Iron Mountain (Tyson *et al*., 2004). Screening and sequencing of these clones indicated the abundance of *Leptospirillum* group II as well as group III. Bacteria of this third group of Leptospirilli have yet to be obtained in pure culture for further analysis.

Morphology: Bacteria belonging to this genus are small, Gram negative, vibroid or spiral shaped cells ranging from 0.9-2.0 μm in length and 0.2-0.5 μm in diameter (Johnson, 2001). Cells are motile by means of a single polar flagellum 18-22 μm in diameter, however, somewhat larger flagella (25 μm) have been reported in moderately thermophilic isolates. Occasionally, in some isolates the position of the flagella may differ, being sub-polar rather than polar, and in certain cases some isolates may even possess two flagella (Goebel and Stackebrandt, 1995). The structure of the cell wall of *Leptospirilli* is similar to that of other Gram-negative bacteria, with the cell membrane consisting of two electron dense layers of 0.6-1.0 μm and 0.35-0.6 μm. Pivovarova *et al*. (1981) reported that nuclear structures can be observed, but the intracellular membrane structures are less obvious. Large numbers of polyribosome may be also present in the cytoplasm, but no β-hydroxybutyrate reserves were observed.

Physiology and Biochemistry: *Leptospirillum* spp. are predominately iron-oxidizing and growth occurs only with ferrous iron or iron-containing sulfide minerals such as pyrite (FeS_2). Norris *et al*. (1988) reported that cells have a high affinity for ferrous iron, which is significantly greater than that of *At. ferrooxidans*. *L. ferrooxidans* was also shown to be more tolerant to ferric iron inhibition than *At. ferrooxidans* (Rawlings *et al*., 1999). Oxidation of mineral sulfides (such as pyrite and arsenopyrite) occurs via the indirect mechanism (Schippers and Sand, 1998). Ferric iron produced by the bacteria chemically oxidizes the mineral, thereby reducing it to ferrous iron, which in turn is reoxidized to ferric iron by the bacteria. This mechanism does not require contact between the bacterium and the mineral; alternatively a direct attachment mechanism has been described (Schippers *et al*., 1996).

Leptospirillum spp. are obligate autotrophs and carbon fixation usually occurs via the Benson-Calvin cycle (Norris *et al*., 1995). It was found that iron-grown cells of *L. ferrooxidans* produce large amounts of a novel red-colored, acid-soluble cytochrome, which is rapidly oxidized by ferrous iron in cell-free extracts showing that it is a crucial component in the

respiratory chain (Blake and Shute, 1997). *Leptospirillum* spp. are obligate aerobes and acidophilic, and grow optimally at pH 1.3-2.0 with a lower limit of 1.1 (Battaglia *et al*., 1994). *L. ferrooxidans* is mesophilic and it is tolerant of higher temperatures (35 ºC) and less tolerant of temperatures lower than 25 ºC. *Leptospirillum* spp. grows very poorly on solid media, which utilizes agar as a gelling agent (*L. ferrooxidans* growth is inhibited by organic materials). Due to the extremely sensitive nature of *Leptospirillum* spp. to organic materials, a specific technique involving a bilayered medium has been developed (Johnson, 1995). This method allows for the pouring of two layers, in which acidophilic heterotrophic bacteria are incorporated into the lower layer, to utilize the organic material present in the media. Growth on solid media is relatively slow and colonies are only visible 7-14 days after incubation at 30 ºC. Colonies are small, round and range from orange to light brown in color. The minimum culture doubling time for *Leptospirillum* spp. in ferrous iron media varies from 10 to 20 hours (Norris *et al*., 1988). Previously, it was reported that *L. ferrooxidans* is capable of fixing nitrogen (Norris *et al*., 1995). Recently, gene expression analysis using DNA micrroarray technology led to the identification of a nitrogen fixation regulon in a *L. ferrooxidans* isolate (Parro and Moreno-Paz, 2003). Nitrogen fixation genes (*nif* genes) were also detected in *Leptospirillum* group III from an acidophilic biofilm. However, *nif* genes were not identified in *Leptospirillum* group II (Tyson *et al*., 2004).

Leptospirillum thermoferrooxidans (Hippe, 2000) (Golovacheva *et al*., 1993)

L. thermoferrooxidans is a Gram-negative, moderately thermophilic, aerobic, chemolithoautotrophic, which grows optimally between temperatures of 40-45 ºC, with an upper limit of 55-60 ºC. Its optimum pH lies between 1.65-1.90 (minimum pH 1.3). *L. thermoferrooxidans* has 26.7 % DNA-DNA relatedness to the type strain. A single strain, which was isolated from acidic hydrothermal springs on Kunashir, has since been lost. The mol% GC of DNA is 56 (T_m).

Leptospirillum ferriphilum (Sand *et al*., 1992)

Cells are small, Gram-negative curved rods or spirilli (0.3-0.6 μm in diameter and 0.9-3.5 μm long). Young cells are vibrio shaped, while cultures older than 4 days are mostly spiral with 2-5 turns. Cells are spore forming and motile by means of a single polar flagellum. Growth is aerobic and chemolithotrophic, with ferrous iron or pyrite as sole energy source. Optimum pH 1.4-1.8. Optimum temperature is between 30-37 ºC; with some isolates able to grow at 45 ºC. Cells are catalase negative and peroxidase positive. The mol% GC is 55-58 %, with two copies

of *rrn* genes. The type strain is ATCC 49881, which is the same as P₃a originally isolated in Peru (Sand *et al*., 1992).

Leptospirillum ferrooxidans (Markosyan, 1972)

Description of *L. ferrooxidans* was obtained from Markosyan (1972).

1.3. Properties of arsenic.

The element arsenic is number 33 in group 15 of the periodic table. The name is synonymous with the word poison and it is believed that it was the supposed cause of death of Napoleon Bonaparte. The element was discovered by Albertus Magnus in 1250 and has been a part of human history ever since. It is naturally present in soil, water and also marine foodstuffs, although in very low concentrations. In addition, it has been widely used in medicine and agriculture (as pesticides, herbicides and animal feed additives). Paul Erlich won the Nobel Prize for medicine in 1908 for the use of Salvarsan (which he nicknamed "Silver bullet") as a chemotherapeutic and antimicrobial agent (Rosen, 1999 and Mukhopadhyay *et al*., 2002)

Arsenic is a metalloid, which shows many metallic properties. It exists in 3+ (arsenite) and 5+ (arsenate) oxidation states with arsenite (Rosen, 2002). The toxicity of arsenite lies in its ability to inhibit metabolic enzymes such as pyruvate dehydrogenase. Arsenic reacts as metal, forming metal-thiol bonds with cysteine residues, thus inhibiting these essential enzymes. Arsenic is also considered to be a carcinogen (Mukhopadhyay *et al*., 2002). Previous research has shown that arsenate is taken up by cells via phosphate transporters and arsenite by aquaglyceroporins (NiDhubhghaill and Sadler, 1991; Sanders *et al*., 1997). In prokaryotes two phosphate transporters (Pit and Pst) are responsible for the uptake of arsenate. Similarly, it has been shown in *S. cerevisiae* that phosphate transporters facilitate the uptake of arsenate (Mukhopadhyay *et al*., 2002). In mammals it is assumed that arsenate would be taken up in a similar fashion, but has yet to be demonstrated. In *E. coli*, GlpF (glycerol facilitator) was identified as a trivalent metalloid transporter, which was believed to be involved in arsenite uptake. GlpF is an aquaglyceroporin, a member of the aquaporin superfamily. Disruption of the *glpF* gene confers resistance to antimonite on an antimonite-sensitive strain, indicating that arsenite is not being transported into the cell (Sander *et al*., 1997). Similarly, Fps1p, a yeast homologue of GlpF was identified as an arsenite transporter (Lui *et al*., 2002). Deletions of this gene also conferred arsenite and antimonite resistance to an arsenite sensitive strain. More

recently, it has been shown that mammalian aquaglyceroporins catalyze the uptake of As(III) and Sb(III). This evidence clearly suggests the involvement of aquaglyceroporins in arsenite transport. Aquaglyceroporins might recognize $As(OH)_3$ as the inorganic equivalent of glycerol, suggesting that it is their likely substrate (Mukhopadhyay *et al*., 2002). Arsenic is continually being added to the environment in many forms. Acid mining industries are of particular concern, releasing soluble arsenic from the ore into the environment in high concentrations. Arsenic is then taken up by a variety of organisms ranging from phytoplankton, algae, crustaceans, mollusks and fish. Microorganisms associated with these and other environments also take up arsenic. Therefore, it is not surprising that these microbes develop resistance to arsenic and other heavy metals (such as antimonite and bismuth), which are associated with it (Mukhopadhyay *et al*., 2002).

1.4. General arsenic resistance in bacteria.

Microorganisms have been discovered to have specific genes for resistance to a variety of toxic metals. These metal-ion resistance systems were initially discovered mainly on plasmids, but through genome sequencing projects they are more frequently also found on chromosomes (particularly in the case of arsenic resistance). The mechanisms from these resistance determinants are general efflux and enzymatic detoxification systems, which involve converting more toxic to less toxic metal-ion species and transporting them out of the cell. However, arsenic resistance is one of the best characterized efflux systems, where less toxic As(V) is converted to more toxic As(III), which is then removed from the cell. The arsenic resistance operon (*ars* operon) has been extensively studied and reviewed for many years and is the topic of this discussion (Cervantes *et* al., 1994; Silver, 1996; Silver and Phung, 1996; Xu *et al*., 1998; Rosen, 1999; Rosen, 2002a, Rosen, 2002b and Mukhopadhyay *et al*., 2002)

Two common forms of the arsenic resistance operon exist. The one is a 5 gene operon (with the arrangement *arsRDABC*), which has only so far been found in Gram-negative bacteria. The other consists of 3 genes (*arsRBC*), which is the more common and most basic form (Rosen, 2002a and Rosen, 2002b).

O/P   arsR   arsB   arsC

104 aa   429 aa   131 aa   *S. aureus pI258*

O/P   arsR   arsD   arsA   arsB   arsC

117 aa   120 aa   538 aa   429 aa   141 aa   *E. coli R773*

O/P   arsR   arsB   arsC

117 aa   436 aa   141 aa   *E. coli* chromosome
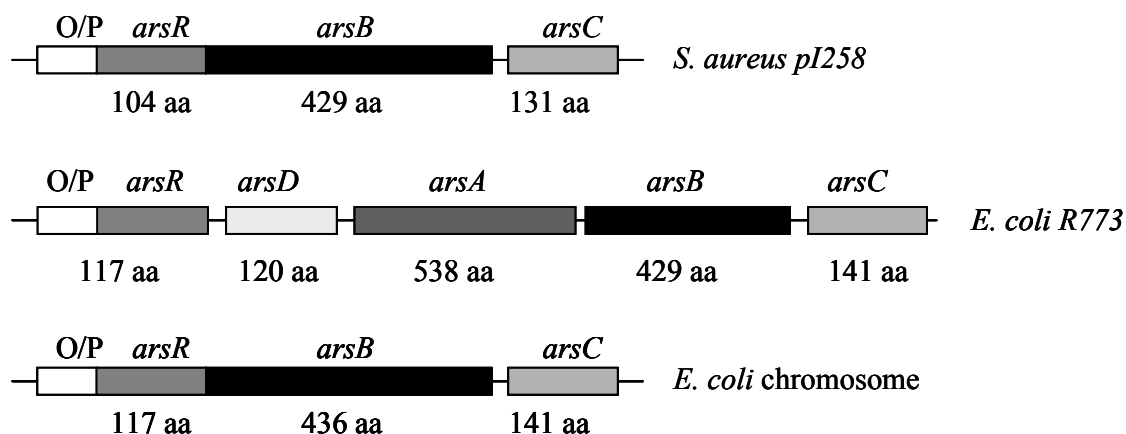
Figure 1.2 The structural organization of arsenic resistance (*ars*) operons from *Escherichia coli* R773, *E. coli* chromosome and *S. aureus* pI258. Alignments of arsenic resistance genes with aa sizes. (Adapted from Silver *et al.*, 1996)

Both operons have the *arsRBC* genes in common. The first gene from both the operons, the *arsR*, encodes an arsenite responsive regulator controlling basal level transcription of the operon. The *arsB* gene encodes a membrane spanning efflux pump. The *arsC* gene encodes an enzyme that reduces arsenate to arsenite (arsenate reductase), which can then be transported to the outside of the cell by the *arsB* (Ji and Silver, 1992b)

The 5 gene operon contains two extra genes, *arsA* and *arsD*, that are missing from the 3 gene operon. The *arsA* gene encodes an arsenite-stimulated ATPase, which interacts with the product of *arsB* gene to form an ATP-driven arsenite efflux pump (ArsB on its own, as in the case of the 3 gene operon, is a chemiosmotic membrane pump). Finally, the *arsD* gene also encodes a trans-acting repressor that controls upper level expression of the 5 gene operon (Wu and Rosen, 1993 and Chen and Rosen, 1997). In addition to being a regulator, ArsD also has a secondary function. Rosen and co-workers have observed that the *arsD* and *arsA* always occur together (B.P Rosen, personal communication), which begged the question whether ArsD had an additional function (discussed later in the chapter). Together the products of these genes create the arsenic efflux system.

How did different operons evolve? Initially, resistance probably arose in the primordial anaerobic environment of the earth, where arsenite would have been a major chemical species. Microorganisms would have evolved arsenite specific transporters to extrude these ions from

the cell. A change in the environment caused the atmosphere to become more oxidizing; therefore arsenate became the more predominant species in solution. It was speculated that *arsB* evolved first, because it is sufficient to provide resistance to arsenite. The addition of a regulatory gene would have occurred early in evolution, because the ArsR repressor controls most *ars* operons. They belong to a large family of metalloregulatory transcriptional repressors called the SmtB/ArsR family (Busenlehner *et al*., 2003). This would constitute a two-gene operon, which is still in existence today. To confer broad-range resistance to arsenate, *arsC* would have evolved. At this point divergence occurred, giving raise to two prokaryotic *arsC* genes (Rosen, 1999). The acquisition of *arsA* and *arsD* genes is speculated to be a relatively recent addition to *ars* operons, conferring high-level resistance. This can be seen in Gram-negative bacteria, in which three 5 gene operons have been identified. Recently, the arsenic resistance operon of *Ferroplasma acidarmanus* was isolated, containing the unusual configuration of *arsB*, *arsR* and *arsA* genes (Gihring *et al*., 2003). Normally, the *arsA* gene is associated with the *arsD* and is found in a pair (mentioned earlier), in this case the *arsA* gene occurred on its own. Another irregularity is that the *arsA* gene is abnormally short. Phylogenetic analysis of a variety of *arsA* genes revealed that it lacked the arsenite-binding cysteine residues (discussed later in the chapter), resulting in a non-functional ArsA protein. The diversity of the different *ars* operons can be seen in Table 1.1.
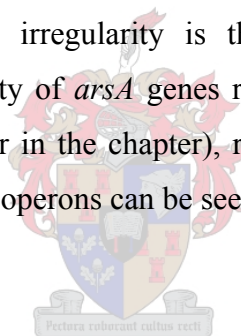
Table 1.1: Examples of different bacterial arsenic resistance operons that have been cloned and sequenced.

Bacterium	Location	Arrangement	Reference
Acidithiobacillus ferrooxidans	Chromosome	*arsCRBH*	Butcher *et al*., 2000
Acidithiobacillus caldus	chromosomal transposon chromosome	*arsRCDADAorf7orf8B* *arsCRB*	Tuffin *et al*., 2005 Unpublished data
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
Ferroplasma acidarmanus	Chromosome	*arsBRA*	Gihring *et al*., 2003
Staphylococcus aureus	plasmid, pI258	*arsRBC*	Ji and Silver, 1992a
Staphylococcus xylosus	*plasmid*, pSX267	*arsRBC*	Rosenstein *et al*., 1992
Bacillus subtilis	*skin* element	*ars*R *ORF2 arsBC*	Sato and Kombayashi 1998
Pseudomonas aeruginosa	chromosome	*arsRBC*	Cai *et al*., 1998
Acidiphilium multivorum	plasmid, pKW301	*arsRDABC*	Suzuki *et al*., 1998
Yersinia enterocolitica	plasmid, pYV	*arsRBC* and divergent *arsH*	Neyt *et al*., 1998

1.5. The proteins of the *ars* operon.

    1.5.1. The efflux pump

       1.5.1.1. ArsA ATPase

Chen and colleagues first discovered the *arsA* gene on the *E. coli* R-factor R773 in 1985, as a member of an arsenic resistance operon consisting of the *arsRDABC* genes. The ArsA is a 63 kDa, cytoplasmically located inner membrane protein, which is normally associated with ArsB. Together they form the arsenic efflux pump, where ArsA is the catalytic subunit (ATPase) and ArsB the membrane spanning subunit that extrudes arsenite ions (Chen *et al*., 1986). When expressed in the absence of ArsB, ArsA is soluble and is found in the cytosol (Rosen, 1999). The ArsA protein is composed of two homologous halves, A1 and A2, which are separated by a flexible linker region. Each half has a consensus nucleotide-binding domain (NBD), of which both are required for ATPase activity. Genetic studies indicated that the mutation of a glycine residue in the A1 NBD (NBD1) resulted in a substantial reduction in arsenic resistance. Arsenic resistance was restored by a second mutation in A2 NBD (NBD2) (Li *et al*., 1996). This result suggested that the two residues are in close proximity, which supports the model that the two NBDs interact to promote catalysis.

Further investigation indicated that the NBDs participate in unisite and multisite catalysis (Kaur, 1999). In the absence of metalloid, a basal level of ATP hydrolysis was reported (Walmsley *et al*., 1999). Whereas in the presence of metalloid, accelerated catalysis was reported (multisite catalysis), where one NBD was more catalytic than the other. The catalytic properties of the NBDs were investigated using the fluorescence of tryptophan residues, which change in response to nucleotide binding (Zhou and Rosen, 1997). Tryptophan derivatives of ArsA indicated that NBD1 participated in both unisite and multisite catalysis. However, NBD2 only participated in multisite catalysis. Moreover, NBD1 hydrolyzed ATP 250 fold faster than NBD2 (Walmsley *et al*., 1999 and 2001, Zhou *et al*., 2001). Speculation based on these results suggested that the NBDs have intrinsic differences and further suggested that they have dissimilar catalytic functions.

A 12- residue consensus sequence $D_{142}TAPTGHTIRLL_{153}$ (termed the DTAP motif) was identified in each of the ArsA halves as well as in other ArsA homologues. Zhou and Rosen (1997) investigated the function of the DTAP motif by constructing ArsA derivatives containing only single tryptophan residues on either side of the DTAP domain. Then relying on the intrinsic fluorescence properties of the tryptophan residues (depending on the environment

in which they are found), it was shown that upon ATP hydrolysis the C-terminal tryptophan residue moves from a relatively hydrophilic environment to a less polar environment. On the other hand, the N-terminal tryptophan is located in a non-polar environment and subsequently moves to a more polar environment as product gets formed. The authors hypothesized that the DTAP motif is a transduction domain facilitating the transport of energy from ATP hydrolysis to parts of the anion pump (Zhou and Rosen, 1997).

The ArsA protein was found to have four cysteine residues (Cys-26, Cys-113, Cys-172 and Cys-422). In order for arsenic or antimony to be bound successfully, the cysteine residues have to be in close proximity to each other in the folded protein. Investigation into the function of these cysteine residues by site-directed mutagenesis showed that three of the cysteines (Cys-113, Cys-172 and Cys-422) are involved in allosteric activation of the ArsA in the presence of As(III) (Bhattacharjee *et al.*, 1995). To investigate the possibility of As(III)/Sb(III) coordination to the thiolates of the cysteine residues, the homobifunctional cross-linker dibromobimane was used. The distance between cysteine pairs were mapped and found to be within 6Å from one another in the native ArsA, suggesting that a As(III)/Sb(III)-thiol complex in the tertiary structure is involved in allosteric activation (Bhattacharjee and Rosen, 1996). Rosen *et al*. (1999) proposed a model where binding of As(III) to the allosteric site physically pulls the two halves of the ArsA together thereby accelerating ATP hydrolysis (Fig.1.4). The energy released from the hydrolysis of ATP is transduced to ArsB, thereby fueling the transport of As(III).
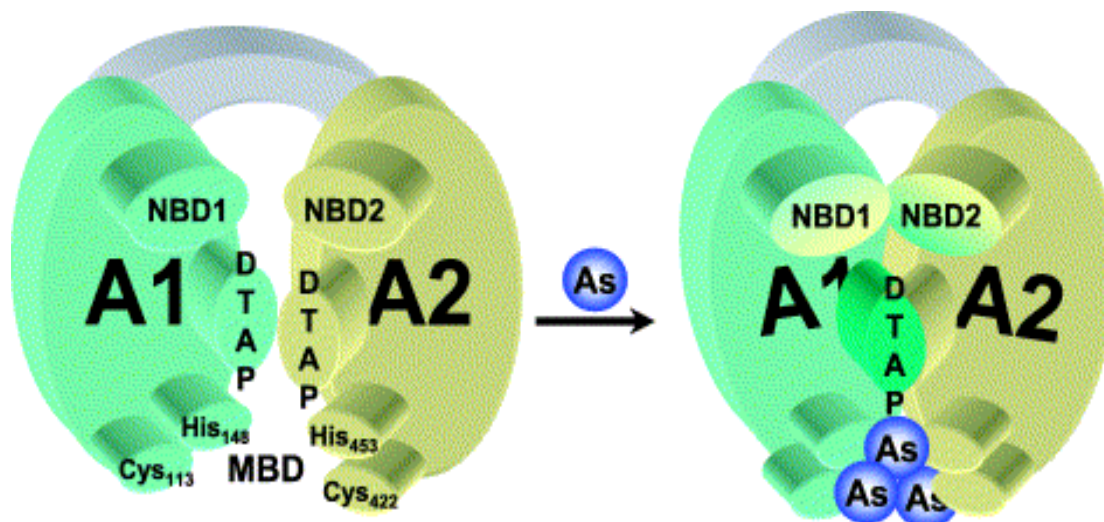
Figure 1.3: The model for the structure of ArsA from *E. coli* R773 during allosteric activation. ArsA protein consists of two homologous halves (A1 and A2), each consisting of a nucleotide binding domain (NBD) and signal transduction domain (DTAP motif). A flexible linker region separates the two halves of ArsA. The cysteines shown in the diagram form part of the ArsA metal-binding site (MBD). In the absence of As(III) the two halves are loosely bound, with only A1 NBD participating in unisite catalysis. Binding of As(III) to the MBD brings the two halves together accelerating catalysis. (taken from Rosen, 2002)

As mentioned previously, a flexible linker region separates the A1 and A2 domains of the ArsA protein (Li and Rosen, 2000). It was observed that the sequence of the linker region is not well conserved among ArsA homologues. However, the linkers all had the same length. This led the authors to believe that the function of the linker might be structural rather than catalytic (Li and Rosen, 2000). The effect of lengthening and shortening the linker by adding or deleting residues was investigated. Li and Rosen (2000) found that insertion of residues had no effect on resistance or catalysis, whereas deleting of residues caused sensitivity as well as decreasing the enzymes affinity for ATP and antimonite (antimonite may also be exported via the *ars* system). These results led the researchers to the conclusion that the length of the linker was important and not the sequence, speculating that the linker had evolved to the shortest length necessary for efficient interaction of the two halves of the ArsA. Further investigation into the role of the linker was done using complementation and mutational experiments Jai and Kaur, 2001). They constructed and expressed various clones of the A1 N-terminal and C-terminal A2 domains with and without a linker. Since each domain is a separate polypeptide, changes to the linker region should only have an effect if it is required for functioning of ArsA. Cross complementation of different clones suggested that the C-terminal half could only

complement the N-terminal clone when the linker was present. Mutational analysis showed that certain residues were crucial for ArsA function (resulting in arsenite sensitivity as well as loss of ATPase activity). It was also shown, based on trypsin proteolysis experiments, that the mutations induced conformational changes in ArsA protein. This evidence indicated that certain linker residues as well as the length of the linker are essential for ArsA function. The evidence also suggested interaction between the linker and the two NBDs (Jai and Kaur, 2001).

The recent elucidation of the ArsA crystal structure revealed that ATP is found at the A2 NBD, while ADP is found at the A1 NBD (Zhou *et al*., 2001). A possible explanation for this observation would be that the A1 NBD is catalytic, while the A2 NBD is not (Zhou *et al*., 2002). However, when ArsA crystals were incubated in the presence of the non-hydrolysable analog, AMP-PNP, the compound was found at the A2 NBD, instead of at the A1 NBD. An attempt to crystallize ArsA in the presence of the non-hydrolysable analog, was unsuccessful, suggesting that crystals can only be formed if the A1 NBD contains ADP. The authors speculated that the crystal lattice does not allow the conformational changes associated with the binding and hydrolysis of ATP at the A1 NBD. They also suggested that restrictions imposed on catalysis by the crystal structure could be the primary reason why ATP and not ADP is found at the A2 NBD. Although evidence provided by presteady state kinetics suggested that the release of ADP from the A2 NBD is associated with the release of As(III)/Sb(III) ions and that binding of ATP favors uptake of such ions (Walmsley *et al*., 2001).

A number of other observations were made regarding the crystal structure of ArsA (Zhou *et al*., 2000). Firstly, the two NBDs were located at the interface between the A1 and A2 halves, in close proximity to each other and are formed by residues from both domains of the protein. However, one NBD consisted of mostly A1 residues and was designated A1 NBD. Likewise, the other NBD consisted mostly of A2 residues and was accordingly named A2 NBD. The crystal structure also showed that at the interface between the two halves at the opposite end with respect to the NBDs, is the allosteric site where three distinct As(III)/Sb(III) ions bind. Three cysteines, two histidines and one serine, where found to act as ligands for these ions, each ion bound by one residue from the A1 and one from A2 halves. This supports previous reports that binding of the metalloids trigger ATP hydrolysis by bringing the domains together. Previous suggestions (Xu *et* al., 1998; Zhou *et* al., 2001) that the ArsA is similar to NifH, the Fe-protein of bacterial nitrogenases, led researchers to the identification of an aspartic acid

residue that is a Mg^{2+} ligand, located in a strand-loop-helix referred to as the Switch I region. ATPase activity largely depends on Mg^{2+}, which in complex with ATP produces conformational changes in ArsA. Other observations made by the authors identified conformational changes involving helices in the A1 and A2 halves associated with ATP hydrolysis (Zhou et al., 2001). This led the authors to propose the model shown in Fig. 1.5. Helices (H9-H10) play a central role in this mechanism, alternating the positions of A1 and A2 at the interface with ArsB forming a gate for As(III)/Sb(III) ions.
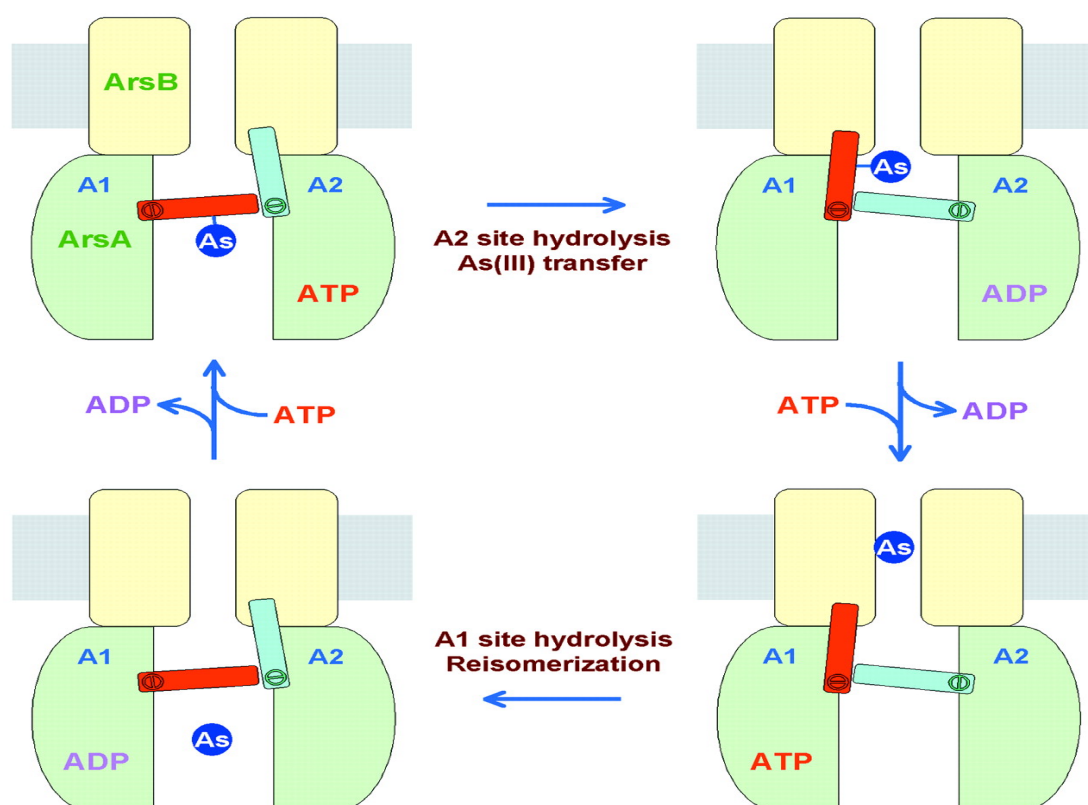
Figure 1.4: Model of the ArsA catalytic cycle. Helices H9-H10 of A1 (red) and A2 (cyan) form the arms of an alternating gate. An arsenite ion is depicted as a blue circle. See text for details. (taken from Zhou et al., 2001)

As(III) interacts with the helices from A1 and moves from the cytosolic inside of the enzyme into the protected pocket at the interface with ArsB. The release of ADP from the A2 NBD results in the liberation of As(III) ions inside this pocket. Hydrolysis of ATP at the A2 NBD brings the ArsA back to its ground state. Based on this model the catalytic cycle of ArsA

resembles a reciprocating engine. Whether the two NBDs look like the cylinders of this engine would still have to be determined. There is evidence that the two sites perform different functions, therefore suggesting non-equivalence of these sites. The crystal structure of the ArsA helped to solve many of the speculations of the different aspects of the ArsA protein, yet many elements of this model remain to be determined (Zhou *et al*., 2000).

1.5.1.2. ArsB

The *arsB* gene of R773 encodes a protein of 429 aa with a molecular mass of 46 kDa. Chen *et al*. (1986) demonstrated the ArsB protein had at least 10 of 19 or more residues with hydropathy values greater than average, suggesting that they could be potential membrane-spanning α-helices. San Francisco *et al*. (1989) constructed an ArsB-β-galactosidase fusion protein. Cells expressing this hybrid were analyzed and it was found that the fusion protein was located in the inner membrane of the cell. To further demonstrate that the ArsB is in fact located in the inner membrane, it was necessary to identify the native ArsB gene product. Membrane proteins are often difficult to identify due to their low levels of expression. San Francisco *et al.* (1989) identified two potential secondary structure locations in the translation initiation region, one immediately upstream of the predicted ribosome-binding site and the other beginning at the third codon. It was thought that these structures interfered with the functioning of the ribosome disrupting translation, thereby limiting the production of ArsB. To overcome this problem the T7 RNA polymerase expression system was used, allowing the authors to visualize the protein on a SDS-PAGE gel. The observed protein found in the membrane fraction was identified as ArsB, with an apparent molecular mass of 36 kDa. However, the predicted size of the ArsB protein was 46 kDa. According to the authors this was due to this basic protein binding abnormally high amounts of SDS, causing rapid migration through the gel. Later, Ji and Silver (1992b) detected the same behavior for the ArsB of a different operon.
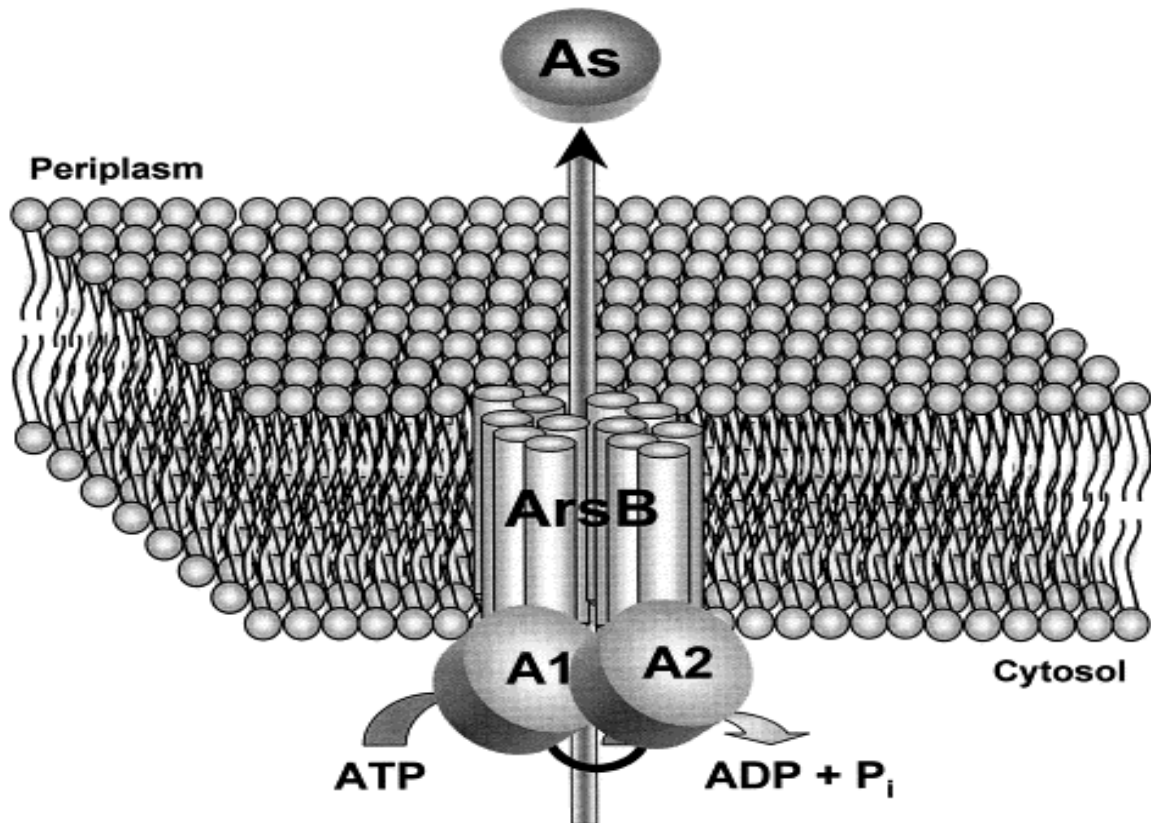
Figure 1.5: Structure of the ArsAB efflux pump. ArsA on its own is an anion-translocating ATPase exhibiting As(III)/Sb(III) stimulated activity and ArsB an inner-membrane protein which serves as an anchor and anion conducting subunit. Together they constitute the arsenite extrusion system coupled to ATP hydrolysis. (taken from Rosen *et al.*, 1999).

The arsenite extrusion system of *E. coli* R-factor R773 consists of the ArsA and ArsB proteins, which form an obligatory ATP-driven pump. Upon comparison of different *ars* operons it was observed that no ArsA homologues were present in many of the other operons. This observation led researchers to investigate whether in the absence of ArsA; energy is supplied by an alternate ATPase from the host cell or whether ArsB functions as a secondary carrier protein (requires electrochemical energy) (Dey and Rosen, 1995). To investigate these two possibilities, an experiment was carried out in an *unc E. coli* strain (lacks H^+ translocating ATPase and is therefore unable to equilibrate chemical and electrochemical energy). The researchers constructed an ArsA deletion of the R773 *ars* operon (pBC101) as well as a clone consisting only of the *arsA* gene (pArsA). These constructs were transformed into the *unc E. coli* strain and grown with only glucose available as an energy source (produces ATP). They found that in the presence of glucose (ATP) only cells with both *arsA* (pArsA) and *arsB*

(pBC101) present could extrude arsenite. When these cells were grown in the presence of low levels of ATP, no arsenite was extruded. This result was indicative that the ArsA/ArsB complex is dependent upon ATP. However, cells containing only ArsB (pBC101) excluded arsenite regardless of the presence of ATP. This exclusion could only be inhibited by the addition of cyanide, indicating that the cells excluded arsenite independent of ATP, which is consistent with the model that arsenite efflux by ArsB is mediated through electrochemical energy (Dey and Rosen, 1995). This observation was later supported by evidence provided by Kuroda and colleagues, suggesting that in the absence of ArsA, ArsB acts as a secondary carrier protein, exporting arsenite from the cell through electrochemical energy (Kuroda *et al*., 1997).

Previous research established that in the absence of ArsA, ArsB functions as a secondary carrier protein coupled to proton motive force (Dey and Rosen, 1995). However, the question remained; what was the chemical nature of the transported species? One possibility was that As(III)/Sb(III) was transported as oxyanions. Another was that As(III)/Sb(III) ions bind to cysteine thiolates in the ArsB and are transported as a soft metal. It is for this reason that the role of a single cysteine residue, predicted to be located in the 11th membrane-spanning region, was investigated (Chen *et al*., 1996). When this cysteine residue was changed to a serine or an alanine residue by site-directed mutagenesis, the levels of As(III)/Sb(III) resistance remained unchanged. The result indicated that the transport of As(III) by the ArsB protein does not involve metal-thiol chemistry. Chen and colleagues suggested that electrophoretic anion transport is the most likely alternative. They proposed two distinct arsenic chemistries: soft metal binding for recognition of arsenic or antimony compounds by the ArsR repressor and for allosteric activation for the ArsA ATPase subunit and non-metal chemistry for oxyanion transport by the ArsB protein (Chen *et al*., 1996). Further investigation by Meng *et al.* (2004) suggested the likely possibility that trivalent As(III) forms oxo-bridged polymers, such as in the case of arsenious oxide (As_4O_6), a six-membered ring $(AS-O)_3$ with the fourth As(III) coordinated to the three axial oxygens. Recently in *S. cerevisiae,* it was found that As(III) was transported by hexose transporters, therefore it was proposed that the substrate for ArsB is a hexose-like six-membered $(AS-O)_3$ ring composed of As(III), Sb(III) and other co-polymers of the two metalloids (Meng *et al*., 2004).

1.5.2. Arsenate Reductase (ArsC)

The ArsC is a small monomeric cytoplasmic enzyme of 16 kDa, originally identified in the *E. coli* R-factor R773. When it was first identified it showed no significant homology to other known proteins in the database (Chen *et al*., 1986). Originally it was thought to be involved in modification of the efflux pump allowing recognition of arsenate by the ArsB protein. However, it was demonstrated that the ArsC protein from the *S. aureus* plasmid pI258 was in fact an arsenate reductase (Ji and Silver, 1992a). Similarly, Oden *et al*. (1994) showed that ArsC from R773 was an arsenate reductase despite the fact that it has less than 20 % homology to the *S. aureus* ArsC protein. Further investigation into the function of the ArsC protein from R773 identified glutathione and glutaredoxin to be required for arsenate reduction (Oden *et al*., 1994). *E. coli* has two genes required for glutathione synthesis, *gshA* (γ-glutamycysteinyl synthase) and *gshB* (glutathione synthetase). Mutant *E. coli* strains defective in these genes containing the R773 *ars* operon indicated wild-type resistance to arsenite, however, they showed increased levels of sensitivity to arsenate. When strains expressed a glutathione reductase gene (*gor*) containing a mutation, they were more sensitive to arsenate than the wild-type cells. Exogenously added glutathione to the *gshA* mutants restored arsenate resistance. Oden *et al*. (1994) also found that strains expressing mutated genes for thioredoxin synthesis, thioredoxin reductase (*trxB*) and thioredoxin (*trxA*), showed no difference in either arsenate or arsenite resistance. These results indicated that the R773 ArsC protein requires glutathione and not thioredoxin for reduction of arsenate. However, the R773 ArsC protein also requires glutaredoxin (*grx*) for reductase activity. Glutaredoxin serves as a hydrogen donor to the ArsC-catalyzed reduction reaction. There are three glutaredoxins in *E. coli* (Grx1, Grx2 and Grx3), which can function in ArsC catalysis (Shi *et al*., 1999). It was shown that Grx2 was the major hydrogen donor for the reduction of arsenate with a catalytic efficiency 2 times greater than Grx3, which in turn is 3 times greater than Grx1. However research done by Ji and Silver, (1992a) indicated that purified *S. aureus* ArsC required thioredoxin as a general disulfide reducing agent and not glutaredoxin as in the case of the R773 ArsC. These differences between the two ArsC proteins led researchers to speculate about the origins of the ArsC protein, suggesting that they evolved independently to carry out the same reaction. Recent phylogenetic studies and sequence analysis of ArsC proteins of different *ars* operons revealed three independently evolved arsenate reductase families, the *E. coli* GSH/Grx family, *S. aureus* Trx family (forming the two prokaryotic ArsC families) and the *Saccharomyces* family (the eukaryotic family which will not be discussed). All three families have common properties, but

differ catalytically. These differences are evident in the crystal structure of these proteins (Zegers *et al*., 2001; Bennett *et al*., 2001 and Martin *et al*., 2001)

The R773 ArsC is the prototype for the *E. coli* GSH/Grx family. The high-resolution crystal structure of the R773 ArsC was determined on its own and in complex with arsenate and arsenite (Martin *et al*., 2001). A structural database search utilizing different algorithms found no considerable similarity to any other protein. However, the R773 ArsC does contain certain elements of super secondary structure that resemble other proteins, such as crambin and glutaredoxin. Analysis of the active site of the R773 ArsC confirmed previous reports that suggested that a single cysteine residue is required for catalytic activity (Liu *et al*., 1995). However, previous studies indicated that the pKa of cysteine was shown to be higher than the pH optimum required for the ArsC-catalyzed reaction (Gladysheva *et al*., 1994). A single histidine residue (His-8) was thought to be involved in ion-pairing with Cys-12 depressing its pKa value (Gladysheva *et al*., 1996). However, the crystal structure indicated that His-8 was 7.2 Å away from Cys-12, stabilizing the active site loop by forming side chain hydrogen bonds with a serine residue (Ser-15). The catalytic Cys-12 residue on the other hand, as indicated by the crystal structure, appears to be activated by hydrogen bonds from Arg-94 and Arg-107. Recent studies reported that Arg-94 and Arg-107 form an anion-binding pocket for non-covalent binding of arsenate, stabilizing the substrate (Shi *et al*., 2003). Previous speculations suggesting that the active site may resemble that of low molecular weight tyrosine phosphatases were not confirmed in the crystal structure. When the catalytic cysteine residues of low molecular weight tyrosine phosphatases bound to sulfate ions and Cys-12 bound to arsenate were superimposed, it indicated that their catalytic centers are different (Martin *et al*., 2001). Martin and colleagues proposed the following reaction mechanism for *E. coli* GSH/Grx ArsC family:

OH
|
As [V] - O
/   \
OH   HO   (arsenate)

ArsC - Cy s12 - SH ────────────**1**────────────▶ ArsC - Cy s12 – S - As [V] = O
                                                                    |
                                                H₂O   OH

4   OH
          |
H₂O   As [III]   GSH
OH⁻   / \
        OH   HO   (arsenite)   H₂O ◀──── **2**

                                        OH
                                        |
              GrxSH   ArsC - Cy s12 – S - As [V] = O
                                        |
ArsC - Cy s12 – S – As (III) ⁺ - As - OH ◀──**3**──   GS
          +
       Grx - SG   H₂O
                            OH⁻
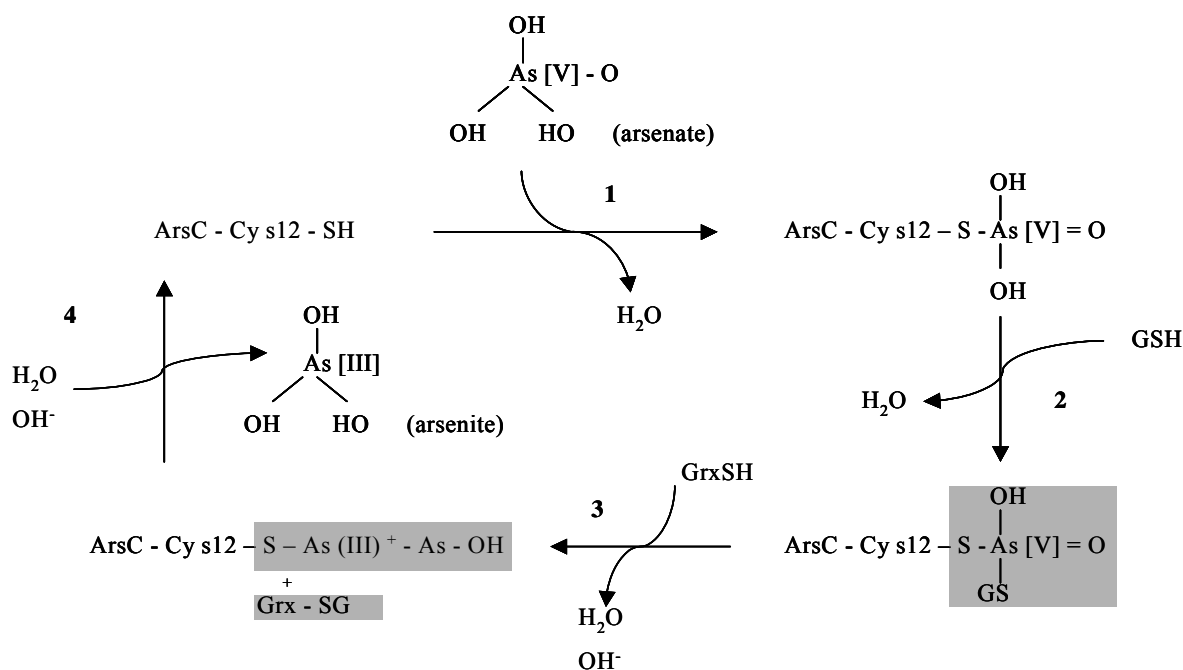
Figure 1.6: Reaction mechanism for GSH/Grx family of arsenate reductases. (taken from Martin *et al.*, 2001). See text for details.

Step I: Involves the formation of a thioarsenate binary adduct (Intermediate I). The native structure binds non-covalently via the thiolate of Cys-12 and the three arginine residues. This changes the overall surface charge from negative to positive, allowing other co-factors and enzymes to bind.

Step II: A {ArsC Cys-12}S-As^v{glutathione} tertiary complex (Intermediate II) was inferred from biochemical studies indicating that glutathione only reacts after arsenate binds and that a free thiol on glutathione and ArsC is required to proceed.

Step III: Arsenate is reduced to arsenite in a quaternary complex involving glutaredoxin, that dissociates into a thioarsahydroxy adduct of ArsC (Intermediate III) and a mixed disulfide complex of glutathione and glutaredoxin.

Step IV: The arsenite-ArsC bond is hydrolyzed releasing arsenite, causing dissociation of the complex and the ArsC returns to its original conformation.

The crystal structures of the ArsC proteins from *S. aureus* plasmid pI258 and *B. subtilis* were solved by (Zegers *et al.*, 2001) and (Bennett *et al.*, 2001), respectively. Bennett *et al.* (2001) described the structure of the *B. subtilis* ArsC protein as a single α/β domain, containing a central four-stranded, parallel open-twisted β-sheet flanked on either side by α-helices.

Analysis of the crystal structures revealed that these proteins unlike the R773 ArsC resemble low molecular weight tyrosine phosphatases (LMP-PTPases). Due to this similarity the active sites of the two enzymes were superimposed, revealing their striking resemblance. It was found that the Trx family of ArsC proteins (to which *S. aureus* ArsC and the *B. subtilis* ArsC belong) shares a conserved CX_5R motif with the LMP-PTPases. The *B. subtilis* crystal structure showed that this region forms an arsenate anion-binding loop (AB loop). This AB loop resembled the PTP loop that is the catalytic site of all classes of PTPases. The AB loop being somewhat larger than the PTP loop and is expected to accommodate the larger arsenate ion (Bennett *et al*., 2001). Research done previously by Messens *et al*. (1999) had shown that three of the four cysteine residues found in this family of ArsC proteins are required for function (Cys-10, Cys-82 and Cys-89). Cys-82 and Cys-89 are believed to form a disulfide bridge upon oxidation. The latter is situated at the end of a flexible region, 11Å away from Cys-82. Cys-89 is expected to move closer to the active site in order to form a disulfide bridge with Cys-82. This was confirmed in the reduced form of the *S. aureus* ArsC crystal structure, where the Cys-82-Cys-89 disulfide bridge is associated with major conformational changes in the ArsC (Zegers *et al*., 2001). A conserved basic arginine residue (Arg-16) was found to be near the cysteine pair Cys-10-Cys-82, where it plays the likely role of lowering the pKa of the cysteine residues, stabilizing the thiolate ions. An aspartate residue (Asp-105) was found quite close to the active site, about 4.1Å away. This residue is absolutely conserved amongst all Trx family ArsC proteins and all classes of PTPases, playing a role as a general acid/base catalyst (Bennett *et al*., 2001). To further strengthen the evidence that Trx family arsenate reductases evolved from LMW-PTPases, researchers observed that ArsC proteins can catalyze dephosphorylation reactions in the presence of the compound (PNPP) para-nitrophenyl phosphate (Zegers *et al*., 2001 and Bennett *et al*., 2001). Based on these observations the following reaction mechanism for the Trx family of ArsC proteins were proposed (Bennett *et al*., 2001 and Zegers *et al*., 2001):
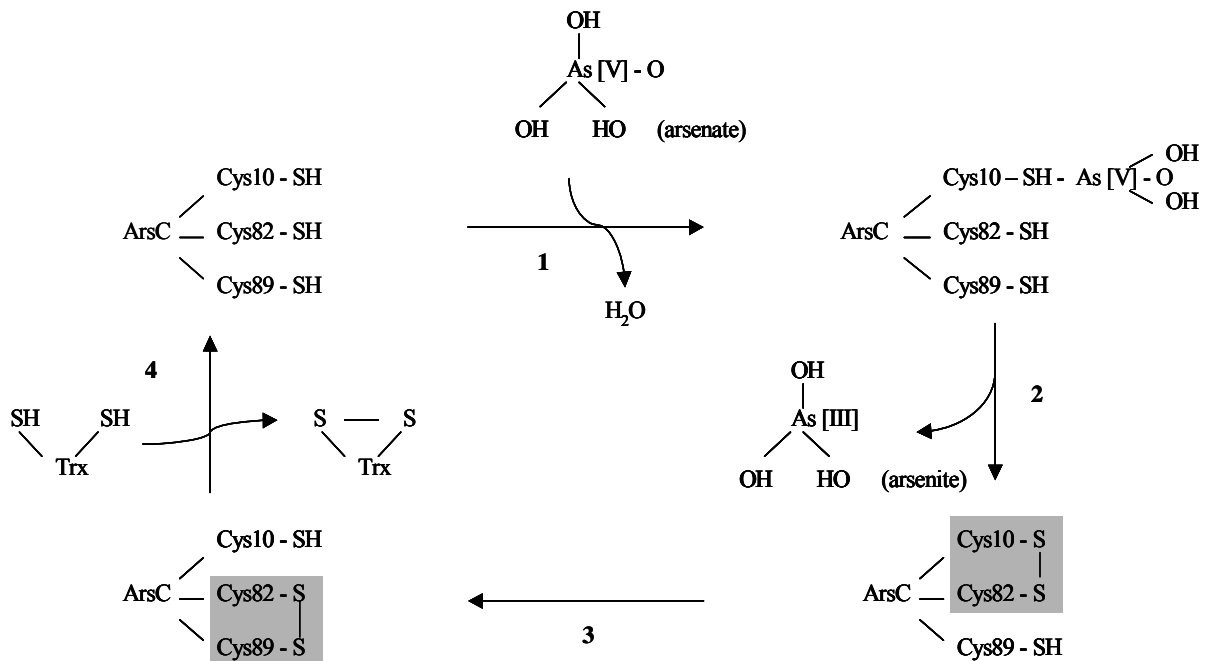
Figure 1.7: The proposed catalytic mechanism for Trx family of arsenate reductases. (taken from Bennett *et al*., 2001). See text for details.

Step I: Nucleophilic attack on arsenate:

Polar residues surround the active site of ArsC, which allows entry of arsenate ions into the active site. Cys-10 then launches a nucleophilic attack forming an arsenylated enzyme substrate.

Step II and III: Reduction of arsenate:

This step involves the three cysteine residues in a triple cysteine redox relay system, producing a Cys-82-Cys-89 disulfide bond and an arsenite ion. Cys-82 attacks the {ArsC Cys-10}S-arsenate bond, forming a disulfide bridge between Cys-82 and Cys-89, regenerating the Cys-10 thiolate. The Cys-10 thiolate is now free to interact with another arsenate ion. An important residue in this mechanism is Arg-16, which is required for stabilizing the AB loop and binding of arsenate. It is also required for lowering the pKa of the cysteine residues for activation.

Step IV: Regeneration of ArsC:

ArsC is regenerated by thioredoxin that reduces the Cys-82-Cys-89 disulfide bond.

29

The findings presented above indicate the major differences of the two families of prokaryotic ArsC proteins, which may have evolved separately to perform the same function. X-ray structure analysis of the R773 ArsC showed its closest relative to be glutaredoxin and glutathione s-transferases (Stevens *et al*., 1999). On the other hand it was shown that *S. aureus* and *B. subtilis* ArsC proteins evolved from low molecular weight tyrosine phosphatases (Bennett *et al*., 2001). Despite different origins the cytoplasmic arsenate reductases share mechanisms based on similar chemistry (cysteine thiol oxidation/reduction cycling coupled to general thiol recycling enzymes Grx/Trx), indicating convergent evolution of these proteins (Mukhopadhyay and Rosen, 2002 and Jackson *et al*., 2003).

1.5.3. Regulation of the *ars* operon.

1.5.3.1. ArsR

The *arsR* is the first gene of the *E. coli* R-factor R773 *ars* operon encoding a 13 kDa dimeric trans-acting repressor. It was shown that expression from the *arsR* gene is autoregulated and that the ArsR protein repressed expression in the absence of inducers such as arsenate (Wu and Rosen, 1991). In subsequent investigations the authors revealed that the ArsR binds to an imperfectly symmetrical dyad sequence close to the –35 site of the *ars* promoter (Wu and Rosen, 1993). Further studies of the ArsR protein binding site involving hydroxyl radical footprinting, which showed that the repressor protected only two small regions of 4 bp long, each separated by 10 bp. The existence of these two short protected areas suggested that the active form of the ArsR protein is a dimer (Wu and Rosen, 1993).

Subsequent research found that the ArsR protein belongs to a new family of DNA binding proteins, which includes the SmtB repressor (a zinc activated repressor which regulates the *smtA* gene encoding for metallothionien). Members of the SmtB/ArsR family contain three distinguishing features, a metal binding domain, a DNA binding domain and a dimerization domain. It has been shown that these proteins possess a helix-turn-helix motif, which contained two conserved cysteine residues (Cys-32 and Cys-34) and either one or two histidine residues (Bairoch, 1993). The helix-turn-helix motif occurs adjacent to the highly conserved region, ELCVCDL, which is present in all members of this family. It was proposed that this sequence forms a metal-binding box and that the cysteine pair is important for metal recognition. Alteration of these cysteine residues to bulky phenylalanine and tyrosine residues indicated that the modified proteins had a reduced metal response suggesting that this cysteine pair is

involved in inducer binding, forming soft metal bonds with trivalent arsenic (Shi *et al.*, 1994). However, the possibility existed that it was the introduction of bulky residues that resulted in these phenotypes. Shi *et al.* (1996) confirmed their previous results by substituting the cysteines with glycine residues instead of bulky aromatic amino acids. Recent comparative spectroscopic and functional analysis involving wild type and mutants of various SmtB/ArsR family members identified a second metal-binding site within the homodimeric repressor (Busenlehner *et al.*, 2003).

Bairoch (1993) proposed that Cys-32 and Cys-34 form part of the DNA binding helix-turn-helix motif of the ArsR protein. However, secondary structure predictions suggested that the cysteine residues are located outside of the helix-turn-helix motif (Shi *et al.*, 1994). Mutations of the histidine residue (mentioned earlier) in the second putative helix of the motif produced constitutive expression, suggesting a decrease in affinity for the operator. This indicated that this residue is somehow involved in DNA binding (Shi *et al.*, 1994).

Since arsenite can form tri-coordinate complexes with sulfur or oxygen atoms, the possibility of the involvement of a third residue was investigated (Shi *et al.*, 1996). In addition to Cys-32 and Cys-34, three other cysteine residues are found in the ArsR protein (Cys-37, Cys-108 and Cys-116). The latter two cysteines were shown to be required for ArsR function (Wu and Rosen, 1993). The remaining cysteine, Cys-37, was investigated by mutational analysis, changing it to an alanine residue. This substitution had no affect on regulation or the DNA binding ability of ArsR, indicating that Cys-37 is not an As(III) ligand. X-ray absorption spectroscopy studies indicated that three sulfur ligands and no oxygen ligands were involved in interactions with arsenite, confirming that serine and threonine residues are not arsenic ligands (Shi *et al.*, 1996). The authors proposed a model for ArsR function, whereby As(III) is bound by the thiolates of the three cysteine residues, Cys-32, Cys-34 and Cys-37, causing a conformational change, resulting in the dissociation of the repressor from the DNA binding domain of the promoter (Shi *et al.*, 1996).

As described above the members of the SmtB/ArsR family contain a metal-binding domain (ELCVCDL), a DNA domain (helix-turn-helix motif) and a dimerization domain. Xu and Rosen (1997), using a yeast two-hybrid system, attempted to isolate the dimerization region of the ArsR protein. The authors were able to show that the region required for dimerization is contained between residues 8 and 89. They also found that there was a clear correlation

between regulation and dimerization of the ArsR protein. Sequence comparisons of the ArsR proteins from *E. coli* chromosome and the R773 plasmid indicated a 75 % similarity along the entire length, whereas an 89 % similarity was observed within residues 8-89. This observation is consistent with the findings from their study that demonstrated that a core region of 80 residues is sufficient to contain all the regulatory properties (such as DNA binding, metal recognition and dimerization) of the ArsR protein (Xu and Rosen, 1997). SmtB from *Staphylococcus aureus* is the closest relative to the ArsR protein. The crystal structure of the SmtB protein predicts that the ArsR will have a similar conformation. Recent nuclear magnetic resonance studies on the *S. aureus* CzrA protein (also a member of the SmtB/ArsR family) found that the homodimer was similar to its predicted form as anticipated from alignments with SmtB, indicating their evolutionary relatedness (Busenlehner *et al.*, 2003).

Recent studies indicated the discovery of a second ArsR family to which the *At. ferrooxidans* ArsR belongs (Butcher and Rawlings, 2002). They found that none of the members of this family of ArsR proteins possessed the metal-binding domain containing the ELCVCDL consensus sequence, however they do have a conserved sequence (GX(L/I)A). These atypical ArsR proteins also contain a cysteine doublet towards their C-termini.

1.5.3.2. ArsD

The *arsD* encodes what was considered to be the second repressor of the *E. coli* R-factor R773 arsenic resistance operon, responsible for regulating upper level expression. When it was initially identified the function of the ArsD protein was unknown and thought to be unnecessary for arsenic resistance. Wu and Rosen, (1993) showed that induced mutations in the ArsD protein caused sensitivity to arsenic, even though the rest of the *ars* operon was intact. Supplying the ArsD protein in *trans* restored arsenic resistance.

Chen and Rosen (1997) demonstrated that ArsD is a homodimer that binds the *ars* promoter at the same position as the ArsR protein and that neither required the other for binding. However, it was also shown that the affinity for the promoter is less than that of the ArsR protein, suggesting that the sufficient concentration of ArsD is required to bind the operator efficiently.

A database search of available genomes revealed only six ArsD homologues, five of which were plasmid-borne (the sixth origin is not known). Sequence alignments of these ArsD

homologues show three conserved vicinal pairs (Cys-12-Cys-13, Cys-112-Cys-113 and Cys-119-Cys-120), of which the first two are only found in four of the six homologues (Li *et al.*, 2001). Previous research indicated that cysteine residues are involved in the function of three proteins involved in arsenic resistance (ArsA, ArsC and ArsR). Recently the role of the three ArsD vicinal pairs was investigated. It was demonstrated through mutagenesis that the Cys-119-Cys-120 is not required for ArsD activity and that vicinal pairs Cys-12-Cys-13 and Cys-112-Cys113 are involved in inducer binding. A third ligand was yet to be discovered, since As(III) is a tri-coordinate. Intrinsic fluorescence of tryptophan residues revealed that binding of inducer to the vicinal pairs produced conformational changes resulting in the dissociation of ArsD from the promoter (Li *et al.*, 2001). Further investigation by Li *et al.* (2002) indicated that positive cooperativity exists between flanking sites on the homodimer. The authors suggesting that the binding sites are formed by and between the equivalent cysteine residues in each subunit (i.e. the Cys-13 residues in each subunit form a binding site).
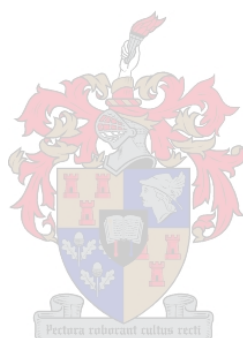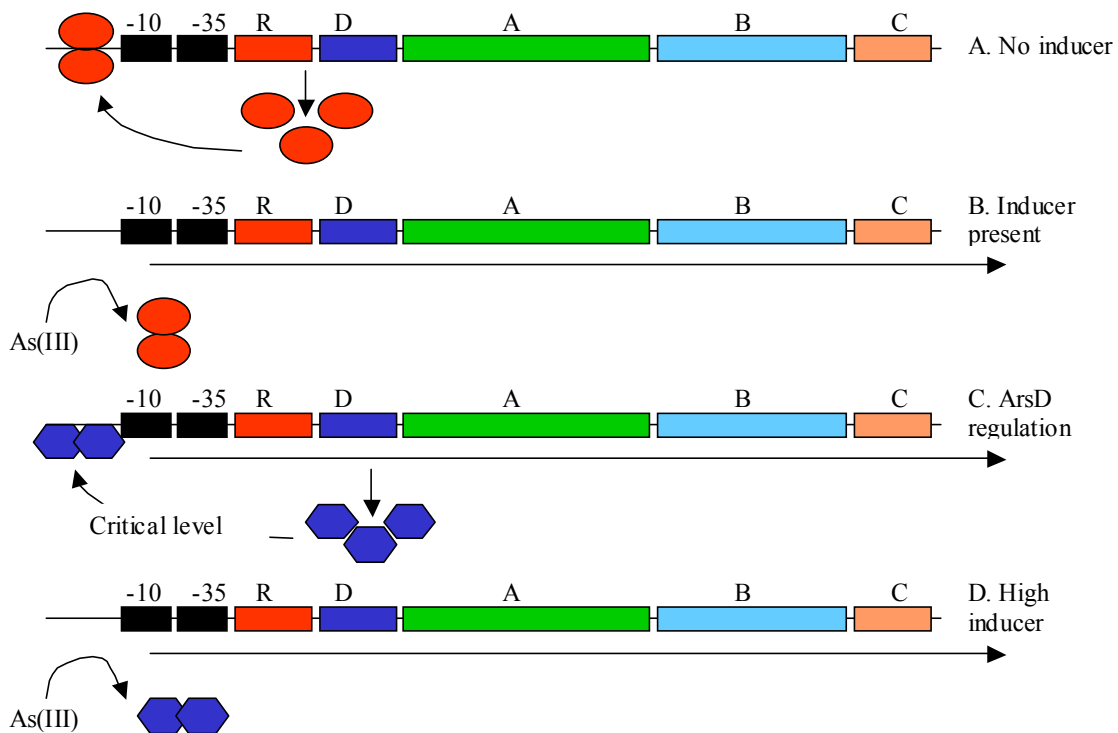
Figure 1.8: The proposed model for the ArsR-ArsD regulatory circuit. (adapted from Chen and Rosen, 1997). See text for details.

Chen and Rosen, (1997) proposed a model for the ArsR-ArsD regulatory circuit (Figure 1.8.). Small quantities of both ArsR and ArsD are produced at basal level in the absence of inducer. ArsR has a higher affinity for the promoter than ArsD; therefore it would prefer binding to the operator, repressing expression of the *ars* operon. Similarly, the ArsR protein also has an increased affinity for inducer. Therefore, a low concentration of inducer would result in dissociation of the ArsR and the *ars* genes including *arsD* would be expressed. The increasing concentration of ArsD results in binding of the operator, shutting down expression of the *ars* genes. ArsD will dissociate from the promoter if the concentration of the inducer increases allowing for further expression of the genes. The action of these two repressors maintain the level of *ars* expression within a narrow range of conditions with ArsR controlling basal level of expression and ArsD controlling maximal expression (Chen and Rosen, 1997). Recently, Rosen and co-workers have noticed that ArsA and ArsD always occur in pairs (mentioned previously) (B.P. Rosen personal communication). This led them to raise the question whether ArsD may have an additional function to being a regulator. Subsequent evidence suggested that the ArsD might indeed function together with ArsA, although the results of this research have yet to be published (B.P. Rosen personal communication).

1.5.6. Other genes associated with *ars* resistance.

The most common forms of *ars* operon are the 3 and 5 gene operons, encoding marginal and high level arsenic resistance, respectively. However, the more operons that are sequenced, the more diversity they seem to contain. Recently, operons containing fairly unusual *ars* gene homologues have been isolated. The first operon belongs to the pYV plasmid from *Yersinia enterocolitica*. This particular *ars* operon has the basic *arsRBC* configuration, but also possesses an additional divergently transcribed *arsH* gene upstream from where the *arsR* gene is located (Neyt *et al*., 1997). The *arsH* gene possesses no hydrophobic domain; no ATP-binding motif and it also contains no other recognizable motifs or domains. The function of the *arsH* is not yet known, however it is required for arsenic resistance. This was determined by constructing a *Y. enterocolitica* strain lacking the *arsH* gene, which exhibited sensitivity to arsenic (Neyt *et al*., 1997). Supplying the *arsH* in *trans* restored resistance to arsenic. This result would have been somewhat surprising, considering that the basic *arsRBC* in other bacteria is sufficient to confer marginal resistance to arsenic. The authors speculated that the function of the *arsH* might be similar to that of *arsD*, acting as some kind of regulator. However, in contrast to these findings, Butcher *et al*. (2002) showed that although the *arsH* homologue of *At. ferrooxidans* was expressed in an *E. coli in vitro* transcription-translation system; it was not absolutely required for arsenic resistance in *E. coli*.

Ryan and Colleran (2002) also discovered another divergently transcribed *arsH* homologue on the IncH12 plasmid R478 from *Serratia marcescens*. This operon was similar to that found on *Y. enterocolitica* plasmid pYV. However, this *arsH* homologue was truncated (16% of the C-terminal region was missing), resulting in slight arsenic sensitivity. When the *arsH* was entirely eliminated, it resulted in complete loss of arsenic resistance.

The other operon mentioned belongs to the *skin* element of *Bacillus subtilis*, consisting of the standard *arsRBC* genes, but also containing an additional fourth gene (*orf2*) (Sato and Kombayashi, 1998). This gene is located between the *arsR* and the *arsB* genes of the operon and showed no significant homology to any of the known *ars* genes. Experiments conducted with this gene did not elucidate much about its function in *ars* resistance.

Even more recently, an arsenite(III)-methyl transferase homologue named *arsM*, was found on the *Halobacterium* spp. plasmid pNRC100. It was thought that this bacterium contained two *ars* operons, the first consisting of the genes *arsADRC*, the second containing the genes *arsR2*

(a second *arsR* gene) and *arsM*, which both occurred on the plasmid with the *arsB* located on the chromosome (Wang *et al*., 2004). In an experiment where the *arsM* was deleted, the sensitivity to arsenite was increased. The authors, therefore, hypothesized that As(III)-methyltransferase methylated intracellular arsenite, thereby creating a concentration gradient to the outside of the cell, although this has yet to be proven.

1.6. Transposons.

A significant percentage of DNA in naturally isolated bacteria exists in the form of plasmids. Many of these plasmids are widely transferable across species and even genus boundaries through a process called conjugation. In addition to carrying genes for their own housekeeping, these plasmids also carry mobile genetic elements called transposons. One of the most recognizable, largest and most widely distributed examples is the multicopy composite transposon Tn*21* (Davies, 1994; Kruse and Sorum, 1994). Transposon Tn*21* in itself is carried on a plasmid, NR1 (R100), which was originally isolated from *Shigella flexneri* in Japan in the late 1950's (Nakaya *et al*., 1960). NR1 is a 95.5 kb, self-transmissible plasmid with multiple antibiotic resistance and is the prototype of a vast collection of R-plasmids that have been discovered around the world (Davies and Rownd, 1972). The Tn*21*-family has been demonstrated to be involved in global dissemination of a variety of resistance determinants, including mercury resistance, among several bacterial species. Many transposons encoding multiple antibiotic resistances isolated from members of the *Enterobacteriaceae* belong to the Tn*21* subgroup of the Tn*3* family of transposable elements. Tn*21* along with its closest relatives carries a potentially independently mobile DNA integron, which encodes an integration system responsible for the acquisition of antibiotic resistance genes. Sequencing of the Tn*21* transposon revealed that it contained four discrete mobile elements: a class 1 integron, the *aadA1* gene cassette, which is contained in the integron and two insertion sequences (IS*1326* and IS*1353*) (Brown *et al*., 1996)

Integrons are also major distributors of multiple antibiotic resistances, which piggyback on transposons such as Tn*21* and related transposons. The integron (In2) located on Tn*21* belongs to the class 1 of integrons and contains an integrated gene cassette, called *aadA1* (Fig.1.9). It is 11 kb long and is flanked by imperfect inverted repeats of 25 bp (IR$_i$ and IR$_j$). In2 does not possess a complete set of transposition (*tni*) genes, therefore cannot self transpose. It is speculated that since relatives of In2 are found in other independent locations, the movement of

In2 into Tn*21* occurred as a result of transposition genes being provided *in trans* from another source. (Brown *et al*., 1996). Three distinct regions makeup the backbone of In2, the 5'-CS (conserved segment), the 3'-CS and the remnants of a *tni* transposition module. (Fig.1.9)
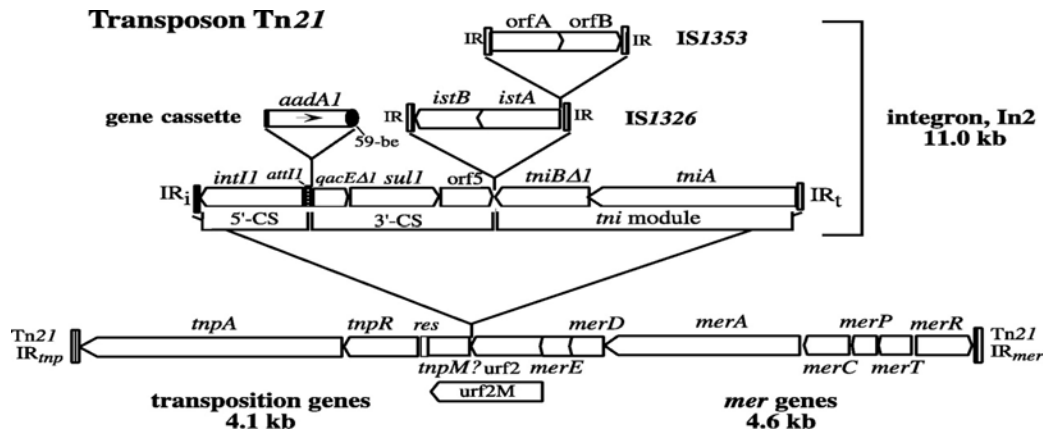
Figure 1.9: Structural layout of transposon Tn*21* and elements. Vertical bars indicate inverted repeats and insertion sequences. The *tnp* region includes the *tnpA* (transposase), *tnpR* (resolvase), *tnpM* (putative transposition regulator) and the *res* site. The 5'-CS of the integron contains the *intI1* (integrase), the *attI1* insertion site and the *aadA1* gene cassette. The 3'-CS includes the genes *qacΔE* conferring resistance to quaternary ammonium compounds and *sul1* conferring sulfonamide resistance. It also includes ORF5, which has no known function and two insertion sequences (IS*1326* and IS*1353*). The *tni* module contains *tniA* and a deleted version of *tniB*. (taken from Liebert *et al*., 1999).

The 5'-CS or *int*I module contains three repeats at its left end, which are expected to bind the *tni* transposase. It is comprised of the following: the *int*I1 gene, encoding a class 1 integrase; promoter regions for the *intI1* and *aadA1* genes; and the majority of the *attI1* site, and the IntI1-specific recombination site into which the cassette is inserted (Radstrom *et al*., 1994). IntI1 is a site-specific recombinase that catalyzes recombination between different parts of the recombination site. The IntI1 also has additional functions, which includes catalyzing the integration of additional gene cassettes at the *attI1* site as well as excision and reorganization of cassettes (Collis *et al*., 1993).

The In2 has a single cassette possessing the *aadA1* gene, determining resistance to streptomycin and spectinomycin. The AAD(3") protein is an aminoglycoside adenyltransferase causing inactivation of these antibiotics via addition of adenyl or nucleotidyl groups. This cassette also has an *IntI1*-specific recombination site, which is described as the *aadA1* 59-base

element situated downstream of the gene. Incorporation of the *aadA1* cassette into In2 supposedly took place at the *attA1* site via IntI1-mediated recombination (Liebert *et al*., 1999).

The 3'-CS (Fig.1.9) was identified as a conserved region in class 1 intregrons, containing three ORFs (*qacEΔ1*, *sul1* and ORF5) (Stokes and Hall, 1989). The *qacE* gene encodes an inner membrane protein, which functions via an efflux mechanism determining resistance to quaternary ammonium compounds (such as antiseptics and disinfectants) (Neyfakh *et al*., 1991; Rouch *et al*., 1984). The *qacEΔ1* on the other hand confers only limited resistance to these compounds due to the last 66 bp being different to the native *qacE*. The *sul1* gene is a dihydropteroate synthase gene conferring sulfonamide resistance. The function of ORF5 is not known (Worley *et al*., 1995).

In2 also contains two insertion sequences, IS*1326* and IS*1353*. IS*1326* has a terminal 26 bp IRS (inverted repeats) and two overlapping ORFs (Brown *et al*., 1996), which have been designated *istA* and *istB*, due to their homology with the IstA and IstB proteins from IS*21* (Reimmann *et al*., 1989). These genes share a common promoter located upstream of the *istA* gene, which cannot express due to the insertion of IS*1353* into the promoter region. IS*1353* is a member of the IS*3* family (Brown *et al*., 1996) and is flanked by a direct duplication of 2 bp of the target sequence. It possesses two ORFs, *orfA* and *orfB*, which are transcribed in the opposite direction to *istA* and *istB* from IS*1326*.

The Tn*3*-like transposons are distinguished by flanking inverted repeats of about 38bp and possess two genes, *tnpA* and *tnpR* (Grinsted *et al*., 1990; Sherratt, 1989) (Fig.1.9) Respectively these genes encode a transposase and a resolvase, which are responsible for transposition. The *tnpR* and the *res* site (site where the resolvase acts) is situated upstream from the *tnpA*. The *tnp* region of the Tn*21* begins at the left IR (IR$_{tnp}$) ending at the right IR (IR$_{mer}$); this region also includes the *res* site. Encoding 988 aa, the *tnpA* is the largest structural gene in the transposon. The *tnpR* gene is 561 bp in length and lies to the upstream of the *tnpA* gene. The predicted promoter region for these genes is situated within the *res* site, which has three resolvase binding sites (I, II, and III), with the recombination crossover point positioned in *resI*. Tn*21* also contains the putative *tnpM*, a gene with unknown function. Early studies involving the *tnpM* showed that it was not necessary for transposition. However, further studies reported that *tnpM* produces a protein, which enhances Tn*21* transposition (Hyde and Tu, 1985)

Transposition of Tn*21* is usually performed by the transposase (TnpA). The TnpA insertion site is recognized as a 5 bp duplication in an AT-rich region of the target DNA. Transposition of Tn*21* involves specific recognition and binding to terminal inverted repeats by the TnpA, which then joins the donor and recipient replicons. The action of the TnpA leads to the formation of a co-integrate intermediate containing both the donor and recipient replicons with two replicated copies of the transposon. The resolvase (TnpR) catalyzes site-specific recombination, resulting in one copy of the transposon being inserted into each of the two elements. Resolvase subunits bind to three subsites of the *res* site, within both recombination loci. These protein-DNA complexes interact with one another at the *res* sites forming a synaptic complex. The resolvase separates these complexes giving rise to a new recombinant replicon containing one copy of the transposon and the original donor replicon (Stark *et al*., 1995).

Transposons carry many traits that can be advantageous to a host cell in a variety of environments. They do not only carry resistance determinants, but can also carry genes useful for substrate utilization. It is for these specific reasons that transposons and other transposable elements are important for evolution of bacterial host cells. Transposons have the ability to travel around collecting different genes/operons even other transposable elements. Tn*21* and its relatives are such transposons, traveling around the world disseminating varieties of different genes to bacteria across species and genus borders. It was observed that Tn*21* and related transposons were found in soil bacteria from three different locations (Pearson *et al*., 1996). This is testament to the widespread nature of these elements. Recently, a Tn*21* derivative was isolated in our lab containing genes for arsenic resistance (de Groot *et al*., 2003), from the biomining bacterium *Acidithiobacillus caldus*. This bacterium is a sulfur-oxidizing acidophile occurring in biooxidation tanks with low pH and high concentrations of arsenic. This environment provided the perfect setting for acquisition of arsenic resistance genes, which happened to be on a Tn*21*-like transposon.

1.7. Aim of Thesis

Cultures of *Leptospirillum* were obtained from various origins, which were used in an earlier study. Some had previously been exposed to high levels of arsenic, while other strains had little or no history of exposure to the toxic metal. A preliminary investigation into the arsenic resistance genes of various *Leptospirillum* strains indicated that *L. ferriphilum* Fairview was the most likely candidate to contain *ars* genes. Therefore, the aim of this study was to characterize the arsenic resistance genes of *L. ferriphilum* Fairview and to investigate how widely distributed these genes are among members of the genus.
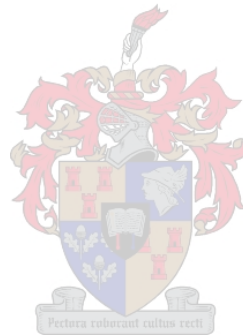
Chapter Two: Results

TABLE OF CONTENTS

2.1. INTRODUCTION

2.2. METHODS

2.3. RESULTS and DISCUSSION

2.1. INTRODUCTION

Leptospirillum ferriphilum is an acidophilic (pH 1.4-1.8), obligately chemolithotrophic bacterium, that uses ferrous iron or pyrite as its sole energy source (oxidizes ferrous to ferric iron). It was isolated from the arsenic rich environment of the industrial biooxidation tanks at the Fairview mine (Barberton, South Africa), where it is a member of a consortium of bacteria used in the recovery of gold from refractory ore (Rawlings and Silver, 1995). Previously, it was shown that *L. ferrooxidans* was the dominant iron oxidizer in the continuous-flow biooxidation process (Rawlings and Silver, 1995; Rawlings *et al*., 1999). This bacterium is now known to be *L. ferriphilum* (Coram and Rawlings, 2002). Arsenic levels in these tanks are usually in excess of 13 g/L and therefore bacteria present in this environment require mechanisms of resistance to arsenic (Dew *et al*., 1997).

Arsenic efflux mechanisms are widely distributed amongst microorganisms and are well studied and reviewed (Cervantes *et al*., 1994; Silver and Phung, 1996; Silver, 1996; Xu *et al*., 1998; Rosen, 1999; Rosen, 2002a; Rosen, 2002b and Mukhopadhyay *et al*., 2002). Initially, arsenic resistance determinants were mainly found on plasmids, but due to the increase in genomic sequencing projects they have now been discovered on chromosomes. The arsenic resistance operon has two basic forms, one consists of a 3 gene arrangement (*arsRBC*), which are most commonly found on chromosomes, such as in the case of *Pseudomonas aeruginosa* (Cai *et al*., 1998). The other operon contains 5 genes consisting of *arsRDABC*, which has thus far only been found on plasmids, for example *E. coli* R773 (Chen *et al*., 1986).

Here we report the isolation and analysis of the evolutionary relationship of arsenic genes from *L. ferriphilum*. We wanted to know, how these bacteria had become so resistant to arsenic. Once we found that this was due to additional *ars* genes that were not present in *Leptospirillum* strains which had not been previously exposed to arsenic, we asked where these *ars* genes could have come from? We further showed that the *ars* genes were located on a transposon occurring only in *L. ferriphilum* strain Fairview, that the genes were functional in *E. coli* and that the transposon on which the genes are located was active in *E. coli*.

2.2. METHODS

Bacterial strains, media and growth. Strains used in this study are listed in Table 2.1. Bacterial strains were grown at 30 ˚C in 800 ml mineral salts medium containing (% w/v): $(NH_4)_2SO_4$, 0.2; K_2HPO_4, 0.05; $MgSO_4.7H_2O$, 0.05; KCl, 0.01; Na_2SO_4, 0.1; and $Ca(NO_3)_2$, 0.001, which was supplemented with $FeSO_4.7H_2O$ (500 mM) and 100 x trace elements solution (1μl/10 ml) containing (% w/v) ($FeCl_3.6H_2O$, 0.011; $CuSO_4.5H_2O$, 0.0005; H_3BO_3, 0.02; $MnSO_4.H_2O$, 0.02; $Na_2MoO_4.2H_2O$, 0.0008; $CoCl_2.6H_2O$, 0.0006; $ZnSO_4.7H_2O$, 0.0009, adjusted to pH 1.6 with concentrated H_2SO_4. Purity of the strains was checked by PCR as described by Rawlings *et al*. (1999).

Construction of the *L. ferriphilum* genebank. Bacterial strains were harvested by centrifugation at 5000 rpm for 30 min in a Beckman JA10 rotor and washed with acid water (pH 1.8) to remove the ferric iron precipitate and resuspended in TE buffer (pH 7.6). Cell lysis was achieved in the presence of 10 % SDS (Sodium Dodecyl Sulfate) and Proteinase K (1mg/ ml) at 37 ˚C for 30 min. Genomic DNA was precipitated with 100 % ethanol, extracted via spooling and washed twice with 70 % ethanol. DNA was then resuspended in TE buffer by gentle shaking overnight. Subsequent cleaning of the DNA occurred with chloroform isoamylalcohol. The prepared *L. ferriphilum* Fairview genomic DNA was partially digested with *Sau*3A for optimization of the digest. Once optimal conditions were achieved a bulk digest of the genomic DNA was performed (80 μg DNA/4 ml digest) for 30 min at 37 ˚C. The partially digested *Sau*3A DNA fragments were precipitated, resuspended in TE buffer (pH 7.6) and separated on 5-40 % sucrose gradient, which was centrifuged at 23000 rpm for 24 h in a SW28 rotor. Gradient fractions were removed in 200 μl increments (26 fractions) and divided into pools according to the sized fragments they contained. The larger fragments were cloned into the *Bgl*II site of the suicide cloning vector pEcoR252. Ligated DNA was transformed into *E. coli* DH5α and growth was selected for on Luria agar (LA) plates containing ampicillin (100 μg/ml). Approximately 9000 colonies were obtained. These colonies were scraped off the plates and inoculated in 800 ml LB containing ampicillin (100 μg/ml) and grown overnight at 37 ˚C. Cells were harvested by centrifugation at 5000 g for 5 min in a JA10 rotor and prepared for plasmid preparation by CsCl ultracentrifugation. The *L. ferriphilum* genebank was screened in the *ars* deletion mutant *E. coli* ACSH50I[q] and clones were selected for their ability to complement the mutant on LA plates containing 0.5 mM sodium arsenite. In order to obtain the entire *ars* operon on one continuous fragment, a 4.4 kb *Bam*HI-*Spe*I fragment from pEco-*ars*11

containing a fraction of the *arsB* gene and complete *tnpA* gene was ligated to a 9.8 kb *Bam*HI-*Spe*I fragment from pEco-*ars*9, resulting in the 14.2 kb hybrid plasmid p*L.fars* (Figure 2.1), later named pTn*LfArs*.

Table 2.1: Strains, plasmids and primers

Stains, plasmids and primers	Description	Reference or source
Strains		
Escherichia coli		
ACSH50I$_q$	*rspL* Δ(*lac-pro*) [F′ *traD36 pro lacI^q* ΔM15] CSH50I$_q$ Δ*ars::cam*	Butcher and Rawlings, 2002
DH5α	ø80d*lacZ*ΔM15 *endA*1 *recA*1 *gyrA*96 *thi*-1 (r$_k$⁻, m$_k$⁺) *relA*1 *supE*44 *deoR*Δ (*lacZYA-argF*) U169	Promega Corp. USA
Leptospirillum ferrooxidans		
DSM2705	Wild type	DSMZ, Braunschweig, Germany, Markosyan strain, Cu mine
ATCC49879	Wild type	Wolfgang Sand, Romania
P$_3$a	Wild type	Barrie Johnson, Coal mine North Wales, UK
Parys	Wild type	Barrie Johnson, Parys Mountain, Anglesey Cu mine, Wales
Chil-Lf2	Wild type	Barrie Johnson, Cu mine, Chile
Leptospirillum ferriphilum		
Fairview	Wild type	Ellen Lawson, South Africa
ATCC49881	Wild type	Wolfgang Sand, Romania
Warwick	Wild type	Paul Norris, Warwick,UK
Mont.4	Wild type	Peggy Arps, Pyrite column, Montana, USA.
Acidithiobacillus caldus		
#6	Wild type	Shelly Deane, Fairview plant, Billiton, South Africa
MNG	Wild type	This laboratory
Acidithiobacillus ferrooxidans		
ATCC33020	Wild type	Rockville, Md.
ATCC27230	Wild type	Rockville, Md.
Plasmids		
pUCBM21	Ap^r *lacZ*′, cloning vector	Stratagene, USA
pBluescript KS	Ap^r *lacZ*′, cloning vector	Stratagene, USA
pUC19	Ap^r *lacZ*′, cloning vector	Stratagene, USA
pEcoR252	Ap^r, cloning vector	Stratagene, USA
pEco-*ars*9	Ap^r (from *L. ferriphilum* plasmid library – 6.6 kb *Sau*3A1 fragment cloned into pEcoR252 digested with *Bgl*II)	This study

pEco-*ars*11	Ap^r (from *L. ferriphilum* plasmid library – 11kb *Sau*3A1 fragment cloned into pEcoR252 digested with *Bgl*II)	This study
pTn*LfArs*	Ap^r (4.4 kb *Bam*HI-*Spe*I fragment of pEco-*ars*11 cloned into pEco-*ars*9 cut with *Bam*HI and *Spe*I)	This study
pUCTn*LfArs*	Ap^r (13 kb *Hind*III fragment from Tn*LfArs* cloned into pUC19)	Marla Tuffin (in press)
Primers		
Sequencing primers		
Nco-Stu.1 Fwd	5´-TATAAGGCCTACCGGAATCATCACC-3´	This study
Nco-Stu.1 Rev	5´-TATAGGATCCTGACCACAGAAGTGC-3	This study
Nco-Stu.2 Fwd	5´-TATAGCACTGATTCTCACG-3´	This study
Nco-Stu.2 Rev	5´-TATATAAGAGGTCAATCCG-3´	This study
Stu-Eco.1 Fwd	5´-TATAGGATCCATCTTGAGCAGGCGG-3´	This study
Stu-Eco.1 Rev	5´-TATAAGGCCTACTGGATTTACTCGG-3´	This study
Stu-Eco.2 Fwd	5´-TATAACGATCTGCTCAAGC-3´	This study
Stu-Eco.2 Rev	5´-TATACGCAGATAGCAGAGC-3´	This study
Eco-Bam Fwd	5´-CGTCAAGATCACAGAATTGC-3´	This study
Eco-Bam Rev	5´-AGCCCAGCAGGTGCATCAGC-3´	This study
TnpA Fwd	5´-GCACGCTCAACATCAAGC-3´	This study
PCR primers		
Ferro *arsB* Fwd	5´-GTTIGCCAACGAIGGIGCGGC-3´	Shelly Deane
Ferro *arsB* Rev	5´-ACATGCAICCAGAGCAGIGTIGC-3´	Shelly Deane
Probes		
L.f arsB	850 bp PCR product from *L. ferriphilum* amplified using Ferro *arsB* Fwd and Ferro *arsB* Rev degenerate primers	This study
*Eco*RI-*Eco*RV	2.0 kb *Eco*RI-*Eco*RV fragment from pTn*LfArs*	This study
*Eco*RV-*Sph*I	1.0 kb *Eco*RV-*Sph*I fragment from pTn*LfArs*	This study
*Nco*I-*Eco*RI	3.7 kb *Nco*I-*Eco*RI fragment from pTn*LfArs*	This study
*Nco*I- *Spe*I	1.7 kb *Nco*I- *Spe*I fragment from pTn*LfArs*	This study
Tn*AtcArs arsDA*	2.2 kb *Ksp*I fragment from pTn*AtcArs*	Marla Tuffin

DNA sequencing and analysis. DNA sequencing was done by the dideoxy chain method (Sanger *et al*., 1977). Sequencing reactions were run on an ABI PRISM automated DNA sequencer by the Stellenbosch University sequencing service. Sequencing primers used for primer walking were synthesized by Inqaba Biotechnological Industries and are listed in Table 2.1. Sequence analysis was done using the PC-based program DNAMAN (version 4.13) from Lynnon Biosoft. Homology 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). Sequence alignments, phylogenetic and homology trees were constructed using the Multiple Sequence Alignment tool in DNAMAN.

DNA techniques. Plasmid preparation, restriction endonuclease digestion, gel electrophoresis, and ligations were carried out using standard methods described by Sambrook *et al*., (1989). For Southern hybridization experiments, DNA probes were labelled with the dioxygenin-dUTP non-radioactive labelling and detection kit (Roche Biochemicals). DNA probes used for Southern hybridization are listed in Table 2.1. Hybridization of probes was carried out at 40 ˚C Hybridization solution (Roche Biochemicals). Stringency washes (**A**: 2 X SSC, 0.1 % SDS) were for 20 min at room temperature followed by another (Stringency wash **B**: 0.1 % SSC, 0.1 % SDS) for 20 min at 65˚C. Membrane detection performed as per manufacturers instruction (Roche Biochemicals).

Polymerase chain reaction. PCR amplification of the *L. ferriphilum arsB* gene was carried out using degenerate primers (Ferro *arsB* fwd and Ferro *arsB* rev.) described in Table 2.1. Approximately 100 ng of genomic DNA was subjected to amplification in total volume of 50 μl containing 20 mM $(NH_4)_2SO_4$; 75 mM Tris-HCl (pH 8.8); 0.1 % (v/v) Tween20; 3 mM $MgCl_2$; 2.5 μM dUTPs (dATP, dCTP, dGTP and dTTP); 0.25 μM of each primer and 2u of *Taq* polymerase (Bioline USA Inc.). The PCR reaction was performed in a PCR sprint (Hybaid). Denaturation occurred at 94 ˚C for 30 s, followed by 30 amplification cycles of 30 s at 94 ˚C, 30 s at 60 ˚C and 90 s at 72 ˚C, with an additional final extension step at 72 ˚C for 120 s before cooling to 4 ˚C completed the reaction.

Arsenic assays. Arsenic assays were performed in *E. coli* ACSH50I[q] containing the plasmids pTn*LfArs*, pTn*AtcArs* and pEcoBlunt as a control. Assays were carried out in LB media containing ampicillin (100 μg/μl) and increasing concentrations of sodium arsenite. Assays for resistance to arsenate were carried out on cells grown in low phosphate media (Oden *et al*., 1994), which was supplemented with 2 mM K_2HPO_4. Overnight cultures were diluted 100-fold into fresh media containing the appropriate antibiotics and increasing concentrations of sodium arsenate and arsenite. Cultures were incubated at 37 ˚C for 5 h and absorbance was determined at 600 nm. For the purpose of the graph absorbance values were converted to percentages. The incubation period corresponds to the middle of the log phase of growth of a controlled culture under the same conditions.

Assay for transposition. This experiment was performed by Dr Marla Tuffin (in press) as follows: To investigate if the *L. ferriphilum* transposon is active, the entire insert from pTn*LfArs* was cloned into the non-mobilizable vector pUC19, to construct pUCTn*LfArs* (Ap[r]). pUCTn*LfArs* was transformed into *E. coli* XL1-blue containing the conjugative plasmid pSa (Km[r]). If the transposon was functional, it would be able to transpose into pSa, which would then be conjugated to the recipient cell. To establish if the transposition event took place, gel electrophoresis of plasmid DNA prepared from the transconjugants should indicate a gain in size for plasmid pSa. Two transformants were selected and incubated in 5 ml LB at 37 ˚C overnight. The transformants were mated with *E. coli* ACSH50I[q] (Rif[r], Marla Tuffin), by mixing equal volumes of donor and recipient strains on Luria agar plates and incubated 2 h at 37 ˚C. The growth was scraped off and resuspended in LB. Various dilutions were spread onto selective media consisting of LA plus rifampicin, kanamycin and 1 mM arsenate (selection for the transposon). Individual colonies (transformants LF D1 and LF D6) were selected and prepared for plasmid DNA isolation. The plasmid DNA was digested, subjected to Pulsed-field Gel Electrophoresis (TAFE Geneline) and Southern hybridization. A 2.2 kb *Ksp*I DNA fragment containing the *arsDA* genes from pTn*AtcArs* was used as a probe.
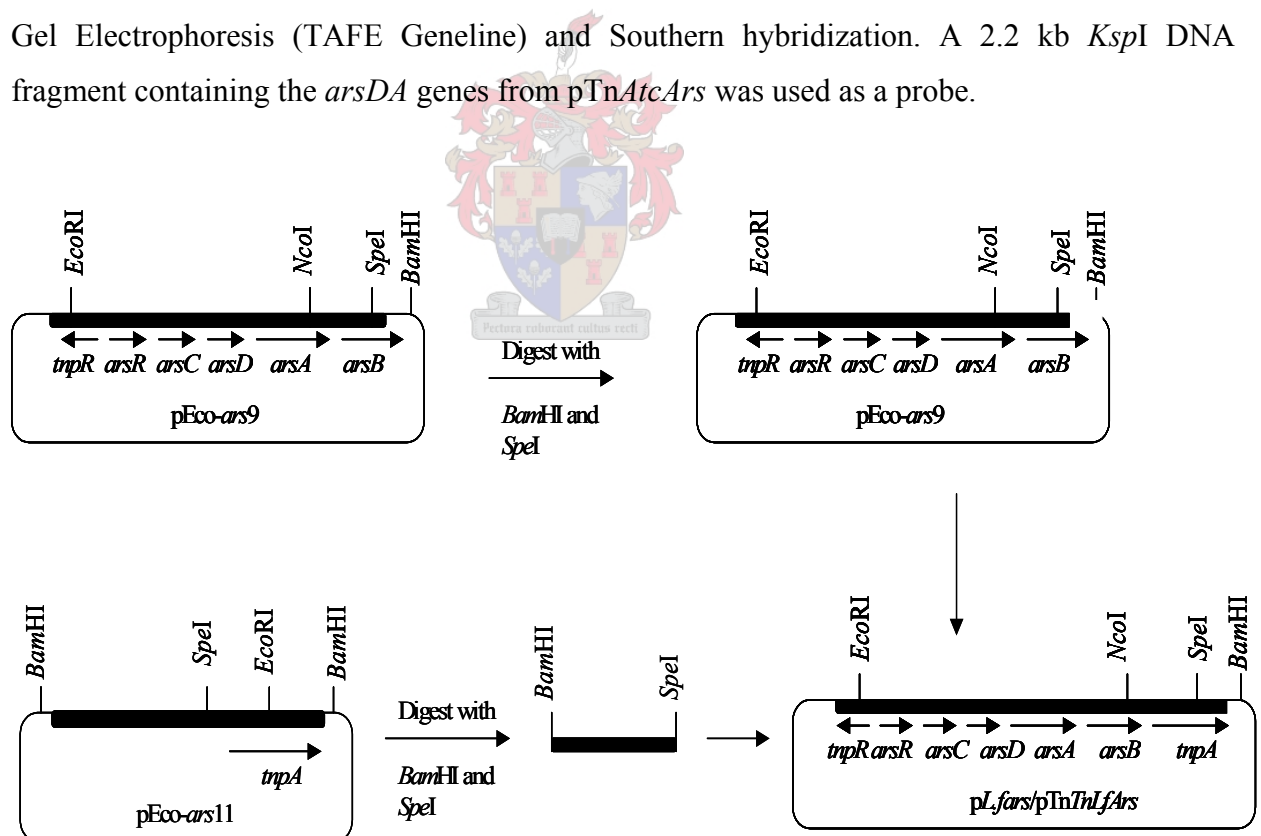
Figure 2.1: Diagram showing the strategy of assembling of entire *L. ferriphilum ars* operon. The plasmid pEco-*ars*11 was digested with *Bam*HI and *Spe*I, the resulting 4.4 kb fragment, which was cloned into pEco-*ars*9 digested with the same restriction endonucleases.
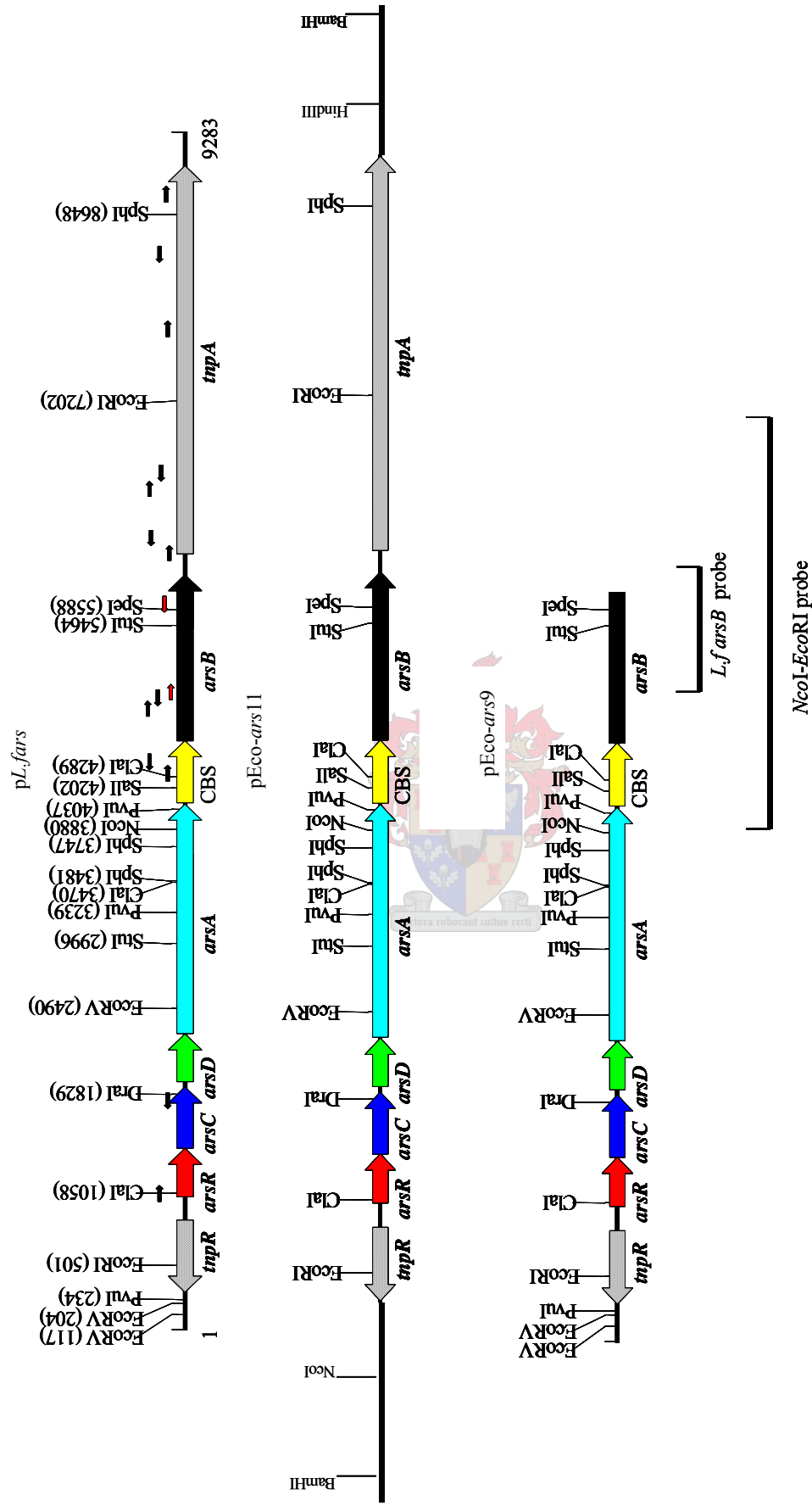
Figure 2.2: Restriction map of Tn*LfArs*. This diagram depicts the restriction endonuclease sites, genebank clones used to construct pTn*LfArs*, probes used in this study and primers used for sequencing and PCR reactions. Black arrows = sequencing primers; Red arrows = PCR primers. Probes and clones are indicated by horizontal black lines.

2.3. RESULTS and DISCUSSION

2.3.1. Cloning of the *ars* genes from *L. ferriphilum*.

It was previously discovered that *At. ferrooxidans* and *At. caldus* possess genes for arsenic resistance (Butcher *et al*., 2000; De Groot *et al*., 2003). These bacteria, as previously mentioned, occur in the same environment as some *L. ferriphilum* and *L. ferrooxidans* strains. Therefore we wished to establish which *Leptospirillum* species in our possession contained *ars* genes. Before attempting to isolate the *ars* genes, from what we presumed to be the highly arsenic resistant *Leptospirillum* strain from the Fairview plant, preliminary experiments were done to detect the presence of *ars* genes within these bacteria.

Sequence alignments of *arsB* proteins from *At. caldus*, *At. ferrooxidans* and *E. coli* R773 indicated that there are a number of conserved regions amongst all three proteins (Figure 2.3). A set of degenerate primers (Table 2.1) were designed (S. Deane, pers. comm.) from the nucleotide sequence in the region of these conserved amino acid sequences in order to amplify the *arsB* genes from *Leptospirillum* strains believed to be arsenic resistant. PCR amplification using these primers resulted in products being obtained for *At. ferrooxidans*, *At. caldus* and *E. coli* (used as positive controls) as well *L. ferriphilum* Fairview (all products were 850 bp in length) (Figure 2.4). The PCR product from *L. ferriphilum* Fairview was cloned and sequenced. BLAST results of this sequence indicated highest hits to an *arsB* homologue from *Serratia marcescens*. In order to detect the presence of *ars* genes in other *Leptospirillum* strains, Southern hybridization analysis was performed using this PCR product as a homologous *arsB* probe. The probe hybridized only to *L. ferriphilum* Fairview genomic DNA (the amplified *L. ferriphilum arsB* was used as a control) (Figure 2.5). This result indicated that either none of the other *Leptospirilli* possessed *ars* genes or the *ars* genes they contained were different to that of *L. ferriphilum* Fairview. From the Southern hybridisation we also noticed that the *L. ferriphilum* Fairview fragment that was detected appeared to be a doublet, which could imply that there were two *L. ferriphilum* Fairview *ars* operons or the DNA could have been partially digested. A *L. ferriphilum* Fairview plasmid bank was then constructed using the plasmid cloning vector pEcoR252 and transformed into the *ars* operon deletion mutant *E. coli* ACSH50I^q. Colonies were selected for their ability to complement the mutant on LA plus 0.5 mM sodium arsenite plates and resulted in 15 arsenite resistant transformants with insert sizes ranging from 3-16 kb. Two clones (pEco-*ars*9 and pEco-*ars*11), which showed both arsenite and arsenate resistance, upon retransformation into *E. coli* ACSH50I^q, was selected for further study. Preliminary mapping and sequencing indicated that these clones had similar restriction

endonuclease sites, but one clone (pEco-*ars*9) appeared to be truncated in the *arsB* gene and the other clone (pEco-*ars*11) appeared to be rearranged upstream of the *arsA* gene. This new joined construct was called p*L.fars*. In order to obtain the entire *ars* operon on a single plasmid a 4.4 kb insert from pEco-*ars*11 was added to pEco-*ars*9 using matching restriction endonuclease sites (as illustrated in the methods). Southern hybridization was used to confirm the source of the insert DNA. An internal *Nco*I-*Eco*RI fragment of 3.4 kb was labelled and used to probe genomic DNA from *L. ferriphilum* Fairview and p*L.fars* digested with the same restriction endonucleases (*Sph*I-*Spe*I, *Nco*I-*Sal*I and *Sal*I-*Spe*I) (Figure 2.6). Corresponding fragments of sizes 1.0 kb (*Sph*I-*Spe*I), 300 bp (*Nco*I-*Sal*I, detected following longer exposure time of Southern hybridization) and 1.4 kb (*Sal*I-*Spe*I) were obtained for both plasmid and genomic DNA digests, confirming that the insert DNA originated from *L. ferriphilum* strain Fairview (discussed further in section 2.3.3). We were confident from sequencing results that the upstream region (upstream from the *arsA* gene) containing the rest of the *ars* operon had no sequence rearrangements, due to the entire region being present on one clone (pEco-*ars*9).

```
At_caldus:  MLT A FL IF IA TIV LVIWQPKGLGV GWSA LL GA AL AL LA GV VHLGDIPT VW
At_ferroox: .ML A VA IF LV TLI LVIWQPKGLSI GWSA MG GA VV AL AT GV ITWSDIPI VW
E_coli_R773: MLL A GA IF IL TIV LVIWQPKGLGI GWSA TL GA VL AL AS GV IHIADIPV VW

At_caldus:  G IVW N ATI TF I A V II T SLL LDE A GFF E WAALHV A RWG N GN GR R LF
At_ferroox: H IVW D ATF TF V A L II I SLI LDE A GFF H WAALHV V RWG G GR GR L LF
E coli_R773: N IVW N ATA TF I A V II I SLL LDE S GFF E WAALHV S RWG N GR GR L LF
```

                              Ferro arsB Fwd.
                         ─────────────────────►

```
At_caldus:  AFF V L LGA AV AA L FANDGAALI LTPIV MAM LL A LR FSPAA TL AFV M AA GF
At_ferroox: PLI V V LGA MI AA V FANDGAALL LTPIV MAI LL R LD FTPAA TF AFV I AT GF
E_coli_R773: TYI V L LGA AV AA L FANDGAALI LTPIV IAM LL A LG FSKST TL AFV M AA GF

At_caldus:  IA DT A SLPL VV SNLVNIVSADF F KIG F TE YA AI MVPVNIV A VA A L AMLMW
At_ferroox: VA DT T SLPL MI SNLVNIVSANY F NIS F DR YA TV MVPVDLV A LA A T L GVLWL
E_coli_R773: IS DT A SLPL IV SNLVNIVSADF F KLG F TE YA SV MVPVDIA A II A T L VMLHL

At_caldus:  F FRRDIP KD YDVMQ LKN PREA IL DSAT F KAGWW VL A LLLVGF FGLEHVGV P I S
At_ferroox: Y FRRVVP RR YLTRD LPE PASV IM DHVV F RASFP VL I LLLVAY FVTAQWQV P V S
E_coli_R773: F FRKDIP PT YELAR LKE PAKA IK DPAT F RTGWV VL L LLLVGF FVLEPMGI P V S

At_caldus:  AVASA GA LI LLGV A AR GHVISTR...... KVL KE APWQ IV I FSLGMYLVV YGLRN
At_ferroox: VVTGT GA LI LLAL A GR WLQGGRGARIPVR KVL RE APWA IV V FSLGMYLVV YGLRN
E_coli_R773: AIAAV GA AV LFAV A KK GHGINTG...... KVL RG APWQ IV I FSLGMYLVI YGLRN

At_caldus:  E GLT EHIAGL L DY FA QG GVWGA AF GTG I LT ALLSS V MNNMPT VL V GAL S I
At_ferroox: A GLT SYVAMA L HW FA GH GTVAA AL GTG F LA ALLSS I MNNMPA VL V GAL A I
E_coli_R773: A GLT DYLSDV L NE LA DK GLWAA TL GTG F LT ALLSS I MNNMPT VL I GAL S I
```

                                  Ferro arsB Rev.
                              ◄─────────────────

```
At_caldus:  DATSATGVVKDAMI... YANVIG S DLGPK V TPIGSLATLLWLHVL AR KG MT I TWG Y Y
At_ferroox: HQLSSVATDPLLREIMV YANVIG C DLGPK F TPIGSLATLLWLHVL SR KG MT V TWG Q Y
E_coli_R773: DGSTATGVIKEAMI... YANVIG C DLGPK I TPIGSLATLLWLHVL SQ KN MT I TWG Y Y

At_caldus:  FKV G VVL T V PVL AA TLAALA LR L SLT.
At_ferroox: MRT G LLI T P PVL LV TLLALA WW L PLA.
E_coli_R773: FRT G IVM T L PVL FV TLAALA LR L SVTL
```

Figure 2.3: Multiple sequence alignment of the *arsB* proteins from *At. caldus* (unpublished), *At. ferrooxidans* (AAF69238) and *E. coli* R773 (ARB1_ECOLI). Black arrows indicate location and direction of the ArsB proteins over which degenerate primers (Table 2.1) were designed.

Figure 2.4: PCR amplification of *arsB* gene from *L. ferriphilum*. Lane 1: λ *Pst*I marker. Lanes 2: *L. ferriphilum* Fairview. Lanes 3-5: *At. ferrooxidans, At. caldus* and *E. coli* as positive controls.

.
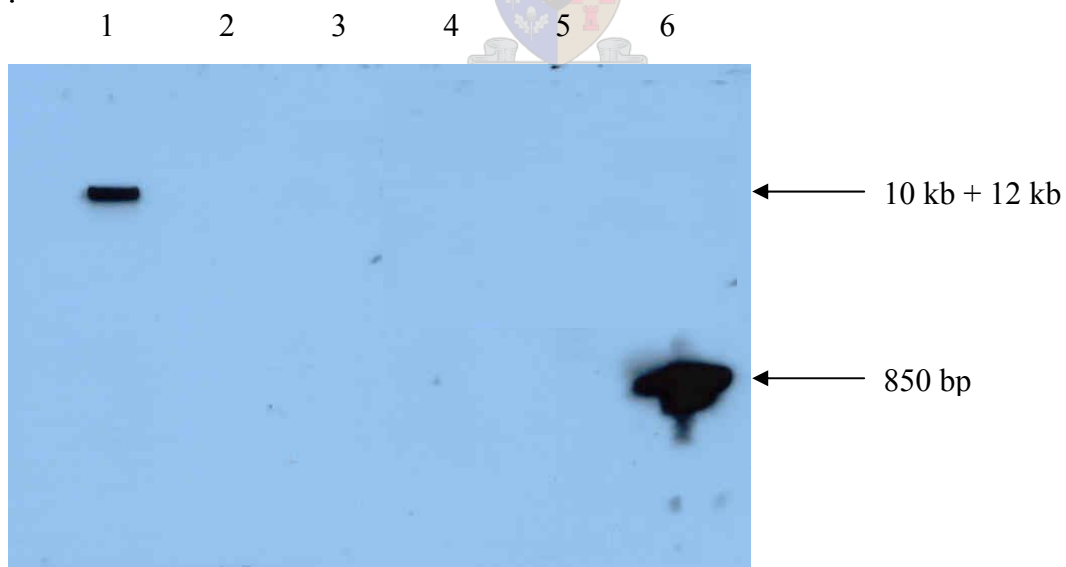
Figure 2.5: Southern hybridization of *Leptospirillum* strains cut with *Bam*HI and probed with *L. ferriphilum arsB* PCR product. Lane 1: *L. ferriphilum* Fairview. Lane 2: *L. ferriphilum* ATCC49881, Lane 3: *L. ferrooxidans* DSM2705 Lane 4: *L. ferrooxidans* ATCC49879, Lane 5: *L. ferriphilum* Warwick and Lane 6: *L. ferriphilum arsB* PCR product.
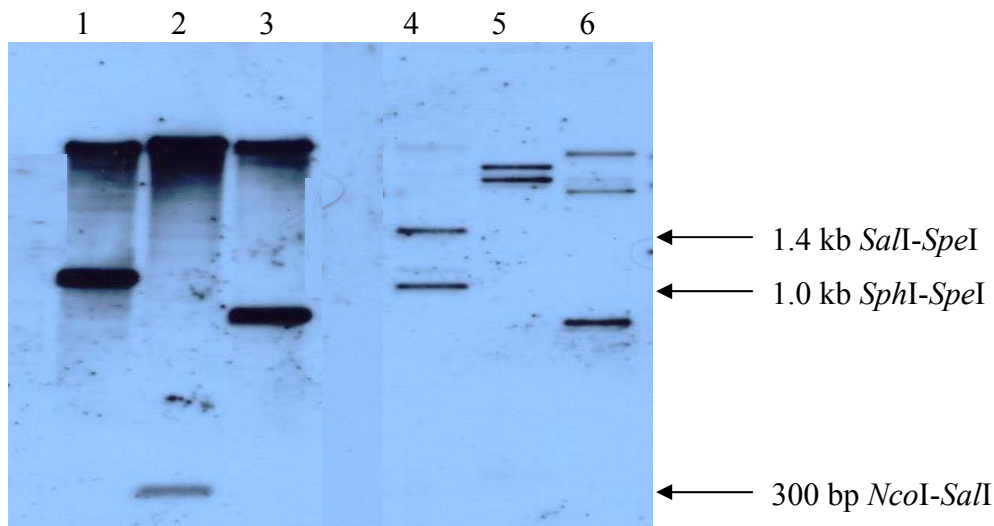
Figure 2.6: Southern hybridization using 3.4 kb *Nco*I-*Eco*RI probed against Tn*Lfars* (lanes 1-3) and *L. ferriphilum* Fairview genomic DNA digested with *Nco*I-*Sal*I, *Sph*I-*Spe*I and *Sal*I-*Spe*I (lanes 4-6), respectively.

2.3.2. Sequencing analysis of the *L. ferriphilum ars* operon.

The entire insert DNA from *L. ferriphilum* Fairview was completely sequenced from both strands, which led to the identification of 8 open reading frames (ORFs) (Table 2.2). The completely annotated sequence can be seen in appendix 2.

BLAST results identified various homologues to *arsRCDAB* genes, all transcribed in the same direction. However, unlike other conventional 5 gene operons, the *L. ferriphilum ars* operon was located between divergently transcribed *tnpR* and *tnpA* genes and further analysis revealed that these genes belong to the Tn*3* family of transposons with features that are most similar to the Tn*21* subfamily. Similarly, to the Tn*3* family, the *tnpR* (resolvase) and *tnpA* (transposase) genes are divergently transcribed, while in the Tn*21* subfamily these genes are transcribed in the same direction as a single unit (Liebert *et al.,* 1999) (Figure 2.7). Analysis of the inverted repeats (IR) indicated that both the left IR and the right IR are 44 bp in length, sharing an overall identity of 84 % to each other. The left and right IR is most similar to that of the Tn*21* subgroup with 36/38 bp and 34/38 bp, respectively (Figure 2.8). Results from the BLAST search (Table 2.2) indicated that the TnpR is most closely related to a resolvase of transposon Tn*5037* from *At. ferrooxidans*, while the TnpA is most closely related to that found on plasmid R100 from *Shigella flexineri* (this TnpA is also a member of the Tn*21* subgroup).

Table 2.2: Summary of ORFs identified in pTn*LfArs*

ORF	Position	No. of aa	Size (kDA)	Closest relationship to known Proteins	E-value
tnpR	854 - 290	187	21.7	93/96 % identity/similarity to resolvase of Tn*5037* from *Acidithiobacillus ferrooxidans* **CAC69253.1**	8e - 82
arsR	1036 - 1409	124	13.6	86/91 % identity/similarity to arsenic operon regulator of *Alcaligenes faecalis* **AAS45114.1**	7e - 54
arsC	1410 – 1885	158	16.9	93/94 % identity/similarity to cytoplasmic arsenate reductase of *Alcaligenes faecalis* **AAS45115.1**	7e - 70
arsD	1930 - 2294	121	12.9	97/98 % identity/similarity to hypothetical protein Mflag03002035 of *Methylobacillus flagellatus* **ZP_00172246.2**	7e - 65
arsA	2304 - 4078	591	63.3	97/98 % identity/similarity to ATPase subunit of *Alcaligenes faecalis* **AAS45117.1**	0.0
CBS	4090 - 4568	159	17.3	92/97 % identity/similarity to CBS domain protein of *Methylobacillus flagellatus* **ZP_00172244.2**	1e - 66
arsB	4569 - 5850	427	45.6	97/98 % identity/similarity to arsenite membrane pump of *Alcaligenes faecalis* **AAS45119.1**	6e - 166
tnpA	6019 - 8983	988	110.7	95/97 % identity/similarity to transposase of plasmid R100 of *Shigella flexneri* **NP_052901.1**	0.0

Another unusual feature of this operon is the presence of an ORF (159 aa) between the *arsA* and *arsB* genes with homology to CBS (cystathione-β-synthase) binding domains of *Methylobacillus flagellatus* and *Geobacter sulfurreducens*, with E-values of $3e^{-67}$ and $6e^{-19}$, respectively. Its function is not known, but the CBS domain usually occurs in two to four copies per protein, which dimerise to form a stable globular structure (Bateman, 1997).
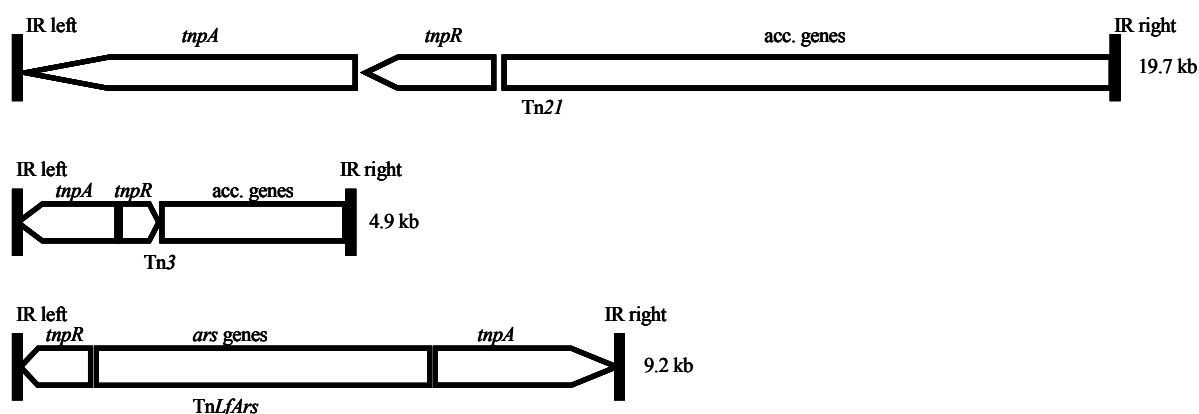
Figure 2.7. Depiction of the Tn*LfArs* compared to Tn*21* and Tn*3* transposons. Vertical bars left and right indicate flanking inverted repeats. acc. genes refers to accessory genes. (Drawing not to scale)

The first ORF upstream of the *tnpR* gene was identified as an *arsR* homologue with highest homology to the ArsR protein from *A. faecalis* (E-value = $7e^{-54}$) and *M. flagellatus* (E-value = $8e^{-50}$). Butcher *et al.* (2000) found that the ArsR protein from *At. ferrooxidans* and related proteins do not contain the putative metal-binding domain (ELCVCDL) typical of other ArsR proteins, which is required for binding of the arsenite inducer (Shi *et al.*, 1994). Sequence alignments of ArsR from *L. ferriphilum* with 12 different ArsR proteins (including closely related and other known ArsR proteins) (Figure 2.9) indicated that the *L. ferriphilum* ArsR protein along with its closest relatives also lacks the metal-binding domain. Subsequent phylogenetic analysis (Figure 2.10) revealed that it is more closely related to a subgroup of ArsR proteins that was described in an earlier study by Butcher and Rawlings (2002). A feature of this subgroup includes the apparent conserved sequence GX(L/I)A downstream of the region corresponding to the metal-binding domain in putative ArsR proteins. An additional feature evident from sequence alignments (Butcher and Rawlings, 2002) is the presence of a cysteine doublet towards the C-terminal ends of these ArsR proteins. When secondary structure predictions of the *L. ferriphilum* ArsR protein were compared to that of proven ArsR regulators

from *E. coli* R773, *At. ferrooxidans* and a member of the ArsR family, SmtB (from *Synechococcus*) (Figure 2.11), it was found that the overall structure of the ArsR proteins from *L. ferriphilum* and *At. ferrooxidans* were similar, comprising of at least 5 helices and 2 β-strands (Figure 2.11).

```
Tn LfArs_left_IR   GGGGTCGTCTCAGA AAACGGAAA AC AA AG CACG T TAAGCT
    Tn21_left_IR   GGGGTCGTCTCAGA AAACGGAAA AT AA AG CACG C TAAGCT
     Tn3_left_IR   GGGGTCTGACGCTC AGTGGAACG AA AA CT CACG T TAAGCT

Tn LfArs_right_IR: TA G CGTG CT TT A T TC T C C GTT T TCTGAGAC G A CCCC
    Tn21_right_IR: TA G CGTG CT TT A T TT T C C GTT T TCTGAGGT G C CCCC
     Tn3_right_IR: TA A CGTG AG TT T T CG T T C CAC T GAGCGTCA G A CCCC
```

Figure 2.8: Sequence alignment of left and right transposon inverted repeats, indicating that the inverted repeats Tn*LfArs* is more similar to the Tn*21* subgroup than to Tn*3*. The alignment was performed over a 38 bp region. (See text for details)

```
                            *           20           *           40
   E_coli_R773 :..................................MLQLTPLQLFKNL...........
  E_coli_chrom :..................................MSFLLPIQLFKIL...........
          SmtB :..................MTKPVLQDGETVVCQGTHAAIASELQAIAPEV
          CadC :MKKKDTCEIFCYDEEKVNRIQGDLQTVDISGVSQILKAI...........
B_subtilis_skin:.......................MDETKSELLRKYEQKFKAL...........
Y_enterocolitica:...............................MLPVQLFKIL..ASPVRLEIFR
 At_ferrooxidans:.......................MEPLQDPAQIVARLEAL.ASPVRLEIFR
 B_parapertussis:.....................METNAAIKALVVL.GHASRLTAFR
B_bronchiseptica:.....................METNAAIKALVAL.GHASRLTAFR
   M_capsulatus :.....................MESKSVVIALAAL.AQESRLAVER
     L.f_Tn_ArsR:.......................MEKQAATSIFESL.SSGLRLDVFR
     Atc_Tn_ArsR:.......................MEKQAATSIFESL.SSGLRLDVFR
      A_faecalis:.......................MEKQAATSIFESL.SSGLRLDVFR


                            *           60           *           80           *
   E_coli_R773 :..........SDETRLGIVLLLREMGHLCVCDLCMALDQSQPKISRHIA
  E_coli_chrom :..........ADETRLGIVLLLSELGHLCVCDLCTALDQSQPKISRHIA
          SmtB :AQSLAEFFAVLADPNRLRLLSLLARS.HLCVGDLAQAIGVSESAVSHQIR
          CadC :..........ADENRAKITYALCQDEHLCVCDIANILGVTIANASHHIR
B_subtilis_skin:..........ADQKRLEIMYELCQRGKTCVCDLTEIFEVTQSKLSYHLK
Y_enterocolitica:..........SDETRLAIVMLLRESGHMCVCDLCGATSESQPKISRHMA
 At_ferrooxidans:LLVEQEPTGLVSGDIAEHLGQPHNG.................ISFHLK
 B_parapertussis:LLVQAGAEGVAAGQLAQAMGIAPSS................LSFHLK
B_bronchiseptica:LLVQAGAEGVAAGQLAQAMGIAPSS................LSFHLK
   M_capsulatus :LLIQAGPEGFSAGRIAEALGIAPSS................LSFHLK
     L.f_Tn_ArsR:LLVKKGPGGMVAGEIANALDIPPAN.................LSFHLK
     Atc_Tn_ArsR:LLVKKGPDGMVAGAIANALDIPPAN.................LSFHLK
      A_faecalis:LLVKKGPDGMVAGEIANALDIPPAN.................LSFHLK


                            *           100          *           120          *
   E_coli_R773 :MIRESGILLDRKQGKWVHYRLSPHIPSWAAQIIEQAWLSQQDDVQVIARK
  E_coli_chrom :LIRESGILLDRKQGKWVHYRLSPHIPAWAAKIIDEAWRCEQEKVQAIVRN
          SmtB :SIRNLRLVSYRKQGRHVYYQLQDHHIVALYQNALDHLQECR.........
          CadC :TIYKQGVVNFRKEGKLALYSLGDEHIRQIMMIALAHKKEVKVNV......
B_subtilis_skin:ILLDANLITKETKGTWSYYDLNDEEVNGLLSEELCCIFRKKGEGDCC...
Y_enterocolitica:ILREAELVLDRREGKWVHYRLSPHMPAWAAETITTSWHCCGKMFVSGWIN
 At_ferrooxidans:NLQHAGLVTVQREGRYQRYRAAMPVVRALVAYLTENCCHGTRDCALSGET
 B_parapertussis:ELVQAGLLEARQDGRFVIYAARYDTMAGLIAYLTDNCCQGAPCDAAPCTR
B_bronchiseptica:ELVQAGLLEARQDGRFVIYAARYDTMAGLIAYLTDNCCQGAPCDAAPCTR
   M_capsulatus :ELVHAGLVRSRQESRYVIYSANFEVMNELIGFLTENCCRGAPCTTSPPIR
     L.f_Tn_ArsR:ALSQAHLVTVQEGRFQRYRADIPLMLDLIAYLTEECCSGSPDQCLELRT
     Atc_Tn_ArsR:ALSQARLVTVEQEGRFQRYRADIPLMLDLIAYLTDECCTGSPAKCVEMVA
      A_faecalis:ALSQARLVTVEQEGRFQRYRADIPLMLDLIAYLTDECCTGSPAKCVETVA
```

Figure 2.9: Multiple sequence alignment of *L. ferriphilum* ArsR compared to a selection of known ArsR proteins and members of the ArsR family (SmtB from *Synechococcus* PCC7942: accession number: S31197 and CadC from *S. aureus* pI258: accession number: AAA25636). Known ArsR proteins are labeled as follows: *E. coli* R773, *Escherichia coli* (CAA34168): *E. coli* chrom, *Escherichia coli* chromosome (AAC76526); *B. subtilis* skin, *Bacillus subtilis* skin element (BAA06967); *Y. enterocolitica*, *Yersinia enterocolitica* pYVe227 (AAD16860). The six closest matches are as follows: *At. ferrooxidans*, *Acidithiobacillus ferrooxidans* chromosome (AAF69241.1); *B. parapertussis*, *Bordetella parapertussis* 12822 (NP_883301); *B. bronchiseptica*, *Bordetella bronchiseptica* RB50 (NP_887739); *M. capsulatus*, *Methylococcus capsulatus* (YP_113846); *Atc* Tn ArsR, *Acidithiobacillus caldus* (unpublished); *A. faecalis*, *Alcaligenes faecalis* (AAS45114.1). Black shading highlights amino acids with 90-100 % conservation between *L. ferriphilum*, *At. caldus* and *A. faecalis*, dark grey shading highlights 70-90 % conservation and light grey 60 % and less across all *arsR* proteins except *L. ferriphilum*, *At. caldus* and *A. faecalis*. White highlighted amino acids indicate conservancy of single amino acids compared to *L. ferriphilum*, *At. caldus* and *A. faecalis* . The metal-binding domain (ELCVCDL) has been boxed using dark gray shading outlined by solid black lines, while conserved regions of the ArsR subgroup has been boxed using broken black lines.
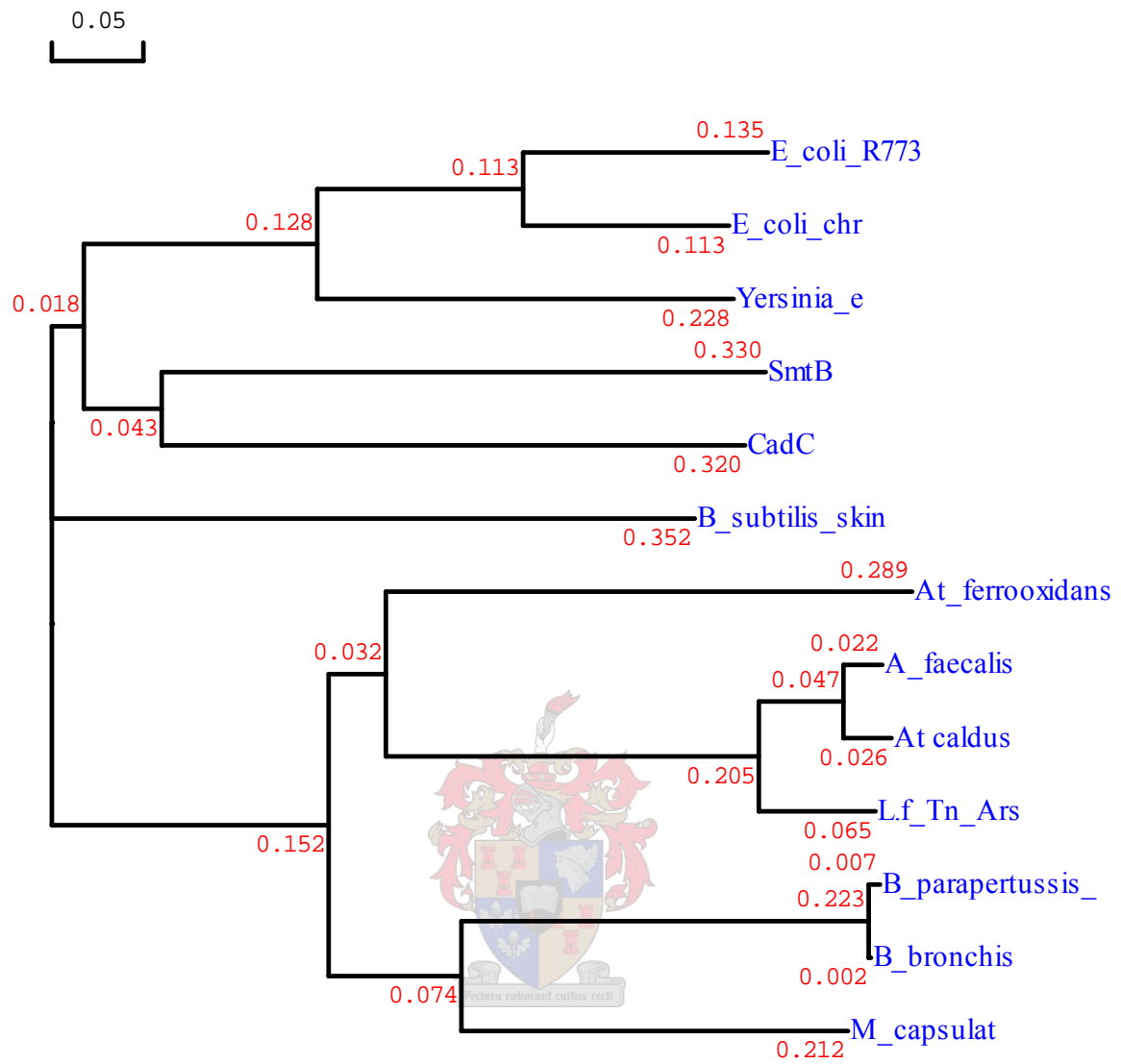
Figure 2.10: Phylogenetic tree of ArsR proteins used for multiple sequence alignment, showing the grouping of *L. ferriphilum* ArsR in a separate subgroup of ArsR proteins. Accession numbers are as previously described (Figure 2.9). Values indicated on phylogenetic tree represent phylogenetic distance.
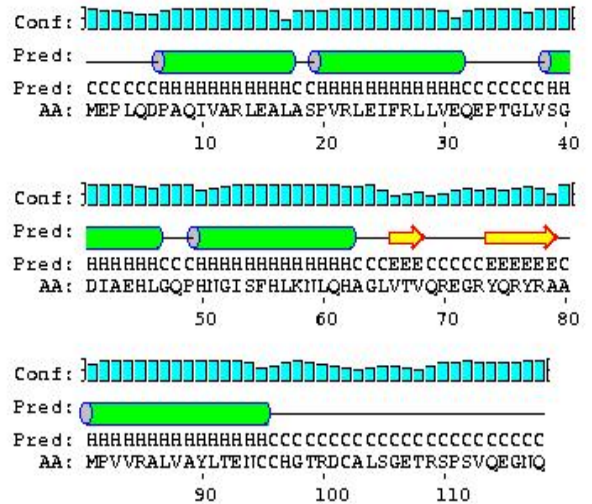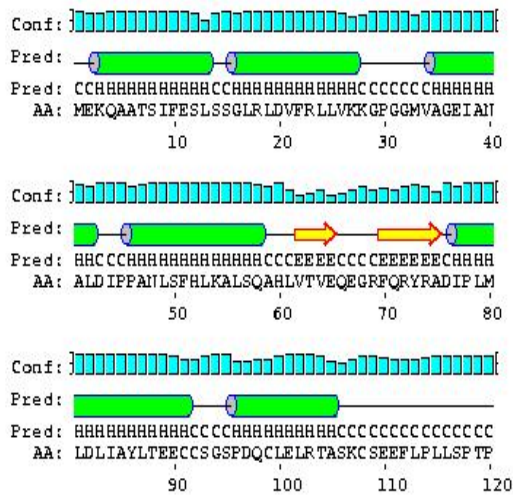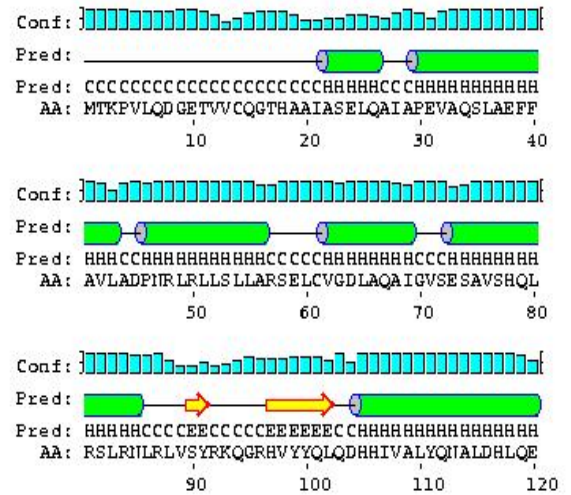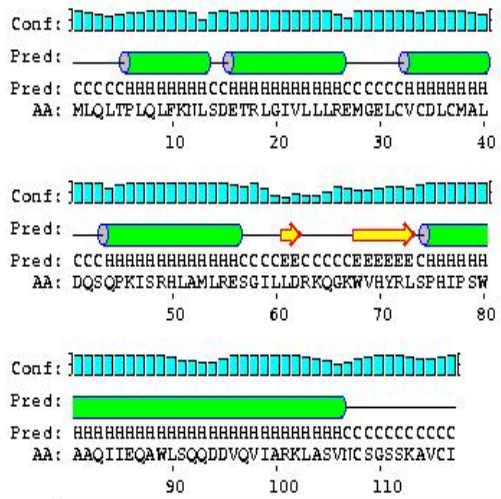
Figure 2.11: Secondary structure predictions of ArsR proteins from *L. ferriphilum* (**A**), *At. ferrooxidans* (**B**), *E. coli* R773 (**C**), and SmtB (**D**) from *Synechococcus* using the web-based program PSIPRED v2.4.

As discussed previously, bacterial ArsC is an arsenate reductase that can be divided into two families, the GSH/Grx family that requires glutathione (GSH) and glutaredoxin (Grx) to function and the Trx family, which requires thioredoxin (Oden *et al*., 1994; Ji and Silver, 1992a). Based on the predicted crystal structures of the two ArsC families (Martin *et al*., 2001; Bennett *et al*., 2001; Zegers *et al*., 2001; Messens *et al*., 2002), it was concluded that certain residues are important in the functioning of these proteins and therefore are conserved (discussed in chapter1).

In the case of the GSH/Grx family, a single catalytic cysteine (Cys-12) is surrounded by an arginine triad (Arg-60, Arg-94 and Arg-107) in the active site. The Trx family, on the other hand, has three conserved cysteine residues at positions 10, 82 and 89, a conserved arginine residue (Arg-16) and an asparagine at position 105. From sequence alignments it can be seen that the two families vary tremendously in structure (Figure 2.12), although they perform the same function. Phylogenetic analysis (Figure 2.13) of different Trx and GSH/Grx ArsC proteins indicate that the ArsC from *L. ferriphilum* groups with ArsC proteins from the Trx family. Similar results have been obtained for *P. aeruginosa* (Cai *et al*., 1998), *At. ferrooxidans* (Butcher *et al*., 2000) and *At. caldus* (Tuffin *et al*., 2004) (Gram-negative bacteria), where their ArsC proteins grouped with that from the Trx family (Mukhopadhyay and Rosen, 2002). The ArsC protein from *L. ferriphilum* contains all the features described above for the Trx family of ArsC proteins, with the cysteine residues being found at positions 10, 82 and 89.

A. GSH/Grx family

```
                            *                 10              *           I     H     YLE    20            *
        A_multivorum:  MSNITIYHNP  A  CGTSRNTLEMIRNSG  N  EPT  V  I  H  YLE  N  PP  SRD  ELVKLIAD
         E_coli_R773:  MSNITIYHNP  A  CGTSRNTLEMIRNSG  T  EPT  I  I  L  YLE  N  PP  SRD  ELVKLIAD
        E_coli_chrom:  MSNITIYHNP  A  CGTSRNTLEMIRNSG  T  EPT  I  I  H  YLE  T  PP  TRD  ELVKLIAD
           K_oxytoca:  MSNITIYHNP  A  CGTSRNTLEMIRNSG  N  EPT  V  I  H  YLE  N  PP  SRD  ELVKLIAD
     Y_enterocolitica: MSNITIYHNP  T  CGTSRNTLEMIRNSG  N  EPT  V  I  Y  YLE  T  PP  THD  ELVKLIAD
         S_marcescens: MSNITIYHNP  A  CGTSRNTLEMIRNSG  T  EPT  V  I  H  YLE  T  PP  SRA  ELVKLIAD
     IncN_plasmid_R46: MSNITIYHNP  H  CGTSRNTLEMIRNSG  I  EPT  V  I  L  YLE  T  PP  SRD  ELLKLIAD

                           40               *              60              *              80
        A_multivorum:  MGIS  VRALLRKNVEP  YE  E  LGL  EE  DK  FT  DDQ  LI  D  FML  QH  P  ILINRPIVVTP
         E_coli_R773:  MGIS  VRALLRKNVEP  YE  Q  LGL  AE  DK  FT  DDQ  LI  D  FML  QH  P  ILINRPIVVTP
        E_coli_chrom:  MGIS  VRALLRKNVEP  YE  E  LGL  AE  DK  FT  DDR  LI  D  FML  QH  P  ILINRPIVVTP
           K_oxytoca:  MGIS  VRALLRKNVEP  YE  E  LGL  AE  DK  FT  DGE  LI  D  FML  QY  P  ILINRPIVVTP
     Y_enterocolitica: MGIT  VRALLRKNVEP  YE  E  LGL  AE  GT  FS  DEQ  LI  G  FML  EH  P  ILINRPIVVTP
         S_marcescens: MGIT  VRALLRKNVEP  FE  A  LGL  AE  DR  FT  DEQ  LI  D  FML  QH  P  VLINRPIVVTP
     IncN_plasmid_R46: MGIS  VRALLRKNVEP  YE  E  LGL  AE  DK  FT  DDQ  LI  D  FML  QH  P  ILINRPIVVTP

                            *                100            *             120
        A_multivorum:  LGTR  LCRPSEVVL  D  IL  PDA  Q  KGAFA  KEDGEK  VV  D  EAGKRLK
         E_coli_R773:  LGTR  LCRPSEVVL  D  IL  QDA  Q  KGAFT  KEDGEK  VV  D  EAGKRLK
        E_coli_chrom:  LGTR  LCRPSEVVL  E  IL  PDA  Q  KGAFS  KEDGEK  VV  D  EAGKRLK
           K_oxytoca:  LGTR  LCRPSEVVL  D  IL  PDA  Q  KGAFA  KEDGEK  VV  D  ETR....
     Y_enterocolitica: LGTR  LCRPSEVVL  D  IL  PEP  Q  QGAFT  KEDGEK  IT  D  ESGKRLK
         S_marcescens: LGTR  LCRPSEVVL  D  IL  YDA  Q  KSAFT  KEDGEK  VV  D  EKGNRLN
     IncN_plasmid_R46: LGTK  LCRPSEVVL  D  IL  PDA  Q  KAAFT  KEDGEK  VV  D  DSGKRLK
```

B. Trx family

```
                           *                 20              *             40
       B_halodurans:   MNKKVI  YF  LC  TG  NSCRS  QMA  E  GWGKKYLGDEWDVY  SAG  IEAHGVNPNAV
         B_subtilis:   MENKII  YF  LC  TG  NSCRS  QMA  E  GWAKQYLGDEWKVY  SAG  IEAHGLNPNAV
       P_aeruginosa:     MRV  LF  MC  TA  NSCRS  ILS  E  AMFNHLAPPGFEAC  SAG  SQPSGRVHPRS
           P_putida:     MKV  LF  MC  TA  NSCRS  ILS  E  AMFNHLAPDGFEAI  SSG  SFPKGLVLPRS
    Sinorhizobium_sp:  MENKKTI  YF  LC  TG  NSCRS  QMA  E  AWGKKYLGDKWNVL  SAG  IEAHCVNPNAI
           S_aureus:   MDKKTI  YF  IC  TG  NSCRS  QMA  E  GWGKEILGEGWNVY  SAG  IETHGVNPKAI
          S_xylosus:   MDKKTI  YF  IC  TG  NSCRS  QMA  E  GWGREILGEDWNVY  SAG  IETHGVNPKAI
    Synechocystis_sp:    MKKV  MF  VC  KR  NSCRS  QMA  E  GFAKTLGAGKIAVT  SCG  LESSRVHPTAI
      At_ferrooxidans: MKTPEI  LF  LC  TG  NSCRS  ILA  E  VTFNALAGPGMHAT  SAG  SHPAGYVHTRS
         L.f_Tn_ArsC:     MNI  LF  LC  TG  NSCRS  ILA  E  ATFNHLAPTGWKAM  SAG  SQPTGTVHPRS
          A_faecalis:     MNI  LF  LC  TG  NSCRS  ILA  E  ATFNHLAPTGWKAM  SAG  SQPTGTVHPRS
         Atc_Tn_ArsC:     MNI  LF  LC  TG  NSCRS  ILA  E  ATFNHLAPTGWKAM  SAG  SQPTGTVHPRS

                           *                 60              *             80             *
       B_halodurans:   KAMKEI.  GID  ISEQT  S  DTIDQELLQKA  DL  VVTLC  GH  AADVCPATPSNKER
         B_subtilis:   KAMKEV.  GID  ISNQT  S  DIIDSDILNNA  DL  VVTLC  GD  AADKCPMTPPHVKR
       P_aeruginosa:   LATLEQA  GIA  THGLY  S  KGSEAFEGAPP  DI  VITVC  DA  AAGEACPLYLGAAL
           P_putida:   LSTLQAA  GIR  TEGLY  S  KGNDVFEGSPP  DV  VITVC  DK  AAGEACPVYFGPAV
    Sinorhizobium_sp:  KAMNEV.  DID  ITDQT  S  VTIDRDILDKA  DL  VVTLC  GH  ANDVCPTTPPHVKR
           S_aureus:   EAMKEV.  DID  ISNHT  S  DLIDNDILKQS  DL  VVTLC  SD  ADNNCPILPPNVKK
          S_xylosus:   EAMKEV.  DID  ISNHT  S  DLIDNHILKQS  DL  VVTLC  SD  ADDNCPILPPNVKK
    Synechocystis_sp:  AMMEEV.  GID  ISGQT  S  DPIENFNADDY  DV  VISLC  GC  GVNLPPEWVTQEIF
      At_ferrooxidans: INLLERE  GFR  TDGLH  S  KSWEDLKET.P  DI  VITVC  AD  AAGETCPAYLGPAI
         L.f_Tn_ArsC:  LSLLTHE  GID  TGGLH  S  KSWDNLPLA.P  DV  VITVC  AN  AAGETCPAYLGPVL
          A_faecalis:  LALLARE  GID  TDGLH  S  KSWDNLPLA.P  DV  VITVC  AN  AAGETCPAYLGPVL
         Atc_Tn_ArsC:  LALLARE  GID  TNGLH  S  KSWDNLPLA.P  DV  VITVC  SN  AAGETCPAYLGSVL

                      100              *               120            *             140
       B_halodurans:   VHWGFDD  PAK  AEGTDEEKWA  VFRRVRDEIG  KRIKTF  A  ETG  K... . ......
         B_subtilis:   EHWGFDD  PAR  AQGTEEEKWA  FFQRVRDEIG  NRLKEF  A  ETG  K... . ......
       P_aeruginosa:   KAHWGLA  DPS  ALDGDEALRD  AAFHATLARI  EQRCRA  F  LGL  PFA  T  L..DRD
           P_putida:   KAHWGLE  DPS  DVQGDDARVQ  AAFDATLKTI  ATRCRA  F  FIL  PFA  E  L..SPT
    Sinorhizobium_sp:  VHWGFDD  PAG  QEWS......  VFQKVRDEIG  IRKKF  A  ETG  E... . ......
           S_aureus:   EHWGFDD  PAG  KEWSEFQRVR  DEIKLAIEKF  KLR... . .. ... ... .......
          S_xylosus:   EHWGLED  PAG  KEWSEFQRVR  DEIKLAIENF  KLR... . .. ... ... .......
    Synechocystis_sp:  EDWQLED  PDG  QSLEVFRTVR  GQVKERVENL  IAKIS.. . .. ... ... .......
      At_ferrooxidans: RTHWGVE  DPA  KVTGTEAQIE  AAFDTAYHIL  RHRIEAL  LQL  PVA  E  LLEKDP
         L.f_Tn_ArsC:  RAHWGVD  DPA  HATGTEAEIE  AAFPTAYRIL  RRRIEA  L  LAL  PLA  Q  LAHDRT
          A_faecalis:  RAHWGVD  DPA  HATGTEAEIE  AAFTAAYRIL  RRRIEA  F  LAL  PLV  E  LARDRT
         Atc_Tn_ArsC:  RAHWGVD  DPA  HATGTEAEIE  AAFTTAYRIL  RRRIEA  F  HAL  PLA  Q  LAHDRT
```

Figure 2.12: Multiple sequence alignment of ArsC proteins from the GSH/Grx (**A**) and Trx (**B**) families. The ArsC protein from *L. ferriphilum* groups with the Trx family along with the ArsC proteins from *P. putida* and *P. aeruginosa*. *Sinorhizobium* sp., AF178758_3; *Klebsiella oxytoca*, AF168737_3. See Figure 2.13 for accession numbers of the rest of the proteins.
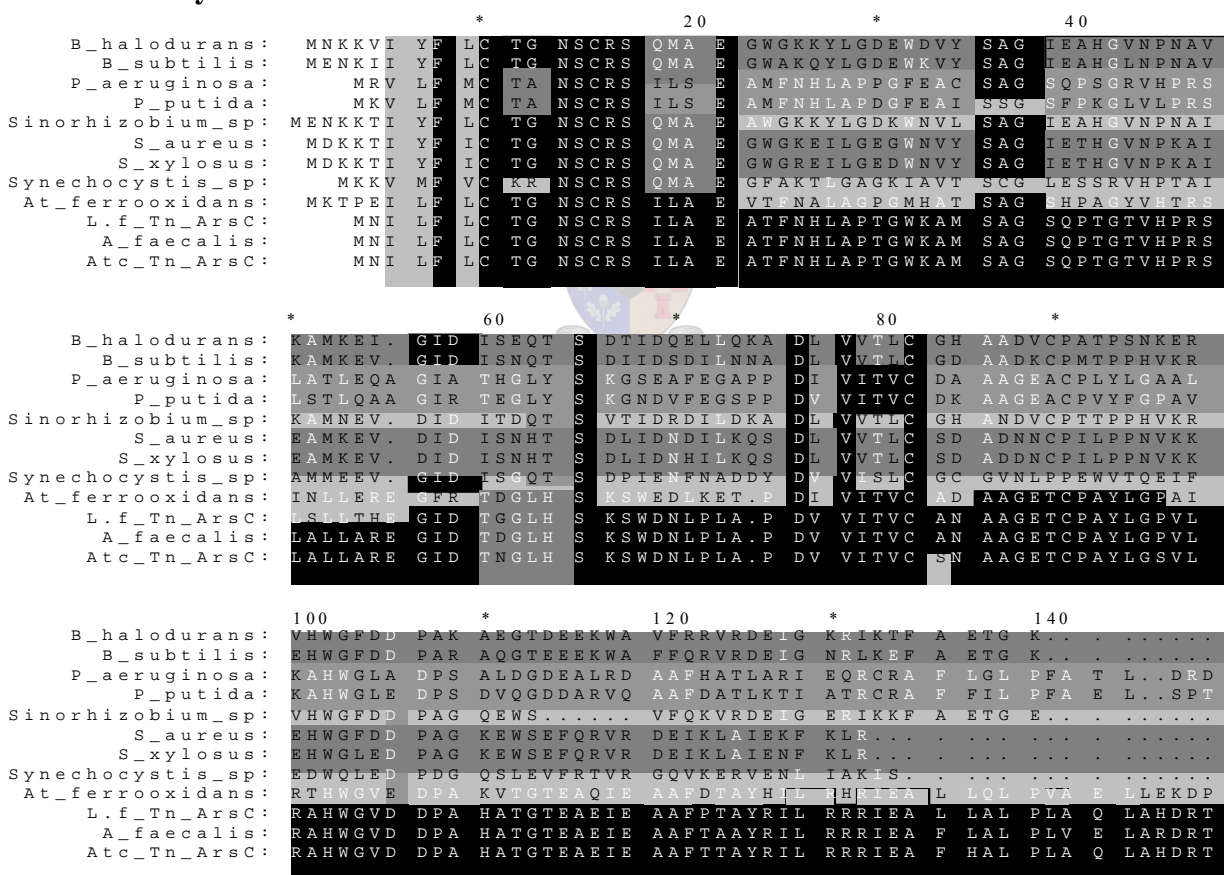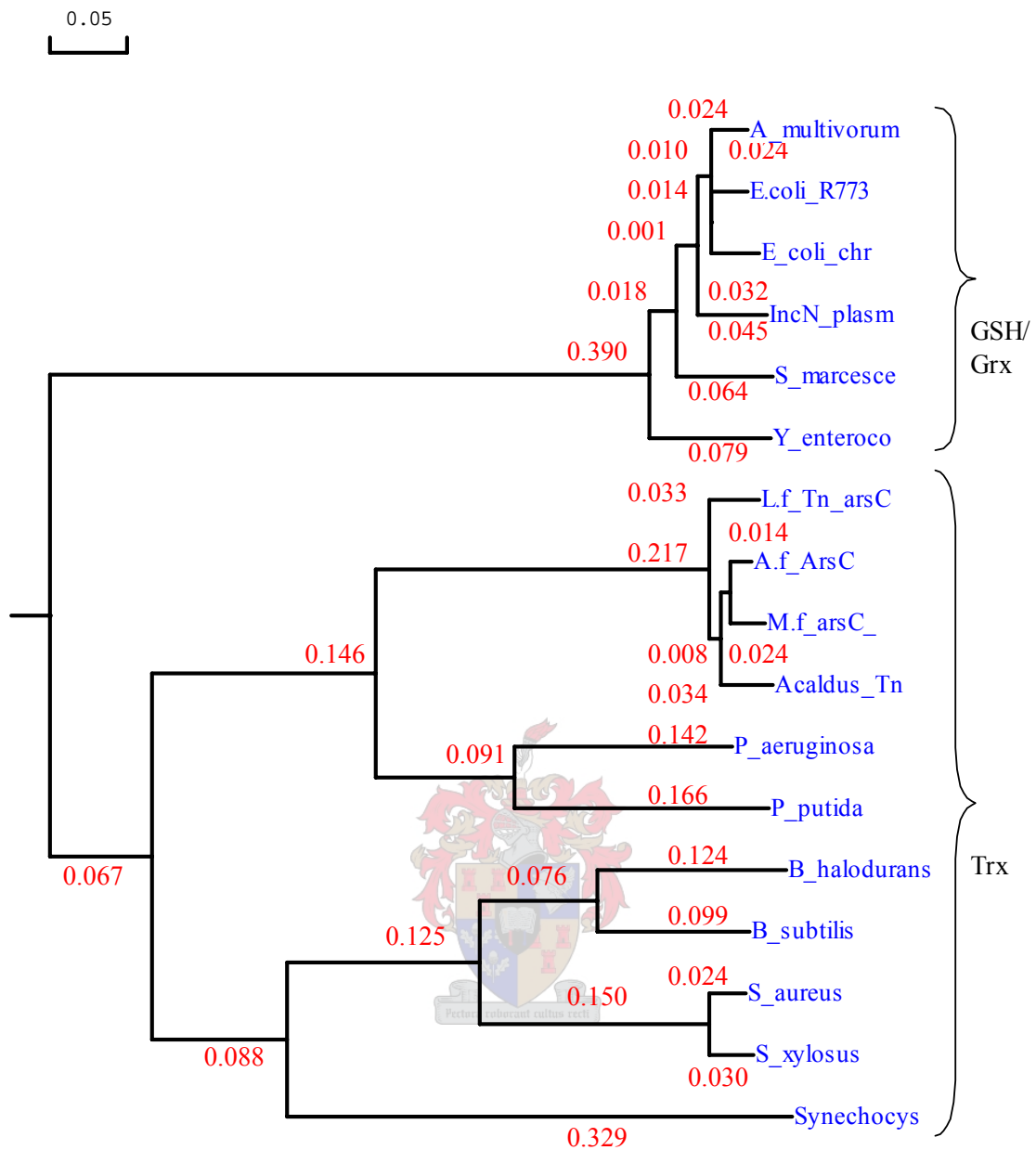
Figure 2.13: Phylogenetic tree of the ArsC from *L. ferriphilum* and other known ArsC proteins including putative ArsC proteins from the NCBI database. Accession numbers: *Acidithiobacillus ferrooxidans* ATCC3020, AAF69239.1; *Synechocystis* sp., NP_441727; *Acidithiobacillus caldus* (unpublished) *Alcaligenes faecalis*, AAS45115.1; *Methylobacillus flagellatus*, ZP_00172247.2; *Pseudomonas aeruginosa*, NP_250969; *Bacillus halodurans*, NP_243864; *Bacillus subtilis* skin element, ARSC_BACSU; *Staphylococcus aureus* pI258, ARSC_STAAN; *Yersinia enterocolitica* pYVe227, np_025438; *S. marcescens*, CAB88404; *Escherichia coli* R46, NP_511240; *Escherichia coli* R773, AAA21096; *Acidiphilium multivorum* pKW301, BAA24824; *Escherichia coli* K12 chromosome, NP_417960. Values on phylogenetic tree represent phylogenetic distance.

An ORF with strong homology to *arsD* (arsenic operon regulator) from *M. flagellatus* (E-value = 7e^{-65}) and *A. faecalis* (E-value = 6e^{-63}) was identified. Previously, Li *et al*. (2001) showed through sequence alignments that putative ArsD proteins contain three conserved vicinal pairs (Cys-12-Cys-13, Cys-112-Cys-113 and Cys-119-Cys-120). However, further investigation by this group demonstrated that only vicinal pairs Cys-12-Cys-13 and Cys-112-Cys-113 are involved in As(III) inducer binding and that the third pair (Cys-119-Cys-120) is not required for ArsD activity (see previous chapter sub-section 1.3.2.1). Sequence alignments of the *L. ferriphilum* ArsD protein and its closest relatives, as indicated by BLAST results, revealed that it contained all three vicinal pairs (Figure 2.14). Phylogenetic analysis indicated that the ArsD of *L. ferriphilum* clusters with that of *A. faecalis*, *M. flagellatus* and *At. caldus* (as indicated by BLAST results), with the ArsD from *At. caldus* being the least similar of the group (Figure 2.15).

The identification of an *arsD* gene coincided with the finding of an *arsA* homologue immediately downstream from where the *arsD* gene was located (the two genes typically co-exist as a pair). The ArsA protein on the other hand usually associates with the ArsB where the two proteins together form the membrane located ATP-driven arsenite efflux pump (Chen *et al*., 1986). The predicted amino acid sequence of the putative ArsA from *L. ferriphilum* was compared to ten of its closest matches from the BLAST search (data not shown). Several key observations were made from this alignment. Firstly, the *L. ferriphilum* ArsA protein comprises of two halves, A1 and A2, which was first identified by Chen *et al*. (1986) in the ArsA protein from *E coli* R773. Secondly, it was observed that the A1 half of the ArsA proteins was highly homologous to other A1 domains, whereas the A2 half of the ArsA proteins had relatively poor homology. Subsequent phylogenetic studies of these ArsA proteins showed that the *L. ferriphilum* ArsA groups with *A. faecalis* and *M. flagellatus* (Figure 2.17). Therefore, a second alignment was done with the *L. ferriphilum* ArsA protein versus selected homologous ArsA proteins from the phylogenetic results, which indicated that amongst these proteins the A1 as well as the A2 half is highly homologous to each other (Figure 2.16). Other observations made from these alignments were that each half contained a DTAP motif (discussed previous chapter) as seen in other ArsA proteins.

```
                     *         ↓      20          *              40        *
A_faecalis:        M KKIE VFDPSL CCSTGVCGVE VDQ A LVT F AA DV DW AK QS GARIER Y NLA Q
D_aromatica:       M SVIQ IFDPAL CCSTGVCGVD VDQ Q LVS F SA DV DW AK QQ GAQVER F NLA Q
E_coli_R773:       M KTLM VFDPAM CCSTGVCGTD VDQ A LVD F ST DV QW LK QC GVQIER F NLA Q
S_typhimurium_R46: M KMLT VFDPAM CCSTGVCGSD VDQ V LVN F SA DV QW LK GR GVQIER Y NLA H
K_oxytoca:         M KTLT VFDPAM CCSTGVCGSD VDQ V LVD F SA DV QW LK GR GVQVER Y NLA Q
M_flagellatus:     M KKIE VFDPSL CCSTGVCGVE VDQ A LVT F AA DV DW AK QN GAHIER Y NLA Q
A_multivorum:      M KTLT VFDPAM CCSTGVCGSD VDQ V LVD F SA DM WK LK GR GVQVER Y NLA Q
S_typhimurium:     M KTLM VFDPAM CCSTGVCGTD VDQ A LVD F SA DV QW LK QC GVQIER F NLA Q
L.f_Tn_ArsD:       M KKIE VFDPSL CCSTGVCGVD VDQ A LVT F AA DV DW AK QN GAHIER Y NLA Q
Atc_Tn_ArsD:       M KKIE VFDPSL CCSTGVCGVE VDQ A LVT F AA DV DW AK QN GAHIER Y NLA Q

                        *              60          *              80
A_faecalis:        Q P QM FAD N ATV K GF LQR SG QDA LPLI LV DGE VALAGRYP K RA EL SL WI
D_aromatica:       Q P MA FAE N AVV K GF LER SG QEA LPLI LV DGE FALAGRYP N RA EL AR WA
E_coli_R773:       Q P MS FVQ N EKV K AF IEA SG AEG LPLL LL DGE TVMAGRYP K RA EL AR WF
S_typhimurium_R46: E P MS FVE N EKA K AF LEA SG AEG LPLL LL DGE TVMAGRYP K RA EL AR WF
K_oxytoca:         Q P MS FVQ N EKA K AF LEA SG AEG LPLL LL DGE TVMAGRYP K RA EL AR WF
M_flagellatus:     Q P QM FAD N ATV K GF LQR SG QGA LPLI LV DGE VALAGRYP K RA EL AL WI
A_multivorum:      Q P MS FVH N EKA K AF LDA SG AEG LPLL LL DGE TVMAGRYP K RA EL AR WF
S_typhimurium:     Q P MS FVQ N EKV K AF IEA SG AEG LPLL LL DGE TVMAGRYP K RA EL AL WF
L.f_Tn_ArsD:       Q P QM FAD N ATV K GF LQR SG QDA LPLI LV DGE VALAGRYP K RA EL AL WI
Atc_Tn_ArsD:       Q P QM FAE N ATV K GF LQR SG QDA LPLI LV DGE VALAGRYP T RD EL VR WA

                   *       100         ↓         *     ↓
A_faecalis:        GIDQPADAAIPATGG CC S G PRGCC
D_aromatica:       KMEAAAVDAVSPAKS CC S G SKCC.
E_coli_R773:       GIPLDKVGLAPS..G CC G G NTSCC
S_typhimurium_R46: GIPLEKVGLAST..S CC G G KTSCC
K_oxytoca:         GIPLEKVGLAPI..S CC G G NTSCC
M_flagellatus:     GIDQPADAAKPATGG CC S G PRGCC
A_multivorum:      GIPLEKVGLAPT..S CC G G NTSCC
S_typhimurium:     GIPLDKVGLAPS..G CC G G NTSCC
L.f_Tn_ArsD:       GIDQPADAAKPTTGG CC S G PRGCC
Atc_Tn_ArsD:       GLAAAPSMKTPS.GG CC G G GSRTC
```

Figure 2.14: Multiple sequence alignment of a selection of closely related ArsD proteins (from database) compared to the ArsD from *L. ferriphilum*. Accession numbers: *Alcaligenes faecalis*, AAS45116.1; *Dechloromonas aromatica*, ZP_00151005.2; *Salmonella typhimurium* R46, AAG43230.1; *Klebsiella oxytoca*, AAF89639.1; *Methylobacillus flagellatus*, ZP_00172246.2; *Acidiphilium multivorum*, BAA24821.1; *Salmonella typhimurium*, BAB91570.1; *Acidithiobacillus caldus* (unpublished). Black arrows point to conserved vicinal pairs (see text).
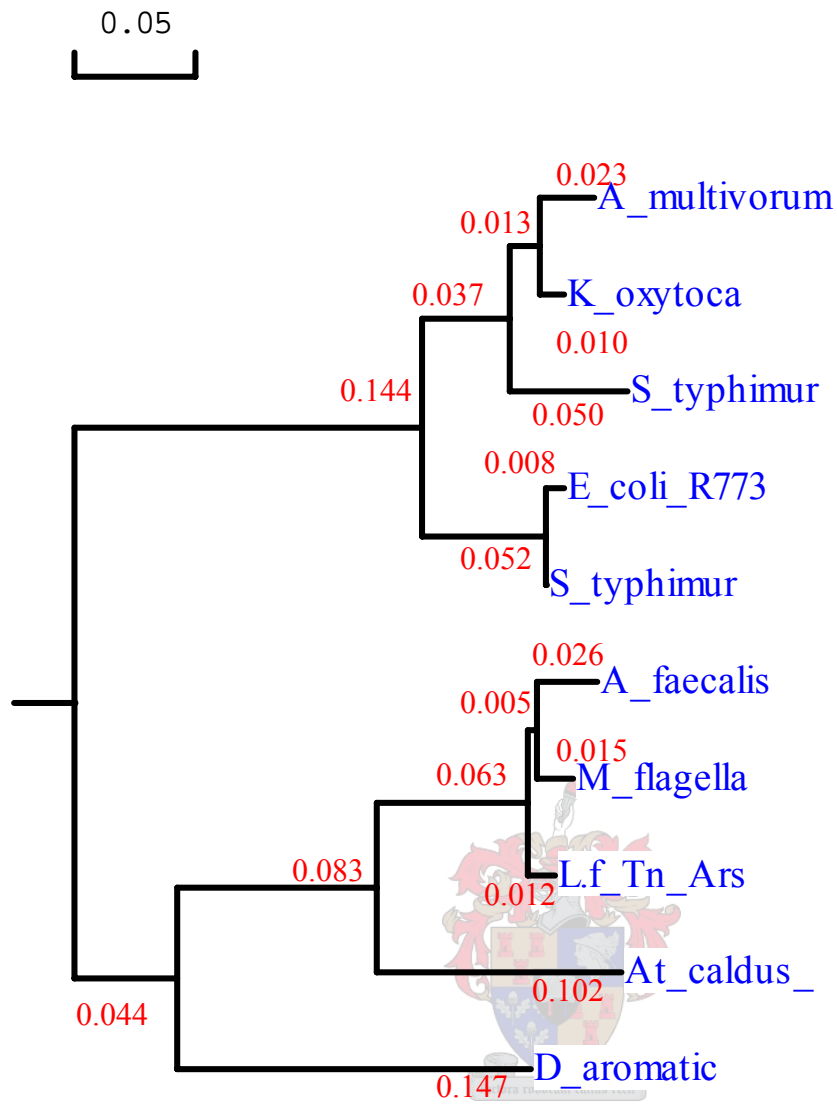
Figure 2.15: Phylogenetic tree of *L. ferriphilum* ArsD and other putative ArsD proteins obtained from the database. Accession numbers are listed in figure 2.14. ArsD from *L. ferriphilum* groups with *A. faecalis* and *M. flagellatus* indicated by the BLAST results. Values represent phylogenetic distance.

A1 domain

```
                                    *            20          *            40          *
D_aromatica:         MK FL EQE PRY LFFTGKGGVGKTS IA CATAI Q LAEAGQRV LLVSTDPASNVGQVF G
E_coli_R773:         MQ FL QNI PPY LFFTGKGGVGKTS IS CATAI R LAEQGKRV LLVSTDPASNVGQVF S
M_flagellatus:       MK FL QLE PRF LFFTGKGGVGKTS IA CATAI Q LAEAGKRV LLVSTDPASNVGQVF G
L.f_Tn_ArsA_: MEKMM MI FL QLE PRF LFFTGKGGVGKTS IA CATAI Q LAEAGKRV LLVSTDPASNVGQVF G
A_faecalis:   MEKTM MK FL QLE PRF LFFTGKGGVGKTS IA CATAI Q LAEAGKRV LLVSTDPASNVGQVF G
Atc_Tn_ArsA:         ML FL QNE PAF LFFTGKGGVGKTS LS CATAI H LAGQGKKI LLVSTDPASNVGQVF S


               *            80          *            100         *
D_aromatica:   VD IGN RIVG IEA V PGLF ALEIDP QA AA QA YR D R I VG PVR G V LPEA VV KG
E_coli_R773:   QT IGN TIQA IAS V PGLS ALEIDP QA AA QQ YR A R I VD PIK G V LPDD VV SS
M_flagellatus: VD IGN RVTP IPA V PRLS ALEIDP EA AA SA YR E R L VG PVR G V LPDD VV KG
L.f_Tn_ArsA_:  VD IGN RVTP IPA V PRLS ALEIDP EA AA SA YR E R L VG PVR G V LPDD VV KG
A_faecalis:    VD IGN RVTP IPA V PRLS ALEIDP EA AA SA YR E R L VG PVR G V LPDV VV KG
Atc_Tn_ArsA:   QE IGN KITT IST V AGLS ALEIDP QQ AA QQ YR E R I VG PVR G A LPDD VV KG


               120         *            140         *            160         *
D_aromatica:   IEEQ LSGACTTEIAAFDEFT A LLT DSA L V.AGYD HIIFDTAPTGHTIRLLQLPGAW TG F LE
E_coli_R773:   INEQ LSGACTTEIAAFDEFT G LLT DTS L L.TRFD HIIFDTAPTGHTIRLLQLPGAW SS F ID
M_flagellatus: IEES LSGACTTEIAAFDEFT A LLT NAA L T.ADYE HIIFDTAPTGHTIRLLQLPGAW SG F LE
L.f_Tn_ArsA_:  IEES LSGACTTEIAAFDEFT A LLT NVA L T.ADYE HIIFDTAPTGHTIRLLQLPGAW SG F LE
A_faecalis:    IEES LSGACTTEIAAFDEFT A LLT NAA L T.ADYE HIIFDTAPTGHTIRLLQLPGAW SG F LE
Atc_Tn_ArsA:   IEEQ LSGACTTEIAAFDEFT A LLT DAT L I.ADYD HIIFDTAPTGHTIRLLQLPGAW SG F IE


               180         *            200         *            220         *
D_aromatica:   EGKGD ASCLGP LAGLEKQR TQ YKS AV EALA DP VR TRL I LVARA QAATL R EVSRTHE
E_coli_R773:   SNPEG ASCLGP MAGLEKQR EQ YAY AV EALS DP KR TRL V LVARL QKSTL Q EVARTHL
M_flagellatus: AGKGD ASCLGP LAGLEKQR TQ YKA AV EALA DP LQ TRL V LVARA QQAAL R EVARTHE
L.f_Tn_ArsA_:  AGKGD ASCLGP LAGLEKQR TQ YKA AV EALA DP LQ TRL V LVARA QQATL R EVARTHE
A_faecalis:    AGKGD ASCLGP LAGLEKQR TQ YKA AV EALA DP LQ TRL V LVARA QQATL R EVARTHE
Atc_Tn_ArsA:   KNPEG ASCLGP LAGLEKQR QR YAE AV KALS DP ER TRL I LVARA QKTTL D EVARTHE


               240         *            260         *            280
D_aromatica:   ELA GIGL KQQYLVING V LP ESAVGS D L LAKAIYN REQA ALLEI PA VL KA LP ID
E_coli_R773:   ELA AIGL KNQYLVING V LP KTEAAN D T LAAAIWE REQE ALANL PA DL AG LP TD
M_flagellatus: ELA AIGL KQQHLVING I LP HVEAAT D P LAAAIHE REQT ALKNI PA TL TA LP RD
L.f_Tn_ArsA_:  ELA AIGL KQQHLVING I LP HIEAAT D P LAAAIHE REQT ALKNI PA TL TA LP CD
A_faecalis:    ELA AIGL KQQHLVING I LP HIEAAT D P LAAAIHE REQT ALKNI PA TL TA LP RD
Atc_Tn_ArsA:   ELA AIGL SRQNLVING V LP ESEAVH D A LAAAIHR REQE AIANM PA VL RD LP LD
```

Linker region

```
               *            300         *            320         *            340
D_aromatica:   QVG L KPF N L VGLD AL RK L LVASVFDAASDIPAATVLPEVPSLAVLVD GIA ADGHGLV
E_coli_R773:   TLF L QPV N M VGVS AL SR L LSTQPVASPSSDEYLQQRPDIPSLSALVD DIA RNEHGLI
M_flagellatus: HVE L KPF N L VGLE AL RQ L LTDLPPQAPAAVDSPIELDEPSVAELI.D GIA ADGHGLI
L.f_Tn_ArsA_:  HVE L KPF N L VGLD AL RQ L LTDLPPQAPVAVDSPIELDEPGVADLI.D GIA ADGHGLV
A_faecalis:    HVE L KPF N L VGLE AL RQ L LTDLPPQAPAAVDSPIELDEPSVAELI.D GIA ADGHGLV
Atc_Tn_ArsA:   QLP L KAF N L VGVE AL SS L FSDRDEAAPVVAGAPAKPVDLPPLSTLVD EIA ETGHGLI
```

A2 domain

```
               *            360         *            380         *            400
D_aromatica:   ML MGKGGVGKTT LAAA VAV E LA SRGLP VHLTTSDPAAHL SE TL AG SL EHLT VSRIDP
E_coli_R773:   ML MGKGGVGKTT MAAA IAV R LA DMGFD VHLTTSDPAAHL SM TL NG SL NNLQ VSRIDP
M_flagellatus: ML MGKGGVGKTT LAAA IAV E LA HRGLP VHLTTSDPAAHL TD TL DS SL DNLT VSRIDP
L.f_Tn_ArsA_:  ML MGKGGVGKTT LAAA IAV E LA HRGLP VHLTTSDPAAHL TD TL DS SL DNLT VSRIDP
A_faecalis:    ML MGKGGVGKTT LAAA IAV E LA HRGLP VHLTTSDPAAHL TD TL DS SL DNLT VSRIDP
Atc_Tn_ArsA:   MM MGKGGVGKTT LAAA VAV A LA ERGLP VHLTTSDPAAHL TD TL AG SL DNLE VSRIDP
```

```
                           *        420        *        440      *        460
D_aromatica:    QA ETERYRQHVL ES KGAT LD AQ G RAL LEEDLRSPCTEEIAVFQAFSR V IREAGKK FVVM
E_coli_R773:    HE ETERYRQHVL ET KGKE LD EA G KRL LEEDLRSPCTEEIAVFQAFSR V IREAGKR FVVM
M_flagellatus:  HA ETERYRQHVL ET KGAQ LD AE G RAL LEEDLRSPCTEEIAVFQAFSR I IREAGKK FVVM
L.f_Tn_ArsA_:   HA ETERYRQHVL ET KGAQ LD AE G RAL LEEDLRSPCTEEIAVFQAFSR I IREAGKK FVVM
A_faecalis:     HA ETERYRQHVL ET KGAQ LD AE G RAL LEEDLRSPCTEEIAVFQAFSR I IREAGKK FVVM
Atc_Tn_ArsA:    QA ETERYRQHVL VT KGKD LD AE G RAM LEEDLRSPCTEEIAVFQAFSR V IREAGKK FVVM

                           *        480        *        500      *        520
D_aromatica:    DTAPTGHTLLLLDATGAYHRE IARQMSGKDMH Y TTPMM QLQDAKQT KV LIA TL A ETTPVL
E_coli_R773:    DTAPTGHTLLLLDATGAYHRE IAKKMGEKG.H F TTPMM LLQDPERT KV LLV TL P ETTPVL
M_flagellatus:  DTAPTGHTLLLLDATGAYHRE VSRQMGKTGVH F TTPMM QLQDPKQT KV LVV TL A ETTPVL
L.f_Tn_ArsA_:   DTAPTGHTLLLLDATGAYHRE VTRQMGKTGMH F TTPMM QLQDPKQT KV LIV TL A ETTPVL
A_faecalis:     DTAPTGHTLLLLDATGAYHRE VSRQMGKTGAH F TTPMM QLQDPKQT KV LVV TL A ETTPVL
Atc_Tn_ArsA:    DTAPTGHTLLLLDATGAYHRE VDRHAESNVRY . TTPMM QLQDPARA KV MIV TL A ETTPVL

                          *        540        *        560       *
D_aromatica:    EAANLQ S DL RRAGI EPW A WV INN SVAAVQSS SPLLRQ RA VN EL AQ I DL VAHTHA
E_coli_R773:    EAANLQ A DL ERAGI HPW G WI INN SLSIADTR SPLLRM RA QQ EL PQ I ES VKRQHA
M_flagellatus:  EAANLQ A DL RRAGI EPW A WI INT SVAAASAK SPLLRQ RA AN EL RE I SA VANQHA
L.f_Tn_ArsA_:   EAANLQ A DL RRAGI EPW A WI INT SVAAASAK SPLLRQ RA AN EL RE I SA VANQHA
A_faecalis:     EAANLQ A DL RRAGI EPW A WI INT SVAAASAK SPLLRQ RA AN EL HE I SA VANQHA
Atc_Tn_ArsA:    EAANLQ E DL RRAGI EPW A WL INN SLSAAPTA SPLLKR RA AF EL TQ I EA VRTRYT

                        580        *        600
D_aromatica:    V R Y AV VP LIQD EP V GVERL RA LANCH PVTA
E_coli_R773:    S R V AL VP VLAS EP T GIDKL KQ LAG
M_flagellatus:  D R Y AV VP LLKE EP I GTERL RA LIHPQ A....
L.f_Tn_ArsA_:   D R Y AV VP LLKE EP I GADRL RA LIHPQ T....
A_faecalis:     D R Y AV VP LLKE EP I GAERL RA LIHPQ A....
Atc_Tn_ArsA:    K R V AL VP MQAE EP I GIEPL LA LVEPR NGFGL
```

Figure 2.16: Multiple sequence alignment of ArsA protein homologues including *L. ferriphilum* ArsA. The alignment shows A1 and A2 domains as well as the linker region in-between the two halves as well as DTAP motifs highlighted in dark grey. See figure 2.17 for accession numbers for proteins. Numbering of amino acids is relevant to the *L. ferriphilum* Tn ArsA.

Further analysis of the *L. ferriphilum* Tn ArsA protein sequence revealed that like other ArsA proteins, it contains the four conserved cysteine residues of the allosteric site (Cys-30, Cys-119, Cys-180 and Cys-435), although slightly shifted from their original positions in the *E. coli* R773 ArsA protein (Cys-26, Cys-113, Cys-172 and Cys-422). Other conserved residues identified include Gly-19 (Gly-15), Ala-352 (Ala-344) also located in the allosteric site and Phe-149 (Phe-141) and Trp-167 (Trp-159), which are at the boundaries of the A1 DTAP motif (Figure 2.16) A flexible linker region, which was also previously identified in other ArsA proteins (Li *et al*., 2000), separates the two halves of the *L. ferriphilum* ArsA protein.
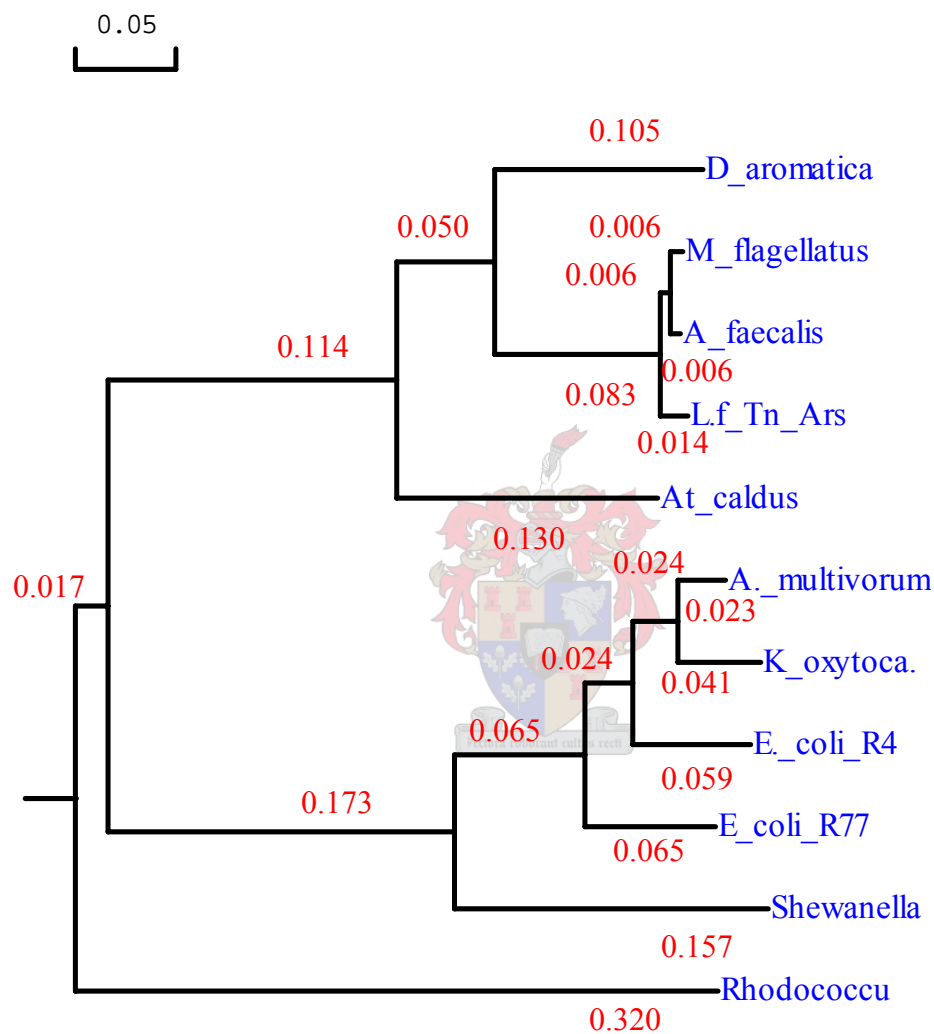
Figure 2.17: Phylogenetic tree of *L. ferriphilum* ArsA protein and other known ArsA proteins. Accession numbers are as follows: *Dechloromonas aromatica*, ZP_00348748.1; *Methylobacillus flagellatus*, ZP_00172245.2; *Alcaligenes faecalis*, AAS45117.1; *Acidithiobacillus caldus* (unpublished); *Acidiphilium multivorum*, BAA24822.1; *Klebsiella oxytoca*, AAF89640; *Escherichia coli* R46; *Escherichia coli* R773, A25937; *Shewanella* sp., AAO031589; *Rhodococcus erythropolis*, AAP74026. Values indicated represent phylogenetic distance.

BLAST results for the *L. ferriphilum arsB* homologue indicated that it has highest homology to the ArsB proteins from *A. faecalis* and *M. flagellatus* with E-values of $6e^{-166}$ and $2e^{-165}$, respectively. Subsequent phylogenetic analysis of various ArsB proteins, show that the *L. ferriphilum* ArsB groups with the ArsB proteins from other Gram-negative bacteria (Figure 2.18), with *A. faecalis* and *M. flagellatus* as its closest relatives. However, two possible clusters exist within the Gram-negative group, one consisting of *L. ferriphilum* ArsB and related proteins and the other containing ArsB proteins from most of the *Enterobacteriaceae.* Sequence alignments indicate that ArsB proteins are highly homologous, even between Gram-negative and Gram-positive ArsB groups. All ArsB proteins are hydrophobic, because they are located in the inner membrane of the cell (Chen *et al.*, 1986). To investigate whether the *L. ferriphilum* ArsB exhibited hydrophobic properties like other ArsB proteins, the web-based program TMPRED (http//www.ch.embnet.org/software/TMPRED_form.html) was used, which predicted 11 membrane-spanning domains for the *L. ferriphilum* ArsB. This hydropathy profile was compared to that of *E. coli* R773, *A. faecalis* and *S. aureus* ArsB proteins, indicating that the *L. ferriphilum* ArsB is very similar to both Gram-negative and Gram-positive ArsB proteins (Figure 2.19).
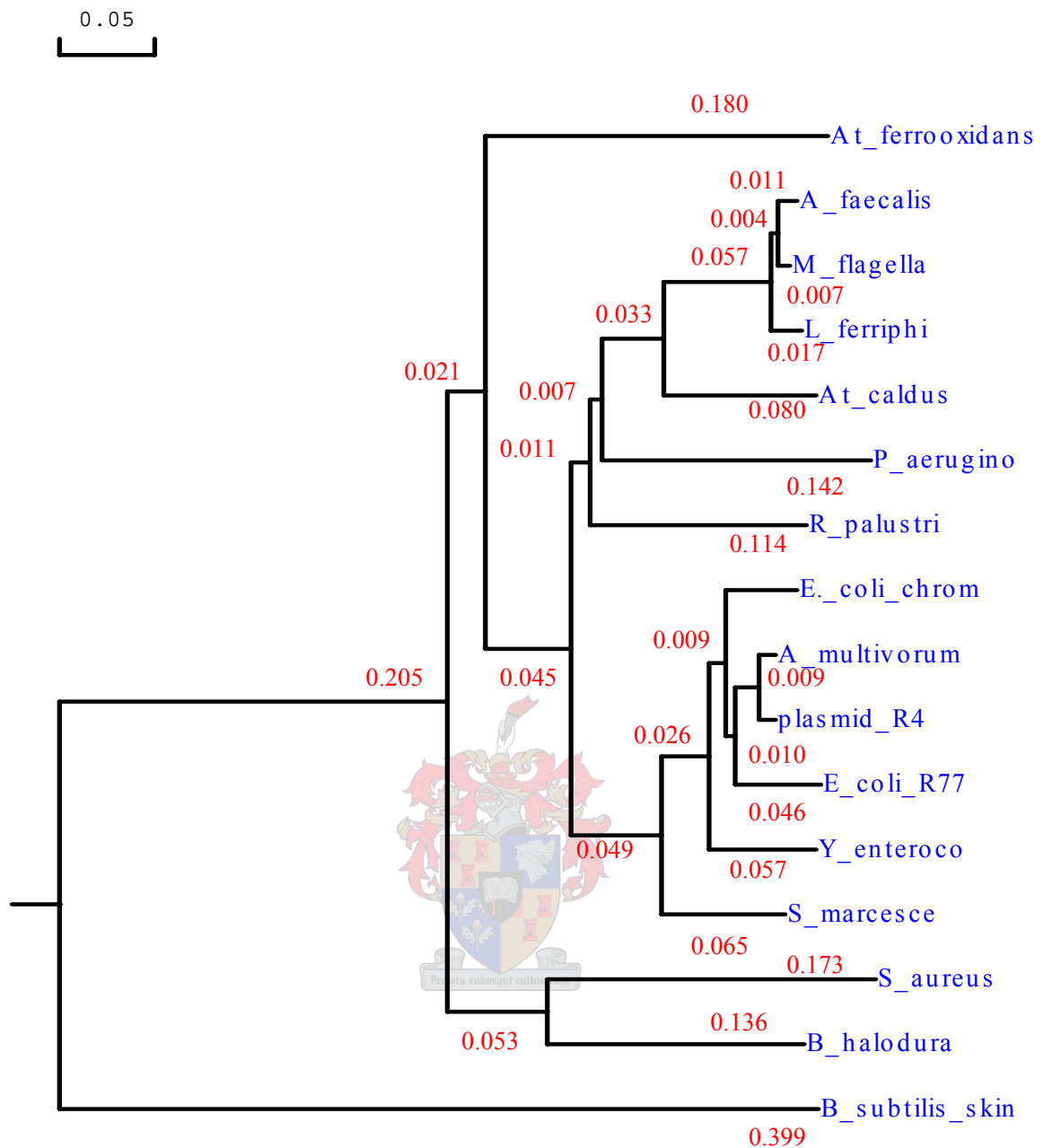
Figure 2.18: Phylogenetic tree of ArsB protein from *L. ferriphilum* and known putative ArsB homologues. Accession numbers: *Acidithiobacillus ferrooxidans*, AAF69238; *Alcaligenes faecalis*, AAS45119.1; *Acidithiobacillus caldus* (unpublished); *Pseudomonas aeruginosa*, NP_250968; *Rhodopseudomonas palustris*, NP_947603; *Escherichia coli* R773, ARB1_ECOLI, *Acidiphilium multivorum*, pKW301, BAA24823; *Salmonella typhimurium* R46, BAB91586; *Escherichia coli* K12 chromosome, NP_417959; *Yersinia enterocolitica* pYVe227, NP_052439; *Serratia marcescens*, CAB88405; *Bacillus halodurans*, NP_243865; *Staphylococcus aureus* pI258, P30329; *Bacillus subtilis* skin element, BAA0696. Values indicated on the tree represent phylogenetic distance.

70

A. *L. ferriphilum* ArsB
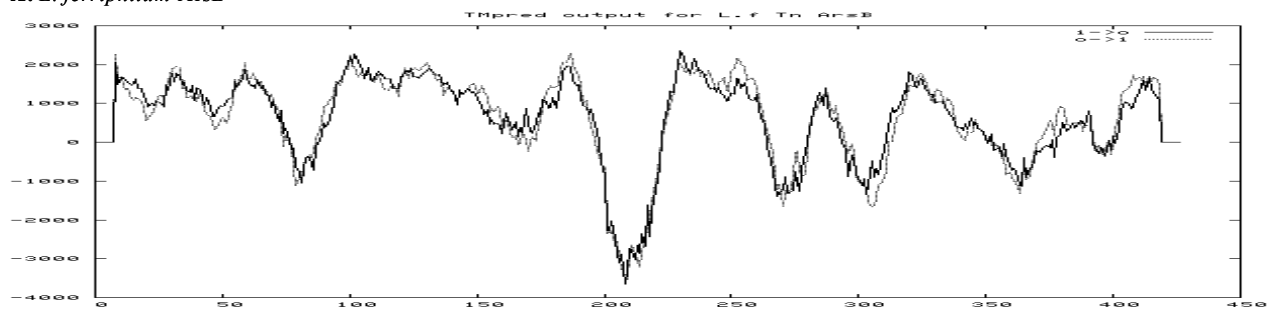
B. *A. faecalis* ArsB
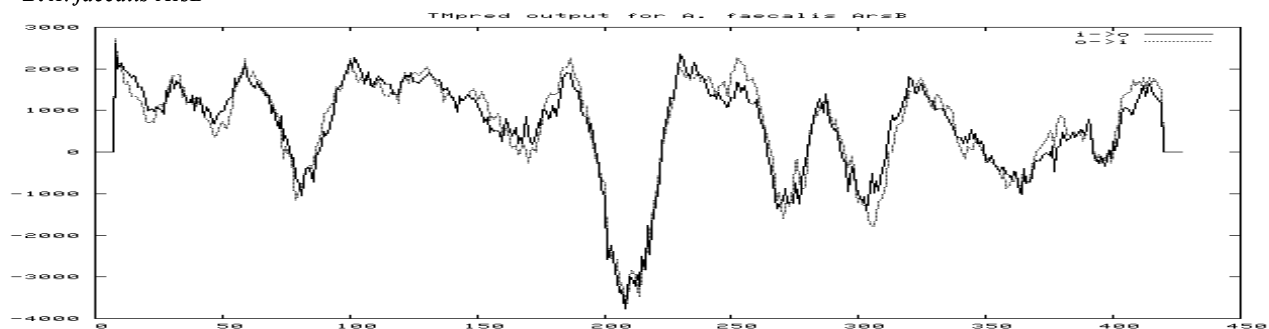
C. *E. coli* R773 ArsB
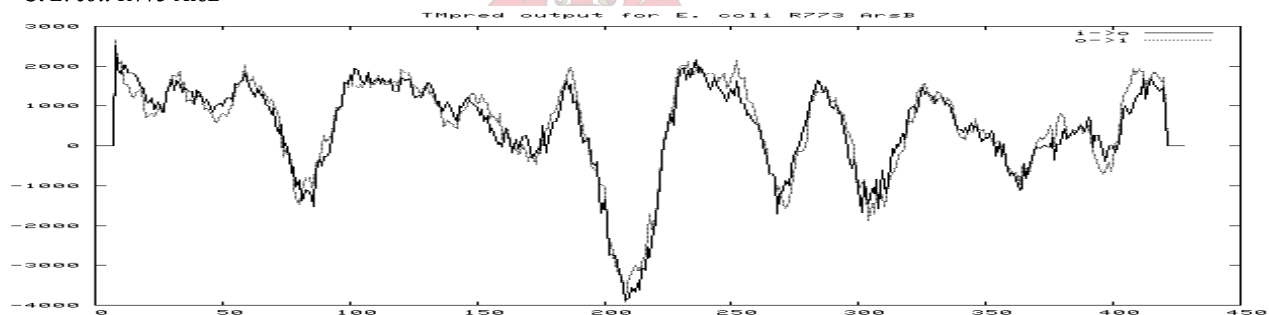
D. *S. aureus* ArsB
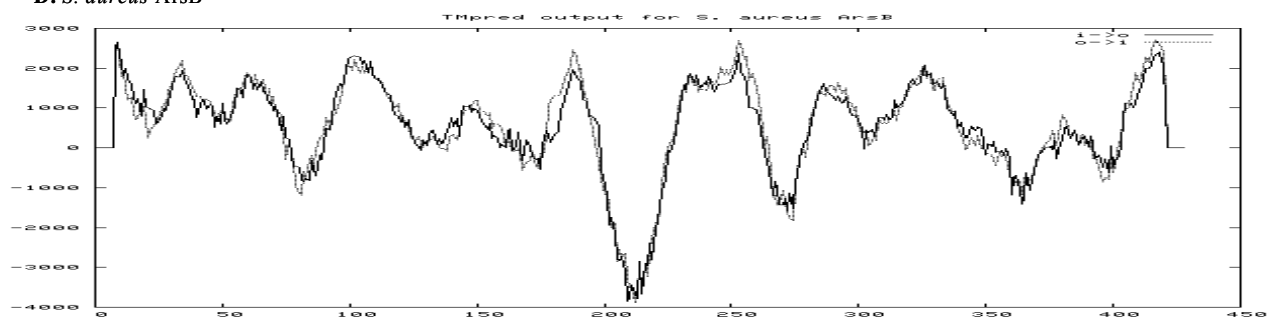
Figure 2.19: Prediction of trans membrane-spanning domains of ArsB proteins from *L. ferriphilum* (**A**), *A. faecalis* (**B**), *E. coli* R773 (**C**), and *S. aureus* (**D**) using the web-based program TMpred. Predictions for *A. faecalis*, *E. coli* R773 and *S. aureus* were included for comparisons.

71

2.3.3. The Tn*LfArs* is unique to *L. ferriphilum* Fairview.

To establish how widely distributed the *L. ferriphilum* Tn *ars* operon was throughout the genus of *Leptospirillum* or whether it was limited to the *L. ferriphilum* isolates from the high arsenic containing tanks at the Fairview mine, a Southern hybridization experiment was performed. An internal fragment (*Eco*RI- *Eco*RI) containing the *ars* genes of the Tn*LfArs* was used to probe genomic DNA from four *L. ferriphilum* strains (Fairview, ATCC49881, Warwick and Mont4) and four *L. ferrooxidans* strains (DSM2705, P₃A, Parys and Chil-LF) digested with *Bam*HI, respectively. As controls, two *At. ferrooxidans* strains (ATCC27230 and ATCC33020) and two *At. caldus* strains (#6 and MNG) were used. The labelled probe gave a positive signal from only the *L. ferriphilum* Fairview and none of the other *Leptospirillum* strains (Figure 2.20). This indicated that *L. ferriphilum* Fairview was the only *L. ferriphilum* strain containing the transposon located *ars* operon. The signal corresponded to a doublet with the approximate sizes of 10 and 12 kb, which hinted at the existence of two possible *ars* operons in *L. ferriphilum* Fairview. Further investigation of this possibility led to the remapping of restriction enzyme endonuclease sites of both genebank clones (pEco-*ars*9 and pEco-*ars*11) from which the p*L.fars* was assembled. This confirmed that the enzyme sites within the inverted repeats were similar, concurring with previous restriction maps of the same region. Further mapping revealed that the restriction enzyme sites within the flanking regions of both clones are different (Figure 2.2). If these clones were from the same region on the chromosome they would have had matching flanking regions, thus providing evidence for this claim. The discovery of two *Bam*HI sites on either side of the pEco-*ars*11 clone, flanking the Tn *ars* operon, could explain the presence of a 10 kb *Bam*HI fragment on the Southern hybridization. This region on the pEco- *ars*11 clone (10 kb *Bam*HI) corresponds exactly to a 10 kb *Bam*HI fragment on the *L. ferriphilum* Fairview chromosome (Southern hybridization), indicating the one *ars* operon. Therefore this would mean that the Tn *ars* operon contained on the pEco-*ars*9 clone represents the 12 kb *Bam*HI fragment, seeing that pEco-*ars*9 and pEco-*ars*11 are different. The recent discovery of the two *L. ferriphilum* Fairview *ars* operons also suggested that the p*L.fars* was unknowingly assembled from two different, but as far as we know, identical *ars* operons. We therefore realise that further experimentation involving the assembled Tn *ars* operon (p*L.fars*) may not exactly represent the natural situation in *L. ferriphilum* Fairview.

Initial analysis of the Southern hybridization result (Figure 2.6) was inconclusive due to the presence of unexplainable fragments obtained for the different restriction digestions.

Regardless of the corresponding fragments, additional fragments of various sizes were obtained, which only became clear in the light of the recent discovery of the two *L. ferriphilum* Fairview *ars* operons. The restriction digestions of the *L. ferriphilum* genomic DNA represented on the Southern hybridization pertaining particularly to *Nco*I-*Sal*I and *Sal*I-*Spe*I indicates two additional fragments for each respective digest (*Nco*I-*Sal*I: 11 kb and 8.0 kb; *Sal*I-*Spe*I: 12 kb and 7.0 kb). These additional fragments are relatively large and include part of the Tn *ars* and part of the flanking chromosome. The Tn *ars* region present on these fragments is partly covered by the probe, which explains their detection on the Southern hybridization, which validates that both the cloned Tn *ars* operons originate from the chromosome of *L. ferriphilum* Fairview.
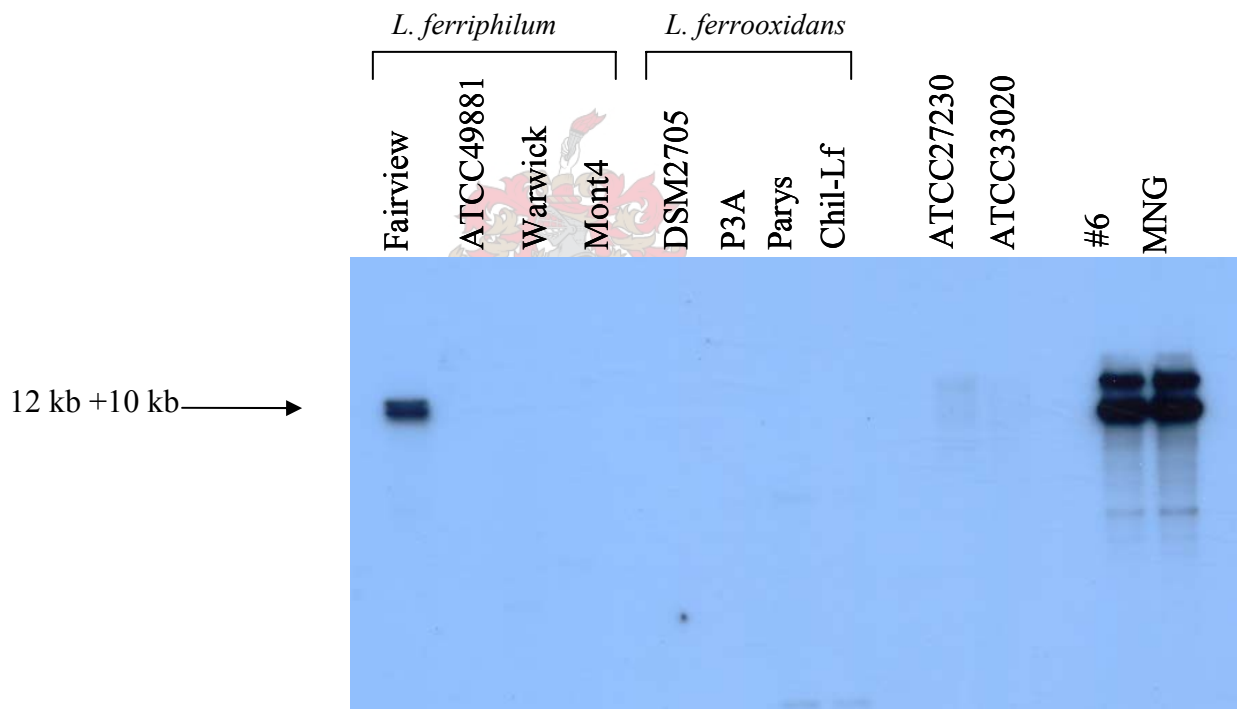
Figure 2.20: Hybridization of *L. ferriphilum* Tn *ars* operon probe, to genomic DNA from four *L. ferriphilum* species and four *L. ferrooxidans* species. Two *At. ferrooxidans* (ATCC27230 and ATCC33020) and two *At. caldus* (#6 and MNG) species were included as controls. All DNA was digested with *Bam*HI.

A. Arsenite
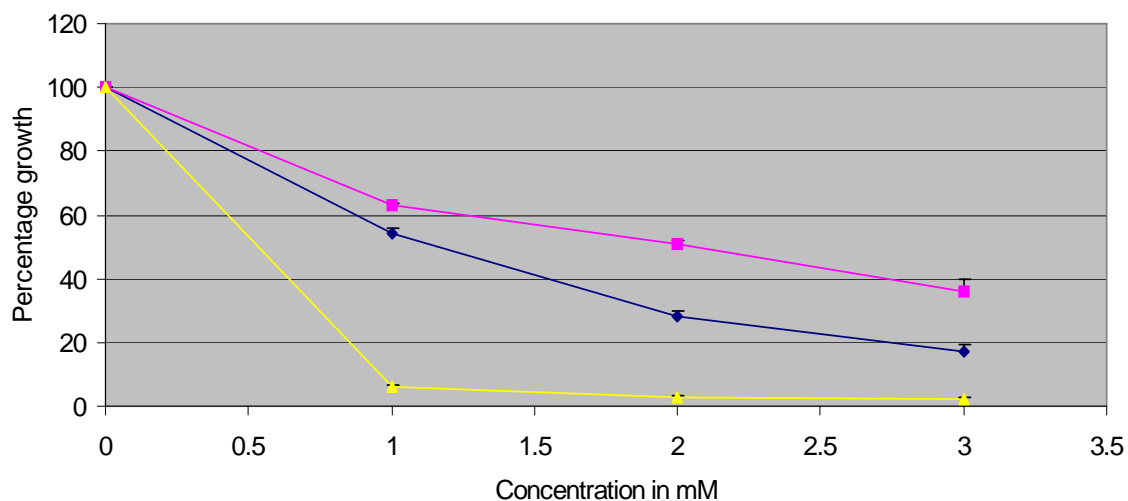
B. Arsenate
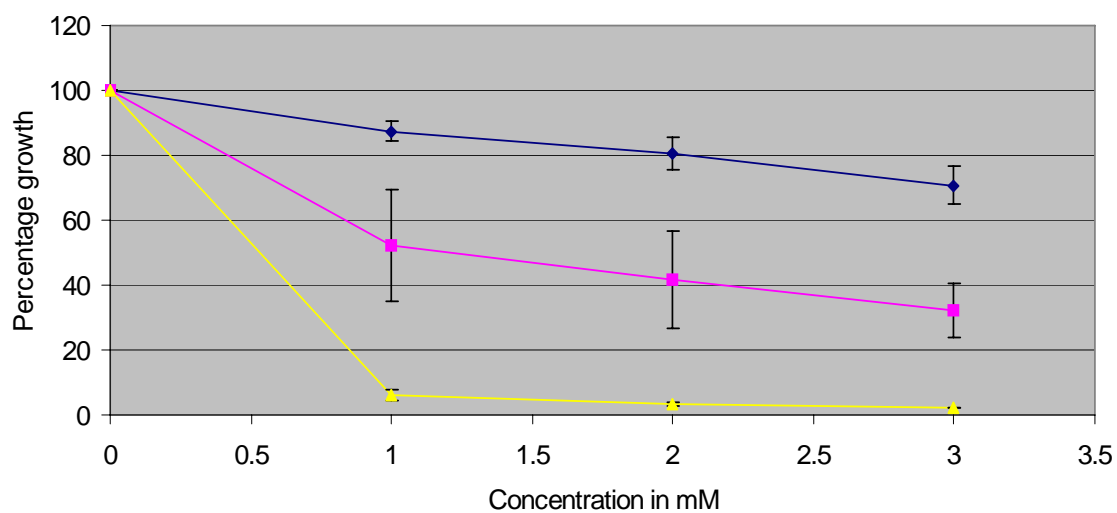
Figure 2.21: Growth of *E. coli* ACSH50 containing pTn*Lf*A*rs* (♦), pTn*AtcArs* (■), and pEcoBlunt (▲) in increasing concentrations of arsenite (**A**) and arsenate (**B**).

2.3.4. As(V) and As(III) Assays.

The ability of the *L. ferriphilum* Tn *ars* operon to confer arsenic resistance to the *ars* deletion mutant *E. coli* ACSH50I[q] was tested as shown in Figure 2.21. The level of arsenic resistance conferred to *E. coli* ACSH50I[q] by the Tn*LfArs* genes on plasmid pEcoR252 (pEcoR252 was digested with *Bgl*II, blunted and religated to inactivate the *Eco*RI enzyme, served as a negative control) was compared to that conferred by the *At. caldus* Tn *ars* genes (also on pEcoR252). Results indicated that the addition of Tn*LfArs* genes to *E. coli* ACSH50I[q] achieved high level arsenite resistance at 3 mM As(III), while the Tn*AtcArs* conferred moderate arsenite resistance at the same concentration (Figure 2.21). However, the opposite results to the previous experiment were obtained when the experiment was repeated with arsenate. These results indicated that *E. coli* ACSH50I[q] containing the Tn*AtcArs* genes exhibited higher levels of arsenate resistance than the *L. ferriphilum ars* genes. This could be due to the Tn*AtcArs* containing a duplication of the *arsDA* genes, which could give an organism possessing it an added advantage in arsenate rich media, suggesting that the duplicated genes (*arsDA*) function together. In the context of both experiments it indicated that the *L. ferriphilum ars* genes are capable of conferring resistance to an *E. coli ars* mutant.

2.3.5. Transposon of *L. ferriphilum* Fairview is functional in *E. coli*.

This work was performed by Dr. Marla Tuffin as mentioned previously. The *L. ferriphilum ars* operon is situated on a transposon. The insert containing the whole Tn*LfArs* operon was cloned into the non-mobilizable plasmid vector pUC19 (pUCTn*LfArs*, pers comm.) and transformed into *E. coli* XL1-Blue containing the conjugative plasmid pSa, to test whether the transposon could function in *E. coli*. The cells containing pUCTn*LfArs* and the conjugative plasmid pSa were mated with an *E. coli* ACSH50I[q]-Rif recipient and plated on selective media. *E. coli* plus pUCTn*LfArs*, but without pSa served as a negative control. The reasoning behind this was that if the *L. ferriphilum* transposon were active it would be able to transpose into pSa, which would then be able to be mated across to the *E. coli* recipient. The pUCTn*LfArs* being non-mobilizable would not be able to conjugate to the recipient. Plasmid DNA prepared from As[r] Km[r] and As[s] Km[r] transconjugants was digested for analysis using Pulsed-field Gel Electrophoresis (TAFE), followed by Southern hybridization. From comparisons of the restriction enzyme digestion patterns (Figure 2.22) of the transconjugants (lanes 3 and 4) and pSa (lane 5) it can be seen that there is an increase in fragment size of the top band. The increase in size of this fragment suggests that the transposon jumped into pSa, expanding its

size. For the Southern hybridization experiment, a 2.2kb *Ksp*I fragment containing the *arsDA* genes from Tn*AtcArs* was labelled and used as a probe, in order to detect the presence of the *ars* gene-carrying transposon contained in the putative transconjugants. The plasmid DNA from the As[r] Km[r] colonies hybridized to the *arsDA* probe, suggesting that the transposon had jumped into pSa, which was subsequently mated across to the recipient cells (Figure 2.22). The possibility that these transconjugants were co-integrates was eliminated due to the transconjugants testing negative on ampicillin (transconjugants didn't grow on ampicillin). If these transconjugants were co-integrates they would have been both ampicillin and arsenic positive.
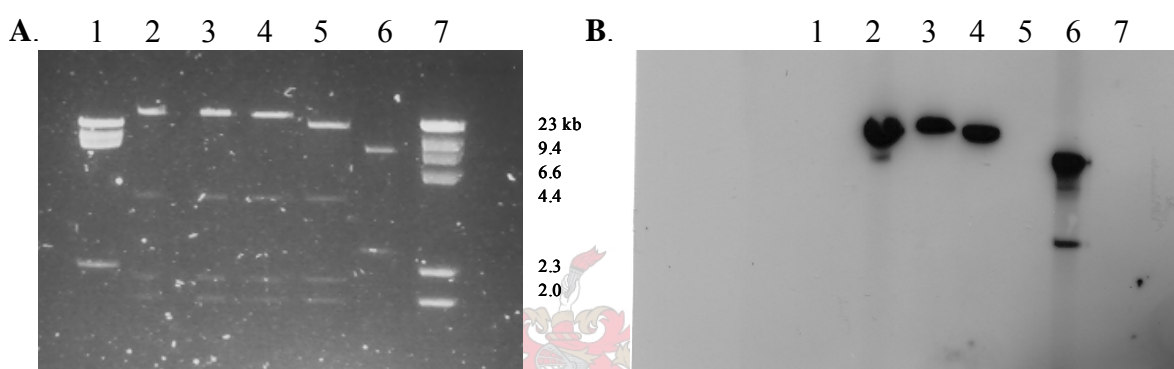
Figure 2.22: TAFE gel (**A**) and Southern hybridization of TAFE gel (**B**) probed with 2.2kb *arsDA* from *At. caldus*. Lane 3 and 4: transconjugants (LF D1 and LF D6), lane 6: pUCTn*LfArs*. Controls: Atc D6 (a transconjugant from *At. caldus* transposon jumping experiment) (lane 2) and pSa (lane 5). A positive signal was detected for the transconjugants as well as the *At. caldus* control. Lane 1 and 7 are DNA markers.

2.3.6. Minimum inhibitory arsenic concentrations in *Leptospirillum*.

L. ferriphilum Fairview was tested for its ability to grow in the presence of As(V) and As(III) and compared with the *L. ferriphilum* type strain ATCC49881 and two *L. ferrooxidans* strains, DSM2705 and ATCC49879 (Figure 2.23). It was found that *L. ferriphilum* Fairview grew at concentrations up to 60 mM for both As(V) as well as As(III), while ATCC49881 grew at concentrations up to 40 mM As(V) and As(III), respectively. The two *L. ferrooxidans* strains DSM2705 grew at concentrations up to 50 mM As(III) and 40 mM As(V), while ATCC49879 grew concentrations up to 30 mM As(III) and 20 mM As(V). Although *L. ferriphilum* Fairview was possibly capable of growing to a much higher concentration of arsenic, precipitation of the arsenic at higher concentrations was a limiting factor. Therefore the effective arsenic concentration above 60 mM was uncertain.

A. Arsenite

	0	10	20	30	40	50	60	80
Fairview	Yes	Yes	Yes	Yes	Yes	Yes	Yes	——
ATCC49881	Yes	Yes	Yes	Yes	Yes	——	——	——
DSM2705	Yes	Yes	Yes	Yes	Yes	Yes	——	——
ATCC49879	Yes	Yes	Yes	Yes	——	——	——	——

B. Arsenate

	0	10	20	30	40	50	60	80
Fairview	Yes	Yes	Yes	Yes	Yes	Yes	Yes	——
ATCC49881	Yes	Yes	Yes	Yes	Yes	——	——	——
DSM2705	Yes	Yes	Yes	Yes	Yes	——	——	——
ATCC49879	Yes	Yes	Yes	——	——	——	——	——

Figure 2.23: Minimum inhibitory arsenic concentrations in *L. ferriphilum* strains Fairview and ATCC49881 compared to *L. ferrooxidans* strains DSM2705 and ATCC49879 in arsenite (**A**) and arsenate (**B**). Concentrations are in mM. Yes indicates growth at a particular concentration.

TABLE OF CONTENTS

3. 1. INTRODUCTION

3. 1. INTRODUCTION

Prior to this study, arsenic resistance in acidophiles was already well known and documented (Butcher *et al*., 2000; Butcher and Rawlings, 2002; De Groot *et al*., 2003; Tuffin *et al*., 2005). Researchers have shown that the genes responsible for arsenic resistance (*ars* genes) occur in a wide variety of different operons (Chen *et al*., 1986; Diorio *et al*., 1995; Sato and Kombayashi, 1998; Butcher *et al*., 2000; Gihring *et al*., 2003; Tuffin *et al*., 2005). However, these *ars* genes are commonly organized into two types of general arrangement, one consisting of 5 genes (*arsRDABC*, such as in the *E. coli* R-factor R773) and the other consisting only of three genes as found on the *E. coli* chromosome and the *Staphylococcal* plasmids pI258 and pSX267. Initially *ars* operons were isolated mainly on plasmids, however in the light of recent sequencing projects, they are now more frequently found on chromosomes. Recently, technological advancement of techniques such as crystallography allowed for the structure of a number of proteins associated with *ars* resistance to be studied in detail, such as the ArsA protein from *E. coli* R773 (Zhou *et al*., 2001) and the ArsC proteins from *S. aureus* pI258 (Zegers *et al*., 2001), *E. coli* R773 (Martin *et al*., 2001) and *B. subtilis* skin element (Bennett *et al*., 2001). This provided much needed insight into the mechanisms of how they function in arsenic resistance.

3.2. Origin of the Tn *ars*.

The *L. ferriphilum* strain Fairview together with *At. caldus*, occur in the arsenic rich environment of the biooxidation tanks at the Fairview mine. During the initial stages of the process development, it was found that the bacteria present in the continuous-flow biooxidation tanks (which operated at 40 °C) were sensitive to the high concentrations of arsenic (12 g/l arsenate and 3-6 g/l arsenite), which was released as a result of mineral oxidation (Dew *et al*., 1997). During the last 15 years of operation, when samples were taken, the culture had adapted to tolerate high concentrations of arsenic, resulting in a selection of highly arsenic resistant microbes. As previously mentioned, the Biox® culture is dominated by *L. ferriphilum* and *At. caldus* (Rawlings *et al*., 1999). However, these two bacteria are very different phylogenetically, with *L. ferriphilum* being a predominately iron-oxidizing bacterium and *At. caldus* sulphur-oxidizing. The question was then asked; what genetic changes had to take place in the Biox® culture to permit arsenic resistant growth?

Studies to answer this question first began with *At. caldus*. Tuffin *et al*. (2005) found 3 different types of arsenic resistance systems in *At. caldus*; the first was a 12.2 kb transposon

with Tn*21*-like inverted repeats and transposase and resolvase genes (Figure 3.1). It contained ORFs homologous to *arsR*, *arsB*, *arsC* and duplicate copies of the *arsD* and *arsA* genes. It also contained two additional ORFs, an *orf7* which had homology to a 447 aa NADH-dependent dehydrogenase from *Thermoanaerobacter tencongensis* and *orf8*, which had homology to a cystathione-β-synthase (CBS) of a membrane protein from *Desulfitobacterium hafiense*. The second operon of 11.9 kb was also located on a Tn*21*-like transposon, which contained only the *arsC* and *arsR* genes as well as a single copy of *arsDA*. It had no *arsB*, *orf8 tnpA* or a second inverted repeat. The genes that were present were identical to their counterparts found previously on the 12.2 kb transposon. A third system was also discovered, which consisted of an *arsR*, *arsC* and an *arsB* (not shown). The genes from this system appear to be completely different from the other Tn*21*-like transposon associated *ars* genes and are thought to be on the chromosome.
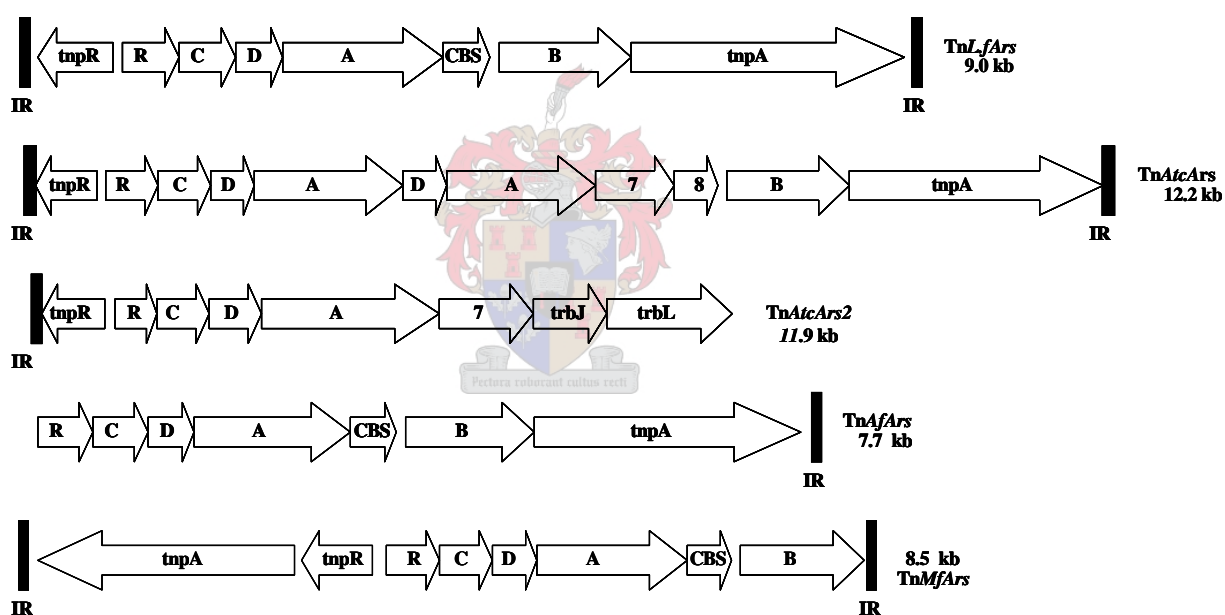
Figure 3.1. Comparison of the transposon operons *At. caldus*, *L. ferriphilum*, *A. faecalis* and *M. flagellatus*. As can be seen above all five transposon operons look remarkably similar, with the Tn*AtcArs* from *At. caldus* showing the most variation i.e. an *arsDA* duplication and the presence of *orf7*. The CBS *orf* of Tn*LfArs* is similar to *orf8* of Tn*AtcArs* and the *trblJ* and *trbL* genes present on the second *At. caldus* operon are not part of the *ars* operon.

This study addressed the question of arsenic resistance in *L. ferriphilum*. The Tn*LfArs* was isolated from a *L. ferriphilum* Fairview genebank as described in Chapter 2. BLAST results from individual ORFs of the Tn*LfArs* identified homologues to *arsR*, *arsC*, *arsD*, *arsA*, *arsB* as well as a cystathione-β-synthase (CBS) domain and it was also discovered that these *ars* genes

were located on a transposon similar to that found in *At. caldus*. At first inspection the arrangement of the *ars* and transposon genes appeared to be remarkably alike. This was not surprising as transposons are very effective at horizontal gene transfer (HGT). This apparent similarity was consistent with the suggestion of Woese (2004), which described genes acquired from the environment that allow organisms to adapt to change, as "cosmopolitan genes", suggesting they are more characteristic of the environment than they are of organismal lineage. However, on closer examination it was surprising to discover that the two transposons are not sufficiently closely related for transfer to have taken place within the last 15 years given the rate of base pair substitutions that would be expected over that time. A nucleotide sequence comparison was done of the transposon *ars* operon genes from *At. caldus*, *L. ferriphilum*, *A. faecalis* and *M. flagellatus* (Table 3.1). Given these sequence differences it is not likely that the *L. ferriphilum* transposon carrying the *ars* genes was inherited from *At. caldus* or vice versa within the last 15 years. Furthermore, it was found that the Tn*LfArs* genes are much more closely related to the Tn *ars* of *Methylobacillus flagellatus* and *Alcaligenes faecalis* than to the *At. caldus* Tn *ars*. As the *At. caldus* and *L. ferriphilum* are the most divergent of the arsenic gene-containing transposons, this suggests that they must have acquired the arsenic resistant transposon independently.

The paths by which the two related but different *ars* gene containing transposons have arrived in two different types of bacteria in the same arsenic-laden biooxidation at the Fairview mine are unknown. However, as indicated above, what is clear is that horizontal gene transfer of the *ars* transposons between the two bacteria has not taken place within the biooxidation tanks during the 15 years from when the Fairview bioleaching tanks began operating until the samples of bacteria used in this study were collected. Woese (2004) suggested that the environment rather than the bacterial lineage would select "cosmopolitan genes" of which *ars* genes are an example. The arsenic rich environment of the biooxidation tanks therefore appears to have selected for similar *ars* gene containing transposons. The observations made in this study suggest that Woese is correct. If the prediction of Woese is interpreted as implying that "cosmopolitan genes" would be passed between bacteria of different lineages within an environment that selects for them, this is not what has happened in the case of the biooxidation tanks. Each set of *ars* genes present in *At. caldus* and *L. ferriphilum* must have been passed along a different chain of organisms to arrive at their final destination.

Table 3.1. Comparison of transposon *ars* operons on nucleotide level.

	L. ferriphilum			
	tnpR	*tnpA*	*arsB*	*arsRCDA*
M. flagellatus	79 %	86 %	96 %	94 %
A. faecalis	–––	90 %	94 %	93 %
At. caldus	95 %	90 %	76 %	75 %

*Nucleotide sequence comparisons were done as follows: *tnpR* and *tnpA* gene comparisons were done separately because not all the genes lie adjacent to each other. Arsenic resistance genes all read in the same direction and therefore were treated as a block. The percentage identity was done over a region that all four transposons have in common.

A comparison of the four transposon *ars* operons from *L. ferriphilum*, *A. faecalis*, *At. caldus* and *M. flagellatus* revealed that they were all variations of the same configuration, which was indicative of their evolutionary relatedness (Figure 3.1). Further evidence supporting this claim was provided by phylogenetic analysis of the products of the Tn*LfArs*, which showed that the products from the four operons always grouped together. These results suggests that all four of the *ars* gene-containing transposons have diverged from the same ancestor, with the *L. ferriphilum*, *M. flagellatus* and *A. faecalis* having diverged more recently than the *At. caldus ars* transposon. Tuffin *et al.* (2005) suggested that the *At. caldus ars* transposon evolved for the sole purpose of distribution of the *ars* genes it contains, enclosing the *ars* genes by flanking the operon with the *tnpR* and *tnpA* genes. Closer inspection of the transposon comparisons, however, indicated that the four *ars* transposons are also sufficiently different, with three out of the four being complete, of which two possessed the configuration described by Tuffin *et al.* (2005). The transposon from *M. flagellatus*, on the other hand, is slightly different to the transposons from *L. ferriphilum* and *At. caldus*, due to the *tnpR* and *tnpA* genes being adjacent to one another. The *A. faecalis* Tn *ars*, although having the same gene order as the *L. ferriphilum* and *At. caldus*, is incomplete missing the *tnpR* and left inverted repeat, suggesting that it is inactive. Tuffin *et al.* (2005), also reported that the Tn *ars* from *At. caldus* is active in *E. coli*. Similarly, we have shown through conjugation experiments that the Tn*LfArs* is still active and capable of transposition in *E. coli*.

3.3. Highly arsenic resistant *L. ferriphilum* strain possesses the Tn*LfArs*.

The *Leptospirillum* spp. used in this study were isolated from different geographical locations around the world including various mining environments (Table 2.1). The possibility, therefore, existed that some of these strains possessed arsenic resistance genes; especially those isolated from mining environments. A survey was done using Southern hybridization, to investigate the geographical distribution of the transposon-located *ars* operon in several strains of the genus, using the *L. ferriphilum* and *L. ferrooxidans* strains from a previous study (Coram and Rawlings, 2002). What we discovered was that the *ars* transposon was unique to *L. ferriphilum* Fairview and it was not detected in any of the other *Leptospirilli*. This result raised an important question, whether any of the other *Leptospirillum* strains had other *ars* genes that were not sufficiently related to give a positive hybridization signal?

Tyson *et al*. (2004) obtained 95 % of the genome sequence of an *L. ferriphilum* isolate during a metagenome sequencing project. We searched this sequence and found an *ars*-like operon. Upon initial inspection of this operon, it appeared only to have the *arsB* and *arsC* genes, but closer examination revealed that the *arsC* had an *arsR* gene fused to it (Figure 3.2). When the sequences of these *ars* genes and their products were analysed and compared to the Tn*LfArs* it was found that the *arsB* gene shares 57 % identity on nucleotide level and 58 % identity on amino acid level to the transposon version. If the fused *arsR* and *arsC* genes and their predicted products were separated, they had only 35 % and 45 % nucleotide similarity respectively to their transposon counterparts and they had only 44 % (ArsR) and 27 % (ArsC) identity on amino acid level with their respective Tn*LfArs* versions. The 58 % or less nucleotide sequence identity would explain why these particular *ars* genes were not detected in the Southern hybridization experiments when probing with an homologous *arsB* probe (Chapter 2, section 2.3.1).

In a PCR-based experiment, using specifically designed primers to the regions flanking the *L. ferriphilum arsRC* and *arsB* genes, a product was only obtained for *L. ferriphilum* (Fairview and ATCC49881), but not for *L. ferrooxidans* (DSM2705, ATCC49879 and CHIL-Lf) Tuffin *et al*. (2005). The amplified *arsRCB* genes were then used as a probe in Southern hybridisation experiments, which confirmed that the genes were only detected in *L. ferriphilum* (Fairview, ATCC49881 and Warwick). When the amplified *arsRCB* genes were tested for activity in the *ars* mutant *E. coli* ACSH50I[q], it was discovered that these genes were not functional in *E. coli*. This scenario where all four *L. ferriphilum* strains possessed chromosomal *ars* genes, but only

the highly arsenic resistant *L. ferriphilum* Fairview contains the Tn *ars* is similar to the situation found in *At. caldus* (Tuffin *et al*., 2004). When six *At. caldus* isolates, three of which had previous exposure to arsenic (#6, MNG and "f") and three which had not been exposed to the toxic metal (BC13, KU and C-SH12), were investigated by Southern hybridization, it was found that only the three strains previously exposed to arsenic contained the Tn*AtcArs*. We tested whether the levels of *ars* resistance in *L. ferriphilum* with the chromosomal copy or *L. ferriphilum* with the chromosomal and transposon copies were different. We found that even though the Fairview culture has been cultivated in our laboratory for more than 6-7 years, it was still more resistant to arsenic than the other strains, indicating that only the *L. ferriphilum* Fairview strain, which was exposed to high levels of arsenic, possessed the Tn*LfArs*.
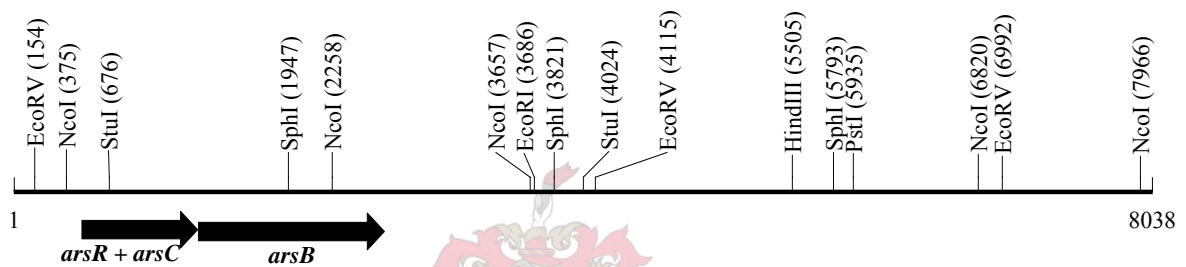
Figure 3.2. Depiction of *L. ferriphilum* chromosomal *ars* genes identified on an 8 kb contig from sequence deposited in the NCBI database (AADL01000040) (see text for detail).

3.4. Conclusion

In conclusion, we succeeded in identifying and sequencing the *ars* genes from *Leptospirillum ferriphilum* Fairview. We found homologues to genes *arsRCDAB* and proved via Southern hybridization that the genes originated from *L. ferriphilum* Fairview. We showed that the genes were situated on a transposon that is active in *E. coli* and that it has significant resemblance to three other transposon-located arsenic resistance systems. We also showed that this Tn *ars* was unique to a highly arsenic resistant *L. ferriphilum* strain and that the other *L. ferriphilum* strains, able to grow in lower amounts of arsenic, possessed chromosomal versions of *ars* genes, which were undetected in Southern hybridization experiments (chapter 2, section 2.3.1).

The situation in Biox® plants is different to laboratory environments as the plants are non-sterile and would be continually selecting for the most efficient *ars* systems. It would, therefore, be interesting to examine whether the same Tn *ars* system is present in Biox® plants operating in other countries besides South Africa, or whether bacteria in these tanks have selected for other, possibly more effective *ars* resistance systems. It would also be interesting to investigate Biox® tanks that have been operating for a number of years in non-South African environments that were not inoculated with the Fairview cultures to see whether the bacteria present in these tanks have similar *ars*-gene containing transposons as the Fairview bacteria.
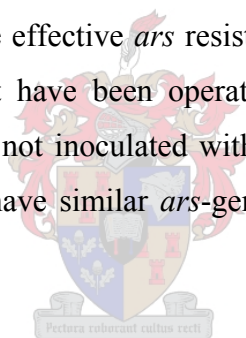
Appendix One: Table of constructs used for sequencing of pTn*LfArs*.

Plasmids	Construction
pH-Ec	Ap^r (*Hind*III-*Eco*RI fragment from pEco-*ars*9 cloned into pBluescript digested with *Hind*III and *Eco*RV)
pEc-Bg	Ap^r (*Eco*RI-*Bgl*II fragment from pEco-*ars*9 cloned into pBluescript digested with *Bam*HI and *Eco*RI)
pUCBg-Ev	Ap^r (*Bgl*II-*Eco*RV fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Bam*HI and *Eco*RV)
pUCEv-Ec	Ap^r (*Eco*RV-*Eco*RI fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Eco*RI and *Eco*RV)
pUCEc-Ev	Ap^r (*Eco*RI- *Eco*RV fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Eco*RI and *Eco*RV)
pUCEv-St	Ap^r (*Eco*RV-*Stu*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Eco*RV and *Stu*I)
pUCEv-Sp	Ap^r (*Eco*RV- *Sph*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Eco*RV and *Sph*I)
pUCSt-Sp	Ap^r (*Stu*I-*Sph*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Stu*I and *Sph*I)
pUCSt-N	Ap^r (*Stu*I-*Nco*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Stu*I and *Nco*I)
pUCSp-Sp	Ap^r (*Sph*I-*Sph*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Sph*I)
pUCSp-St	Ap^r (*Sph*I-*Stu*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Stu*I and *Sph*I)
pUCN-Sa	Ap^r (*Nco*I-*Sal*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Sal*I and *Nco*I)
pUCN-St	Ap^r (*Nco*I-*Stu*I fragment from pEco-*ars*9 cloned into pUCBM21 digested with *Stu*I and *Nco*I)
pSt-Ec	Ap^r (*Stu*I-*Eco*RI fragment from pEco-*ars*11 cloned into pUCBM21 digested with *Stu*I and *Nco*I)
pEc-Ba	Ap^r (*Eco*RI- *Bam*HI fragment from pEco-*ars*11 cloned into pUCBM21 digested with *Eco*RI and *Bam*HI)

Appendix Two: Annotated sequence obtained from pTn*LfArs*

Shown below is double stranded sequence of the *Leptospirillum ferriphilum* transposon arsenic operon, which was obtained during the course of this study. Protein translations of the eight open reading frames are located below the DNA sequence. Restriction endonuclease cleavage sites, inverted repeats and stop codons are highlighted in italics.

```
1       TTTTTCATTGAAATTCCTTTCTGTCAATAGGGGTTACGACGGAAACATCCGGCGAATATT
        AAAAAGTAACTTTAAGGAAAGACAGTTATCCCCAATGCTGCCTTTGTAGGCCGCTTATAA

                                                                 EcoRV
61      CTCCAGTTCCGGAACAGATCGGGAAATGGAGAGGACGCACTGTTTATCGACCTTGATATC
        GAGGTCAAGGCCTTGTCTAGCCCTTTACCTCTCCTGCGTGACAAATAGCTGGAACTATAG


121     GGAGTGTTCCTGGAGAAATCGCAGGGCATCCTTGATCGACCATGCGTCGCGATAGGGACG
        CCTCACAAGGACCTCTTTAGCGTCCCGTAGGAACTAGCTGGTACGCAGCGCTATCCCTGC

                          EcoRV
181     ATTGCTGGTGAGCGCGTCAAAGATATCTGCGACAGAGACAATGCGGGATTCGATCGGAAT
        TAACGACCACTCGCGCAGTTTCTATAGACGCTGTCTCTGTTACGCCCTAAGCTAGCCTTA
                                             Left inverted repeat
                                             _____

241     ACGGGGTCCATTCAGGGGTCGTCTCAGAAAACGGAAAACAAAGCACGTTAAGCTCGAATT
        TGCCCCAGGTAAGTCCCCAGCAGAGTCTTTTGCCTTTTGTTTCGTGCAATTCGAGCTTAA
                                                           *   A   R   I

301     TCGCGCAGGTACTGGTACAACGTTTCGCGGCTGATGCCAAATTCACGGGCCAGCTTTGCC
        AGCGCGTCCATGACCATGTTGCAAAGCGCCGACTACGGTTTAAGTGCCCGGTCGAAACGG
        E   R   L   Y   Q   Y   L   T   E   R   S   I   G   F   E   R   A   L   K   A

361     TTTTGCTCGCCGGCATTGACGCGTTGCAGAAGTTCAGCGGCGCGTTCGGGCGAAAGCGCT
        AAAACGAGCGGCCGTAACTGCGCAACGTCTTCAAGTCGCCGCGCAAGCCCGCTTTCGCGA
        K   Q   E   G   A   N   V   R   Q   L   L   E   A   A   R   E   P   S   L   A

421     TTCTTGCGGCCCCTGTAAGCTCCGCGCTGTTTGGCGAGCGCAATACCTTCTTTCTGTCGC
        AAGAACGCCGGGGACATTCGAGGCGCGACAAACCGCTCGCGTTATGGAAGAAAGACAGCG
        K   K   R   G   R   Y   A   G   R   Q   K   A   L   A   I   G   E   K   Q   R

                        EcoRI
481     TCACGGATCAAGGCCCGCTCGAATTCGGCGAACGCGCCCATGACCGACAACATCAAGTTC
        AGTGCCTAGTTCCGGGCGAGCTTAAGCCGCTTGCGCGGGTACTGGCTGTTGTAGTTCAAG
        E   R   I   L   A   R   E   F   E   A   F   A   G   M   V   S   L   M   L   N

541     GCCATCGGTGAATCCTCGCCGGTGAAGGTCAGGCATTCCTTGACGAACTCGATGCGCACA
        CGGTAGCCACTTAGGAGCGGCCACTTCCAGTCCGTAAGGAACTGCTTGAGCTACGCGTGT
        A   M   P   S   D   E   G   T   F   T   L   C   E   K   V   F   E   I   R   V
```

601 CCGCGCTTGGTCAGCTTTTGCACGAGGCGACGCAGGTCATCGAGGTTGCGGGCCAGGCGA
    GGCGCGAACCAGTCGAAAACGTGCTCCGCTGCGTCCAGTAGCTCCAACGCCCGGTCCGCT
    G R K T L K Q V L R R L D D L N R A L R

661 TCCATGCTGTGCACCACCACCGTATCGCCTTCGCGCACGAAGGCCAGTAGTGAGTCAAGT
    AGGTACGACACGTGGTGGTGGCATAGCGGAAGCGCGTGCTTCCGGTCATCACTCAGTTCA
    D M S H V V V T D G E R V F A L L S D L

721 TCGGGCCGCTGGGTGTCCTTGCCCGATGCCTTGTCGGTGAATACCTTTCCGACTTCGGCC
    AGCCCGGCGACCCACAGGAACGGGCTACGGAACAGCCACTTATGGAAAGGCTGAAGCCGG
    E P R Q T D K G S A K D T F V K G V E A

781 TGCTCCAGTTGCCGCTCCGGGTTTTGGTCGAAGCTGCTGACCCGGACGTAGCCGATACGT
    ACGAGGTCAACGGCGAGGCCCAAAACCAGCTTCGACGACTGGGCCTGCATCGGCTATGCA
    Q E L Q R E P N Q D F S S V R V Y G I R

Start *tnpR*

841 TGACCTT*GC*AAGATGCCTCCAAATAGAAAGTGTCAGGAAGAAATCTATGACCTTGCGTAG
    ACTGGAA*CGT*TCTACGGAGGTTTATCTTTCACAGTCCTTCTTTAGATACTGGAACGCATC
    Q G Q L

901 CTTGTGTCAATAAATTAAGTACGCGAGTCTATCCGGACGTTTCGGGATAGGCTGTCCTGA
    GAACACAGTTATTTAATTCATGCGCTCAGATAGGCCTGCAAAGCCCTATCCGACAGGACT

961 CGCCCGGTTAGGGTATAGCCTAATCGGACGCGACGGCATGATTGAATCATTTTGATAGTT
    GCGGGCCAATCCCATATCGGATTAGCCTGCGCTGCCGTACTAACTTAGTAAAACTATCAA

Start *arsR*

                                              *Cla*I
1021 GTAGTACAATTTGAGC*ATG*GAAAAGCAAGCCGCAAC*ATCGAT*CTTCGAATCCCTCTCCTC
     CATCATGTTAAACTCG*TAC*CTTTTCGTTCGGCGTTG*TAGCTA*GAAGCTTAGGGAGAGGAG
                     M E K Q A A T S I F E S L S S

1081 TGGCCTGCGCTTGGATGTGTTCAGGCTGTTGGTCAAGAAGGGGCCGGGCGGCATGGTGGC
     ACCGGACGCGAACCTACACAAGTCCGACAACCAGTTCTTCCCCGGCCCGCCGTACCACCG
      G L R L D V F R L L V K K G P G G M V A

1141 AGGCGAGATTGCTAACGCGCTGGACATACCGCCGGCCAACCTTTCCTTTCACCTCAAGGC
     TCCGCTCTAACGATTGCGCGACCTGTATGGCGGCCGGTTGGAAAGGAAAGTGGAGTTCCG
      G E I A N A L D I P P A N L S F H L K A

1201 GCTGTCACAAGCTCATCTGGTGACGGTTGAGCAGGAGGGTCGCTTCCAACGTTACCGAGC
     CGACAGTGTTCGAGTAGACCACTGCCAACTCGTCCTCCCAGCGAAGGTTGCAATGGCTCG
      L S Q A H L V T V E Q E G R F Q R Y R A

1261   GGACATCCCGTTAATGTTGGATTTGATTGCGTACCTCACCGAAGAATGCTGTTCGGGTAG
             CCTGTAGGGCAATTACAACCTAAACTAACGCATGGAGTGGCTTCTTACGACAAGCCCATC
         D I P L M L D L I A Y L T E E C C S G

1321   TCCCGATCAATGCCTTGAACTACGTACCGCCTCGAAGTGTTCCGAGGAGTTCTTGCCGCT
             AGGGCTAGTTACGGAACTTGATGCATGGCGGAGCTTCACAAGGCTCCTCAAGAACGGCGA
         P D Q C L E L R T A S K C S E E F L P L

[Start *arsC*]

1381   CCTGTCACCCACTCCAACGAAAGCGACAAC*ATGA*ATATTCTGTTTCTTTGTACCGGTAAC
             GGACAGTGGGTGAGGTTGCTTTCGCTGTTG*TACT*TATAAGACAAAGAAACATGGCCATTG
         L S P T P T K A T T M N I L F L C T G N

1441   TCCTGCCGTTCCATCCTGGCGGAAGCGACCTTCAACCACTTGGCTCCGACAGGCTGGAAA
             AGGACGGCAAGGTAGGACCGCCTTCGCTGGAAGTTGGTGAACCGAGGCTGTCCGACCTTT
         S C R S I L A E A T F N H L A P T G W K

1501   GCGATGAGTGCGGGAAGTCAGCCGACAGGGACGGTCCATCCGCGTTCCCTGTCGCTCCTG
             CGCTACTCACGCCCTTCAGTCGGCTGTCCCTGCCAGGTAGGCGCAAGGGACAGCGAGGAC
         A M S A G S Q P T G T V H P R S L S L L

1561   ACACATGAAGGTATCGACACGGGCGGACTGCACAGCAAATCATGGGACAACCTGCCGCTG
             TGTGTACTTCCATAGCTGTGCCCGCCTGACGTGTCGTTTAGTACCCTGTTGGACGGCGAC
         T H E G I D T G G L H S K S W D N L P L

1621   GCCCCGGATGTGGTCATCACTGTGTGCGCAAATGCTGCCGGGGAAACCTGCCCGGCTTAT
             CGGGGCCTACACCAGTAGTGACACACGCGTTTACGACGGCCCCTTTGGACGGGCCGAATA
         A P D V V I T V C A N A A G E T C P A Y
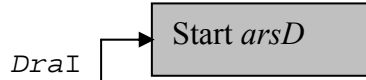
1681   CTTGGTCCGGTACTGCGTGCGCATTGGGGTGTTGACGATCCCGCTCATGCCACAGGGACC
             GAACCAGGCCATGACGCACGCGTAACCCCACAACTGCTAGGGCGAGTACGGTGTCCCTGG
         L G P V L R A H W G V D D P A H A T G T

1741   GAAGCGGAAATCGAAGCCGCGTTCCCGACGGCCTACCGTATCTTGCGGCGTCGCATCGAA
             CTTCGCCTTTAGCTTCGGCGCAAGGGCTGCCGGATGGCATAGAACGCCGCAGCGTAGCTT
         E A E I E A A F P T A Y R I L R R R I E

1801   GCCCTCCTCGCGCTGCCGTTGGCGCAACTTGCGCACGACCGGACACGCTTGAAGGCGGAA
             CGGGAGGAGCGCGACGGCAACCGCGTTGAACGCGTGCTGGCCTGTGCGAACTTCCGCCTT
         A L L A L P L A Q L A H D R T R L K A E

1861   CTGGATCGGATTGGGGCGCTTTCGGCG*TAG*TTTTTTTGACTGATATTTGAATACTTGATG
             GACCTAGCCTAACCCCGCGAAAGCCGC*ATC*AAAAAAACTGACTATAAACTTATGAACTAC
         L D R I G A L S A *
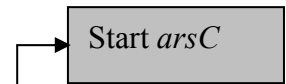
```
1921   GAGATTTTAAATGAAGAAGATCGAAGTGTTTGACCCATCGCTGTGTTGCAGCACCGGCGT
       CTCTAAAATTTACTTCTTCTAGCTTCACAAACTGGGTAGCGACACAACGTCGTGGCCGCA
             M   K   K   I   E   V   F   D   P   S   L   C   C   S   T   G   V


1981   CTGCGGCGTGGATGTTGACCAAGCCTTGGTGACTTTTGCTGCTGATGTCGATTGGGCGAA
       GACGCCGCACCTACAACTGGTTCGGAACCACTGAAAACGACGACTACAGCTAACCCGCTT
         C   G   V   D   V   D   Q   A   L   V   T   F   A   A   D   V   D   W   A   K


2041   GCAGAACGGCGCGCACATCGAGCGCTACAACCTGGCGCAGCAGCCACAGATGTTTGCCGA
       CGTCTTGCCGCGCGTGTAGCTCGCGATGTTGGACCGCGTCGTCGGTGTCTACAAACGGCT
         Q   N   G   A   H   I   E   R   Y   N   L   A   Q   Q   P   Q   M   F   A   D


2101   CAATGCGACCGTAAAGGGATTTTTGCAGCGCTCCGGTCAGGATGCGCTGCCACTGATCCT
       GTTACGCTGGCATTTCCCTAAAAACGTCGCGAGGCCAGTCCTACGCGACGGTGACTAGGA
         N   A   T   V   K   G   F   L   Q   R   S   G   Q   D   A   L   P   L   I   L


2161   GGTCGATGGTGAAGTGGCCCTCGCGGGGCGTTATCCAAAGCGGGCTGAATTGGCCCTCTG
       CCAGCTACCACTTCACCGGGAGCGCCCCGCAATAGGTTTCGCCCGACTTAACCGGGAGAC
         V   D   G   E   V   A   L   A   G   R   Y   P   K   R   A   E   L   A   L   W


2221   GATCGGCATTGATCAACCGGCAGACGCCGCCAAACCGACGACGGGCGGTTGCTGCTCCGG
       CTAGCCGTAACTAGTTGGCCGTCTGCGGCGGTTTGGCTGCTGCCCGCCAACGACGAGGCC
         I   G   I   D   Q   P   A   D   A   A   K   P   T   T   G   G   C   C   S   G
```

Start *arsA*

```
2281   CCCGCGTGGCTGCTGCTGAACAACATGGAGAAAATGATGATGATATTCCTGCAACTTCCT
       GGGCGCACCGACGACGACTTGTTGTACCTCTTTTTACTACTACTATAAGGACGTTGAAGGA
         P   R   G   C   C   *       M   E   K   M   M   M   I   F   L   Q   L   P


2341   CCCCGCTTTCTGTTTTTTCACGGGCAAGGGCGGCGTCGGCAAAACCTCGATTGCCTGTGCC
       GGGGCGAAAGACAAAAAGTGCCCGTTCCCGCCGCAGCCGTTTTGGAGCTAACGGACACGG
         P   R   F   L   F   F   T   G   K   G   G   V   G   K   T   S   I   A   C   A


2401   ACGGCTATTCAATTGGCCGAGGCCGGAAAACGCGTCCTCCTGGTCAGTACCGACCCAGCA
       TGCCGATAAGTTAACCGGCTCCGGCCTTTTGCGCAGGAGGACCAGTCATGGCTGGGTCGT
         T   A   I   Q   L   A   E   A   G   K   R   V   L   L   V   S   T   D   P   A
```

*Eco*RV

```
2461   TCCAACGTCGGGCAGGTATTTGGTGTTGATATCGGTAATCGCGTCACACCGATTCCGGCG
       AGGTTGCAGCCCGTCCATAAACCACAACTATAGCCATTAGCGCAGTGTGGCTAAGGCCGC
         S   N   V   G   Q   V   F   G   V   D   I   G   N   R   V   T   P   I   P   A
```

```
2521   GTTCCACGTCTTTCTGCCTTGGAAATTGATCCCGAGGCAGCGGCCAGTGCCTATCGGGAG
       CAAGGTGCAGAAAGACGGAACCTTTAACTAGGGCTCCGTCGCCGGTCACGGATAGCCCTC
        V  P  R  L  S  A  L  E  I  D  P  E  A  A  A  S  A  Y  R  E


2581   CGCCTGGTCGGCCCTGTGCGCGGTGTGCTTCCTGATGACGTGGTGAAGGGCATCGAAGAA
       GCGGACCAGCCGGGACACGCGCCACACGAAGGACTACTGCACCACTTCCCGTAGCTTCTT
        R  L  V  G  P  V  R  G  V  L  P  D  D  V  V  K  G  I  E  E


2641   TCGTTGTCCGGCGCGTGTACCACCGAAATTGCCGCATTTGACGAGTTCACCGCGCTACTA
       AGCAACAGGCCGCGCACATGGTGGCTTTAACGGCGTAAACTGCTCAAGTGGCGCGATGAT
        S  L  S  G  A  C  T  T  E  I  A  A  F  D  E  F  T  A  L  L


2701   ACCAACGTGGCACTCACGGCTGATTACGAGCACATCATCTTTGATACTGCGCCCACCGGC
       TGGTTGCACCGTGAGTGCCGACTAATGCTCGTGTAGTAGAAACTATGACGCGGGTGGCCG
        T  N  V  A  L  T  A  D  Y  E  H  I  I  F  D  T  A  P  T  G


2761   CACACCATCCGCTTGCTGCAACTGCCGGGCGCATGGAGCGGTTTCCTGGAAGCTGGCAAG
       GTGTGGTAGGCGAACGACGTTGACGGCCCGCGTACCTCGCCAAAGGACCTTCGACCGTTC
        H  T  I  R  L  L  Q  L  P  G  A  W  S  G  F  L  E  A  G  K


2821   GGCGATGCCTCGTGCCTCGGCCCGCTGGCCGGTCTGGAAAAGCAGCGTACTCAGTACAAG
       CCGCTACGGAGCACGGAGCCGGGCGACCGGCCAGACCTTTTCGTCGCATGAGTCATGTTC
        G  D  A  S  C  L  G  P  L  A  G  L  E  K  Q  R  T  Q  Y  K


2881   GCGGCTGTTGAAGCCTTGGCTGATCCGCTGCAAACCCGTCTGGTGCTGGTCGCTCGCGCC
       CGCCGACAACTTCGGAACCGACTAGGCGACGTTTGGGCAGACCACGACCAGCGAGCGCGG
        A  A  V  E  A  L  A  D  P  L  Q  T  R  L  V  L  V  A  R  A


                                                                 StuI
2941   CAGCAGGCGACTTTGCGCGAGGTAGCCCGAACCCACGAAGAACTGGCCGCCATAGGCCTC
       GTCGTCCGCTGAAACGCGCTCCATCGGGCTTGGGTGCTTCTTGACCGGCGGTATCCGGAG
        Q  Q  A  T  L  R  E  V  A  R  T  H  E  E  L  A  A  I  G  L


3001   AAACAGCAACATCTCGTCATCAACGGCATCCTGCCGCACATCGAAGCCGCCACCGACCCA
       TTTGTCGTTGTAGAGCAGTAGTTGCCGTAGGACGGCGTGTAGCTTCGGCGGTGGCTGGGT
        K  Q  Q  H  L  V  I  N  G  I  L  P  H  I  E  A  A  T  D  P


3061   CTGGCCGCAGCAATCCACGAACGGGAACAAACGGCGTTGAAGAACATCCCGGCTACGTTA
       GACCGGCGTCGTTAGGTGCTTGCCCTTGTTTGCCGCAACTTCTTGTAGGGCCGATGCAAT
        L  A  A  A  I  H  E  R  E  Q  T  A  L  K  N  I  P  A  T  L


3121   ACTGCGCTCCCGTGTGATCATGTTGAACTCAAGCCCTTCAATCTCGTCGGTCTCGACGCG
       TGACGCGAGGGCACACTAGTACAACTTGAGTTCGGGAAGTTAGAGCAGCCAGAGCTGCGC
        T  A  L  P  C  D  H  V  E  L  K  P  F  N  L  V  G  L  D  A
```

91

```
3181   CTGAGGCAGTTGCTGACCGACCTTCCACCACAAGCACCTGTCGCCGTTGACTCCCCGATC
       GACTCCGTCAACGACTGGCTGGAAGGTGGTGTTCGTGGACAGCGGCAACTGAGGGGCTAG
        L  R  Q  L  L  T  D  L  P  P  Q  A  P  V  A  V  D  S  P  I

3241   GAACTCGACGAGCCCGGCGTGGCCGACCTGATCGACGGCATCGCGGCGGATGGACACGGG
       CTTGAGCTGCTCGGGCCGCACCGGCTGGACTAGCTGCCGTAGCGCCGCCTACCTGTGCCC
        E  L  D  E  P  G  V  A  D  L  I  D  G  I  A  A  D  G  H  G

3301   CTGGTCATGTTGATGGGCAAAGGTGGTGTAGGCAAGACGACCCTGGCCGCCGCCATCGCG
       GACCAGTACAACTACCCGTTTCCACCACATCCGTTCTGCTGGGACCGGCGGCGGTAGCGC
        L  V  M  L  M  G  K  G  G  V  G  K  T  T  L  A  A  A  I  A

3361   GTCGAACTGGCACATCGCGGCTTGCCCGTGCATCTGACGACCTCCGATCCTGCTGCCCAC
       CAGCTTGACCGTGTAGCGCCGAACGGGCACGTAGACTGCTGGAGGCTAGGACGACGGGTG
        V  E  L  A  H  R  G  L  P  V  H  L  T  T  S  D  P  A  A  H

                                                         ClaI
3421   TTGACCGACACCCTGGATTCCTCGCTTGACAATCTGACCGTGAGCCGAATCGATCCGCAT
       AACTGGCTGTGGGACCTAAGGAGCGAACTGTTAGACTGGCACTCGGCTTAGCTAGGCGTA
        L  T  D  T  L  D  S  S  L  D  N  L  T  V  S  R  I  D  P  H

       SphI
3481   GCCGAGACCGAGCGCTATCGCCAGCACGTGCTGGAAACCAAGGGCGCTCAACTCGATGCC
       CGGCTCTGGCTCGCGATAGCGGTCGTGCACGACCTTTGGTTCCCGCGAGTTGAGCTACGG
        A  E  T  E  R  Y  R  Q  H  V  L  E  T  K  G  A  Q  L  D  A

3541   GAAGGTCGCGCGCTGCTGGAAGAGGATTTGCGTTCGCCTTGCACGGAAGAGATTGCGGTC
       CTTCCAGCGCGCGACGACCTTCTCCTAAACGCAAGCGGAACGTGCCTTCTCTAACGCCAG
        E  G  R  A  L  L  E  E  D  L  R  S  P  C  T  E  E  I  A  V

3601   TTCCAGGCGTTCTCCCGAATCATTCGCGAGGCTGGGAAAAAGTTCGTCGTCATGGACACG
       AAGGTCCGCAAGAGGGCTTAGTAAGCGCTCCGACCCTTTTTCAAGCAGCAGTACCTGTGC
        F  Q  A  F  S  R  I  I  R  E  A  G  K  K  F  V  V  M  D  T

3661   GCCCCGACAGGGCACACCTTGCTTCTGCTCGACGCGACGGGTGCGTATCACCGCGAAGTG
       CGGGGCTGTCCCGTGTGGAACGAAGACGAGCTGCGCTGCCCACGCATAGTGGCGCTTCAC
        A  P  T  G  H  T  L  L  L  L  D  A  T  G  A  Y  H  R  E  V

                              SphI
3721   ACGCGGCAAATGGGCAAGACCGGCATGCACTTCACGACGCCGATGATGCAATTGCAGGAC
       TGCGCCGTTTACCCGTTCTGGCCGTACGTGAAGTGCTGCGGCTACTACGTTAACGTCCTG
        T  R  Q  M  G  K  T  G  M  H  F  T  T  P  M  M  Q  L  Q  D

3781   CCGAAGCAAACCAAGGTGCTGATCGTCACGTTGGCAGAGACGACGCCGGTACTGGAGGCC
       GGCTTCGTTTGGTTCCACGACTAGCAGTGCAACCGTCTCTGCTGCGGCCATGACCTCCGG
        P  K  Q  T  K  V  L  I  V  T  L  A  E  T  T  P  V  L  E  A
```

```
                                                    NcoI
3841   GCCAACCTGCAAGCTGATTTGCGCCGTGCCGGGATCGAGCCATGGGCCTGGATCATCAAC
       CGGTTGGACGTTCGACTAAACGCGGCACGGCCCTAGCTCGGTACCCGGACCTAGTAGTTG
        A  N  L  Q  A  D  L  R  R  A  G  I  E  P  W  A  W  I  I  N


3901   ACCAGCGTAGCGGCGGCCTCGGCCAAGTCGCCATTACTGCGTCAGCGTGCGGCCAACGAG
       TGGTCGCATCGCCGCCGGAGCCGGTTCAGCGGTAATGACGCAGTCGCACGCCGGTTGCTC
        T  S  V  A  A  A  S  A  K  S  P  L  L  R  Q  R  A  A  N  E


3961   CTACGCGAAATCAGCGCCGTGGCCAATCAGCACGCGGACCGTTACGCGGTTGTCCCGCTT
       GATGCGCTTTAGTCGCGGCACCGGTTAGTCGTGCGCCTGGCAATGCGCCAACAGGGCGAA
        L  R  E  I  S  A  V  A  N  Q  H  A  D  R  Y  A  V  V  P  L


4021   CTGAAGGAAGAACCGATCGGTGCAGATCGACTGCGTGCGCTCATCCATCCCCAAACATAG
       GACTTCCTTCTTGGCTAGCCACGTCTAGCTGACGCACGCGAGTAGGTAGGGGTTTGTATC
        L  K  E  E  P  I  G  A  D  R  L  R  A  L  I  H  P  Q  T  *
```

```
                                  ┌─────────────────┐
                              ───▶ │  Start CBS      │
                                  └─────────────────┘
4081   GAGATCGACCATGTTGATCCGGAATTTCATGACGCCCGATCCGGTAACCATTCAACCTGA
       CTCTAGCTGGTACAACTAGGCCTTAAAGTACTGCGGGCTAGGCCATTGGTAAGTTGGACT
                M  L  I  R  N  F  M  T  P  D  P  V  T  I  Q  P  E
```

```
4141   AACACCGGTTGAGGACATTGCGAGGCTGCTGCTTGCCCATCGCATCAACGGCGTTCCGGT
       TTGTGGCCAACTCCTGTAACGCTCCGACGACGAACGGGTAGCGTAGTTGCCGCAAGGCCA
        T  P  V  E  D  I  A  R  L  L  L  A  H  R  I  N  G  V  P  V
```

```
       SalI
4201   GGTCGACGGCGCTGGTCGGCTGATCGGCGTTGTGACTGCGGAGGACTTGATTCACCGGGG
       CCAGCTGCCGCGACCAGCCGACTAGCCGCAACACTGACGCCTCCTGAACTAAGTGGCCCC
        V  D  G  A  G  R  L  I  G  V  V  T  A  E  D  L  I  H  R  G
```
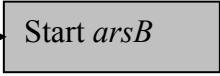
```
                                      ClaI
4261   AGCGGACGAACGACTTGAACCCCGTGAATCGATCTGGAAGGAGAACTTCTGGGTTTCCTT
       TCGCCTGCTTGCTGAACTTGGGGCACTTAGCTAGACCTTCCTCTTGAAGACCCAAAGGAA
        A  D  E  R  L  E  P  R  E  S  I  W  K  E  N  F  W  V  S  F
```

```
4321   TCTTGGCCCAAAGGGGACGCAGCGTGACAAGGCCGAGGGACGTACTGCCGCAGAGGTAAT
       AGAACCGGGTTTCCCCTGCGTCGCACTGTTCCGGCTCCCTGCATGACGGCGTCTCCATTA
        L  G  P  K  G  T  Q  R  D  K  A  E  G  R  T  A  A  E  V  M
```

```
4381   GACCACAGAAGTGCACAGCGTCACGCCAGCCATGCATCCCTCCGTTGCAGCCCGGCTGAT
       CTGGTGTCTTCACGTGTCGCAGTGCGGTCGGTACGTAGGGAGGCAACGTCGGGCCGACTA
        T  T  E  V  H  S  V  T  P  A  M  H  P  S  V  A  A  R  L  M
```

```
4441   GGTGGATCATCACTTGACGGCCCTTCCGGTAGTGGATGATGGGAAAGTGATCGGCGTCAT
       CCACCTAGTAGTGAACTGCCGGGAAGGCCATCACCTACTACCCTTTCACTAGCCGCAGTA
        V  D  H  H  L  T  A  L  P  V  V  D  D  G  K  V  I  G  V  I
```

```
4501   CTCCCGGATTGACCTCTTAAGCCTTCTTAAAGAACTCGAAAACCCGCTGAAAAGAGAGAA
       GAGGGCCTAACTGGAGAATTCGGAAGAATTTCTTGAGCTTTTGGGCGACTTTTCTCTCTT
        S   R   I   D   L   L   S   L   L   K   E   L   E   N   P   L   K   R   E   N
```


Start *arsB*

```
4561   TTGACGTGATTGCTGCCCTACCGATTTTTCTTGCCACTATCGTCCTGGTGATCTGGCAAC
       AACTGCACTAACGACGGGATGGCTAAAAGAACGGTGATAGCAGGACCACTAGACCGTTG
        *   V   I   A   A   L   P   I   F   L   A   T   I   V   L   V   I   W   Q

4621   CGCGCGGGCTGGGCATTGGCTGGAGTGCGACGCTTGGCGCTGTCGTGGCGTTACTGTCCG
       GCGCGCCCGACCCGTAACCGACCTCACGCTGCGAACCGCGACAGCACCGCAATGACAGGC
        P   R   G   L   G   I   G   W   S   A   T   L   G   A   V   V   A   L   L   S

4681   GTGTCGTTCACTTCGGTGACATTCCTGTCGTCTGGCAGATCGTCTGGAACGCCACAGCCA
       CACAGCAAGTGAAGCCACTGTAAGGACAGCAGACCGTCTAGCAGACCTTGCGGTGTCGGT
        G   V   V   H   F   G   D   I   P   V   V   W   Q   I   V   W   N   A   T   A

4741   CATTCATTGCCATCATCATCATTAGCCTGTTGCTCGATGAGGCAGGGTTTTTCGAATGGG
       GTAAGTAACGGTAGTAGTAGTAATCGGACAACGAGCTACTCCGTCCCAAAAAGCTTACCC
        T   F   I   A   I   I   I   I   S   L   L   L   D   E   A   G   F   F   E   W

4801   CAGCCTTGCACGTCGCACGCTGGGGCGGAGGAAAAGGACGTCTGCTGTTTGCACTGATCA
       GTCGGAACGTGCAGCGTGCGACCCCGCCTCCTTTTCCTGCAGACGACAAACGTGACTAGT
        A   A   L   H   V   A   R   W   G   G   G   K   G   R   L   L   F   A   L   I

4861   TCCTGCTAGGTGCAGCCGTCGCTGCGCTTTTTGCCAACGATGGCGCAGCACTGATTCTCA
       AGGACGATCCACGTCGGCAGCGACGCGAAAAACGGTTGCTACCGCGTCGTGACTAAGAGT
        I   L   L   G   A   A   V   A   A   L   F   A   N   D   G   A   A   L   I   L

4921   CGCCAATCGTCATGGCCATGTTGCTGGCACTGGGGTTCAGTCCGGCGGCAACGCTCGCCT
       GCGGTTAGCAGTACCGGTACAACGACCGTGACCCCAAGTCAGGCCGCCGTTGCGAGCGGA
        T   P   I   V   M   A   M   L   L   A   L   G   F   S   P   A   A   T   L   A

4981   TCGTCATGGCGGCGGGTTTCATCGCAGATGCCGCCAGCCTGCCGCTTGTCGTATCCAACC
       AGCAGTACCGCCGCCCAAAGTAGCGTCTACGGCGGTCGGACGGCGAACAGCATAGGTTGG
        F   V   M   A   A   G   F   I   A   D   A   A   S   L   P   L   V   V   S   N

5041   TGGTCAACATCGTCTCCGCCGACTTTTTCAATATCGGGTTCAATGACTACGCTGCGGTGA
       ACCAGTTGTAGCAGAGGCGGCTGAAAAAGTTATAGCCCAAGTTACTGATGCGACGCCACT
        L   V   N   I   V   S   A   D   F   F   N   I   G   F   N   D   Y   A   A   V

5101   TGATTCCGGTGAATATTGTCGCCATCATTGCCTCTCTGGCCGTACTCTCGATCTATTTCC
       ACTAAGGCCACTTATAACAGCGGTAGTAACGGAGAGACCGGCATGAGAGCTAGATAAAGG
        M   I   P   V   N   I   V   A   I   I   A   S   L   A   V   L   S   I   Y   F
```

```
5161    GCCGCAGCATCCCCGCACACTACGACGTGAATCAGTTGAAACAGCCGAATGAGGCCATTC
        CGGCGTCGTAGGGGCGTGTGATGCTGCACTTAGTCAACTTTGTCGGCTTACTCCGGTAAG
        R  R  S  I  P  A  H  Y  D  V  N  Q  L  K  Q  P  N  E  A  I


5221    GCGATGTGGCCACGTTCCGTTTCGGCTGGGTGGTTCTGGCCCTGCTGCTGGTCGGATTCT
        CGCTACACCGGTGCAAGGCAAAGCCGACCCACCAAGACCGGGACGACGACCAGCCTAAGA
        R  D  V  A  T  F  R  F  G  W  V  V  L  A  L  L  L  V  G  F


5281    TCGGCCTTGAGCCGCTAGGCGTACCGGTTAGCGCCGTGGCTGCCGCAGGCGCCCTACTGT
        AGCCGGAACTCGGCGATCCGCATGGCCAATCGCGGCACCGACGGCGTCCGCGGGATGACA
        F  G  L  E  P  L  G  V  P  V  S  A  V  A  A  A  G  A  L  L


5341    TGCTGGCAGTCGCTGCGCGCGGGCATGTCATCAGCACCCGCAAAGTGATCCGTGAAGCGC
        ACGACCGTCAGCGACGCGCGCCCGTACAGTAGTCGTGGGCGTTTCACTAGGCACTTCGCG
        L  L  A  V  A  A  R  G  H  V  I  S  T  R  K  V  I  R  E  A


5401    CTTGGCAAATTGTCGTTTTCTCGCTGGGAATGTATCTCGTGGTCTATGGGTTGCGTAACC
        GAACCGTTTAACAGCAAAAGAGCGACCCTTACATAGAGCACCAGATACCCAACGCATTGG
        P  W  Q  I  V  V  F  S  L  G  M  Y  L  V  V  Y  G  L  R  N


           StuI
5461    AAGGCCTGGCCGAGCATATCGCAAGGCTGCTGGATTACTTCGCACAAGGCGGTGTATGGG
        TTCCGGACCGGCTCGTATAGCGTTCCGACGACCTAATGAAGCGTGTTCCGCCACATACCC
        Q  G  L  A  E  H  I  A  R  L  L  D  Y  F  A  Q  G  G  V  W


5521    GGGCCGCCTTTGGAACAGGCTTCCTCACAGCGCTGCTGTCCTCGGCTATGAACAACATGC
        CCCGGCGGAAACCTTGTCCGAAGGAGTGTCGCGACGACAGGAGCCGATACTTGTTGTACG
        G  A  A  F  G  T  G  F  L  T  A  L  L  S  S  A  M  N  N  M


           SpeI
5581    CTACGGTACTAGTGGGCGCCCTATCCATTGATGCCACAAGCGCAACGGGCGTGGTGAAAG
        GATGCCATGATCACCCGCGGGATAGGTAACTACGGTGTTCGCGTTGCCCGCACCACTTTC
        P  T  V  L  V  G  A  L  S  I  D  A  T  S  A  T  G  V  V  K


5641    ACACGATGATCTACGCCAATGTGATTGGTAGTGATCTAGGACCGAAGATCACCCCCATCG
        TGTGCTACTAGATGCGGTTACACTAACCATCACTAGATCCTGGCTTCTAGTGGGGGTAGC
        D  T  M  I  Y  A  N  V  I  G  S  D  L  G  P  K  I  T  P  I


5701    GCTCGCTGGCGACGCTGCTTTGGTTGCATGTGCTGGCGCGCAAGGGAATGACAATCACAT
        CGAGCGACCGCTGCGACGAAACCAACGTACACGACCGCGCGTTCCCTTACTGTTAGTGTA
        G  S  L  A  T  L  L  W  L  H  V  L  A  R  K  G  M  T  I  T


5761    GGGGGTATTACTTCAAAGTGGGCGTTGTGCTGACCGTCCCCGTACTCGCGGCGACCCTGG
        CCCCCATAATGAAGTTTCACCCGCAACACGACTGGCAGGGGCATGAGCGCCGCTGGGACC
        W  G  Y  Y  F  K  V  G  V  V  L  T  V  P  V  L  A  A  T  L
```
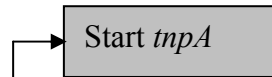
```
5821   CAGCGCTGGCACTGCGCCTGAGTCTCACTTGATGCGAGCGGAGCGCTGTTTGCGGACATG
       GTCGCGACCGTGACGCGGACTCAGAGTGAACTACGCTCGCCTCGCGACAAACGCCTGTAC
        A   A   L   A   L   R   L   S   L   T   *


5881   CAGTATTTGAGTAAAGAAGGGTGAGCGGCTTGGCGTATGGCTATTCTCTTGAGTGCGGTC
       GTCATAAACTCATTTCTTCCCACTCGCCGAACCGCATACCGATAAGAGAACTCACGCCAG


5941   ATCGGGATACCGCTCTATATCCGCGCCTTTCTTTAGTGGAACTGGATTTACTCGGCCACA
       TAGCCCTATGGCGAGATATAGGCGCGGAAAGAAATCACCTTGACCTAAATGAGCCGGTGT
```

```
                                          ┌─────────────────┐
                              ───────────▶│  Start tnpA      │
                                          └─────────────────┘
```

```
6001   GAAAGAAAAAGTGCATCCATGCCACGCCGTTCGATCCTGTCCGCCGCCGAGCGTGACAA
       CTTTCTTTTTCACGTAGGTACGGTGCGGCAAGCTAGGACAGGCGGCGGCTCGCACTGTT
                          M   P   R   R   S   I   L   S   A   A   E   R   D   N


6061   CCTGCTGGCATTGCCGGACGCCAAGGAAGAGCTGATCCGTCACTACACGTTCAGCGACTC
       GGACGACCGTAACGGCCTGCGGTTCCTTCTCGACTAGGCAGTGATGTGCAAGTCGCTGAG
         L   L   A   L   P   D   A   K   E   E   L   I   R   H   Y   T   F   S   D   S


6121   CGATCTGTCCATCATCAGGCAGAGGCGCGGCCCGGCCAACCGCCTGGGCTTCGCCGTGCA
       GCTAGACAGGTAGTAGTCCGTCTCCGCGCCGGGCCGGTTGGCGGACCCGAAGCGGCACGT
         D   L   S   I   I   R   Q   R   R   G   P   A   N   R   L   G   F   A   V   Q


6181   GCTCTGCTATCTGCGCTTTCCCGGCGTCATCCTTGGCGCCGATGAACCGCCGTTTCCGCC
       CGAGACGATAGACGCGAAAGGGCCGCAGTAGGAACCGCGGCTACTTGGCGGCAAAGGCGG
         L   C   Y   L   R   F   P   G   V   I   L   G   A   D   E   P   P   F   P   P


6241   CTTGCTGAAACTGGTCGCTGACCAGCTCAAGATTGGCATCGAAAGCTGGGGTGAATATGG
       GAACGACTTTGACCAGCGACTGGTCGAGTTCTAACCGTAGCTTTCGACCCCACTTATACC
         L   L   K   L   V   A   D   Q   L   K   I   G   I   E   S   W   G   E   Y   G


6301   GCAGCGGGGGCAGACCCGGCGCGAGCACCTAGTCGAGCTGCAAACGGTGTTCGGCTTCCA
       CGTCGCCCCCGTCTGGGCCGCGCTCGTGGATCAGCTCGACGTTTGCCACAAGCCGAAGGT
         Q   R   G   Q   T   R   R   E   H   L   V   E   L   Q   T   V   F   G   F   Q


6361   GCCGTTCACCATGAGCCACTACCGGCAGGCTGTCCAGTTGCTGACCGAGTTGGCCATGCA
       CGGCAAGTGGTACTCGGTGATGGCCGTCCGACAGGTCAACGACTGGCTCAACCGGTACGT
         P   F   T   M   S   H   Y   R   Q   A   V   Q   L   L   T   E   L   A   M   Q


6421   GACCGACAAGGGCATTGTGCTGGCCAGCGCCTTGATCGAGCATCTGCGGCGGCAGTCGGT
       CTGGCTGTTCCCGTAACACGACCGGTCGCGGAACTAGCTCGTAGACGCCGCCGTCAGCCA
         T   D   K   G   I   V   L   A   S   A   L   I   E   H   L   R   R   Q   S   V
```

96

```
6481    CATTTTGCCTGCCCTCAATGCCGTCGAGCGGGCGAGCGCCGAAGCAATCACCCGCGCCAA
        GTAAAACGGACGGGAGTTACGGCAGCTCGCCCGCTCGCGGCTTCGTTAGTGGGCGCGGTT
          I  L  P  A  L  N  A  V  E  R  A  S  A  E  A  I  T  R  A  N


6541    CTGGCGCATCTACGACGCCTTGGCCGAACCACTGTCGGACGTGCATCGCCGCCGCCTCGA
        GACCGCGTAGATGCTGCGGAACCGGCTTGGTGACAGCCTGCACGTAGCGGCGGCGGAGCT
           W  R  I  Y  D  A  L  A  E  P  L  S  D  V  H  R  R  R  L  D


6601    CGATCTGCTCAAGCGCCGGGACAACGGCAAGACGACCTGGCTGGCCTGGCTGCGCCAATC
        GCTAGACGAGTTCGCGGCCCTGTTGCCGTTCTGCTGGACCGACCGGACCGACGCGGTTAG
           D  L  L  K  R  R  D  N  G  K  T  T  W  L  A  W  L  R  Q  S


6661    ACCGGCCAAGCCGAACTCGCGGCACATGCTGGAACACATCGAACGCCTCAAGGCGTGGCA
        TGGCCGGTTCGGCTTGAGCGCCGTGTACGACCTTGTGTAGCTTGCGGAGTTCCGCACCGT
          P  A  K  P  N  S  R  H  M  L  E  H  I  E  R  L  K  A  W  Q


6721    GGCGCTCGATCTTCCTTCCGGCGTCGAGCGGTCGGTGCACCAGAACCGCCTGCTCAAGAT
        CCGCGAGCTAGAAGGAAGGCCGCAGCTCGCCAGCCACGTGGTCTTGGCGGACGAGTTCTA
          A  L  D  L  P  S  G  V  E  R  S  V  H  Q  N  R  L  L  K  I


6781    CGCCCGCGAGGGCGGCCAGATGACGCCTGCCGACCTGGCCAAGTTCGAGGCGCAGCGACG
        GCGGGCGCTCCCGCCGGTCTACTGCGGACGGCTGGACCGGTTCAAGCTCCGCGTCGCTGC
          A  R  E  G  G  Q  M  T  P  A  D  L  A  K  F  E  A  Q  R  R


6841    CTATGCCACCCTGGTGGCGCTGGCCATCGAGGGCATGGCGACCGTCACCGACGAAATCAT
        GATACGGTGGGACCACCGCGACCGGTAGCTCCCGTACCGCTGGCAGTGGCTGCTTTAGTA
          Y  A  T  L  V  A  L  A  I  E  G  M  A  T  V  T  D  E  I  I


6901    CGACTTGCACGACCGCATCCTGGGCAAGCTGTTCAACGCCGCCAAGAACAAGCATCAGCA
        GCTGAACGTGCTGGCGTAGGACCCGTTCGACAAGTTGCGGCGGTTCTTGTTCGTAGTCGT
           D  L  H  D  R  I  L  G  K  L  F  N  A  A  K  N  K  H  Q  Q


6961    GCAGTTCCAGGCATCCGGCAAGGCGATCAACGCCAAGGTGCGGCTGTATGGACGCATCGG
        CGTCAAGGTCCGTAGGCCGTTCCGCTAGTTGCGGTTCCACGCCGACATACCTGCGTAGCC
          Q  F  Q  A  S  G  K  A  I  N  A  K  V  R  L  Y  G  R  I  G


7021    CCAGGCGCTGATCGACGCCAAGCAGTCGGGCCGTGATCCGTTCGCCGCTATCGAGGCCGT
        GGTCCGCGACTAGCTGCGGTTCGTCAGCCCGGCACTAGGCAAGCGGCGATAGCTCCGGCA
          Q  A  L  I  D  A  K  Q  S  G  R  D  P  F  A  A  I  E  A  V


7081    CATGTCTTGGGATGCCTTCGCCGAAAGCGTCACCGAGGCGCAAAAGCTCGCGCAACCCGA
        GTACAGAACCCTACGGAAGCGGCTTTCGCAGTGGCTCCGCGTTTTCGAGCGCGTTGGGCT
          M  S  W  D  A  F  A  E  S  V  T  E  A  Q  K  L  A  Q  P  D
```

```
7141   TGACTTCGATTTTTTGCACCGCATCGGTGAGAGCTACGCCACCCTGCGGGGCTATGCACC
       ACTGAAGCTAAAAAACGTGGCGTAGCCACTCTCGATGCGGTGGGACGCCCCGATACGTGG
        D  F  D  F  L  H  R  I  G  E  S  Y  A  T  L  R  G  Y  A  P


       EcoRI
7201   GGGAATTCCTTGCCGTGCTCAAGCTGCGGGCCGCGCCCGCCGCCAAAAACGTGCTGGATGC
       CCTTAAGGAACGGCACGAGTTCGACGCCCGGCGCGGGCGGCGGTTTTTGCACGACCTACG
        E  F  L  A  V  L  K  L  R  A  A  P  A  A  K  N  V  L  D  A


7261   CATTGAGGTACTGCGCGGTATGAACACCGACAACGCCCGCAAGGTGCCCGCCGATGCGCC
       GTAACTCCATGACGCGCCATACTTGTGGCTGTTGCGGGCGTTCCACGGGCGGCTACGCGG
        I  E  V  L  R  G  M  N  T  D  N  A  R  K  V  P  A  D  A  P


7321   GACCGGCTTCATCAAACCGCGCTGGCAGAAGCTGGTGATGACCGACACCGGAATCGACCG
       CTGGCCGAAGTAGTTTGGCGCGACCGTCTTCGACCACTACTGGCTGTGGCCTTAGCTGGC
        T  G  F  I  K  P  R  W  Q  K  L  V  M  T  D  T  G  I  D  R


7381   GCGCTACTACGAACTGTGCGCCCTGTCGGAACTCAAGAACTCCCTGCGCTCGGGCGACAT
       CGCGATGATGCTTGACACGCGGGACAGCCTTGAGTTCTTGAGGGACGCGAGCCCGCTGTA
        R  Y  Y  E  L  C  A  L  S  E  L  K  N  S  L  R  S  G  D  I


7441   CTGGGTGCAGGGTTCGCGCCAGTTCAAGGACTTCGAGGACTACCTGGTGCCGCCCGCGAA
       GACCCACGTCCCAAGCGCGGTCAAGTTCCTGAAGCTCCTGATGGACCACGGCGGGCGCTT
        W  V  Q  G  S  R  Q  F  K  D  F  E  D  Y  L  V  P  P  A  K


7501   GTTCGCCAGCCTCAAGCAGTCCAGCGAATTACCGCTGGCCGTGGCCACCGACTGCGACCA
       CAAGCGGTCGGAGTTCGTCAGGTCGCTTAATGGCGACCGGCACCGGTGGCTGACGCTGGT
        F  A  S  L  K  Q  S  S  E  L  P  L  A  V  A  T  D  C  D  Q


7561   GTACCTGAGCGAGCGCTTGGAGCTGCTGGAAGCGCAGCTCGCCACGGTCAACCGCATGGC
       CATGGACTCGCTCGCGAACCTCGACGACCTTCGCGTCGAGCGGTGCCAGTTGGCGTACCG
        Y  L  S  E  R  L  E  L  L  E  A  Q  L  A  T  V  N  R  M  A


7621   GGCGGCGAACGATCTGCCGGACGCCATCATCACAGAGTCGGGCCTGAAAATCACGCCGCT
       CCGCCGCTTGCTAGACGGCCTGCGGTAGTAGTGTCTCAGCCCGGACTTTTAGTGCGGCGA
        A  A  N  D  L  P  D  A  I  I  T  E  S  G  L  K  I  T  P  L


7681   CGATGCGGCGGTGCCGGACACTGCGCAAGCGTTGATCGACCAGACGGCCATGATCCTGCC
       GCTACGCCGCCACGGCCTGTGACGCGTTCGCAACTAGCTGGTCTGCCGGTACTAGGACGG
        D  A  A  V  P  D  T  A  Q  A  L  I  D  Q  T  A  M  I  L  P


7741   GCACGTCAAGATCACAGAATTGCTGCTTGAAGTCGATGAATGGACGGGCTTCACCCGCCA
       CGTGCAGTTCTAGTGTCTTAACGACGAACTTCAGCTACTTACCTGCCCGAAGTGGGCGGT
        H  V  K  I  T  E  L  L  L  E  V  D  E  W  T  G  F  T  R  H
```

```
7801    CTTCACGCACCTGAAATCGGGCGATCTGGCCAAGGACAAGAACCTGTTGCTGACCACGAT
        GAAGTGCGTGGACTTTAGCCCGCTAGACCGGTTCCTGTTCTTGGACAACGACTGGTGCTA
         F  T  H  L  K  S  G  D  L  A  K  D  K  N  L  L  L  T  T  I

7861    CCTGGCCGACGCGATCAACCTGGGGCTGACGAAAATGGCCGAGTCCTGCCCCGGAACGAC
        GGACCGGCTGCGCTAGTTGGACCCCGACTGCTTTTACCGGCTCAGGACGGGGCCTTGCTG
         L  A  D  A  I  N  L  G  L  T  K  M  A  E  S  C  P  G  T  T

7921    CTACGCCAAGCTCGCCTGGCTGCAAGCCTGGAATACCCGAGACGAAACCTATTCGACGGC
        GATGCGGTTCGAGCGGACCGACGTTCGGACCTTATGGGCTCTGCTTTGGATAAGCTGCCG
         Y  A  K  L  A  W  L  Q  A  W  N  T  R  D  E  T  Y  S  T  A

7981    GCTGGCCGAGTTGGTCAACGCCCAATTCCGGCACCCCTTCGCCGGGCATTGGGGCGACGG
        CGACCGGCTCAACCAGTTGCGGGTTAAGGCCGTGGGGAAGCGGCCCGTAACCCCGCTGCC
         L  A  E  L  V  N  A  Q  F  R  H  P  F  A  G  H  W  G  D  G

8041    CACCACGTCATCGTCGGACGGCCAGAACTTCCGCACCGGCAGCAAGGCCGAGAGTACCGG
        GTGGTGCAGTAGCAGCCTGCCGGTCTTGAAGGCGTGGCCGTCGTTCCGGCTCTCATGGCC
         T  T  S  S  S  D  G  Q  N  F  R  T  G  S  K  A  E  S  T  G

8101    GCACATCAACCCGAAATACGGCAGCAGCCCCAGGACGGACTTTCTACACCCACATTTCCGA
        CGTGTAGTTGGGCTTTATGCCGTCGTCGGGTCCTGCCTGAAAGATGTGGGTGTAAAGGCT
         H  I  N  P  K  Y  G  S  S  P  G  R  T  F  Y  T  H  I  S  D

8161    CCAGTACGCGCCATTCCACACCAAAGTGGTCAATGTCGGCGTGCGCGACTCGACCTATGT
        GGTCATGCGCGGTAAGGTGTGGTTTCACCAGTTACAGCCGCACGCGCTGAGCTGGATACA
         Q  Y  A  P  F  H  T  K  V  V  N  V  G  V  R  D  S  T  Y  V

8221    GCTCGACGGCCTGCTGTACCACGAGTCCGACTTGCGGATCGAGGAGCACTACACCGACAC
        CGAGCTGCCGGACGACATGGTGCTCAGGCTGAACGCCTAGCTCCTCGTGATGTGGCTGTG
         L  D  G  L  L  Y  H  E  S  D  L  R  I  E  E  H  Y  T  D  T

8281    GGCGGGCTTCACCGATCACGTCTTTGCGCTGATGCACCTGCTGGGCTTTCGTTTCGCGCC
        CCGCCCGAAGTGGCTAGTGCAGAAACGCGACTACGTGGACGACCCGAAAGCAAAGCGCGG
         A  G  F  T  D  H  V  F  A  L  M  H  L  L  G  F  R  F  A  P

8341    GCGCATCCGCGACCTGGGCGACACCAAGCTCTACATTCCCAAGGGCGAAACCGCCTATGA
        CGCGTAGGCGCTGGACCCGCTGTGGTTCGAGATGTAAGGGTTCCCGCTTTGGCGGATACT
         R  I  R  D  L  G  D  T  K  L  Y  I  P  K  G  E  T  A  Y  D

8401    CGCCCTCAAGCCGATGATCGGCGGCACGCTCAACATCAAGCATGTACGCGCCCATTGGGA
        GCGGGAGTTCGGCTACTAGCCGCCGTGCGAGTTGTAGTTCGTACATGCGCGGGTAACCCT
         A  L  K  P  M  I  G  G  T  L  N  I  K  H  V  R  A  H  W  D
```

```
8461   CGAAATCCTGCGGCTGGCCACCTCGATCAAGCAGGGCACCGTGACTGCCTCGCTGATGCT
       GCTTTAGGACGCCGACCGGTGGAGCTAGTTCGTCCCGTGGCACTGACGGAGCGACTACGA
         E   I   L   R   L   A   T   S   I   K   Q   G   T   V   T   A   S   L   M   L


8521   GCGCAAGCTCGGCAGCTACCCACGCCAGAACGGCCTGGCCGTGGCCCTGCGCGAGCTGGG
       CGCGTTCGAGCCGTCGATGGGTGCGGTCTTGCCGGACCGGCACCGGGACGCGCTCGACCC
        R   K   L   G   S   Y   P   R   Q   N   G   L   A   V   A   L   R   E   L   G


8581   CCGCATCGAGCGCACATTGTTCATCCTTGACTGGCTACAAAGCGTGGAACTGCGCCGCCG
       GGCGTAGCTCGCGTGTAACAAGTAGGAACTGACCGATGTTTCGCACCTTGACGCGGCGGC
         R   I   E   R   T   L   F   I   L   D   W   L   Q   S   V   E   L   R   R   R


                   SphI
8641   CGTGCATGCCGGGTTGAACAAGGGCGAAGCCCGCAACGCGCTCGCCAGGGCGGTGTTCTT
       GCACGTACGGCCCAACTTGTTCCCGCTTCGGGCGTTGCGCGAGCGGTCCCGCCACAAGAA
         V   H   A   G   L   N   K   G   E   A   R   N   A   L   A   R   A   V   F   F


8701   CAACCGGCTGGGCGAAATCCGCGACCGCAGTTTCGAGCAGCAGCGTTACCGGGCCAGCGG
       GTTGGCCGACCCGCTTTAGGCGCTGGCGTCAAAGCTCGTCGTCGCAATGGCCCGGTCGCC
         N   R   L   G   E   I   R   D   R   S   F   E   Q   Q   R   Y   R   A   S   G


8761   CCTCAACCTGGTGACGGCAGCCATCGTGCTGTGGAACACAGTTTATCTGGAGCGAGCCGC
       GGAGTTGGACCACTGCCGTCGGTAGCACGACACCTTGTGTCAAATAGACCTCGCTCGGCG
         L   N   L   V   T   A   A   I   V   L   W   N   T   V   Y   L   E   R   A   A


8821   GAATGCCCTGCGTGACCACGGCAAGCCCGTTGATGACTCTCTGTTGCAGTATCTGTCGCC
       CTTACGGGACGCACTGGTGCCGTTCGGGCAACTACTGAGAGACAACGTCATAGACAGCGG
         N   A   L   R   D   H   G   K   P   V   D   D   S   L   L   Q   Y   L   S   P


8881   GCTGGGCTGGGAGCACATCAACCTGACCGGCGATTACCTCTGGCGCAGCAGCGCCAAGAT
       CGACCCGACCCTCGTGTAGTTGGACTGGCCGCTAATGGAGACCGCGTCGTCGCGGTTCTA
         L   G   W   E   H   I   N   L   T   G   D   Y   L   W   R   S   S   A   K   I

                                                          Right inverted repeat
8941   CGGCGCGGGCAAGTTCAGGCCGCTACGGCCGCTGCAACCGGCTTAGCGTGCTTTATTCTC
       GCCGCGCCCGTTCAAGTCCGGCGATGCCGGCGACGTTGGCCGAATCGCACGAAATAAGAG
         G   A   G   K   F   R   P   L   R   P   L   Q   P   A   *


9001   CGTTTTCTGAGACGACCCCTGAAAGAGGCTGGGTGCGGTGAGTTCCATTGGGCCGGGATT
       GCAAAAGACTCTGCTGGGGACTTTCTCCGACCCACGCCACTCAAGGTAACCCGGCCCTAA


9061   TCCTTATCCGTTTTACCAACCACATCTCCAGGGAATGATTCACGG
       AGGAATAGGCAAAATGGTTGGTGTAGAGGTCCCTTACTAAGTGCC
```

REFERENCES

1. **Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**: 3389-3402.

2. **Bairoch, A.** (1993). A possibility for the mechanism for the metal-ion induced DNA-protein dissociation in a family of prokaryotic transcriptional regulators. *Nucleic Acids Res*. **21**: 2515.

3. **Bateman, A.** (1997). The structure of a domain common to archaebacteria and homocystinuria disease protein. *Trends Biochem. Sci.* **22**: 12-13.

4. **Battaglia, F.D., D. Morin, J-L. Garcia and P. Olivier.** (1994). Isolation and study of *Leptospirillum*-like bacteria from a natural mixed population cultured on a coralliferous pyrite substrate. *Antonie von Leeuwenhoek* **66**: 295-302.

5. **Bennett, M.S., Z. Guan, M. Laurberg and X.D. Su.** (2001). *Bacillus subtilis* arsenate reductase is structurally similar to low molecular weight protein tyrosine phosphatase. *Proc. Natl. Acad. Sci USA* **98**: 13577-13582.

6. **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.

7. **Bhattacharjee, H. and B.P. Rosen.** (2001). Structure and functional analysis of ArsA ATPase: contribution of histidine residues. *Bioenerget. Biomem. J.* **33(6)**: 459-468.

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

9. **Blake, II, R.C. and Shute.** (1997). Purification and characterization of a novel cytochrome from *Leptospirillum ferrooxidans*, pp. PB 31-PB 310. In International Biohydrometallurgy Symposium, IBS.97. Glenside, South Aust.: Aust. Mineral. Found.

10. **Bond, P.L., G.K. Druschell and J.F. Banfield.** (2000b). Comparison of acid mine drainage communities in physically and chemically distinct ecosystems. *Appl. Environ. Microbiol.* **66**: 4962- 4971.

11. **Bond, P.L., S.P. Smirga and J.F. Banfield.** (2000a). Phylogeny of microorganisms populating a thick, subaerial predominantly lithotrophic biofilm at an extreme acid mine drainage site. *App. Environ. Mirobiol.* **65**:3627- 3632.

12. **Brown, H.J., H.W. Stokes, and R.M. Hall.** (1996). The integrons InO, In2 and In5 are defective transposition derivatives. *J. Bacteriol.* **178**: 4429- 4437.

13. **Bruhn, D.F., J. Li, S. Silver and B.P. Rosen.** (1986). The arsenical resistance operon of IncN plasmid R46. *FEMS. Microbiol. Lett.* **261**: 15030- 15038.

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

15. **Butcher, Butcher, B.G.** (2003). Arsenic resistance of the biomining bacterium *Acidithiobacillus ferrooxidans*. Ph.d Thesis. University of Stellenbosch.

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

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

18. **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.

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

20. **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.

21. **Cervantes, C., G. Ji, J.L. Ramirez and S. Silver.** (1994). Resistance to arsenical compounds in microorganisms. *FEMS Microbiol. Rev.* **15**: 335- 367.

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

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

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

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

26. **Collis, C.M., G. Grammaticopoulos, J. Briton, H.W. Stokes and R.M. Hall.** (1993). Site specific insertion of the gene cassettes into integrons. *Mol. Microbiol.* **9**: 41-52

27. **Coram, N.U. and D.E. Rawlings.** (2002). Molecular relationship between two groups of the genus *Leptospirillum* and the finding that *Leptospirillum ferriphilum* sp. nov. dominates South African commercial biooxidation tanks that operate at 40°C. *Appl. Environ. Microbiol.* **68**: 838- 845.

28. **Davies, J.** (1994). Inactivation of antibiotics and the dissemination of resistance genes. *Science* **264**: 375-382.

29. **Davies, J.E. and R. Rownd.** (1972). Transmissible multidrug resistance in *Enterobacteriaceae. Science* **176**: 758- 768.

30. **De Groot, P., S.M. Deane and D.E. Rawlings.** (2004). A transposon-arsenic resistance mechanism from a strain of *Acidithiobacillus caldus* isolated from commercial, arsenopyrite biooxidation tanks. *Hydrometallurgy* **71**: 115-123..

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

32. **Dey, S. and B.P. Rosen.** (1995). Dual mode energy coupling by the oxyanion-translocating ArsB protein. *Biochem. J.* **178**: 195-200.

33. **Dou D., J.B. Owolabi, S. Dey and B.P. Rosen** (1994). ATP-dependent arsenite transport in everted membrane vesicles of *Escherichia coli. J. Biol. Chem.* **177**: 2050-2056.

34. **Diorio, C., J. Cai, R. Schinder 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.

35. **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.

36.	**Edwards, K.J., T.M. Gihring and J.F. Banfield.** (1999). Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Appl. Environ. Microbiol.* **65**: 3627- 3632.

37.	**Erlich, H.L.** (1995). *Geomicrobiology* 3rd edition. New York: Marcel Dekker.

38.	**Gardener, M.N. and D.E. Rawlings.** (2000). Production of rhodanese by bacteria present in biooxidation plants used to recover gold from arsenopyrite ores. *J. Appl. Microbiol.* **89**: 185- 190.

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

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

41.	**Gladysheva, T., K.L. Oden and B.P. Rosen.** (1994). Properties of arsenate reductases plasmid R773. *Biochemistry* **33**: 7288- 7293.

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

43.	**Goebel, B.M. and E. Stackebrandt.** (1995). Molecular analysis of microbial diversity in a natuaral acidic environment. P43-52. *In* C.A. Jerez, T. Vargas, H. Toledo and J.V. Weirtz (eds.) *Biohydrometallurgical Processing*, Volume II. *University of Santiago Press*. Santiago.

44.	**Golovacheva, R.S., O.V. Golyshina, G.L. Karavaiko, A.G. Dorofeev, T.A. Pivovarova and N.A. Chernykh.** (1993). A new iron-oxidizing bacterium, *Leptospirillum thermoferrooxidans* sp. nov. *Microbiology* (English translation *Microbiologiya*) **61**: 744-750.

45.	**Grinsted, J., F. de la Cruz and R. Schmitt.** (1990). The Tn*21* subgroup of bacterial transposable elements. *Plasmid* **24**: 163-189.

46.	**Hallmann, R., A. Friedrich, H.P. Koops, A. Pommerening-Röser, K. Rohde, C. Zenneck and W. Sand.** (1992). Physiological characteristics of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans* and physiochemical factors influence in microbial metal leaching. *Geomicrobiol. J.* **10**: 193-206.

47. **Harrison, A.P., Jr. and P.R. Norris.** (1985). *Leptospirillum ferrooxidans* and similar bacteria: some characteristics and genomic diversity. *FEMS Mircobiol. Lett.* **30**: 99-102.

48. **Hippe, H.** (2000). *Leptospirillum* gen. nov. (ex Markosyan 1972), nom. rev., including *Leptospirillum ferrooxidans* sp. nov. (ex Markosyan 1972), nom. rev. and *Leptospirillum thermoferrooxidans* sp. nov. (Golovacheva *et al*. 1992). *Int. J. Sys. Evol. Microbiol.* **50**: 501-503.

49. **Hyde, D.R. and C.P.D. Tu.** (1985). *tnpM*: a novel regulatory gene that enhances Tn*21* transposition and suppresses cointegrate resolution. *Cell* **42**: 629-638.

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

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

52. **Ji, G. and S. Silver.** (1992b). 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.

53. **Jia, H and P. Kuar.** (2001). The role of the linker region of the anion-stimulated ATPase ArsA. *J. Biol. Chem*. **276**: 29582- 29587.

54. **Johnson, D.B.** (1995). Selective solid media for isolating and enumerating acidophilic bacteria. *J. Microbiol. Meth*. **23**: 205-218.

55. **Johnson, D.B.** (1998). Biodiversity and ecology of acidophilic organisms. *FEMS Microbiol. Ecol.* **27**: 307-317.

56. **Johnson, D.B.** (2001). Genus II *Leptospirillum* Hippe 2000 (ex Markosyan 1972, 26). In G. Garitty (ed). *Bergey's Manual of Comparitive Bacteriology*, Vol. I, pp 443- 447.

57. **Kruse, H. and H. Sorum.** (1994). Transfer of multiple drug resistance between bacteria of diverse origins in natural microenvironments. *Appl. Environ. Microbiol.* **60**: 4015- 4021

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

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

60. **Lane, D.J., A.P. Harrison, Jr., D. Stahl, B. Pace, S.J. Giovannoni, G.J. Olsen and N.R. Pace.** (1992). Evolutionary relationships among sulfur- and iron-oxidizing eubacteria. *J. Bacteriol.* **174**: 269- 278.

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

62. **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.

63. **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.

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

65. **Liebert, C., R.M. Hall and A.O. Summers.** (1999). Transposon Tn*21*, flagship of the floating genome. *Microbiol. Mol. Biol. Rev.* **63**: 507- 522.

66. **Liu, 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.

67. **Liu, Z., J. Shen, J.M. Carbrey, R. Mukhopadhyay, P. Agre and B.P. Rosen.** (2002). Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proc. Natl. Acad. Sci. USA*. **99**: 6053-6058.

68. **Lundgren, D.G. and M. Silver.** (1980). Ore leaching bacteria. *Annu. Rev. Microbiol.* **34**: 263-283.

69. **Markosyan, G.E.** (1972). A new iron-oxidizing bacterium *Leptospirillum ferrooxidans* nov. gen. nov. sp. *Biol. J. Armenia* **25**: 26-29. (in Russian).

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

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

72. **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.

73. **Messens, J., J. Martins, E. Brosens, K. Van Belle, D.M. Jacobs, R. Willem and L. Wyns.** (2002). Kinetics and active site dynamics of *Staphylococcus aureus* arsenate reductase. *J. Biol. Chem.* **7**: 146- 156.

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

75. **Mukhopadyay, 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.

76. **Nakaya, R., A. Nakamura and Y. Murata.** (1960). Resistance transfer agents in *Shigella*. *Biochem. Biophys. Res. Commun.* **3**: 654-659.

77. **Neyfakh, A.A., V.E. Bidnenko and L.B. Chen.** (1991). Efflux-mediated multidrug resistance in *Bacillus subtilis*: similarities and dissimilarities with the mammalian system. *Proc. Natl. Acad. Sci. USA.* **88**: 4781- 4785.

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

79. **NiDhubhghail, O.M. and P.J. Sadler.** (1991). The structure and reactivity of arsenic compounds: biological activity and drug design. *Struct. Bond.* **78**: 129-190.

80. **Norris, P.R.** (1983). Iron and mineral oxidation with *Leptospirillum*-like bacteria, p. 83- 96. In G. Rossi and A.E. Torma (eds.), *Recent Progress in Biohydrometallurgy*. Iglesias, Italy: Associazione Mineraria Sarda.

81. **Norris, P.R., D.W. Barr and D. Hinson.** (1988). Iron and mineral oxidation by acidophilic bacteria: affinities for iron and attachment to pyrite. p. 43- 59. In P.R. Norris and D.P. Kelly (eds.), *Biohydrometallurgy, Proceedings of the International Symposium*, Science and Technology Letters: Kew (UK).

82. **Norris, P.R., J.C. Murrell and D. Hinson.** (1995). The potential for diazotrophy in iron- and sulfur-oxidizing acidophilic bacteria. *Arch. Microbiol.* **164**: 294- 300.

83. **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.

84. **Parro, V. and M. Moreno-Paz.** (2003). Gene function analysis in environmental isolates: The *nif* regulon of the strict iron-oxidizing bacterium *Leptospirillum ferrooxidans*. *PNAs* **100**: 7883- 7888.

85. **Pearson, A.J., K.D. Bruce, A.M. Osborn, D.A. Ritchie and P. Strike.** (1996). Distribution of class II transposase and resolvase genes in soil bacteria and their association with *mer* genes. *Appl. Environ. Microbiol.* **62**: 2961- 2965.

86. **Pivovarova, T.A., G.E. Markosyan and G.I. Karavaiko.** (1981). Morphogenesis and fine structure of *Leptospirillum ferrooxidans*. *Mikrobiologiya* **50**: 482- 486.

87. **Radstrom, P., O. Skold, G. Swedberg, J. Flensburg, P.H. Roy and L. Sundstrom.** (1994). Transposon Tn*5090* of the plasmid R751, which carries an integron is related to Tn*7*, Mu, and retroelements. *J. Bacteriol.* **176**: 3257-3268.

88. **Rawlings, D.E.** (1995). Restriction enzyme analysis of 16S rRNA genes for the rapid identification of *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans* and *Leptospirillum ferrooxidans* strains in leaching environments. p. 9-17. *In* C.A Jerez, T. Vargas, H. Toledo and J.V. Wiertz (eds.) *Biohydrometallurgical Processing*. Volume II University of Chile Press: Santiago.

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

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

91. **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 biooxidation of pyrite and related ores. *Microbiology* **145**: 5- 13.

92. **Rawlings, D.E., D. Dew and C. du Plessis.** (2003). Biomineralizion of metal-containing ores and concentrates. *Trends in Biotech.* **21**: 38- 44.

93. **Reimmann, C., R. Moore, S. Little, A. Savioz, W.S. Willets and D. Haas.** (1989). Genetic structure, function and regulation of the transposable element IS*21*. *Mol. Gen. Genet.* **215**: 416- 424.

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

95. **Rosen, B.P.** (2002). Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Compar. Biochem. Physiol.* PartA **133**: 689- 693.

96. **Rosen, B.P.** (2002). Biochemistry of arsenic detoxicification. *FEBS Lett.* **529**: 86- 92.

97. **Rosen, B.P., H. Bhattacharjee, T. Zhou and A.R. Walmsley.** (1999). Mechanisms of ArsA ATPases. *Biochem. Biophys. Acta.* **1461**:207- 215.

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

99. **Rouch, D.A., D.S. Cram, D. DiBenardino, T.G. Littlejohn and R.A. Skurray.** (1990). Efflux mediated antiseptic resistance gene *qacA* from *Staphylococcus aureus*: common ancestry with tetracycline- and sugar-transport proteins. *Mol. Microbiol*. **4**: 2051- 2062.

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

101. **Sambrook, J., E.F. Fritsch and T. Maniatis.** (1989). Molecular cloning: a laboratory manual, 2nd ed. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, N.Y.

102. **Sand, W., K. Rohde, B. Sobotke and C. Zenneck.** (1992). Evaluation of *Leptospirillum ferrooxidans* for leaching. *Appl. Environ. Microbiol.* **58**: 85- 92.

103. **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.

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

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

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

107. **Schippers, A., P-G. Jozsa and W. Sand.** (1996). Sulfur chemistry in bacterial leaching of pyrite. *Appl. Environ. Microbiol.* **62**: 3424-3431.

108. **Schippers, A. and W. Sand.** (1998). Bacterial leaching of metal sulfides proceeds by two direct mechanisms via thiosulfate or via polysulfides and sulfur. *Appl. Environ. Microbiol.* **65**: 319- 321.

109. **Schloter, M., B. Aßmus and A. Hartmann.** (1995). The use of immunological methods to detect and identify bacteria in the environment. *Biotech. Adv.* **13**: 75-90.

110. **Sherratt, D.** (1989). Tn*3* and other related transposable elements: site specific recombination and transposition, p. 163- 184. In D.E. Berg and M.M. Howe (eds.), Mobile DNA. *American Society for Microbiology*, Washington, D.C.

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

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

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

114. **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.

115. **Silver, S.** (1996). Bacterial resistance to toxic metals. *Gene* **179**: 9- 19.

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

117. **Stark, W.M., and M.R. Boocock.** (1995). Topological selectivity in site-specific recombination, p. 101-125. *In* D.J. Sherratt (ed.), Mobile genetic elements. Oxford University Press, New York, N.Y.

118. **Stokes, H.W. and R.M. Hall.** (1989). A novel family of potential mobile DNA elements encoding a site-specific gene-integration function: integrons. *Mol. Microbiol.* **3**: 1669- 1683.

119. **Stevens, S.**, **W. Hu, T. Gladysheva, B.P. Rosen, E.R.P. Zuiderweg and L. Lee.** (1999). Secondary structure and fold homology of the ArsC protein from the *Escherichia coli* arsenic resistance plasmid R773. *Biochemistry* **38**: 10178-10186.

120. **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 pKW301 in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**: 411- 418.

121. **Tisa, L.S. and B.P. Rosen.** (1990). Molecular characterization of an anion pump. *J. Biol. Chem.* **265**: 190- 194.

122. **Tuffin, M., P. de Groot, S.M. Deane and D.E. Rawlings.** (2005). An unusual Tn*21*-like transposon containing an *ars* operon is present in highly arsenic-resistant strains of biomining bacterium *Acidithiobacillus caldus*. *Microbiology* **151**:

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

124. **Tyson, G.W., J. Chapman, P. Hugenholtz, E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, E.M. Ruben, D.S. Rokhsar and J.F. Banfield.** (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37- 43.

125.	**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.

126.	**Walmsey, A.R., T. Zhou, M.I. Borges-Walmsey and B.P.Rosen.** (2001). Antimonite regulation of the ATPase activity of the ArsA, the catalytic subunit of the arsenical pump. *J. Biochem.* **360**: 589- 597.

127.	**Wang, G., S.P. Kennedy, S. Fusiludeen, C. Rensing and S. Das Sarma.** (2004). Arsenic resistance in *Halobacterium* sp. strain NRL-1 examined by using an improved gene knockout system. *J. Bacteriol.* : **186**: 3187- 3194.

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

129.	**Worley, K.C., B.A. Wiese and R.F. Smith.** (1995). BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequences similarity search results. Genome Res. **5**:173-184.

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

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

132.	**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.

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

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

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

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

137.	**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.

138. **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.