

**Analysis of an 18kb accessory region of
plasmid pTcM1 from *Acidithiobacillus
caldus* MNG**



**This thesis is presented in partial fulfillment of the requirements for the
degree of master of science at the University of Stellenbosch**

**Department of Microbiology
Promoter: Prof Douglas E. Rawlings
March 2009**

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 24 February 2009

Copyright © 2009 Stellenbosch University

All rights reserved

ABSTRACT

Biomining organisms are generally found in metal-rich, inorganic environments such as iron and sulfur containing ores; where they play a vital role in mineralization and decomposition of minerals. They are typically obligatory acidophilic, mesophilic or thermophilic, autotrophic, usually aerobic, iron-or sulfur oxidizing chemolithotrophic bacteria. The most prominent biomining organisms used in bioleaching of metal sulfides are *Acidithiobacillus ferrooxidans*, *At. thiooxidans*, *At. caldus*, *Sulfobacillus* spp. and *Leptospirillum* spp. Biomining enables us to utilize low grade ores that would not have been utilized by conventional methods of mining. Research has focused on the backbone features of plasmids isolated from bacteria of biomining environments. The aim of this study is to sequence and analyze an 18 kb region of the 66 kb plasmid pTcM1 isolated from *At. caldus* MNG, focusing on accessory genes carried by this plasmid.

Fifteen putative genes / open reading frames were identified with functions relating to metabolism and transport systems. The genes are located in two divergently located operons. The first operon carries features related to general metabolism activities and consists of a transcriptional regulator (ORF 2), a succinate / fumarate dehydrogenase-like subunit (ORF 3), two ferredoxin genes (ORF 4 and ORF 7), a putative HEAT-like repeat (ORF 6) which is interrupted by an insertion sequence (ORF 5) and a GOGAT-like subunit (ORF 8). The second operon contains an ABC-type nitrate / sulfonate bicarbonate-like gene (ORF 9), a binding protein-dependent inner membrane component-like gene, another ABC sulfonate / nitrate-like gene (ORF 12ⁱ and 12ⁱⁱ) which is interrupted by an insertion sequence (ORF 13) and two hypothetical proteins with unknown functions (ORF 14 and ORF 15).

Southern hybridization analysis have shown that most of the genes from the two operons are found in other *At caldus* strains #6, "F", C-SH12 and BC13 from different geographical locations. Expression of the GOGAT-like subunit and the succinate / fumarate-like subunit was demonstrated in *At. caldus* MNG showing that these genes are functional and actively transcribed. The transcriptional regulator (ORF 2) has been shown to repress the downstream genes of putative operon 1. The persistence of these genes on plasmids together with the fact that they are being expressed, represents a potential metabolic burden, which begs the question why they have been maintained on the plasmid from geographically separated strains (and perhaps also growing under very different nutrient availability conditions) and therefore what possible role they may play.

SAMEVATTING

Bio-mynwese stel ons instaat om erts wat arm is aan minerale te ontgin, wat andersins met normale mynwese nie benut sou kon word nie. Mynwese omgewings is oligotrofies en bevat anorganiese minerale soos yster en swaël. Bakterieë wat in hierdie omgewings aangetref word het `n voorkeur vir suur omgewings (asidofilies), is mesofielies of termofilies, outotrofies en yster- of swaël- oksiderende chemolitotrofe. Dié tipe bakterieë speel `n belangrike rol in mineralisasie en ontbinding van minerale in die grond. Van die belangrikste bakterieë betrokke by die prosesse is *Acidithiobacillus ferrooxidans*, *At. thiooxidans*, *At. caldus*, *Sulfobacillus* spp. en *Leptospirillum* spp. Navorsing is daarop gemik om die ruggraat eienskappe van plasmiede wat vanuit die mynwese omgewings geïsoleer is te bestudeer. Die doel van hierdie projek is om `n 18 kb deel van plasmied pTcM1 wat uit *At. caldus* geïsoleer is, se DNA volgorde te bepaal en die addisionele gene te bestudeer.

Vyftien veronderstelde gene / ooprees rame met funksies wat gekoppel is aan die metabolisme van bakterieë en vervoer van substrate is geïdentifiseer. Die gene / ooprees rame is geleë in twee moonlike operonne wat in teenoorgestelde rigtings wys. Die eerste operon se gene is gekoppel aan metaboliese aktiwiteite en bestaan uit die volgende: `n transkripsionele regulerings geen (ORF 2), `n suksinaat / fumaraat dehidrogenase subeenheid (ORF 3), twee ferridoksien gene (ORF 4 & 7), `n “HEAT-tipe herhaling” (ORF 6) wat deur `n invoegingselement onderbreek is (ORF 5), en `n “GOGAT-tipe” subeenheid (ORF 8). Die tweede operon is gekoppel aan transport aktiwiteite en bevat die volgende gene: `n “ABC-tipe nitraat / sulfonaat bikarbonaat-tipe” geen (ORF 9), `n “bindingsproteïen-afhanklike binnemembraan-tipe” geen, nog `n “ABC sulfonaat / nitraat-tipe” geen wat deur `n invoegingselement onderbreek is, asook twee hipotetiese proteïene met onbekende funksies (ORF 14 en ORF 15).

“Southern” hibridisasie eksperimente het getoon dat die gene van hierdie twee operonne in verskillende *At. caldus* spesies voorkom, naamlik #6, “F”, C-SH12 en BC13 wat uit verskillende omgewings geïsoleer is. Die “GOGAT-tipe” subeenheid en die suksinaat / fumaraat subeenheid proteïene word wel in *At. caldus* uitgedruk. Die transkripsionele reguleerder het `n onderdrukkende effek op die gene stroom-af in die eerste operon. Die gene wat op hierdie plasmied voorkom word dus in die *At. caldus* spesie gehandhaaf, wat `n vraag plaas op die voordeel wat hierdie addisionele gene vir *At. caldus* inhou?

ACKNOWLEDGEMENTS / BEDANKINGS

I would like to express my sincere gratitude to my promoter Professor Douglas Rawlings. He was always willing to lend an ear and supported this study. Thank you for believing in me and giving me this opportunity to enrich my life.

I would like to thank Lonnie van Zyl for his unwavering faith in me, his encouragement and support. He was a inspiration to me and helped me to persevere, I am privileged to call him my friend.

I would like to thank Dr. Shelly Deane for her support through the years.

I would like to thank the National Research Foundation, BHP Billiton and the University of Stellenbosch for their financial aid during my studies.

Ek wil my ouers bedank vir die geweldige opofferings wat hulle gemaak het om my die geleentheid te gee om verder te studeer. Soveel soos die sand en die see, so lief het ek julle. My ou sussie - my beste vriendin en my medestudent. Dit was wonderlik om terug te gaan Stellenbosch toe en die Bos lewe met jou te kon deel. Aan my toekomstige man Abraham, ek is bly ons het ontmoet terwyl ek studeer het, die opofferings, laat aande en toegewydheid wat dit gekos het om klaar te maak sou ek nie aan jou kon verduidelik het as jy dit nie saam met my deurgemaak het nie. Dankie dat jy my so getrou en geduldig bygestaan het.

Aan my vriende Anneke, Antoinette, Marlize, Maryke, Angela, Sybille, Wesley, Isa, Daleen, Gwen en Arrie wil ek ook my dankbaarheid teenoor julle uitspreek - julle het my bygestaan en aangespoor.

Habakuk 3: 19 (NLV) "Die Here die oppermagtige is my krag! Hy maak my voete soos die van `n ribbok; op hoë plekke laat Hy my veilig loop.

TABLE OF CONTENTS:

CHAPTER ONE: INTRODUCTION	8
1. BACKGROUND TO BIOMINING	8
1.1 A BRIEF INTRODUCTION OF THE GENERAL PRINCIPLES INVOLVED IN BIOMINING	8
1.2 THE CONSORTIA OF BACTERIA FOUND IN BIOMINING ENVIRONMENTS	10
1.3 AN OVERVIEW OF THE ELECTROCHEMICAL PROCESSES INVOLVED IN BIOLEACHING / BIO-OXIDATION	12
2. ACIDOPHILIC, MODERATE THERMOPHILIC BACTERIA	14
2.1 TAXONOMY OF THE <i>ACIDITHIOBACILLUS</i> GENUS	14
2.2 THE NUTRITIONAL AND ENERGY REQUIREMENTS OF <i>ACIDITHIOBACILLI</i>	16
2.3 NITROGEN FIXATION AND AMMONIUM ASSIMILATION	22
3. MOBILE GENETIC ELEMENTS	25
3.1 A SHORT INTRODUCTION INTO PLASMID BIOLOGY	25
3.2. PLASMID BACKBONE	26
3.2.1 REPLICATION	26
3.2.1.1 THETA REPLICATION	26
3.2.1.2 ROLLING CIRCLE REPLICATION	30
3.2.2 COPY NUMBER CONTROL AND HOST RANGE	31
3.2.3 STABILITY SYSTEMS	34
3.3 PLASMID VERSATILITY	38
3.4 PLASMIDS ISOLATED FROM <i>ACIDITHIOBACILLI</i>	43
AIMS OF THIS PROJECT	46
CHAPTER TWO: SEQUENCING RESULTS	48
1. INTRODUCTION	48
2. MATERIALS AND METHODS	50
2.1 MEDIA AND GROWTH CONDITIONS	50
2.2 DNA ISOLATION, PURIFICATION, CLONING AND SEQUENCING	50
2.3 BACTERIAL STRAINS, PLASMIDS AND PCR PRIMERS	51
3. RESULTS	52
3.1 CLONING AND RESTRICTION ENZYME MAP	52
3.2 DNA ANALYSIS	54
4. DISCUSSION	76
CHAPTER THREE: DISTRIBUTION OF PUTATIVE GENES	79
1. INTRODUCTION	79
2. MATERIALS AND METHODS	80
2.1 MEDIA AND GROWTH CONDITIONS	80
2.2 BACTERIAL STRAINS, PLASMIDS AND PROBES	80
2.3 TOTAL DNA ISOLATION FROM <i>ACIDITHIOBACILLI</i>	81
2.4 DETECTION OF PUTATIVE GENES USING SOUTHERN HYBRIDIZATION	81
3 RESULTS	81
3.1 OCCURANCE OF THE PUTATIVE OPEN READING FRAMES IN <i>ACIDITHIOBACILLI</i>	81

3.1.1	TRANSCRIPTIONAL REGULATOR (ORF 2)	82
3.1.2	SUCCINATE DEHYDROGENASE (ORF 3)	83
3.1.3	FREQUENCY OF INSERTION ELEMENTS (ORF 5 AND 13)	84
3.1.4	FERREDOXIN (ORF 7)	86
3.1.5	GLUTAMATE SYNTHASE (ORF 8)	87
3.1.6	ABC-TYPE SULFONATE/BICARBONATE-LIKE GENES (ORF 12 ²)	88
3.1.7	PUTATIVE pRSB101_23 AND pRSB_24 PROTEINS (ORF 14 & 15)	89
4	DISCUSSION	90

CHAPTER FOUR: EXPRESSION AND FUNCTIONALITY ANALYSIS OF PUTATIVE GENES **94**

4.1	INTRODUCTION	94
4.2	MATERIALS AND METHODS	96
4.2.1.	MEDIA AND GROWTH CONDITIONS	96
4.2.2.	BACTERIAL STRAINS, PLASMID CONSTRUCTS AND RT-PCR PRIMERS	96
4.2.3.	TOTAL RNA EXTRACTIONS AND PURIFICATIONS	96
4.2.4.	SLOT BLOT ANALYSIS	96
4.2.5.	PROMOTER- <i>lacZ</i> REPORTER CONSTRUCTS	97
4.2.6.	β -GALACTOSIDASE ACTIVITY ASSAYS	97
4.2.7.	RT-PCR ANALYSIS OF mRNA	97
4.3.	RESULTS	99
4.3.1	ANALYSIS OF PROMOTER ACTIVITY OF SELECTED GENES	99
4.3.2.	DETECTION OF mRNA EXPRESSION	105
4.3.3.	ANALYSIS OF THE TRANSCRIPTION PRODUCTS	106
4.4.	DISCUSSION	107

CHAPTER FIVE: GENERAL DISCUSSION **110**

LITERATURE CITED **116**

APPENDIX 1 **134**

APPENDIX 2 **135**

APPENDIX 3 **136**

APPENDIX 4 **138**

APPENDIX 5 **141**

CHAPTER ONE: INTRODUCTION

1. BACKGROUND TO BIOMINING

1.1 A BRIEF INTRODUCTION OF THE GENERAL PRINCIPLES INVOLVED IN BIOMINING

BIOMINING has become an attractive method for extracting metals since the late 1980's. It can be described by two processes, bioleaching and bio-oxidation. The former process involves the dissolution of various insoluble metal sulfides (FeS, CuS, NiS, ZnS, CoS) to soluble metal sulfates (such CuSO₄, NiSO₄, ZnSO₄ and CoSO₄) and the metals are recovered from the soluble phase. Gold and silver containing ores are inert from chemical attack by ferric iron and acid produced by microbes, these recalcitrant (difficult-to-treat) pyrite and arsenopyrite ores are pretreated with bacteria. The biooxidative action of the bacteria decompose the mineral matrix and expose the entrapped gold (Brierley, 1997; Rawlings *et al.*, 2003). The gold remains in the mineral and is extracted by cyanide. Biomining is implemented in the industry in the form of aerated stirred tanks, bioheaps, dump leaching or in situ leaching, operating at temperatures ranging from ambient to 80°C.

One of the methods for gold recovery is using aerated stirred tank bioreactors. The bioreactor consists of three or more tanks in series and one in parallel to permit a longer retention time. The final stage of discharge is water-washed and treated with limestone to neutralize the arsenic and iron present in the ore. The residue is then washed with water to remove excess acid and soluble metals; followed by treatment with cyanide to recover the gold (Brierley, 1997 and Hallberg *et al.*, 1994). Examples of commercial-scale bioleach/bio-oxidation plants include Fairview (South-Africa), Sao Bento (Brazil) and Sansu (Ghana).

Copper and gold are extracted from chalcocite ores and low-grade, refractory sulfuric gold ores, respectively, using bioheaps reactors. In this process the ore is crushed, stacked on lined pads and supplied with and aeration piping where after the heaped ore is treated with sulfuric acid to prepare the heap for bacterial activity. In the case of chalcocite, the heap is irrigated with the effluent allowing bacteria to catalyze the release of copper in soluble form (Brierley, 1997). In the case of gold, the heap is treated with an acidic, ferric solution containing bacteria. Typically, heaps are aerated and irrigated for several months. Copper is recovered from pregnant leach solutions as they are produced but in the case of gold the bio-oxidized ore is water-washed, the heap disassembled and treated with lime or

cement, restacked on lined pads and cyanide is used to recover the gold. Examples of bioheap plants for copper recovery include Cerro Colorado (Chile), Mt. Leyshon (Australia) and for gold recovery, Newmont-Carlin (USA).

Both uranium and copper have been extracted from exhausted underground mines by *in situ* bioleaching (Brierley, 1997). Largely worked out underground stopes that contain some remaining metal are blasted to fragment the ore and establish permeability. The shafts are left intact to allow air-flow- through. The top surface of the fractured ore is treated with acidified leach solutions allowing bacteria to become established. Leach solutions are collected in sumps and pumped to the surface for metal extraction.

Economically it is more cost effective to make use of biomining processes when metal values are low. These methods do not make use of roasting and smelting procedures. Generally they require lower capital and operational costs and less skilled labor is necessary. Shorter construction periods are needed and the operational setup is much simpler when compared to traditional mining (Brierley, 1997).

Feasibility studies for developing new mines are based on the base/precious metal grade in the ore and concentrate. Typically, traditional mining processes cannot make use of low-grade ores since it would not be lucrative to construct a mine for such a low metal recovery. Biomining offers an alternative for metal recovery from low-grade ores and dumps. In-situ bioleaching is also very cost-effective with very little additional disturbance to the environment (Brierley, 1997).

1.2 THE CONSORTIA OF BACTERIA FOUND IN BIOMINING ENVIRONMENTS

Metal-rich, inorganic environments provide an ideal habitat for obligate acidophilic¹, mesophilic or thermophilic², autotrophic³ or heterotrophic⁴ mostly aerobic, iron- or sulfur-oxidizing chemolithotrophic⁵ bacteria (Hallberg *et al.*, 2003). These bacteria play a vital role in the mineralization and decomposition processes of mining ores. The most prominent biomining organisms belong to the genera *Acidithiobacillus*, *Leptospirillum*, *Sulfobacillus*, *Metallosphaera*, *Acidianus* and *Sulfolobus*. These organisms thrive in oligotrophic and acidophilic environments where they oxidize ferrous iron or mineral sulfide substrates.

Biomining environments are often associated with mixed cultures of iron - and sulfur oxidizing bacteria allowing a type of symbiosis where carbon and energy sources can be shared. Prominent iron-oxidizing bacteria include Leptospirilla species such as *L. ferrooxidans*, *L. ferriphilum*, recently discovered *L. ferrodiazotrophum* and *At. ferrooxidans* (previously *T. ferrooxidans*). Sulfur-oxidizing bacteria include Acidibacilli such as *At. caldus* (formerly *T. caldus*) and *At. thiooxidans* (previously *T. thiooxidans*) (Rawlings D.E., 1995; Norris *et al.*, 1998; Rawlings *et al.*, 1999, Coram and Rawlings 2002,). In some instances the chemolithotrophs and heterotrophs might have a mutualistic relationship where the heterotrophic organisms aid in the “detoxification” of the immediate surroundings by removing excessive organic materials (Johnson *et al.*, 1997). Another benefit might be that substrates is regenerated. Acidophilic heterotrophs may also provide vitamins, cofactors, chelating agents and surfactants that are beneficial to the chemolithotrophic acidophiles.

Competition between various iron-oxidizing or sulfur-oxidizing bacteria is determined by many factors such as temperature, carbon dioxide, nutrients, tolerance for higher acidity, a higher affinity for ferrous iron or their tolerance to heavy metals such as arsenic or mercury (Dew *et al.*, 1997).

Typically *Leptospirillum* species can tolerate high redox potentials and will grow better in low pH (less than pH 1.5) and high temperature (40-45°C) whereas *Acidithiobacillus* species such as *At.*

¹ Acidophilic: Growth optimum between pH 0 and 5.5
² Mesophilic and thermophilic: Growth optimum between 20-45°C and 55-65°C; respectively
³ Autotrophic: Uses CO₂ as a sole carbon source
⁴ Heterotrophs: Uses reduced, preformed, organic molecules as carbon source
⁵ Chemolithotrophic: Uses inorganic molecules such as sulfur or iron as energy source

ferrooxidans favors lower redox potentials (high ferrous conditions), pH of 1.8 to 2.0 and temperatures ranging between 35 and 40°C (Dew *et al.*, 1997).

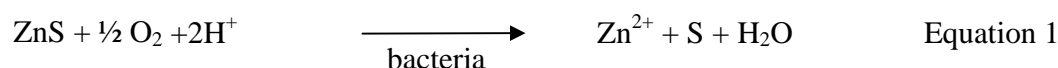
For many years it was thought that *At. ferrooxidans* was the dominant species in pyrite-oxidizing mixed cultures from acid drainage mines, ore leaching dumps and coal spoil sites. However, the latest research showed that *L. ferrooxidans* is the more prominent iron-oxidizing species in processes operating at 40°C (Norris, 1997 and Rawlings, 2005^a). Mixed cultures of the iron-oxidizing bacterium *L. ferrooxidans* and the sulfur-oxidizing *At. caldus*, have been shown to be very effective for the dissolution of chalcopyrite ores.

Very little research has been done on the molecular genetics of acidophilic chemolithotrophs. The genome of *L. ferrooxidans* has been sequenced. At present the 2.7 Mb genome of *At. ferrooxidans* (ATCC23270) is the only strain of the *Acidithiobacillus* genus for which the genome sequence is available, although the genome of *At. caldus* has been sequenced recently this information is not available as yet (personal communication with Professor Rawlings). Several chromosomal, plasmid and transposon genes have been cloned and sequenced (Rawlings, 2002). At this stage the mechanisms of substrate utilization by chemolithotrophs are rather poorly understood and only a few of the proteins involved have been well characterized.

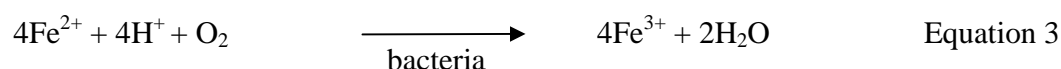
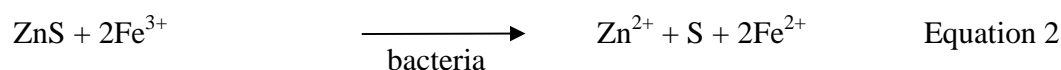
Several attempts have been made to develop gene transfer systems for acidophiles such as *At. ferrooxidans* and *At. caldus* but with limited success. Investigations into the molecular biology of chemolithotrophic acidophiles are difficult as these organisms are not easily cultured on solid media due to their sensitivities to organic matter such as traces of sugar present as impurities in gelling agents (Rawlings, 2002). Even in liquid culture the organisms can be temperamental often yielding highly variable cell densities for example. Successful transfer systems include the exploitation of recombinant plasmids to introduce antibiotic resistance or metal resistance genes into other species. Some successful attempts have included the transfer of broad host range plasmids of different incompatibility groups by conjugation or electroporation into *Acidiphilium*, *Acidocella* and *At. ferrooxidans* species. Recently Van Zyl *et al.*, 2008_b, developed a transfer system for *At. caldus* based on a conjugation system where broad-host range plasmids pSa and R388 were used to transfer suicide vectors from *E. coli* to *At. caldus*. The low frequency rate of transconjugants remains a problem with genetic transfer systems (Rawlings, 2001).

1.3 AN OVERVIEW OF THE ELECTROCHEMICAL PROCESSES INVOLVED IN BIOLEACHING / BIO-OXIDATION

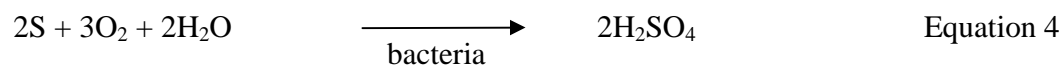
The exact mechanism of leaching is rather controversial, some researchers believe it to be a direct process entailing an enzymatic attack whereas other believe the process is indirect and due to chemical attack on the mineral (Dew *et al.*, 1997). The former constitutes the actual attachment of bacteria to the mineral surface which enhances the rate of mineral dissolution. In this process the bacterial membrane components directly interact with the sulfide and metal moieties of the mineral (Crundwell, 1997 and Rawlings, 2002). This mechanism can be explained by using sphalerite dissolution as an example (Equation 1).



Indirect bioleaching refers to the chemical attack by ferric iron or protons on a mineral sulfide and the dissolution of the mineral into various forms of sulfur and ferrous iron (Rawlings, 2002). Iron-oxidizing microbe's aid in dissolving sulfide minerals by oxidizing ferrous ions to ferric ions and regenerating the reactant as is illustrated by equations 2 & 3. The presence of iron and / or sulfur is a prerequisite (Crundwell, 1997).



This process is often aided by the presence of extracellular polymeric substances (EPS) layers and biofilms. It is thought that the EPS layer encapsulates the iron and serves as a reservoir for ferric iron to mount an attack on the valence bonds of the mineral (Rawlings, 2002). Bacteria are able to use sulfur as a substrate for both the direct and indirect mechanism of bioleaching (Equation 4):



In general, most researchers now believe that the principal role of the microbes is to provide the sulfuric acid for proton attack and maintain iron in the oxidized ferric state for oxidative attack on minerals (Rawlings, 2002). Dew and co-workers (Dew *et al.*, 1997) also reported that the prevalence of attached bacteria in mixed cultures varies depending on the levels of oxidized iron and sulfur intermediates available. Semenza and co-workers (Semenza *et al.*, 2002) proposed two mechanisms for the dissolution process of acid-soluble (FeS_2 , MoS_2 and WS_2) and insoluble metal sulfides (ZnS , CuFeS_2 and PbS). The thiosulphate mechanism is based on the oxidative attack of ferric iron on acid-insoluble metal sulfides involving thiosulphate as the main intermediate. The polysulfide mechanism involves a proton and / or ferric iron attack on acid-soluble metal sulfides with polysulfide and elemental sulfur as the main intermediates (Olsen *et al.*, 2003, Rawlings, 2002 and Schippers *et al.*, 1999). It has been suggested that the role of *At. caldus* in biomining is to increase arsenopyrite-leaching indirectly by utilizing the sulfur compounds that can cause an inhibitory layer on the surface of the mineral. It could contribute to the heterotrophic and mixotrophic growth by releasing organic chemicals and aid in the solubilization of solid sulfur by the production of sulfur-active agents (Dopson and Lindström, 1994).

2. ACIDOPHILIC, MODERATE THERMOPHILIC BACTERIA

2.1 TAXONOMY OF THE *ACIDITHIOBACILLUS* GENUS

Acidithiobacilli are ubiquitous and typically isolated from extreme environments such as sulfur springs and acid mine drainage. Members of this genus were formerly included in the genus *Thiobacillus*. The new genus was created to accommodate the extremely acidophilic members of the thiobacilli. Members of *Acidithiobacillus* include *At. ferrooxidans*, *At. thiooxidans*, *At. albertensis* and *At. caldus*. This genus is placed in the γ -subdivisions of the *Proteobacteria* (Kelly and Woods, 2000; Rawlings, 2002 and Bergamo *et al.*, 2004). *At. ferrooxidans* is able to use ferrous iron and reduced inorganic sulfur compounds as energy source. These Gram-negative bacteria have a G + C ratio of 57% - 59%, a pH optimum of 1.8 - 2.0 and prefer temperatures ranging between 20° - 35°C. *At. thiooxidans* is restricted to reduced sulfur compounds as energy source and has a G+C ratio of 53%. It favors a wider pH range of 0.5 - 5.5 and temperatures of up to 45°C. The DNA-DNA similarity between these two species is about 20% or less.

Members of the genus *Acidithiobacillus* are rod shaped, non-sporulating, aerobic, moderate thermophilic, chemolithotrophic, autotrophic and obligate acidophilic. Their optimum growth temperature is 45°C but they can tolerate up to 55°C and although they can grow well at a pH of around 1.5, have a pH optimum of 2 (Norris, 1997; Kelly and Woods, 2002). *At. caldus* utilizes reduced sulfur compounds for energy sources and it was found that some species can grow mixotrophically⁶ with yeast extract or glucose (Hallberg *et al.*, 1994 and Rawlings, 2002). It has a G+C content is 63.9 mol% and isolates from these species exhibit no significant DNA homology to any other *Acidithiobacillus* species (Hallberg *et al.*, 1994). This species was found to be the dominant sulfur-oxidizing bacterium of arsenopyrite and copper bio-oxidation in pilot plants operating at 35°-50°C (Olsen *et al.*, 2003).

Strain C-SH12 (DSM 9466) was isolated from a continuous bioreactor in Brisbane Australia (Goebel & Stackebrandt, 1994). Two strains were isolated from the United Kingdom, strain BC13 (ATCC51756) from a Birch coppice in Warwickshire and strain KU (DSM 8584) from a coal spoil in Kingsbury (Hallberg and Lindström, 1994). Three *At. caldus* strains were isolated from biomining environments in South Africa. Strains “f” and #6 were isolated from a nickel pilot plant in Billiton and Fairview mine

⁶ Mixotrophs: Utilizes both CO₂ and glucose as carbon source and use inorganic substances as electron sources

in Barberton; respectively (Rawlings, 1999). Strain MNG was isolated from an inoculum obtained from an arsenopyrite pilot plant at the chemical engineering department at the University of Cape Town (Gardner *et al.*, 2001). Previous studies have focused primarily on mixed populations of *At. caldus* and other iron-oxidizing bacteria such as *L. ferrooxidans*

As more DNA sequences become available molecular and genetic information has improved the classification of these bacteria. A phylogenetic tree depicting the relatedness of *Acidithiobacillus*, *Thiobacillus* and other biomining bacteria is illustrated in Figure 1.

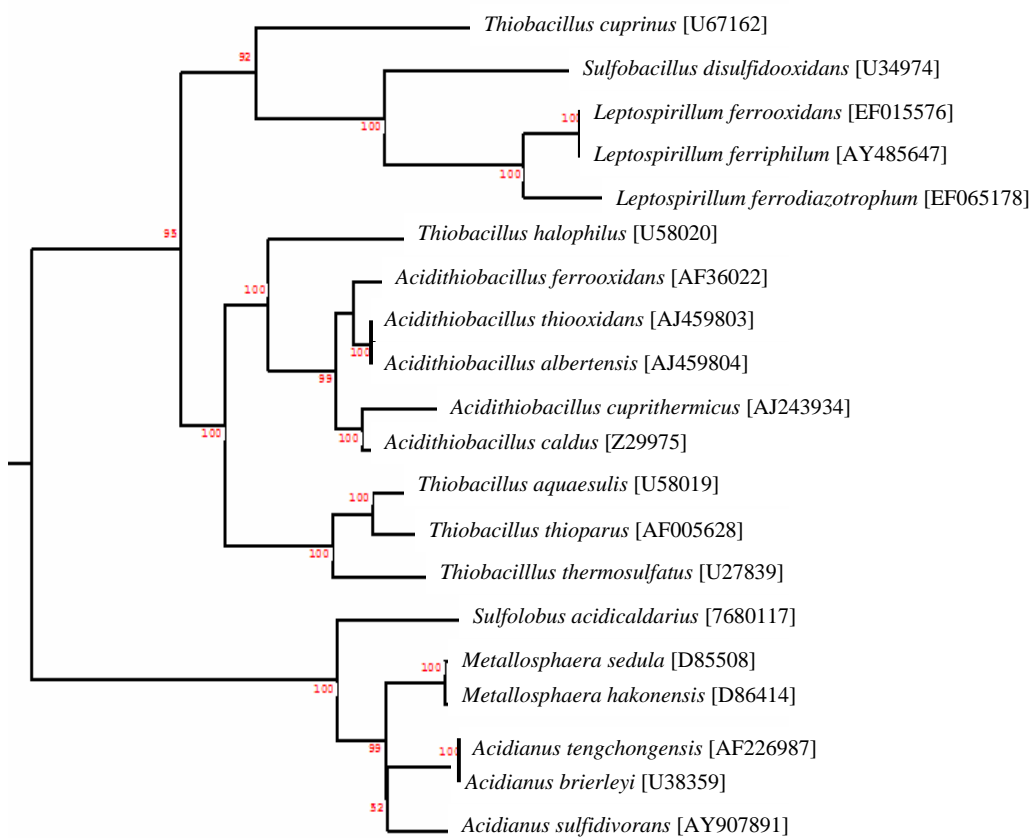


Figure 1: Phylogenetic tree signifying the relative relatedness of bacteria often associated with biomining processes. Their NCBI accession number is given in brackets.

2.2 THE NUTRITIONAL AND ENERGY REQUIREMENTS OF ACIDITHIOBACILLI

Acidophilic bacteria have become well adapted to their oligotrophic environment. Carbon, nitrogen and oxygen sources are provided by air or artificial aeration systems. These bacteria have modest nutrient requirements, occasionally inorganic fertilizer can be added to provide extra nitrogen, phosphate, potassium and other trace elements. Metals, such as iron, are not only a nutritional requirement for some bacteria but also play an important role in energy generation by acting as electron donors or acceptors (Johnson, 2006).

Acidophilic bacteria are capable of chemolithotrophic growth and obtain their energy from the oxidation of reduced inorganic sulfur compounds (RISC's) to sulfate as shown in Figure 2 (Snyders and Champness, 2003 and Rawlings, 2002). Biological oxidation takes place when the RISC's serve as electron donor and oxygen as the electron acceptor which generates more energy compared to iron oxidation. Chemical reactions of sulfide minerals with water, ferric iron and oxygen produces natural RISC's wherever sulfide-containing minerals are exposed to the surface. Some species are capable of utilizing hydrogen as an electron donor. At this point in time many of the exact mechanisms and intermediates involved in sulfur oxidation are not fully understood (Rawlings, 2005^a).

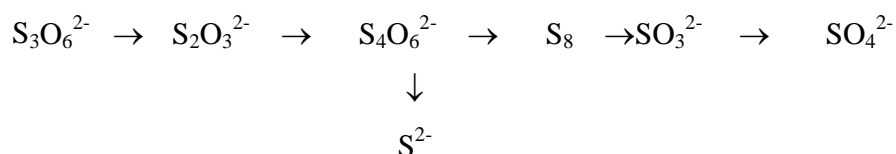


Figure 2: The oxidation of sulfuric compounds⁷.

A model was proposed for the oxidation of elemental sulfur and free sulfide. Thiol groups of specific outer membrane proteins mobilize extracellular elemental sulfur (keeping the zero valence sulfur from precipitating in the periplasm). Sulfur is transferred into the cytoplasm as persulfide sulfane sulfur; which in turn is oxidized to sulfate by a sulfite acceptor oxidoreductase while the electrons are presumably transferred to the cytochromes (the proposed model is illustrated in Figure 3). It is thought that sulfide oxidation requires the catalyzation of glutathione disulfide to form glutathione persulfide. A

⁷ Rawlings, 1997

separate sulfide-quinone oxidoreductase is thought to oxidize free sulfide in the periplasm to elemental sulfur (Rawlings, 2005^a).

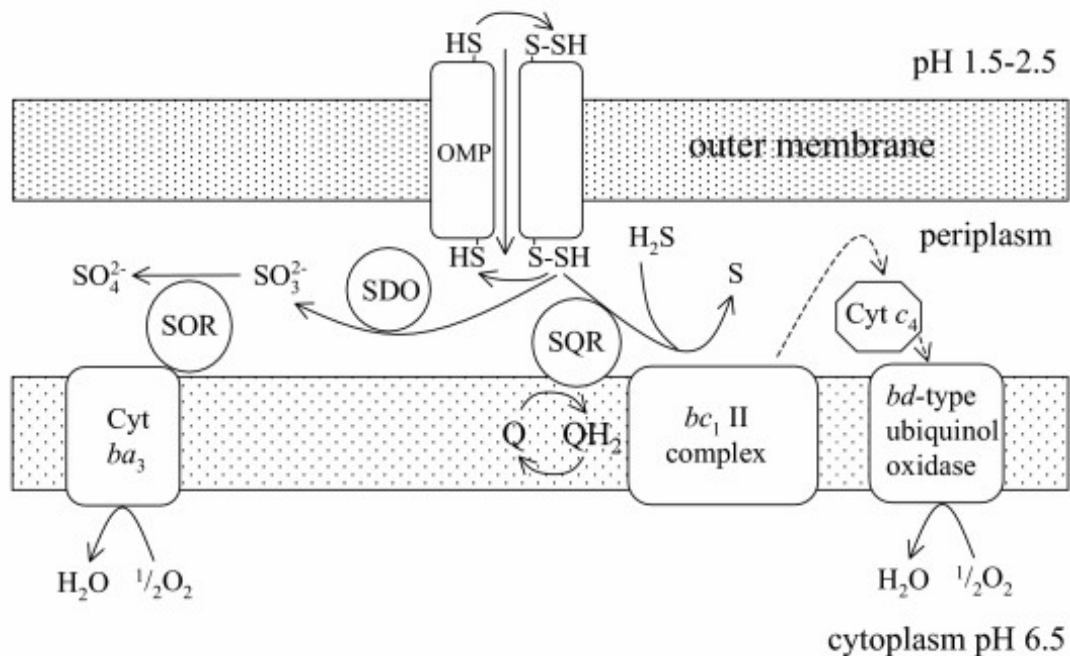


Figure 3: Diagram of sulfur oxidation electron transport. Sulfur is transported to the periplasm by thiol groups of the outer membrane. A periplasmic sulfur dioxygenase (SDO) oxidizes the sulfur to sulfite and the sulfite acceptor oxidoreductase (SOR) oxidizes the sulfite to sulfate. It is thought that the major role players in sulfur oxidation are the ba₃ cytochrome oxidase and a bc₁ II complex together with a bd-type ubiquinol oxidase⁸.

Several genes involved in sulfur/iron oxidation have been identified in *At. ferrooxidans* and *L. ferrooxidans* which could shed some light on the possible energy sources of these organisms. A sulfite ferric iron oxidoreductase and a hydrogen sulphide ferric oxidoreductase (SFORase) were found in these bacteria which are thought to be involved in sulfur oxidation (Rawlings 1997). It is unclear whether it plays such a role in *Leptospirillum* species (Rawlings, 1997 and Rawlings, 2005^a).

Autotrophic sulfur-oxidizing bacteria need extra sources of electrons, NAD(P)H and ATP to reduce their carbon source as much less energy is available from the oxidation of inorganic sulfur molecules as opposed to organic molecules (Rawlings, 2005^a). They generate ATP by oxidative phosphorylation and

⁸ Rawlings, 2005^a

substrate level phosphorylation involving adenosine 5'-phosphosulfate (AMPS); which is a high energy molecule formed from sulfate and adenosine monophosphate (Rawlings, 1997). Dopson and co-workers (Dopson *et al.*, 2002) found that *At. caldus* gains its ATP exclusively from oxidative phosphorylation of reduced inorganic sulfur compounds by means of a membrane-bound F_0F_1 ATPase.

Most bacterial F-type ATPases are similar in structure and consist of two domains made up of 8 subunits. An example of the F_0F_1 ATPase is shown in Figure 4. The globular domain F_1 extending from the membrane is made up of the α 3-, β 3-, γ -, δ - and ϵ - subunits (α - and β -subunits alternate around the central γ -subunit). The intrinsic domain F_0 consists of the a-, b_2 - (not shown) and c_{12} -subunits. The two domains are linked by the central rotor stalk (γ - and ϵ - subunits) and peripheral stalk (δ - and two copies of β -subunits) that presumably keeps the stator subunits from spinning along with the rotor (Sun *et al.*, 2004).

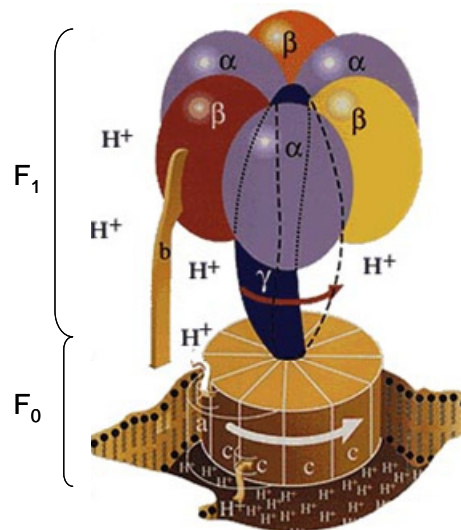


Figure 4: Membrane-bound F_0F_1 ATPase. The globular F_0 domain consist of α 3-, β 3-, γ -, δ - and ϵ -subunits and the intrinsic F_1 domain consists of a-, b_2 - (not shown) and c_{12} -subunits. H^+ : Hydrogen ions, driven by proton motive force (www.millerandlevine.com).

ATP is synthesized in the β -subunits; where the alternating cooperative binding of adenosine diphosphate (ADP) and an inorganic phosphate (Pi) at one of the catalytic β -sites are coupled to the release of adenosine triphosphate (ATP) from the other β -subunit. The γ -subunit is thought to provide the different binding affinities at the β -subunit sites by its rotation in the center of the $\alpha_3\beta_3$ hexamer;

while the c-subunit is the H⁺-catalytic translocation subunit of F_O. Structural changes in the extramembranous loop of subunit c are coupled to protonation/ deprotonation reactions in the center of the membrane and the c-subunit interacts with the γ - and ϵ - subunits. These two subunits seem to move from one c-subunit to the other as ATP is synthesized. The rotation of the γ -subunit within the $\alpha_3\beta_3$ complex is generated by the torque of the movement. Four protons are translocated for each ATP that is synthesized (Sun *et al.*, 2004). Since the cytoplasmic pH of acidophilic bacteria such as *At. ferrooxidans* is close to neutral whereas the external pH is low, the F_OF₁ ATP synthase of these bacteria must have evolved to cope with a large transmembrane pH gradient. Working with the ATP synthase of *At. ferrooxidans*, Brown *et al.*, 1994, showed that the subunits of the F₁ domain could complement those of *E. coli* F₁ mutants but not *E. coli* mutants in the subunits of the F_O domain. Features of the F_O domain that allowed for function at low pH were however, not identified.

Acidithiobacilli are obligate autotrophs and obtain their carbon for cell growth from the air. They employ the Calvin Reductive Pentose Phosphate Cycle (Figure 5) to fix CO₂ as their main carbon source (Rawlings, 2005^a). A number of sulfur-oxidizing bacteria can also grow heterotrophically if supplied with reduced organic carbon sources such as glucose as is the case with *At. caldus*. It is thought that *At. caldus* and *At. thiooxidans* might have the same autotrophic growth requirements and it seems both are able to utilize glucose when grown on sulfur compounds (Norris, 1997).

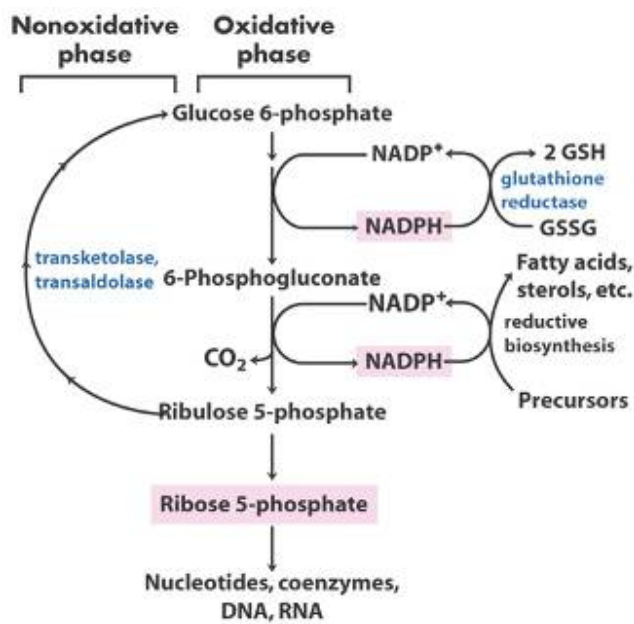
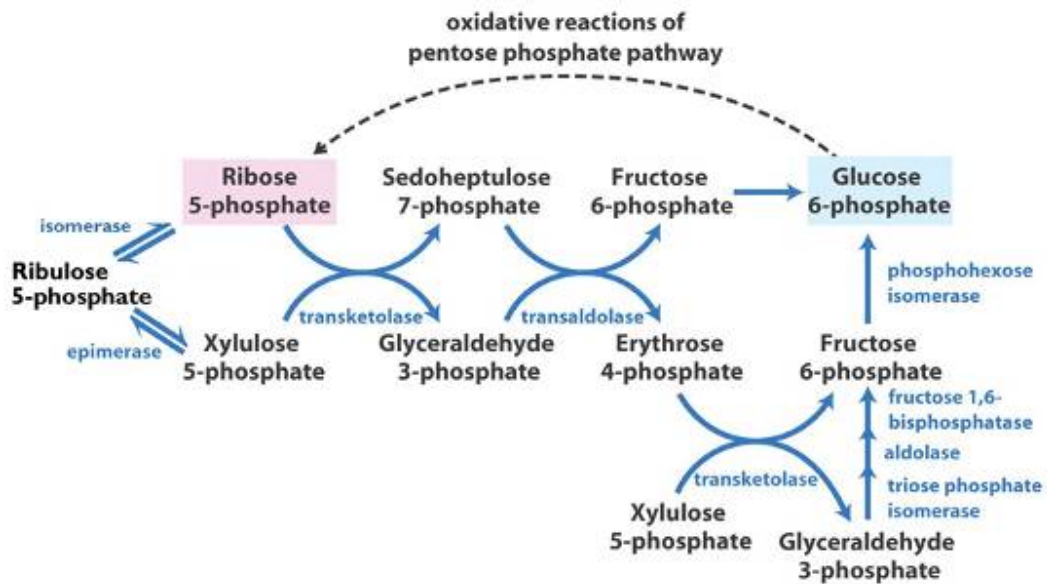
A**B**

Figure 5: The pentose phosphate cycle. A: The pathway can be divided into an oxidative and non-oxidative stage. B: An overview of the reactions. The final products of the pathway yield five carbon sugars for biosynthesis and NADPH (www.biochem.arizona.edu).

Autotrophic bacteria cannot regenerate their NAD(P)H from carbon sources like heterotrophs. The redox couple for the $\text{NAD}^+/\text{NADH}_2$ has a value of -320mV which is lower than most of the energy

sources available to chemolithotrophic bacteria. Therefore autotrophs need a reverse transport mechanism for the synthesis of extra quantities of NADH and / or NADPH which is essential for CO₂ fixation and other processes (Rawlings, 2001 and Rawlings, 2005^a). Autotrophs make use of a process called the reverse proton motive force in which a transmembrane-proton gradient generates the proton motive force essential for the synthesis of NAD(P)H. There are various types of electron carriers that move the necessary electrons and protons to generate NAD(P)H. For instance the flavin mononucleotide (FMN) electron carrier; which carries two electrons and two protons on a complex nicotinamide ring structure of NAD flavoproteins. Ferredoxin (Fd), another type of iron-containing electron carrier is involved in the electron transport system of both the photosynthetic pathway and other electron transport systems and carries only one electron at a time.

2.3 NITROGEN FIXATION AND AMMONIUM ASSIMILATION

Nitrogen fixation involves the reduction of atmospheric nitrogen gas by the enzyme nitrogenase, this process is found only in certain prokaryotes that are either free-living or in symbiotic relationships with higher plants such as the Rhizobia bacteria. Rhizobia infect their host leguminous plants and form nodules in which the nitrogen is fixed and transferred to the plant and in return the plant provides carbon substrate to the bacteria. Cyanobacteria are able to fix nitrogen through an anaerobic or micro-aerobic process.

In acid conditions such as those often associated with biomining bacteria ammonium is highly soluble and most of the nitrogen requirements can be supplied by atmospheric ammonia (Rawlings, 1997). For this reason it is not easy to predict the ability of nitrogen fixation in biomining environments. Several researchers have reported that *L. ferrooxidans*, *L. diazotrophum* and *At. ferrooxidans* (at least fifteen strains) are capable of fixing nitrogen (Rawlings, 2005^a and Tayson *et al*, 2004). This diazotrophic nature (ability to fixate nitrogen compounds) has been attributed to the presence of the *nif* HDK operon (Dew *et al.*, 1997). The presence of these genes in other species for instance *At. thiooxidans* were shown however the functionality of these genes have not been confirmed. One possible explanation for this might be that the high aeration conditions required during the bio-oxidation processes inhibits nitrogen fixation due to the sensitivity of the nitrogenase enzymes to oxygen.

Nitrogen can be incorporated into the cell by ammonia or nitrate. Nitrate is reduced to nitrite by an FAD and a molybdenum containing enzyme (nitrate reductase) in a process called assimilatory nitrate reduction. Nitrate is then reduced to ammonia via nitrite reductase. When nitrogen (ammonia) is abundant it is mostly integrated into the tricarboxylic acid (TCA) cycle by glutamate dehydrogenase as a glutamate intermediate. Ammonia can be incorporated into several amino acids by transamination⁹ reactions.

If the nitrogen concentrations become limiting for bacteria another pathway for nitrogen assimilation is used which takes place via the glutamate synthase (glutamine 2-oxyglutarate amidotransferase [GOGAT]) and glutamine synthetase (GS) system (Garrett and Grisham, 1995). The GOGAT enzyme synthesizes glutamate from glutamine and α -ketoglutarate (α -KG) (Kameya *et al.*, 2007) with the

⁹ Transamination involves the transfer of an α -amino group from an amino acid to the α -keto position of an α -keto acid. The amino donor (such as glutamate) becomes an α -keto acid (α -ketoglutarate).

reduction of electron carriers such as NADPH (bacteria), NADH (yeast) or ferredoxin (plants) (Figure 6).

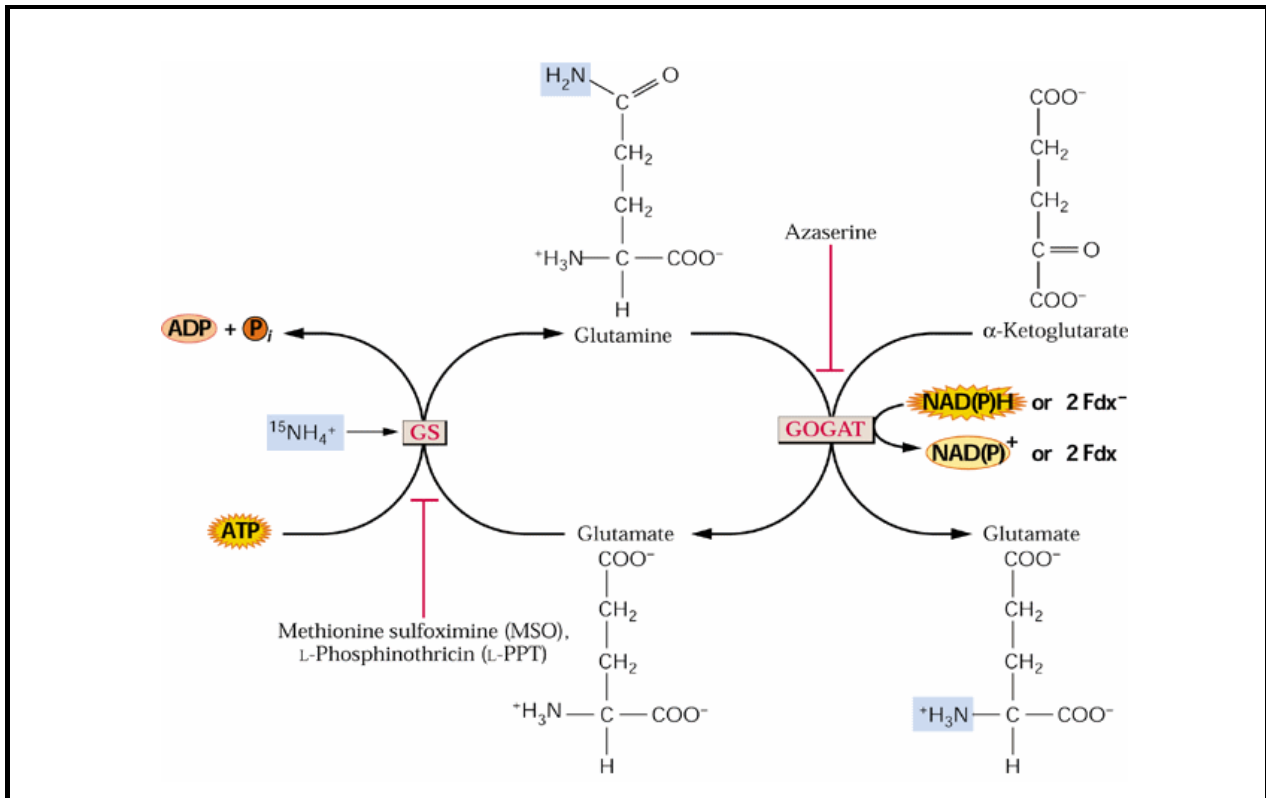


Figure 6: Ammonium incorporation using glutamine synthetase (GS) and glutamate synthase (GOGAT)¹⁰. Two central intermediates are produced, glutamine and glutamate. GS catalyzes glutamine synthesis and glutamate is synthesized by the action of either the GS/GOGAT or GDH (glutamate dehydrogenase).¹¹

Glutamate synthases are typically large complex proteins. The GOGAT found in *E. coli* is an 800 kDa flavoprotein that contains both FMN (flavin mononucleotide) and FAD (flavin adenine dinucleotide) and an iron-sulfur (4Fe-4S) cluster. In the GS/GOGAT pathway GS is responsible for fixing nitrogen from ammonium and GOGAT regenerates glutamate at a cost of 2 equivalents of ATP and 1 NADPH (Garrett and Grisham, 1995).

Glutamate synthases are grouped according to their preference for electron donor types. The Fd-GOGAT utilizes reduced ferredoxin as an electron donor and is found in photosynthetic organisms such as Cyanobacteria and chloroplasts of plants (“plant type GOGAT”) (Kameya, *et al*, 2007). This

¹⁰ <http://www.uky.edu/~dhild/biochem/24/lect24.html>

¹¹ Yan, 2007, PNAS 104, no. 22 p. 9475-9480

protein is monomeric (150 kDa) and it contains a conserved region of 18 amino acid residues called the Fd-loop which is thought to be involved in ferredoxin-Fd-GOGAT binding. In bacteria the GOGAT is typically associated with NADPH as the electron carrier (“bacterial type GOGAT”). This protein is a hetero-octamer consisting of α - and β -subunits of 150 kDa and 50 kDa; respectively.

The Fd-GOGAT and the NADPH-GOGAT have similar reaction kinetics. The Fd-GOGAT and the α -subunit of the NADPH-GOGAT consist of a glutamine amidotransferase (GAT) domain, a central domain, a synthase domain and a β -helical domain. Ammonia is generated by the hydrolysis of glutamine at the GAT domain and then transferred to the synthase domain. The transfer process takes place via an intramolecular ammonia channel that consists of residues of the central and β -helical domains and prevents the leakage of ammonia. In the synthase domain ammonia is converted to glutamate. In the NADPH-GOGAT the β -subunit provides the electrons for the reduction of NADPH (Kameya *et al.*, 2007 and Ceccarelli *et al.*, 2004).

3. MOBILE GENETIC ELEMENTS

PLASMIDS, TRANSPOSONS AND INSERTION SEQUENCES play an important role in bacterial adaptation to environmental pressures. These elements are thought to contribute to the overall fitness of a bacterium or population in specific niches. Plasmids increase the probability of gene transfer between bacteria which, leads to a wide variety of new genetic traits and offer a selective advantage to colonize a specific ecological niche.

Plasmids place a metabolic burden on the host (estimated at 1- 6% per generation). The metabolic burden can be reduced by (i) losing genes not beneficial to the host or the plasmid itself, (ii) or by tight control of gene expression when a gene is not needed or (iii) the number of plasmid copies per host (copy number) can be restricted (Thomas, 2004). Continual efforts are made to try and understand why there are so many of these genetic elements available and what benefit gene carriage on these mobile elements (rather than on the chromosomes) might have. It has been shown that the 1.43 Mb genome of *E. coli* contains more than 600 kb of horizontally transferred DNA and the rate of transfer has been estimated to be 31 kb per million years (Lilley *et al.*, 2000). The increased discovery of related and novel mobile elements in the past few years suggests that this might be a vast underestimation.

3.1 A SHORT INTRODUCTION INTO PLASMID BIOLOGY

Plasmids are extrachromosomal genetic elements capable of autonomous replication (Thomas, 2004). They vary hugely in size from as small as 1.5 kb to larger than 1 Mb. Plasmids can be found as single or multiple copies in a host cell, several hundred copies per cell have been reported (Osborn, *et al.*, 2000). Plasmids are mostly circular although linear plasmids have been found in some species such as *Streptomyces* and *Borrelia* and some Gram-positive bacteria can accumulate plasmids as single stranded DNA (ssDNA).

There have been several philosophical viewpoints concerning the evolution of plasmids. Osborn and his colleagues (Osborn *et al.*, 2000) took a comprehensive look at plasmid evolution. In brief, (i) plasmids can be seen as a group of selfish genes or a gene system co-existing together, (ii) plasmids can be viewed as well-established elements that evolved independently from the chromosome, (iii) they can be considered to be parasitic elements hitch-hiking on microbial genomes ensuring their own survival, (iv) or they can be seen as mutualistic associations with microbes.

Plasmids often display several characteristics such as different types of maintenance functions that include replication machinery, stability systems and various means of mobilization often referred to as the plasmid “backbone” (Smalla *et al.*, 2000).

3.2. PLASMID BACKBONE

3.2.1 REPLICATION

Plasmids contain elements that enable them to replicate independently of the host chromosome. Plasmids generally contain three features (i) a distinct origin of replication (*oriV*), where initiation of replication takes place, (ii) several initiation proteins (Rep proteins) that are plasmid- or host encoded and (iii) replicating controlling mechanisms that control the copy number of plasmids in the host (Osborne *et al.*, 2000).

De Solar and coworkers (De Solar *et al.*, 1998) defined the origin of replication as (i) the minimum-cis acting region of a plasmid able to replicate on its own, (ii) the region where replication is initiated and the DNA strands are separated or (iii) the starting point for leading strand synthesis. Two types of replication mechanisms have been identified namely theta (θ) and rolling circle (RC) replication. Theta replication is the most common method amongst plasmids of gram-negative bacteria.

3.2.1.1 THETA REPLICATION

The replicating plasmid molecule resembles the shape of the Greek letter theta (θ). There are many plasmids that are theta replicating, some contain iterons and carry the necessary genes for their replication (such as pS10, R6K, RK2), while others do not have iterons and require an initiator protein for replication and use host proteins for the rest of the replication process (such as ColE1). Typically the origins of theta replicating plasmids consist of (i) a set of iterons (ii) one or two binding sites for the DnaA initiator protein (dnaA boxes) or other plasmid-encoded initiator proteins, (iii) GATC sequence methylation sites for host Dam methylase and (iv) in the case of iteron containing plasmids an A+T-rich sequence next to the iterons or in the case of non-iteron containing plasmids next to the *oriV* (Espinosa *et al.*, 2000 and Kornberg and Baker, 1992).

Initiation of theta replication in iteron containing plasmids requires a plasmid encoded replication initiation protein (Rep) that specifically recognizes a series of DNA sequence repeats (iterons) in the plasmid origin (*ori*) with the help of the DnaA protein (Figure 7). DNA replication is catalyzed by an enzyme complex namely the replisome and is continuous in the leading strand and discontinuous in the lagging strand (Del Solar *et al.*, 1998, Espinosa *et al.*, 2000, Giraldo *et al.*, 1998 and Kornberg and Baker, 1992). Theta replication involves the distortion of the DNA double helix by the DnaB helicase which separate and unwind the original parental strands at the A+T-rich region and expose the two individual strands. DnaA contributes to the separation of the DNA strands and directs the incorporation of the DnaB protein to one of the strands to form a nucleoprotein complex (a replication fork). Interactions of DnaA and DnaC (helicase-leader protein) are involved in the complex formation. The one strand will act as a template for DNA Polymerase III complex to synthesize DNA from the leading strand in the 5' to 3' direction. The replication bubble is extended by the DnaB helicase and assists DnaG (primase) to enter the replication fork. Continuous DNA synthesis in the leading strand is maintained by the anchoring of the leading strand polymerase on the template strand by the β -sliding clamp (McGlynn and Lloyd, 2002).

In contrast, lagging strand synthesis takes place in segments using pRNA (RNA oligonucleotide) primers which are synthesized by DnaG. These RNA primers allows DNA synthesis to repeatedly initiate on the lagging strand. The DNA polymerase III complex continually associates and dissociates with the lagging strand template to extend each pRNA primer and form Okazaki fragments. In bacteria these fragments can be between 1 to 2 kb in length. In the case of the lagging strand an γ -complex clamp loader continually reload the β -sliding clamp that is associated with the lagging strand DNA polymerase. The pRNA primers are degraded, Polymerase I fills in the gaps and ligation of the 5' end of the Okazaki fragments with the 3' end adjacent fragments takes place to form a single continuous strand (Espinosa *et al.*, 2000, McGlynn and Lloyd, 2002).

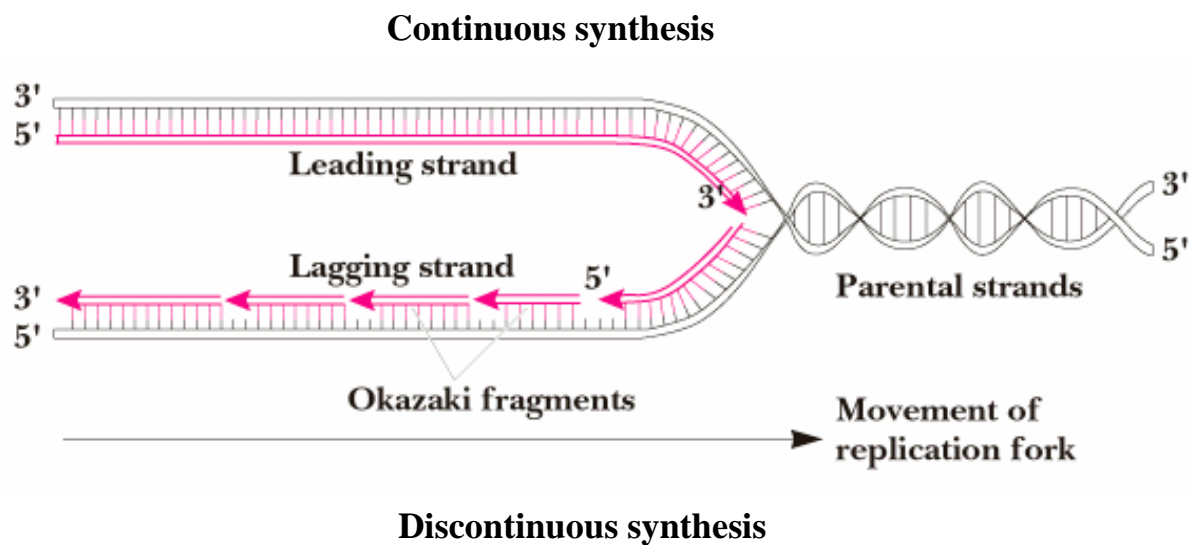


Figure 7: DNA replication¹²: The parental strand is distorted and continuous DNA replication is initiated in the leading strand while discontinuous synthesis takes place in the lagging strand. The arrows represent 3' ends of DNA strands.

Non-iteron containing plasmids for example the ColE1-type plasmids differ from the abovementioned stages. ColE1-type plasmid replication involves (i) transcription across the origin, (ii) the formation of a RNA/DNA hybrid, (iii) the generation of a RNA primer and (iv) initiation of DNA synthesis by a host-encoded DNA polymerase and the extension of the primer RNA (Espinosa *et al.*, 2000). The first step (A) in replication of the ColE1-type plasmids is the initiation of the synthesis of an RNA molecule (RNA II) 555 nucleotides upstream of the *ori* site (Figure 8). The second step (B) involves the extension of RNA II (about 155 nucleotides downstream of the *ori*) and (C) the formation of a duplex with the template plasmid DNA at its 3'-end. A secondary structure is formed at the 5'-end of the RNA II molecule which allows the coupling process between RNA II and the template DNA to take place. This results in an interaction between the RNA section on the *ori* and the template DNA. RNase H recognizes this RNA II- DNA duplex and digests the RNA II molecule and a free 3'-OH group is generated that serves as the primer for DNA Polymerase I and (D) DNA replication continues in one direction only, with the initiation of the lagging strand synthesis at specific ColE1 sites.

¹² <http://www.facstaff.bloomu.edu/gdavis/MoBio/REPLICATION>

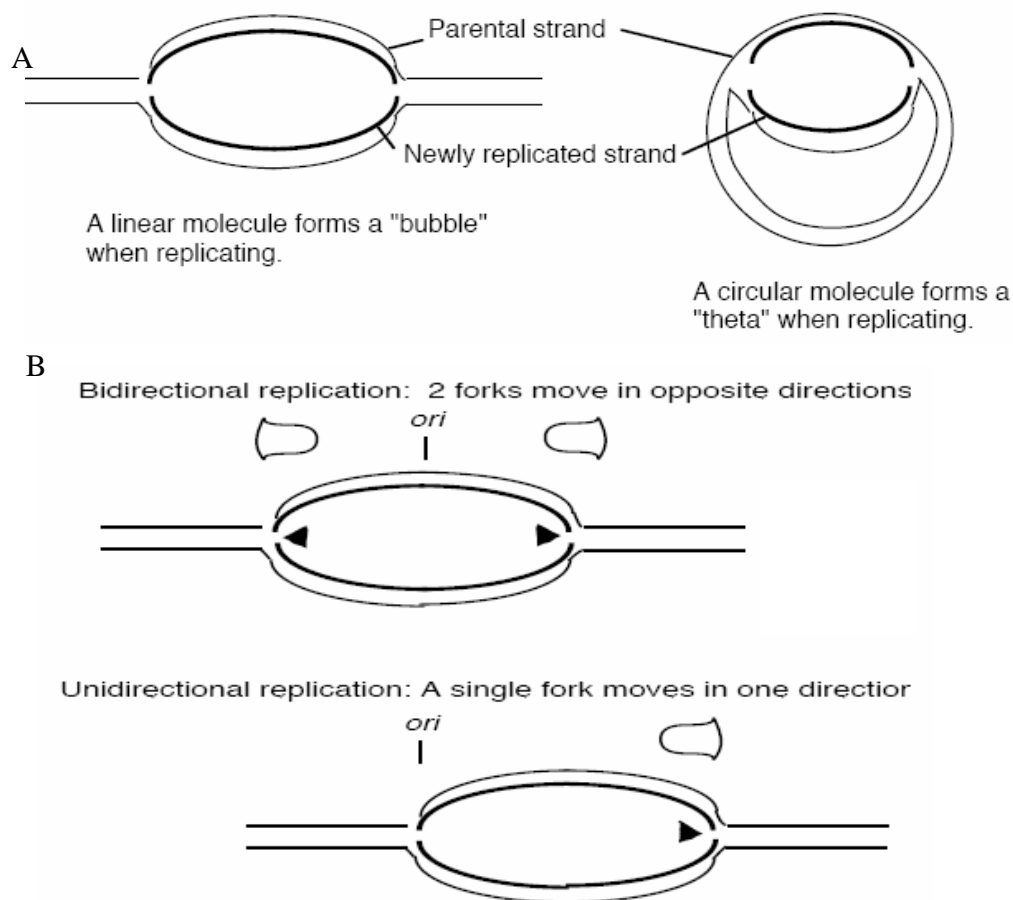


Figure 8: Theta replication. A: A replication bubble is formed when the two parental template chains are separated and copied during replication. Replication bubbles in a circular DNA molecule resemble the Greek letter theta, or θ . B: Illustrates that a replication bubble can result from either unidirectional or bidirectional replication. The origin of replication is labeled *ori*¹³

It seems that host-encoded replication factors appear to play an important role in the theta replication process. For instance, the ColE1-type plasmid requires DNA Polymerase I for the early stages of replication. Other host encoded factors such as DNA polymerase III holoenzyme, DnaB helicase and single strand DNA binding protein (SSB) are required for elongation. In some cases the DnaA host-encoded protein seems to assist in the formation of the plasmid origin open complex. Host encoded DNA gyrase is involved in removing supercoiling. Topoisomerase IV is responsible for the separation of the DNA daughter molecules generated at the end of the replication cycle (Espinosa *et al.*, 2000). All plasmids require host factors although some are less dependent on the host which enables it to broaden

¹³ <http://www.bx.psu.edu/~ross/workmg/Replication1Ch5.pdf>

its host range. The incompatibility group Q plasmids for example RSF1010, carries their own initiator protein (RepC), helicase (RepA) and primase (RepB) (Scherzinger *et al.*, 1991).

3.2.1.2 ROLLING CIRCLE REPLICATION

Rolling circle replication plasmids are more widespread and found in both gram-negative and gram-positive eubacteria and archaea. Replication is initiated when the plasmid-encoded RepC protein that has sequence-specific endonuclease and topoisomerase I-like activities, binds to a specific site (Rep binding site) that contains an inverted repeat. RepC introduces a site-specific nick at the double stranded origin (*dso*) in the leading strand. There is no RNA primer necessary for leading strand synthesis, the exposed open 3'-OH- end then serves as a primer for leading strand synthesis. An initiation complex is formed with DNA Polymerase III, a helicase protein and SSB protein. Replication of the leading strand does not require plasmid-encoded proteins. During rolling circle replication the parental strand (-) is displaced during synthesis of the new strand. During leading strand synthesis the newly synthesized strand (+) is covalently bound to the parental strand (-). Initiation of replication in the lagging strand is initiated at the single stranded origin (*sso*) and not the *dso* as is the case with the leading strand. As a result of the synthesis of the leading strand a displaced single stranded DNA is generated which allows initiation of replication of the lagging strand. Strand transfer is complete when the replisome reaches the reconstituted double stranded origin (*dso*) and replication is terminated (Actis *et al.*, 1999 and Espinosa *et al.*, 2000).

RC replication takes place in one direction and is asymmetric because the leading and lagging strand synthesis is not coupled. Double stranded RC replication leads to a double strand molecule consisting of the parental strand (-) and the newly synthesized ssDNA (+) strand intermediate (identical to the parental stand). Host replicating proteins are responsible for the conversion of the ssDNA intermediate to dsDNA leading to the displacement of the parental strand (-) at the single strand origin (*sso*) (Figure 9) (Espinosa *et al.*, 2000 and Del Solar *et al.*, 1998). The host DNA gyrase enables the final step in RC replication by providing supercoiling of the replication products (Espinosa *et al.*, 2000).

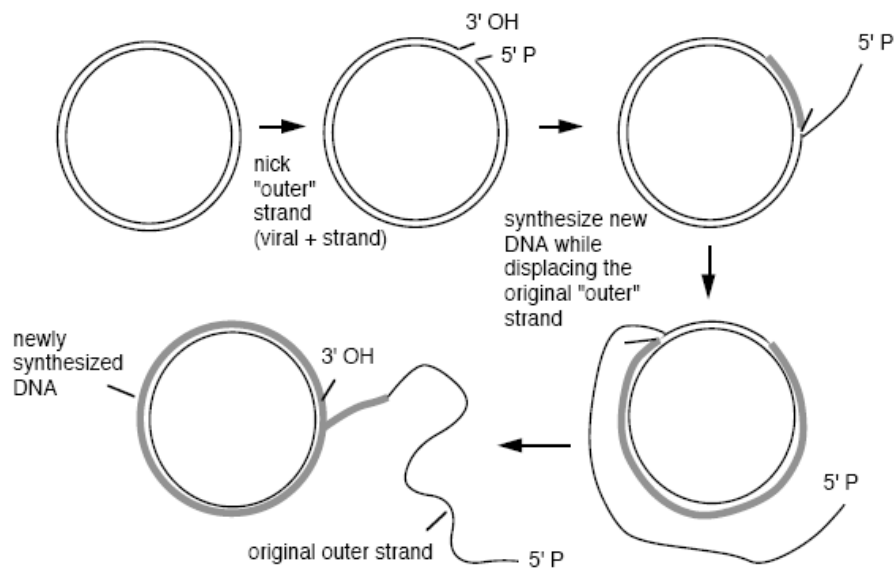


Figure 9: Rolling circle replication, the parental strand (black strand) is nicked by the Rep protein leaving a free 3'-OH end that serves as a primer for discontinuous DNA synthesis¹⁴.

The replicon of a plasmid is used as one of the major classification criteria for plasmids. The replicon also gives rise to other characteristics of plasmids i.e. plasmid incompatibility, copy number control and host range.

3.2.2 COPY NUMBER CONTROL AND HOST RANGE

Single or multiple copies of a plasmid can be found in a host cell. The copy number refers to the number of plasmid copies per chromosome immediately after cell division. Various mechanisms have been identified to maintain the correct copy number by regulating initiation of replication.

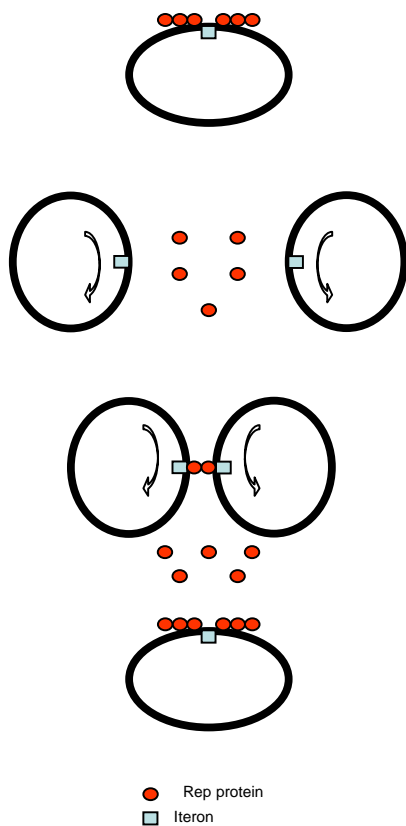
Replication initiation proteins are often involved in copy number control by acting as effectors of replication. Auto-regulation of a replication initiator such as the Rep protein can take place. This helps maintain the optimum protein concentration which is independent of the copy number; which in turn allows for a regulatory loop to prevent uncontrolled initiation of replication when there is an increase in the Rep protein concentration (Espinosa *et al.*, 2000).

¹⁴ <http://www.bx.psu.edu/~ross/workmg/Replication1Ch5.pdf>

Another example of copy number control is dependent on the location of the operator or promoter on the plasmid. The operator(s)/promoter(s) and the origin of replication can be located separately on the plasmid for example on pPS10, pSC101 and R6K. In plasmids such as P1 the operator(s)/promoter(s) are located within the iterons (Park *et al.*, 1998). In the case of P1 the RepA is bound to the iterons and prevents the RNA polymerase from accessing the promoter. The bound RepA is removed from the iterons as the replication fork moves across the iterons, which leads to transcription of more RepA until repression of transcription is instituted again by the binding of RepA (Mukhopadhyay and Chattoraj, 2000).

Iterons have been shown to act as negative effectors. RepA will bind to the iterons until a saturation level has been reached, when this happens the strands separate to form an open complex. Two models for iteron-mediated replication control have been proposed. The “titration model” suggests that the Rep protein is rate limiting. In this model the Rep protein is titrated by the iterons until all iterons have been bound at which point excess Rep protein is available and another round of replication is initiated. However, over-expression of the RepA protein has found to have very little effect on pPS10 and P1 on the copy number (Garcia de Viedma *et al.*, 1995, Chattoraj, 2000). A model that is more consistent with this observation has been suggested, i.e. the “handcuffing” model.

The handcuffing model predicts that the Rep proteins bound to the iterons reaches a saturation level, when this is reached the strand opens and replication is initiated. Replication in turn induces *rep* transcription leading to an increase in cellular Rep concentration. Newly synthesized Rep proteins form dimers while the pre-existing Rep protein monomers bind to the newly formed plasmid’s origin. A protein-protein interaction takes place between the iteron-bound Rep protein monomers located on newly synthesized daughter plasmids. This leads to plasmid “handcuffing” which prohibits a new round of replication and represses Rep operators and promoters as illustrated in Figure 10 (Chattoraj, 2000 and Snyders and Champness, 2003).



Rep protein bind to the iterons of the plasmid origin and a saturation level is reached which promotes strand opening.

Replication takes place thus inducing Rep protein synthesis and a new daughter plasmid is produced.

When the plasmid copy number has reached it's maximum the rep proteins of two adjacent plasmids bind together – “handcuffing model”

When the cell volume increases, during cell division, the handcuffing is reversed.

Figure 10: Iteron-dependent copy number control - “handcuffing model”.¹⁵

ColE1 plasmids exhibit another type of mechanism for plasmid copy number control. They are regulated by a small plasmid encoded RNA I (counter-transcribed RNA – ctRNA) which inhibits DNA replication by forming a complex with the replication initiation primer, RNA II. The plasmid encoded Rop protein helps stabilize the complex and assists in the inhibition of replication. (Snyders and Champness, 2003). Mechanisms for copy number control of plasmids can also combine ctRNA, regulatory proteins, auto-regulation and iterons depending on the different plasmids present in the host.

The host range of a plasmid may be restricted by (i) host encoded replication proteins such as DNA polymerase I and DnaA which are essential proteins for plasmid replication (Caspi *et al.*, 2000). It is thought that some plasmids are unable to form a stable complex with the DNA orthologs of the host.

¹⁵ Reproduced from Chatteraj (2000) and Gardner (2003).

(ii) The structure of the origin and its position relative to the iterons also restricts the host range. Alterations in the helical phasing or the intrinsic DNA curvature can prevent strand opening and plasmid replication (Doran *et al.*, 1999). (iii) Single amino acid substitutions within the Rep protein can alter the host range (Maestro *et al.*, 2003; Fernandez-Tresguerres *et al.*, 1995). (iv) The host cell can inhibit essential plasmid –encoded proteins which would lead to plasmid loss from the population (Del Solar *et al.*, 1996).

3.2.3 STABILITY SYSTEMS

Incompatibility refers to the inability of plasmids to coexist in a host cell in the absence of selection. If plasmids interfere with one another, one plasmid will be lost at a higher rate than the cell divides. These plasmids are incompatible and are considered to belong to the same incompatibility group. There are two reasons why incompatible plasmids cannot coexist in a host cell. Firstly, they share the same replication control mechanism; in which case the control system is unable to distinguish between the two and either one is randomly selected for replication. The second reason why plasmids cannot coexist within a host cell is that the plasmids share partitioning systems (*par*), where one daughter cell will receive all the copies of one type of plasmid, and the other daughter cell the other plasmid.

Plasmid stability during cell growth and division depends on a number of functions that prevents plasmid-loss. Various stability systems exist such as post-segregational killing systems (PSK), multimere resolution systems (*mrs*), active centromere-like partitioning systems, and plasmid- encoded restriction modification systems (RM) (Gerdes *et al.*, 2000).

One of the best known types of stability system is the PSK system. The intrinsic instability of the antitoxin triggers toxin activity in plasmid-free cells leading to killing of the host cell. Post-segregational systems can be divided into two groups. The *hok*-like systems consist of antisense RNA's that serve as the antitoxin and regulate toxin-encoded mRNAs post-transcriptionally. The proteic plasmid system consists of toxin and antitoxin proteins that inhibit toxin activity by direct protein-protein contact (Gerdes *et al.*, 2000).

The *hok/ sok* system was discovered in 1985 from plasmid R1 and represents an example of antisense controlled regulation. The *hok* gene encodes the Hok protein (host killing), the *sok* gene encodes the antisense RNA (suppressor of killing) which is complementary to the *hok* mRNA leader region and the

mop gene (modulator of killing) encodes the Mok protein. The *hok* and *mok* reading frames overlap (therefore the translation of Hok and Mok proteins is coupled) and the Sok-RNA inhibits the *mok* reading frame by regulating the translation of Hok proteins (Gerdes *et al.*, 2000). The *hok* gene is constitutively transcribed from a weak promoter but its high degree of stability allows it to be present in considerable amounts in growing cells. It is present in two configurations in the cell; the full-length RNA is inactive in translation and in antisense RNA binding (present in plasmid-carrying cells). Antisense RNA binding and translation of full-length mRNA is also inhibited by folding of the structure. A low 3' processing rate is necessary to prevent translational activation; which requires folding dependent pairing of the 3'-end and the 5'-end of the full length mRNA. In plasmid-free cells a truncated form of the *hok* mRNA exist (full length RNA shortened by 40 nucleotides at the 3'-end). This truncated mRNA is very stable, translationally active and rapidly binds Sok antisense RNA.

The unstable antisense RNA is transcribed from a strong constitutive promoter and in plasmid-carrying cells the Sok-RNA is more abundant than the *hok* mRNA. The Sok-RNA forms a complete duplex with the *hok* mRNA which is rapidly cleaved by RNase III and degraded. In plasmid-free cells the antisense Sok-RNA is rapidly degraded and the full-length *hok* mRNA is converted to truncated mRNA which unfolds into translationally active configurations and kills the plasmid-free cells (the *hok/sok* model is illustrated in Figure 11). Similar PSK systems have been discovered on several other plasmid systems (refer to Gerdes *et al.*, 2000 for a detailed description of these stability systems).

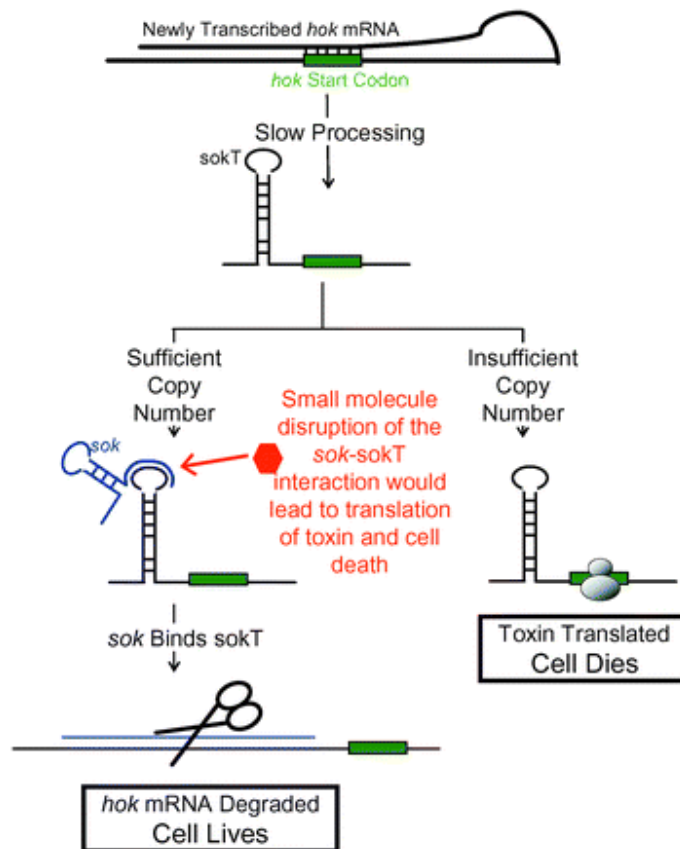


Figure 11: An example of a post-segregational killing system. In cells carrying plasmids transcription of the *hok* gene leads to full-length inert *hok* mRNAs. Full-length mRNA's build up because of the slow 3' processing rate and truncated mRNAs are generated. Sok-RNA forms a duplex with the truncated, refolded mRNA and this prevents its translation. The duplex is cleaved by RNase III and inactivated. The *sok* (suppressor of killing) gene binds to *sokT* (antisense RNA target in *hok* mRNA) and degrades *hok* mRNA. In plasmid-free cells the suppression of *hok* genes are uplifted and the cells are killed¹⁶.

3.2.4 MOBILIZATION AND TRANSFER SYSTEMS

Self-transmissible plasmids contain sufficient conjugation transfer systems to allow the plasmid to move from one host cell to the other. Mobilizable plasmids are not able to conjugate on their own but can be transferred by conjugation with the aid of self transmissible plasmids in the host cell (Zechner *et al.*, 2000). Conjugative gene transfer uniformly involves pair formation between the donor and recipient cell (with the exception of *Streptomyces* transfer systems). The plasmid moves as a linear single stranded molecule to the recipient cell. Transfer systems are encoded by plasmid *tra* genes and initiation takes place at the *oriT* (origin of transfer). Two major complexes are involved in conjugative

¹⁶ DeNap and Hergenrother, 2005

transfer: the mating pair formation system (Mpf) and the DNA transfer and replication system (Dtr). Mpf systems are responsible for donor-recipient associations such as holding the mating cells together and constructing the conjugation-channel and the Dtr systems combines DNA transport and replication functions from the donor and recipient cell preparing the plasmid for transfer (Snyders and Champness, 2003). The transfer gene regions of self transmissible plasmids vary in size among plasmids. Gram negative bacteria need up to 30 kb to encode all the transfer functions (Zechner *et al.*, 2000). Mobilizable plasmids usually carry a 3 kb region containing a set mobilizable (*mob*) genes and a set of Dtr functions enabling them to move to other recipient cells. Several classes of self-transmissible plasmids have been identified in gram negative bacteria. The best studied examples of narrow host-range plasmids include IncF and Inc I while broad-host range plasmids include IncW, IncN, IncP and IncX.

It is thought that close contact is required for DNA transfer during bacterial conjugation in gram negative bacteria. This is established by physical contact of a sex pilus from the donor cell to the cell surface of the recipient cell. The pili vary between different plasmid types, some are short and rigid others can be longer and more flexible. The sex pilus retracts which stabilizes the cell surface association and a mating bridge is formed that serves as a channel for DNA transport. DNA relaxases (origin-cleaving enzymes) introduce nicks at the *oriT* site initiating DNA single strand formation that is designated for transfer to the recipient cell (Byrd and Matson, 1997). Recently Babić *et al.*, 2008 have found that the F-pilus mediates DNA transfer at considerable cell-to-cell distances. These authors have shown that there is at least 1.2µm between the donor and recipient cells, showing that direct close contact is not needed for DNA transfer for some plasmids.

Conjugative transfer systems of gram positive bacteria are not associated with pili but contact is mediated by aggregating substances. Their plasmids can be divided into four groups: (i) broad-host range plasmids such as pGO1 from Staphylococci often carrying resistance genes (Berg *et al.*, 1998); (ii) pheromone responding plasmids such as pAD1 found in Enterococci (An and Clewell, 1997); (iii) conjugative transposons capable of moving from the chromosome to a wide variety of recipients cells (Franke and Clewell, 1981) and (iv) plasmids from mycelium-forming *Streptomyces* for instance plasmid pSG5 (Muth *et al.*, 1995).

Some bacteria (Hfr strains) are capable of transferring their chromosome through a chromosomally integrated self-transmissible plasmid. Integration of the plasmid into the chromosome can be

accomplished by various mechanisms i.e. recombination between related insertion sequences on both the plasmid and the chromosome (Zechner *et al.*, 2000). Promiscuous plasmids can transfer DNA between unrelated species and could play an important role in evolution. This could help explain why genes with related functions are often very similar to each other regardless of the type of host organism. Horizontal gene transfer allows for easy access to the gene pool and new genetic traits can be established. Most plasmids can readily exchange and carry new traits between two different bacterial species.

3.3 PLASMID VERSATILITY

Plasmids have been found to carry accessory genes necessary for survival in competitive or selective environments. Transposable elements contribute to the plasticity of plasmids as well as the bacterial genome and provide a means for the rearrangement of genes between species and kingdom barriers. As illustrated in Figure 12 plasmids are extremely versatile and can consist of several mobile elements.

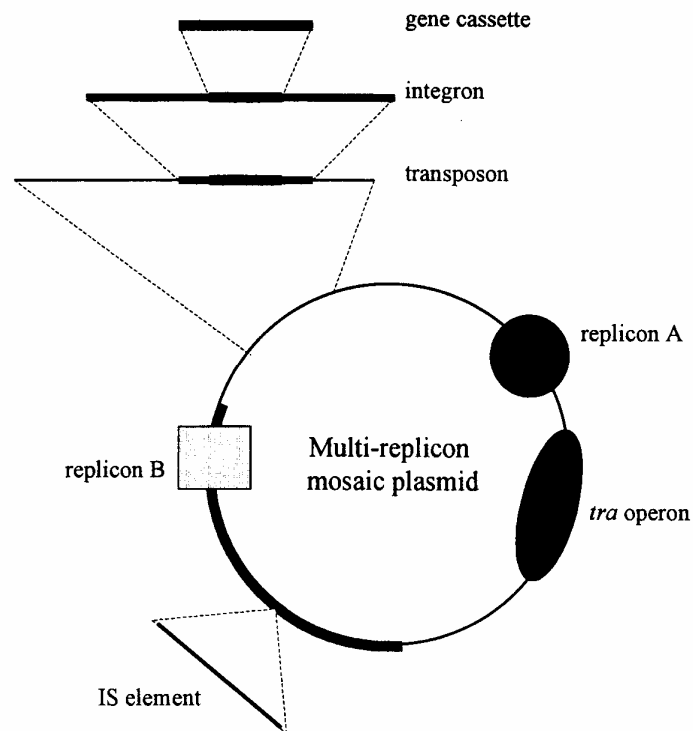


Figure 12: A diagrammatic illustration of the various types of transposable elements found on plasmids¹⁷.

Insertion sequences (IS) range in size from 700-3200 bp (with average length of 1200 bp). They consist of a transposase gene (essential for transposition) that is usually flanked by two almost perfect inverted repeats (IR) of 9 - 41 nucleotides (Merlin *et al.*, 2000; Dale and Park, 2004). Insertion elements often interrupt genes into which they insert which lead to a loss in gene function and can ultimately influence the integrity of plasmids. Most IS are similar in their structure. However some elements such as IS900 from *Mycobacterium paratuberculosis* do not have inverted repeats.

An example of a typical IS can be seen in Figure 13. Insertion sequences often create direct repeats (DR, sequences repeated in the same orientation) because of duplication of the DNA at the insertion site. Therefore different copies of the same IS usually have different target sequence repeats depending on the point of insertion. Several IS have been completely sequenced and it was found that roughly a third of these elements are iso-elements; that is they share more than 90% DNA homology and 95% protein homology (Merlin *et al.*, 2000).

¹⁷ Merlin *et al.*, 2000



Figure 13: The general structure of insertion elements, IS1¹⁸ (768 bp). DR (direct repeat); IR_L (left inverted repeat), IR_R (right inverted repeat). Both IR's consists of 23 nucleotides.

Recombination between two IS *in trans* can lead to deletions and inversions on the plasmid or chromosome which contributes to plasmid instability and genome variation (Dale and Park, 2004). Similarly these homologous insertion sequences can lead to recombination between the chromosome and plasmid and vice versa. Some transposable elements can actually promote the expression of genes neighbouring the insertion site which is due to the promoter activity of the IS which can have a downstream effect on other genes.

It has been speculated that in the case of the biomining bacterium *At. ferrooxidans*, the presence of these mobile IS elements on the chromosome may be associated with phenotypic switching, colony morphology and the ability to oxidize iron (Holmes *et al.*, 2001). However this has not been experimentally confirmed and is merely theory at this stage. This is similar to the phenomena known as phase variation where a reversible genetic change can take place that switches the expression of specific genes on or off (Dale and Park, 2004). Multiple copies of two families of conserved IS elements (ISAfe1, renamed, previously IST1, and IST2) have been found on the chromosome of *At. ferrooxidans*. Several insertion sequences have been found in *At. ferrooxidans* IST2 (Yates, *et al.*, 1988), IST445 (Chakraborty *et al.*, 1997), IST3091 (Chakravarty *et al.*, 1999) and ISAfe (Holmes *et al.*, 2001).

Transposons are similar to insertion sequences and consist of two inverted repeats at their extremities and a transposase enzyme but they also carry a genetic marker (such as a resistance gene) and in some cases accessory genes as well. An example of a transposon is Tn3; which is shown in Figure 14. Tn3 is 5000 bp in length and has short inverted repeat sequences of 38 bp. It carries an ampicillin resistance

¹⁸ Dale and Park, 2004

gene (*bla*, β -lactamase), a transposase gene (TnpA) and a bi-functional protein TnpR that acts as the repressor and is responsible for resolution stage of transposition.

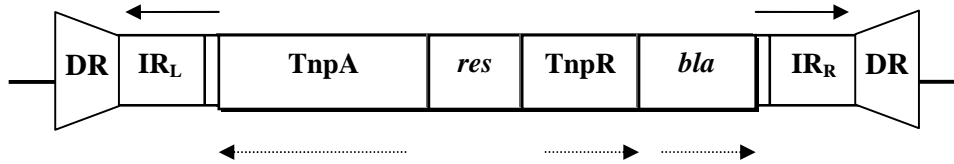


Figure 14: The structure of Tn3, a basic transposon IR's (38 bp inverted repeats), DR (5 bp direct repeats), TnpA (transposase), *res* (resolution site), TnpR (repressor) and *bla* (β -lactamase)¹⁹.

The process of moving DNA (transposition) from a donor to a recipient cell involves DNA recombination mechanisms. Basically this entails the translocation of a DNA segment by the catalytic activity of a recombinase gene. Various types of recombinases exist, which are based on (i) the enzymatic activity and (ii) specificity of transposons (iii) the translocation mechanism utilized by transposons and (iv) the products formed by the translocation process. The three main recombinases involved are transposases, integrases and resolvases (Merlin *et al.*, 2000). Recombination of integrases and resolvases involves the two terminal sites while transposases also bring the target site together. Integrases and resolvases form a covalent intermediate with the recombining DNA and promote site-specific recombination which rearranges existing sequences. Recombination of transposases leads to the synthesis of a new DNA arrangement in the recombination product. These enzymes catalyze the trans-esterification reactions and do not require high-energy co-factors.

Transposition can take place by either a replicative or non-replicative mechanism. In the former mechanism a copy of the element is inserted into a different site on a different plasmid while the original copy is reserved. The insertion site can be random or in some cases transposition can be site-specific. This type of transposition leads to the formation of cointegrate molecule consisting of both plasmids fused together and two copies of the transposon; which is illustrated in Figure 15 (Dale and Park, 2004).

¹⁹ Smalla *et al.*, 2000

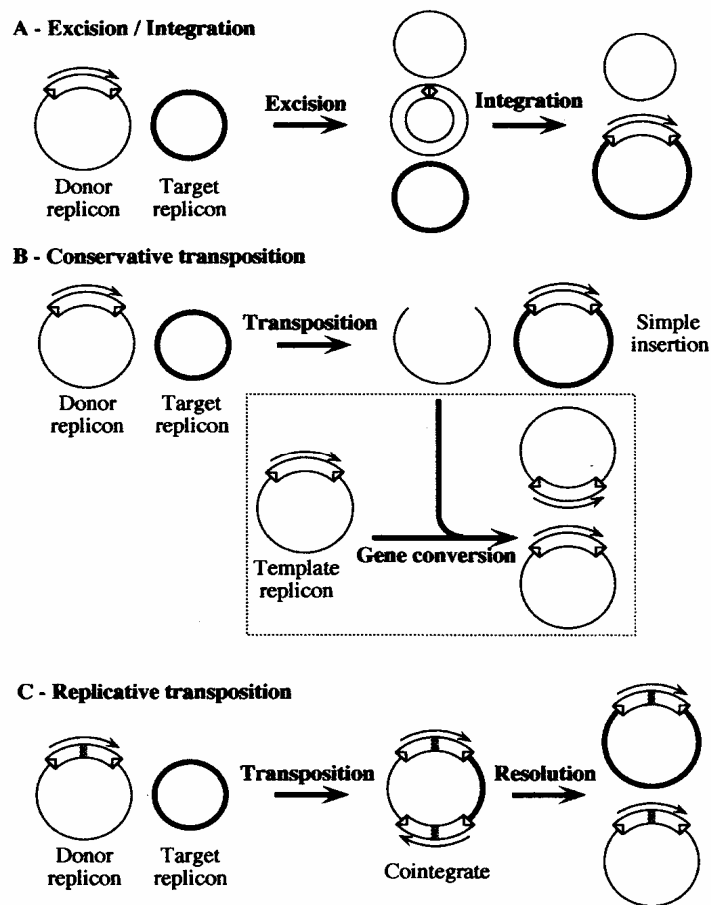


Figure 15: The different types of transposition²⁰. The open box represents the transposable element (TE) and its flanking insertion sequences (open triangles). The donor and recipient is represented by a thin and bold line; respectively. **A.** Excision/integration, two covalently closed circles are generated i.e. the donor replicon without the element and the element itself. The element can then integrate into the target replicon (integrase catalyzed reaction) generating the donor replicon without the element and the target replicon with the element. **B.** Conservative transposition, the TE is excised from the donor replicon (catalyzed by the transposase) and inserted into the target replicon; this leaves a gap that if left un-repaired the donor DNA will be degraded. The intact donor replicon can be used as a template to regenerate the donor. **C:** Replicative transposition, replication of the donor replicon takes place through the TE and generates a cointegrate in which the donor and recipient are linked by two TE elements in the same orientation. The cointegrate is resolved by a site-specific recombinase at the *res* site. The donor and the recipient both end up with a copy of the TE.

²⁰ Merlin *et al.*, 2000

The cointegrate is resolved into two plasmids; each plasmid containing a copy of the transposon while the recipient plasmid will also contain a duplication of the target sequence on either side of the inserted transposon. The resolution process can be accomplished by using the recombination system of the host or the transposon can code for its own *tnpR*, resolvase gene. The transposon- encoded resolvase ensures accurate resolution of the cointegrate through site-specific recombination (Merlin *et al.*, 2000). Several transposons have been identified in *At. ferrooxidans* for example Tn501 (Bennette *et al.*, 1978), Tn5467 (Clennell *et al.*, 1995), Tn5468 (Oppen *et al.*, 1998), as well as TnAtcArs in *At. caldus* (Tuffin *et al.*, 2005 and Kotze *et al.*, 2006).

3.4 PLASMIDS ISOLATED FROM ACIDITHIOBACILLI

Plasmids from various acidithiobacilli have been isolated and sequenced. Most plasmids have been identified from *At. ferrooxidans*. Japanese researchers found that over 73% of the *At. ferrooxidans* strains they examined contained one or more plasmids ranging in size from 2 to 30 kb (Shiratori *et al.*, 1991 and Roberto, 2003). Martin and his coworkers (Martin *et al.*, 1981) demonstrated that 11 to 15 *At. ferrooxidans* strains from the United states and Bulgaria contained at least 1 to 5 plasmids ranging between 7.4 kb and 75 kb. Other plasmids observed from *At. ferrooxidans* were isolated from South Africa (Rawlings *et al.*, 1983), Italy and Mexico (Valenti *et al.*, 1990), Chile (Sanchez *et al.*, 1986) and Canada (Martin *et al.* 1983). The occurrence of plasmids in *At. ferrooxidans* is widespread.

In the last decade we gained considerable knowledge of the maintenance systems of acidithiobacilli plasmids and their expression. Plasmid pTF-FC2 was isolated from *At. ferrooxidans* strain FC in South Africa. The plasmid backbone of this plasmid has been extensively studied. It is a 12.2 kb broad-host range plasmid with an IncQ-like replicon (*repBAC*) (Dorrington *et al.*, 1990). The mobilization system consists of five *mob* genes (*mobABCDE*). pTC-FC2 has a proteic poison-antidote plasmid stability system (plasmid addiction system genes, *pasABC*) (Clennell *et al.*, 1995; Rawlings *et al.*, 1993; Rawlings and Kusano, 1994; Rawlings 2001 and 2005; Rohrer and Rawlings, 1993 and Van Zyl *et al.*, 2003). Accessory genes were identified on a 3.5 kb transposon-like element (Tn 5467) which contains a glutaredoxin and a MerR-like regulator protein, these genes have not been shown to confer phenotypic effects onto their host. The glutaredoxin gene has been shown to complement *E. coli* thioredoxin mutants for several thioredoxin-dependent functions (Clennell *et al.*, 1995). The transposon seems defective in transposition.

The narrow host range plasmid TF1 (6.7 kb) was isolated from *At. ferrooxidans* ATCC33020 (Holmes *et al.*, 1984). This plasmid remains cryptic, the mobilization region has been examined, *mobL* and *mobS* genes with similarity to the mobilization genes of broad-host range IncQ-plasmid RSF1010 have been identified (Drolet *et al.*, 1990 and Drolet and Lau, 1992) and a 1.3 kb replicon related to pSC101 and an open reading frame with similarity to the DNA binding KfrA protein of IncP plasmid PK2 has been identified on pTF1 (Lau *et al.*, 1993).

Broad-host range plasmid pTC-F14 (14 kb) was isolated from *At. caldus* (Gardner *et al.*, 2001). Similar to the mobilization region of pTC-FC2 it also carries a 5 gene *mobABCDE* region, IncQ-plasmid-type replicon (*repBAC*) and a *pasAB* stability system (Rawlings, 2001 and 2005, van Zyl *et al.*, 2003).

Plasmid pTFI91 (9.8 kb) was identified in four strains of *At. ferrooxidans*. The replication system of pTFI91 is not clearly defined although this type of replicon seems to occur frequently in plasmids isolated from *At. ferrooxidans* strains. This replication system has been cloned but it has not been shown to be functional in *E. coli* and *Pseudomonas putida* (Chakravarty *et al.*, 1995).

Plasmid pTF5 is 19.8 kb in size and was isolated from *At. ferrooxidans*. It belongs to the pTFI91 family which has a conserved replicon region. Fourteen open reading frames have been identified with predicted functions relating to plasmid maintenance proteins. A 5.6 kb accessory region containing a [3Fe-4S,4Fe-4S] ferredoxin-like protein, FNR regulator-like protein, prismatic-like protein and a NADH oxidoreductase-like subunit of methane monooxygenase (involved in redox reactions) have been identified (Dominy *et al.*, 1998). A copy of this accessory region was also present on the chromosome. Several incomplete transposon elements and transposon scars were also observed.

Recently our research group isolated four plasmids from various *At. caldus* strains. The largest one of the four is a 66.4kb plasmid pTcM1, which has now been fully sequenced. An 18 kb region of this plasmid has been sequenced as part of this project (Van Zyl *et al.*, 2008^a). A 39.1 kb plasmid pTcF1 was isolated from strain “f” and a 28.9 kb plasmid pC-SH12 was isolated from strain C-SH12 (DSM 9466) (Van Zyl *et al.*, 2008^a). The fourth plasmid pTc6 (40 kb) was identified from *At. caldus* strain #6, unfortunately all attempts to clone this plasmid were unsuccessful. Another 9.5 kb plasmid pTK1 from *At. caldus* strain KU has also been identified (K. Hallberg personal communication, unpublished, Rawlings). The backbone functions of these three plasmids pTcM1, pTcF1 and pC-SH12 were analyzed. The replication/ partitioning region of pTcM1 was found to be related to that of the 47.8 kb

plasmid pRSB101 (Szczepanowski *et al.*, 2004) and pRSB105 (Schlüter *et al.*, 2007). Plasmid pC-SH12 has a conserved MobA domain to the MobA protein of RSF1010 and MobL domain to the MobL protein of *At. ferrooxidans* (Van Zyl *et al.*, 2008^a). Several insertion sequences have been found in pTcM1 and pTcF1. Arsenic resistance genes previously only found on the chromosome of *At. caldus* strains exposed to high levels of arsenic have been identified on some of these plasmids (Tuffin *et al.*, 2005). Plasmid pTcF1 contains *Tnars* genes that differ from *TnAtcArs*, a transposon from *At. caldus* #6, which also differs from the *ars* region of pTcM1. Plasmid pTcM1 contains putative HNH endonuclease- and modification-like open reading frames, it carries the 5 gene region similar to pTC-FC2 and pTC-F14 and a *pasAB* plasmid addiction system and homologues to the *repA* and *repC* of pTC-FC14. pTcF1 was found to occur together with pTC-F14 in strain *At. caldus* “f”.

As described at the beginning of this chapter mobile elements offer a selective advantage to their host cell to out-compete or eliminate their competitors. Most of the plasmids from *Acidithiobacillus* sp. are cryptic and very little research has been conducted on their accessory regions. Most of the research has been focused on the backbone regions of these plasmids and examining backbone gene functions in *E. coli* strains. These plasmids share features that resemble that of the IncQ-like plasmids which could indicate that the plasmids from acidithiobacilli might have evolved to increase the fitness of these acidophiles in biomining environments. The pTF-like plasmids (pTF1 and pTF5) share a common backbone with each other and are more closely related to the pRSB101-like plasmids and their replication systems. The MobA-like domain/gene on pTF1 seems to be related to the RSF1010 mob gene.

The accessory genes identified on these plasmids are related to metabolism functions. They could contribute to the host bacterium’s ability to utilize nutrients in its environment. These genes could form part of the horizontal gene pool as they are located on mobilizable plasmids. Some of the genes have also been identified on the chromosome of their hosts which could indicate that the genes might have originated from the chromosome of an ancestral bacterium and have been incorporated into plasmids through homologous recombination events.

AIMS OF THIS PROJECT

At the time that this project was begun, almost nothing was known about the plasmid and the aim of this study was to sequence an 18 kb region of the plasmid pTcM1, isolated from *At. caldus* strain MNG. As a result of this and other projects, the sequence is now known, see Figure 16. This project was one of several that focused on the determination of the types of plasmid backbone and accessory genes carried on plasmids from biomining bacteria.

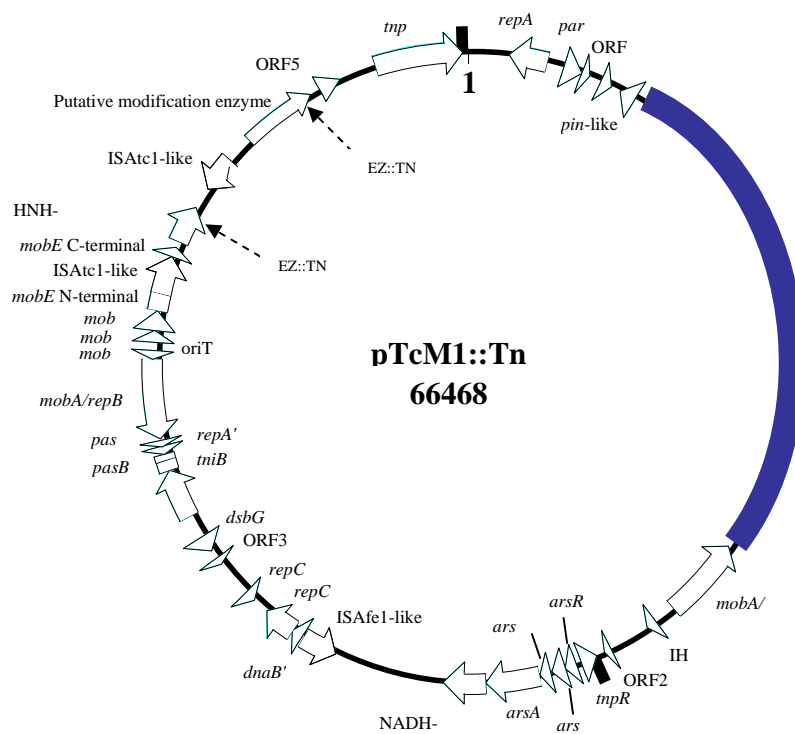


Figure 16: The physical map of plasmid pTcM1 from *At caldus* MNG²¹. The 18 kb region is indicated in blue.

²¹ Van Zyl *et al.*, 2008^a

The objectives of this project were:

1. To sequence the 18 kb region of pTcM1 shown above and identify the genes and other features present.
2. To determine whether these genes are similar to those found on plasmids of related biomining and other bacteria, or if there appear to be any unique traits present only on plasmids of bacteria found in biomining environments.
3. To determine whether the genes present are expressed, including how the genes are grouped in operons and determine what the role of these accessory genes are.
4. To determine whether these genes are also found on the chromosome of the host and other isolates of *At. caldus*, or if they belong to the horizontal gene pool and if possible to determine their origin.

CHAPTER TWO: SEQUENCING RESULTS

1. INTRODUCTION

Our research group found that plasmids pTcM1, pTcF1 and pC-SH12 had a 26 kb region in common. It is thought that the plasmid backbone genes of the three plasmids are similar and located in the common region. Restriction mapping, sequencing and Southern hybridization experiments have shown the putative replicon and partitioning system of these bacteria are related to previously reported plasmids pRSB101 and pRSB105, which were isolated from activated sludge. The plasmid backbone of these plasmids therefore does not seem to be unique to the biomining environment. It is hoped that the accessory genes also thought to be located in the common 26kb region, might be common to these bacteria and possibly unique to biomining bacteria.

Plasmid pTcM1 was isolated from *At. caldus* strain MNG by means of the CsCl/Ethidium bromide gradient centrifugation method and rescued with the Epicentre transposon EZ::TN in vitro transposition system. The transposon contains a replicon capable of replication in *E. coli pir*⁺ and a selectable kanamycin marker (Figure 17). It randomly inserts into the plasmid and the cloned plasmid can be selected from kanamycin resistant transformants. The plasmid can then be sequenced in both directions using specific primers for the transposon. The plasmid can also be sequenced using primer, but this is time consuming and expensive, especially in the case of such a large plasmid.

The transposon used to rescue pTcM1 was inserted at two different locations. In the first clone designated pTcM1::Tn#11, the insertion took place into the HNH endonuclease-like gene. The second clone designated pTcM1::Tn#12 represents the insertion into the modification/methylase gene (Van Zyl *et al.*, 2008^a). Constructs were made from pTcM1 for subcloning and sequencing by cutting the plasmid with various restriction enzymes that would allow more manageable subclones of 10kb and smaller to be generated. A specific 18.15 kb region of Tn#11, flanked by two *Xho*I endonuclease restriction sites, was cloned and is the subject of this study. The first step in analyzing this piece of DNA was to make overlapping subclones ranging in sizes of 0.5 kb and 1.5 kb which could be sequenced in both directions. This chapter focuses on the sequencing results and the open reading frames (ORF's) that were identified on this piece of DNA.

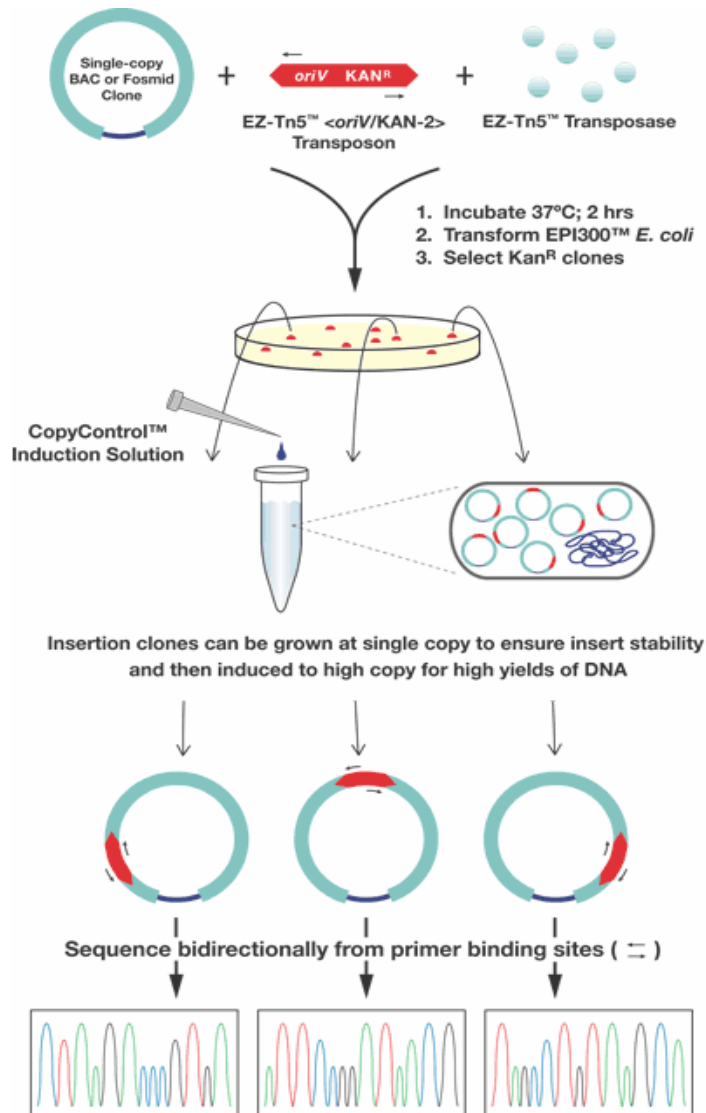


Figure 17: An illustration of the process involved in cloning and sequencing a plasmid, BAC or fosmid clones using the Epicentre EZ::TN™ <R6Kγori / KAN-2> Insertion Kit. The plasmid is mixed in vitro with the transposon, transposase is added and the reaction mix is incubated at 37°C. The transformants are selected from kanamycin containing plates. The plasmid can be sequenced using transposon specific primers.

2. MATERIALS AND METHODS

MEDIA AND GROWTH CONDITIONS

E. coli strains used for transformation reactions were grown at 37°C in Luria-Bertani (LB) (see Appendix 1 & 2) broth and agar medium supplemented with ampicillin (100µg/ml) and kanamycin (50µg/ml) as needed (Sambrook *et al.*, 1989).

2.2 DNA ISOLATION, PURIFICATION, CLONING AND SEQUENCING

Plasmid DNA was cut with various restriction enzymes according to the manufactures recommendations (Roche[®], Jenna[®] and Fermentas[®] enzymes) and separated during electrophoresis on 0.8% agarose gel and occasionally 0.6% gels depending on the size of the fragments. DNA fragments were cut from the gel and purified using the GFX[™] PCR DNA and gel band purification kit (Amersham[®] Biosciences Corporation), CLN Talent kit (Cleanmix TA 050 CLN talent kit) or the Illustra[™] DNA and gel purification kit (GE Healthcare). Plasmids were isolated and purified on a small scale using standard current protocol procedures (Sambrook *et al.*, 1989). For large scale plasmid isolations the alkaline lyses and CsCl/Ethidium bromide gradient centrifugation method or the Nucleobond AX100 kit[™] (Macherey-Nagel) were used.

Clones used for sequencing were isolated and purified using a High Pure Plasmid Isolation Kit[™] (Roche[®], Basel, Switzerland) or the Wizard[®] Plus SV minipreps. Dideoxy-chain-termination method and the ABI PRISM[™] 377 automated DNA sequencer was used for sequencing. Primer walking was used to close the gaps. Software programs such as the PC-based DNAMAN (version 4.1, Lynnon Biosoft) and Glimmer 2.02 were used to analyze and annotate the sequence. The gapped-BLAST program of the National Center of Biotechnology (NCBI, www.ncbi.nlm.nih.gov) was used to determine open reading frames (ORF) and characterize proteins. Phylogenetic trees of protein homologies were created using the multiple alignment and tree output program of DNAMAN. Protein sequences used to construct phylogenetic trees and multiple alignments were obtained from the NCBI website, the European Bioinformatics Institute (www.ebi.ac.uk) and the Swiss Institute of Bioinformatics UNIPROT (www.ebi.ac.uk/uniprot) websites.

PCR reactions were performed using 1U of GoTaq[®] DNA polymerase (catalogue number: M8305) from Promega with a Hybrid[®] PCR Sprint cycler. The primers PWK1, PWK2, PWK3, PWK4 were used to determine nucleotide sequences located between 14256 bp and 16532 bp (Table 5). Template DNA of pTcM1 was used at a concentration of 0.1µg/µl. The initial denaturing step was performed at 94°C for 120s; this was followed by 25 cycles alternating between the denaturing step at 94°C for 30s, the annealing step at 50°C for 45s and the elongation step at 72°C for 30s. The PCR products were cloned into pGEM[®]-T vector system (Promega) and the sequence was determined by using the Dideoxy-chain-termination termination method and the ABI PRISM[™] 377 automated DNA sequencer with the standard M13 forward and reverse primers.

2.3 BACTERIAL STRAINS, PLASMIDS AND PCR PRIMERS

The bacterial strains and plasmids used in this study are shown in Appendix 2. The primer sequences used for primer walking to close the gaps in sequencing is shown in Table 1.

Table 1: Primers used for sequencing.

Primers used for primer walking	Sequence (5'-3')	
pTcM7.5_PWK1	5'-GACAGTAGAGGGATTTATTGGT -3'	This study
pTcM7.5_PWK2	5'-TCACTGCCCTTTTTATAAGGAGG -3'	This study
pTcm7.5_PWK3	5'-TACAAATCTTTGCGGAGAAGTAT-3'	This study
pTcM7.5_PWK4	5'-TAGCTGCCAATAGCATCTGTTTAA-3'	This study
pTcM3.9_PWK5	5' -GCAAGGGGGAAGTCCTCA-3'	This study

3. RESULTS

3.1 CLONING AND RESTRICTION ENZYME MAP

The 18.15 kb *XhoI/XhoI* region was divided into 4 smaller subclones i.e. a 7.7 kb *XbaI/XhoI* clone (24: pTcM7XBXH), a 1.7 kb *XhoI/XbaI* clone (1: pTcM1.7XBXH), a 4.7 kb *XbaI/XbaI* clone (16: pTcM4.7XBXB) and a 3.9 kb *XbaI/XbaI* clone (6: pTcM3.9XBXB); which is shown in Figure 18. Subsequent plasmid constructs were made from these subclones using various restriction digest sites and cloned into the cloning vectors pBluescript SK/KS or pUCBM21. All the plasmid constructs are listed in Appendix 2. All subclones were sequenced in both directions; there was difficulty in sequencing the *XhoI*-end of the 7.7 kb *XhoI/XbaI* clone because of a lack of restriction enzyme sites. To overcome this problem primer walking was used, starting with the *XhoI* end sequence obtained from the 7.7 kb *XhoI/XbaI* clone.

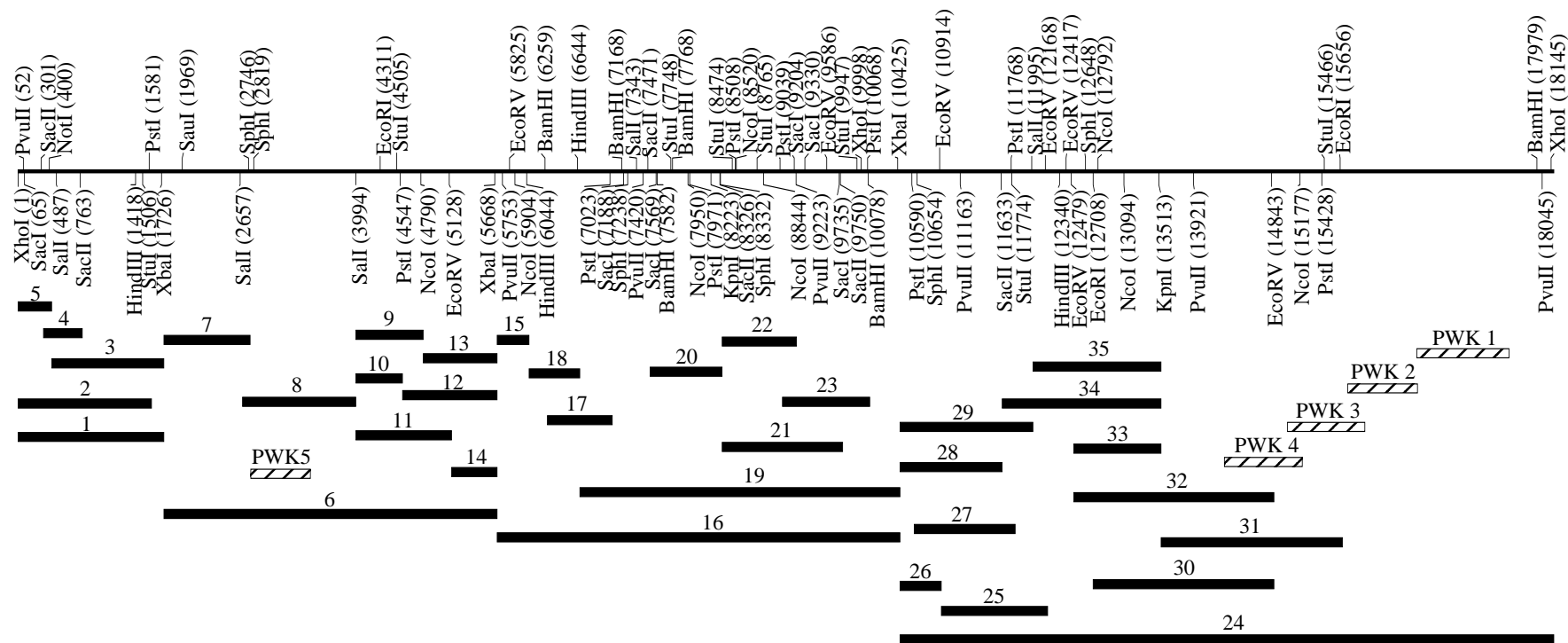


Figure 18: The physical and genetic map of the 18.15 kb *XhoI/XhoI* region located on pTcM1. The black bars represent the various plasmid constructs made for sequencing. The crossed bars represent sequencing obtained by primer walking. The various plasmid constructs made during this study are summarized in Appendix 2 and illustrated with numerical letters.

3.2 DNA ANALYSIS

The nucleotide sequences were assembled and submitted to the NCBI, BLAST results of this region identified 15 putative open reading frames. The proteins that are most closely related to those identified on pTcM1, according to BLAST results, are summarized in Table 2 and the position and orientation of the putative open reading frames (ORF's) are shown in Figure 19. The annotation of this region is shown in Appendix 5.

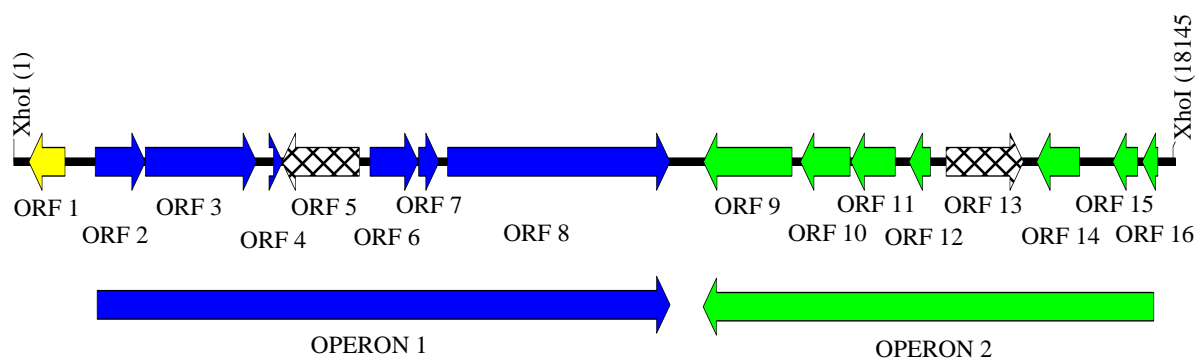


Figure 19: A map of the putative ORF's identified on the region. The best results are summarized in Table 2. The putative ORF's are indicated with the boxed arrows.

The first ORF identified (ORF 1) is located between 796 bp and 229 bp and encodes a possible protein product of 188aa. ORF 1 shares an 85% similarity to an invertase/recombinase protein from *Pseudomonas aeruginosa* and 84% similarity to a invertase recombinase-like protein from *Xanthomonas axonopodis* pv. Citri str. 306. A probable ribosomal binding site (RBS) was identified 8 bp upstream of the transcriptional initiation site (GTG) (Appendix 5).

Site-specific recombinases play a role during transposition events such as the excision or insertion of a piece of DNA. Recombinases can be divided into two families namely the integrase family and the invertase/resolvase family. Integrases generally have low sequence similarity but have four perfectly conserved amino acid residues (HRRY) and introduces staggered 6 – 8 bp cuts to their target sites generating recessive 3'-ends (Crellin and Rood, 1997). This conserved residues could not be identified in ORF 1.

Table 2: Most related protein, proposed function and predicted size.

Putative protein or ORF	Position	Ribosomal Binding site	Start codon	Size amino acids	Most related protein and proposed function and predicted size	Identity/similarity (part of protein)	BLAST E value	Reference NCBI accession no.
ORF 1	796-229	GGAG	GTG	188	Invertase/recombinase (<i>Pseudomonas aeruginosa</i>)	160/188	2e ⁻⁶⁶	YP_245443.1
ORF 2	1304-2056	GGAG	GTG	248	Transcriptional regulator (<i>Pseudomonas fluorescens</i> Pf01)	172/240	9e ⁻⁶⁰	ZP_348893.1
ORF 3	2059-3795		ATG	574	Succinate Dehydrogenase / fumarate reductase, flavoprotein subunit (<i>Pseudomonas aeruginosa</i> C3719)	390/559	0.0	ZP_00969386.1
ORF 4	3814-4047	GGAGAA	ATG	90	Ferredoxin, 4Fe-4S (<i>Ralstonia pickettii</i> 12J)	58/73	2e ⁻¹⁷	ZP_01662086.1
ORF 5	5400-4185	AGGAG	ATG	404	Transposase-like protein of ISAtc1 (<i>At. caldus</i>)	398/404	0.0	NP_835386.1
	5474-4164				Direct repeats of 7 bp of Transposase-like protein of ISAtc1 (<i>At. caldus</i>) [TAACTAT]			
ORF 6	5510-5959			N/A	Putative phycobiliprotein (<i>Pseudomonas putida</i> UCBPP-PA14, 321aa) (Truncated)	56/154	1e ⁻¹³	YP_790943.1
					Putative COG1413 FOG: HEAT repeat (<i>Pseudomonas aeruginosa</i> 2192, 321aa) (Truncated)	56/154	2e ⁻¹³	ZP_00975351.1
ORF 7	6328-6651	GGGG	ATG	107	Ferredoxin, 4Fe-4S (<i>Halorhodospira halophila</i> SL1)	66/106	5e ⁻³⁸	YP_001003224.1
ORF 8	6800-10261	GGGG	ATG	1160	NADPH-dependent glutamate synthase beta chain and related oxidoreductases (<i>Magnetospirillum magnetotacticum</i> MS-1)	509/1121	0.0	ZP_00056229.2

Putative protein or ORF	Position	Ribosomal Binding site	Start codon	Size amino acids	Most related protein and proposed function and predicted size	Identity/similarity (part of protein)	BLAST E value	Reference accession no.	NCBI
ORF 9	12162-10777		TTG	429	Hypothetical protein; putative signal peptide (<i>Acinetobacter</i> sp. ADP1)	102/429	3e ⁻¹⁰	YP_044995.1	
					Or Outer-membrane porin (<i>Shewanella putretaciens</i> CN32)	67/274	1e ⁻¹⁰	YP_001181789.1	
ORF 10	13070-12289		ATG	273	ABC-type nitrate/sulfonate/bicarbonate transport system (<i>Magnetospirillum magneticum</i> AMB-1)	117/214	4e ⁻⁶³	ZP_01615336.1	
ORF 11	13771-13085	GAAG	TTG	275	Binding-protein-dependent transport systems inner membrane component (<i>Nitrobacter winogradskyi</i> Nb-255)	111/227	2e ⁻⁵⁴	YP_317291.1	
ORF 12 ⁱ	14499-13993			N/A	ABC Transporter sulfonate / nitrate transport system substrate binding protein (<i>Nitrobacter winogradskyi</i> Nb255, 474aa) (Truncated)	69/170	3e ⁻²⁸	YP_317292.1	
ORF 13	14572-15786	AGGAG	ATG	404	Transposase-like protein of ISAtc1 (<i>At. caldus</i>)	398/404	0.0	NP_835386.1	
	14491-15801				Direct repeats of 8 bp of Transposase-like protein of ISAtc1 (<i>At. caldus</i>) [AAAATATA]				
ORF 12 ⁱⁱ	16646-15987	GGAG	ATG	N/A	ABC Transporter sulfonate / nitrate transport system substrate binding protein (<i>Nitrobacter winogradskyi</i> Nb-255, 474aa) (Truncated)	112/198	2e ⁻⁵⁶	YP_317292.1	
ORF 14	17568-17173	AGGGA/ AAGGAA	ATG	131	Hypothetical protein pRSB101_23 (uncultured bacterium)	101/131	1e ⁻³⁰	YP_133856.1	
ORF 15	17816-17565	GGAGGG	GTG	83	Hypothetical protein pRSB101_24 (uncultured bacterium)	54/83	6e ⁻²⁵	YP_133857.1	

ⁱ169aa of the end of the ABC transporter-like gene

ⁱⁱ219aa of the beginning of the ABC transporter-like gene

Members of the integrase/resolvase family display a high degree of sequence similarity within their N-terminal regions. These members make staggered 2 bp cuts within the crossover site of the DNA backbone leaving recessed 5'-ends. In Figure 20 the various invertase/recombinase proteins were aligned with that of ORF 1; illustrating the high homogeneity and relatedness between the invertase/recombinase family members.

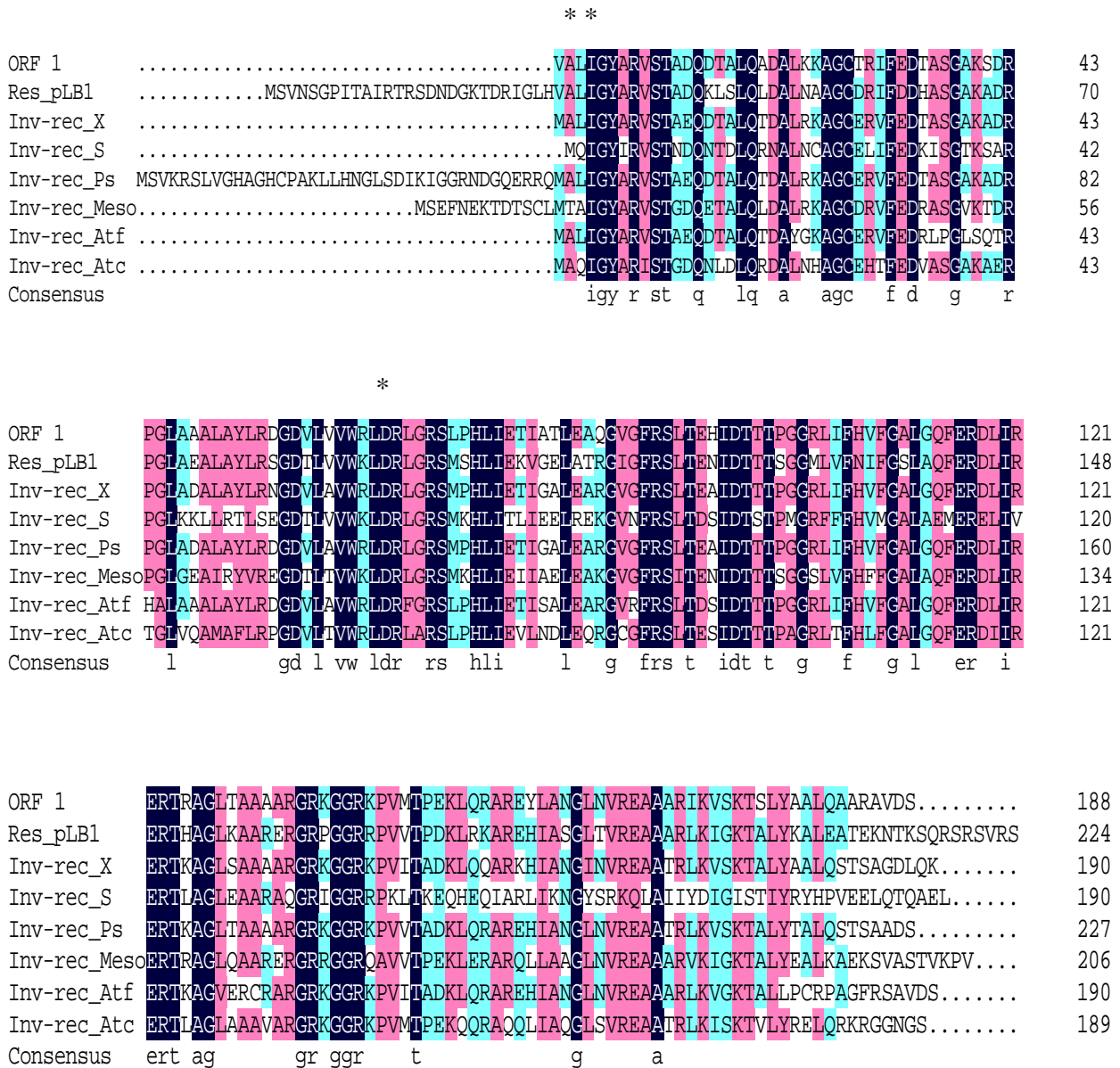


Figure 20: The invertase/ resolvase proteins with the highest homology to that of ORF 1 are shown. Sequence conservation among the invertase/recombinase family is illustrated by the shading patterns. The perfectly conserved residues are indicated with small caps at the bottom. The three highly conserved amino acid residues are indicated with stars. Res_pBL1: Resolvase of plasmid pBL1

(YP_740317.1); Inv-recX: Invertase/recombinase of *Xanthomonas axonopodis* pv. Citri str. 306 (NP_644692.1); Inv-rec_S: Invertase/recombinase of *Sulfitobacter* sp NAS14.1 (ZP_00964570.1); Inv-rec_Ps: Invertase/recombinase of *Pseudomonas aeruginosa* (YP_245443.1); Inv-rec_Meso: Invertase/recombinase of *Mesorhizobium loti* MAFF303099 (NP_085804.1); Inv-rec_Atf: Invertase/recombinase of *Acidithiobacillus ferrooxidans* (NP_863594.1); Inv-rec_*Acidithiobacillus caldus* (NP_835383.1).

ORF 2 is located between 1304 bp – 2056 bp and displays a 75% similarity and 45% identity at nucleotide level (blastx) to the transcriptional regulator GntR family of *Pseudomonas fluorescens*. The ribosomal binding site for this ORF seems to be a tetranucleotide GGAG site located at 1296 bp. The GntR protein family of transcriptional regulators consists of more than 2000 members and is found in a wide range of bacteria and archaea. This family consists of a helix-turn-helix (HTH) motif at the N-terminal called the binding domain (D-b). The binding domain is coupled to a signaling domain called the effector binding/ oligomerization domain (E-b/o) which responds to small stimuli such as carbohydrate effector molecules that activates the AraR repressor protein from *Bacillus subtilis* (Haydon and Guest , 1991, Gorelik *et al.*, 2006, Rigali *et al.*, 2002). It has been shown that these proteins are autoregulatory and mostly act as repressor proteins. They are involved in the regulation of biological processes and have been shown to play a role as environmental sensors controlling gene expression in response to external stimuli (Hillerich and Westpheling, 2006).

The HTH-motif is a highly conserved DNA recognition motif and consists of two α -helices packed at an angle of 120° followed by a turn of four tightly arranged residues. The second α -helix is known as the recognition helix and is inserted into the major groove of DNA and is critical for specificity (Huffman and Brennan, 2002). HTH proteins bind as dimers (2-fold symmetric DNA sequences) with each monomer recognizing a half-site. Based on the specificity of the HTH motifs the GntR family is subdivided into five sub-families namely MocR, YtrR, FadR, HutC and AraR, each subfamily with its unique HTH signature. The signaling domain displays a high degree of versatility (Rigali *et al.*, 2002 and Gorelik *et al.*, 2006).

Both the D-b domain and the E-b/o domain were identified at the terminal ends of ORF 2 (Appendix 5). A conserved domain similar to the PhnF transcriptional regulator was identified. The domain structure of PhnF is composed of a six-stranded antiparallel β -sheet, two-stranded parallel sheet (the core of the protein) and four short α -helices. Similar type structures were identified in ORF 2 as shown

in Figure 22. The *phnF* gene from *E. coli* belongs to a Phn operon; which is involved in the transport and biodegradation of phosphates (Gorelli *et al.*, 2006).

This protein is a member of the HutC subfamily. The HutC family represents 30% of the GntR family (Gorelik *et. al.*, 2006). The HutC-type regulators were named after the *Pseudomonas putida* regulator that inhibits the expression of histidine utilization genes. A phylogenetic relationship of the various subfamilies is shown in Figure 21. The average length of the C-domain of the HutC subfamily is 170aa. In this family the E-b/o topology varies greatly among the members but the core D-b domain is well conserved and the secondary structure elements are in similar relative positions. The topology of the core region is $\alpha_1\alpha_2\alpha_3\beta_1\beta_2$; the HTH motif is formed by the α_2 and α_3 helices (Rigaldi *et al.*, 2002). The α - helices as well as the β -strand secondary structure of ORF 2 was identified by sequence alignments and the PsiPRED program as illustrated in Figures 21, 22 and 23.

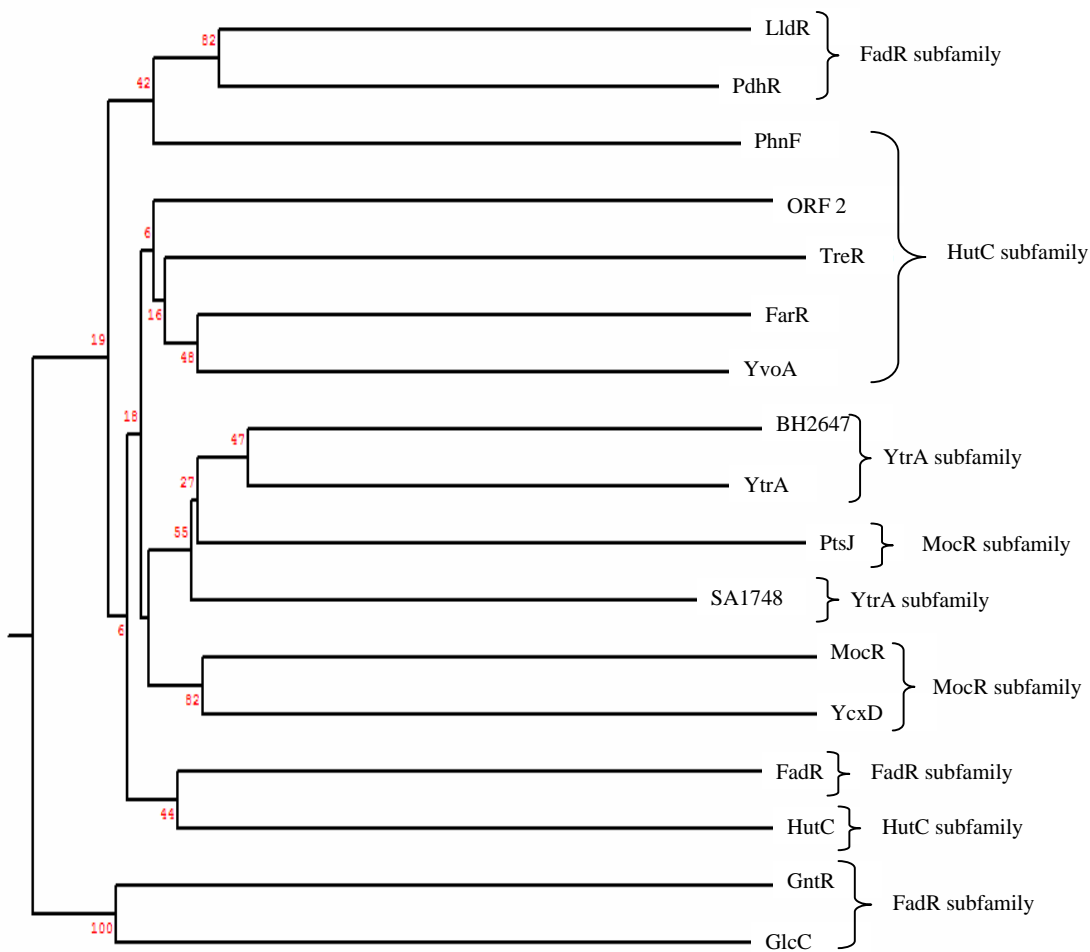


Figure 21: The relation of ORF 2 relative to the four subfamilies of the GntR regulatory proteins are shown. Swiss Prot numbers are shown in brackets: LldR_ *E. coli* (258aa; P33233); PdhR_ *E. coli* (254aa; PO6957); PhnF_ *E. coli* (241aa; P16684); TreR_ *Bacillus subtilis* (238aa; P39796); FarR_ *E. coli* (240aa; P13669); YvoA_ *B. subtilis* (243aa; Q34817); BH2647_ *B. halodurans* (123aa; Q9K9J9); YtrA_ *B. subtilis* (130aa; O34712); PtsJ_ *Salmonella typhimurium* (430aa; P40193); SA1748_ *Thermoplasma acidiphylum* (126; Q99SV4); MocR_ *Rhizobium meliloti* (493aa; P49309); YcxD_ *B. subtilis* (444aa; Q08792); FadR_ *E. coli* (238aa; PO9371); HutC_ *Pseudomonas putida* (248aa; P22773); GntR_ *B. subtilis* (243aa; P10585); GlcC_ *P. aeruginosa* (254aa; P52072);

ORF 2 protMQPLYLQIKDDIRANIVEGTYKADDKLPSEHEMVLMYNASRITIRQSIKDLQNEGLVYTI..P	61
PA01MGKLLPLSPVPLYSQLKELLRGRILDGVYPPLSRMPSENELGKAFDVSRTIVRQALGDLQKEGLIFKI..H	69
Bruc melMNVESDHAPLYAKVEAALAAEIAKGIFFPRGSRLLPEDSLIERFVGSRTIVRKAIENTVMRGLVEIR..R	67
Strep pyrMPKEQPLYLQIVDDLEVKIKKSMTEKLLSERELSDLYGVSRTIIRLTLKLELRLDIYKK..Q	63
Bac halMINKNSPLPIYYQIEEQIKQQIESGVLKPGDMLKSEREYAEYYDVSRTIVRQAINNIVNQGYIYK..K	67
XlnRMANAADDRPKYQRIADTLREGFRSGEYGPDRLPSENDLMATHGVARVTACALSVERDEGVAEARKGA	70
TreRMKVNKFITIIYKDIAQQIEGGRWKAEEILPSEHELTAQYGTSRRTIVRKAALHMLAQNGYIQKIRGK	64
PhnFMHLSTHPTSYPTRYQEIAAKLEQLRQHYRCGDYLPAAEQQLAARFEVNRHFLRRAIDQIVKQGWVQRQGV	71
HutC	MPTPPVSALVAQMGEPPAPLYARVKQMIQQIDNCSWPPHHRVPSSELVNELGFSRVTINRALRELTADGLLRM..Q	77
FarRMGHKPLYRQIADRIREQIARGELKPGDALPTESALQTEFGVSRVTVRQALRQIVEQQILESI..Q	63
YvoAMNINKQSPPIYYQIMEQLKTKIKNGELQDMPPLPSEREYAEQFGISRVTIVRQALSINLVNEGLLYRL..K	68
Consensuse r t l	

	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₄</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₃</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₅</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₄</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₅</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₆</div> </div>	
ORF 2 prot	GKGTFFISRPKVSQTLTSLQ.GFGEAMSPNGHETYSTIIISQRYGIADKLVQNKLRLLTQSQIFELQRIRYLD.REPISLDVTYL	141
PA01	GKGTFFVARPKAFQNVSTLQ.GLGESMTQMGYEVINRHLGLRYVPAGARVAERLRVEEGSPVCEIKRVRLVNRREPVSLEVITYL	150
Bruc mel	GKGTFFVAEPKIRQELTELS.GFVEDMVALGRKPTANLLDKRSVPAACEVAEHLGVAVGTQVYRIERVRLADGVAMSFDEITYL	148
Strep pyr	GKGTIVSGIKEPATDLSSYTFEQMKMGKTPKTEILSFEQYQVTPYLSGLLELDPDTEVFELERLRIADDMPLMFERSYI	145
Bac hal	GSQTYVQEKKIEQALNGLT.SFTEDMRKRGMPESSRLLKLFELIPATAKIAKELNLKENTPVTEIKRIRYGDGVPIAIERNLL	148
XlnR	GVFVREFRPLRRRGIQLRARQWNGRSIWSADIEDRSLDQVTVSEEAAPDGVAAVLDLADGETVCVRRRRFVLDGKPVL	152
TreR	GSVLNRKEMQFPVS.GLV.SFKELAQTGLGKTKTIVHKFLEPPSELIQKQLRANLDDDIWEVIRSKIDGHEVLDKDYF	144
PhnF	GVLVLMRPFDPYPLNAQARFSQNLDDQGSHTPEKLLSVLRPASGHVADALGITEGENVIHLRTRRRVNGVALCLIDHYFADL	153
HutC	GVGTFFVAEPKGRSALFEVN.NIADEIAARGHQHSCQVITLTEEAAGSERALALDMREGQRFVHSLIVHFENGVPVQIEDRYV	158
FarR	GSQTYVKEERVNYDIFQLT.SFDEKLSDRHVDTHSEVLIFEVIPPADDFLQQQLQITPQDRVWHVKKRVRYRKQKPMALLETWM	144
YvoA	CRGTFVSKPKMEQALQGLT.SFTEDMKSRRGMPGSRLLIDYQLIDSTELAAAILGCGHPSSIHKITRVRLANDIPMAIESSHI	149
Consensus	g	

	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₆</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₆</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₉</div> <div style="border: 1px solid black; background-color: #00FFFF; padding: 2px 10px; margin: 2px;">α₉</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₇</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₈</div> </div>	
ORF 2 prot	PSDVGSKLENEEDLTGK.DLFMLLESRLG.....VQLEAILAGN.....LAHLNIEQGSPILRMERLTYTSLGVP	213
PA01	PQALGERLEKADLVTR.DIFLILENDCG.....LAIDAVLADAD.....LTRALAVEEGAPIMIRIERLTHAADGAP	222
Bruc mel	PLDIGEKIASNDLEVE.PIFSLENKYG.....VQLEAVTADDH.....VAQALGIAPGSPIFLIERITTYCKGSEP	220
Strep pyr	PAQPFQGLSIADLKRKALYDIFAKDYQF.....EFYASIALDYE.....AGLLGIKKGDVPLHIIKRYNDKNILI	218
Bac hal	PANLVKGLNEEIIINQS.LYQYIEEELNLI.....V.IEASTASKT.....EADLLEIQKGSPIILLIERKTFADLGTV	219
XlnR	LATSYPALVAGSAITQEDTGKGGTYA.....IGPVHFREERIRSRMPSKAEASQLSVAAGTPVILICRTAFTDEGRA	231
TreR	FRKHVPHLTKEICENSIYEYIEGELGLS.....IIVAEPTDEDR.....ELLDLRGYDHMVVVRNYVFLDTSLFQ	217
PhnF	TLWPTLQRFDSGSLHDFLREQTGIALRR.....RRAQAKECQRLEIPNMSP.....LLCVRTLNRDGE	220
HutC	NAAIAPDYLKQDFTRQ.TPYAYLSQVAP.....VEAI.....IAEPEECRLLQIERGEPCLLIRR	219
FarR	PLALFPDLTWQVMENS.KYHFIEEVKKM.....E.IIPLMPTEE.....MSRLLNISQTKPILEKVSRYLVDGRV	215
YvoA	PFELAGELNESHFQSS.IYDHIERYNSI.....E.LEPSAATE.....EANILGIQKGAPELLIKRITTYLQNGTA	220
Consensus		

	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₈</div> <div style="border: 1px solid black; background-color: #FFDAB9; padding: 2px 10px; margin: 2px;">β₉</div> </div>	
ORF 2 prot	LDFFEYLYYRGDYLYKYSLLL SRHG.....	237
PA01	LDFFEYLYYRGDAFYRLRIDRQKGERA..	249
Bruc mel	VDEKLYYRGDLIKFCTRLSRRTR.....	244
Strep pyr	ELTFSIARADQFRYRVQHPNG.....	240
Bac hal	LELVKSAYRADRYKFMITMQR.....	240
XlnR	VEINEMTLDAASYVLEYDFDA.....	252
TreR	YTESRHRLDKFRFVDFARRGK.....	238
PhnF	SPAEYSVSLTRADMIEFTMEH.....	241
HutC	RTWGRQPVTAAARLIHPGSRHRLEGRFSK	248
FarR	FEYSRNFNNTDYKFTLIAQRKSSR....	240
YvoA	FEHAKSVYRGDRYTFVHYMDRLS.....	243
Consensus		

Figure 22: Multiple alignment of protein sequences of ORF 2 and homologous proteins from the HutC subfamily. Putative predicted α-helices and β-sheets are based on the HutC and PhnF proposed

locations and shown as shaded areas. Accession and Swiss prot numbers are shown in brackets: PA01_ *Pseudomonas aeruginosa* PA01 (249aa; NP 250989); Bruc mel_ *Brucella melitensis* 16M (244aa; NP 541093); Strep pyr_ *Streptococcus pyogenes* M1 GAS (240aa; NP 268947); Bac hal_ *Bacillus halodurans* C-125 (240aa; NP 241285); XlnR_ *S. lividans* (252aa; Q9ACN8); TreR_ *B. subtilis* (238aa; P39796); PhnF_ *E. coli* (241aa; P16684); HutC_ *Ps. putida* (248aa; P22773); FarR_ *E. coli* (240aa; P13669); YvoA_ *B. Subtilis* (243aa; Q34817)

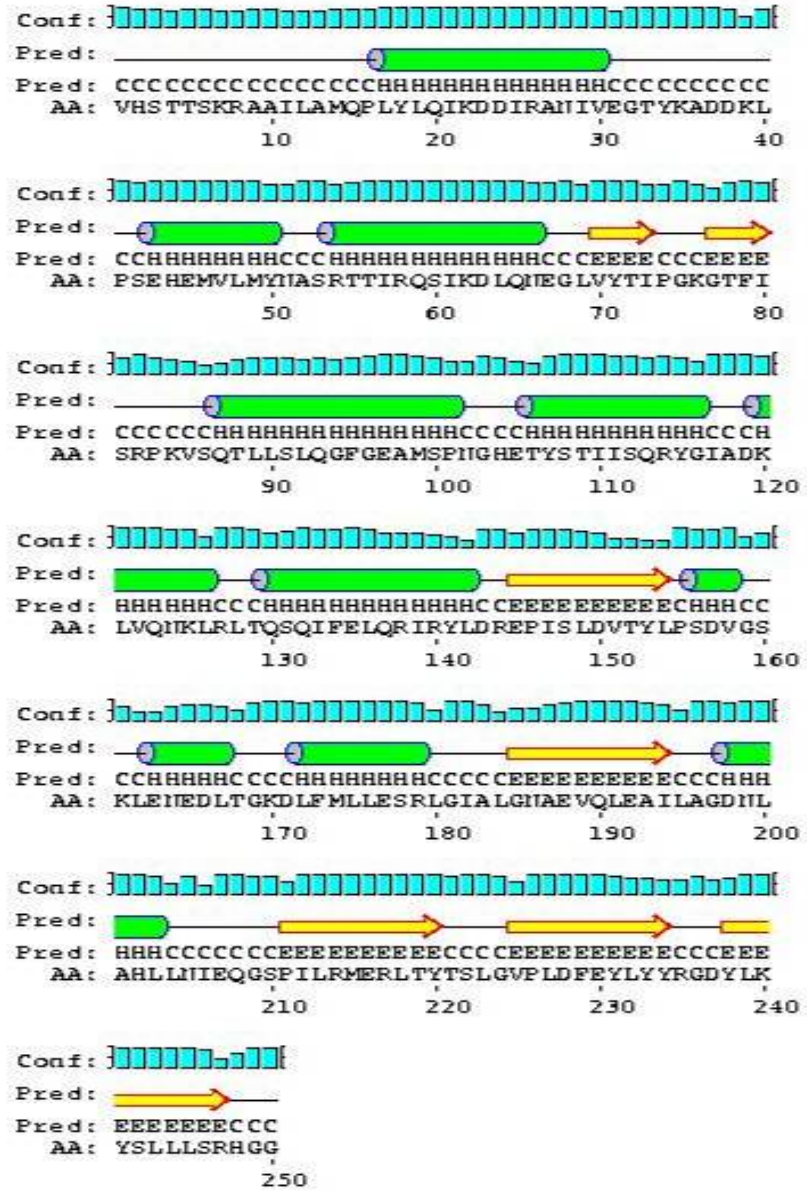


Figure 23: The PsiPRED web-based secondary structure prediction for ORF 2. The green bar represents the predicted α -helix, the yellow arrow represents the predicted β -strand.

The third ORF (ORF 3) was identified between 2059 bp – 3795 bp and displays 68% identity to a succinate dehydrogenase/ fumarate reductase (SdhA/ FrdA), flavoprotein subunit of *Pseudomonas aeruginosa*. Succinate is oxidized by the catalytic activity of succinate dehydrogenase (Sdh, or succinate/ quinone oxidoreductase) to produce fumarate as part of the TCA cycle (Figure 24). The heterotetramer enzyme has peripheral and membranous parts that has a catalytic domain (SdhA) and electron transfer subunit (SdhB) and a dimeric membrane anchor (SdhC/ SdhD).

The peripheral part is hydrophilic and the SdhA subunit contains a flavin adenine dinucleotide cofactor (FAD). The SdhA has an active site for succinate. The peripheral component also has an iron sulfur protein (SdhB subunit) which carries three iron sulfur clusters (Schirawski and Uden, 1998). The peripheral part shows great homology between a wide range of gram-positive and gram-negative bacteria as well as archaea and eukaryotes. There is heterogeneity among the sequences, subunit compositions and heme-B contents of the membranous component. The membrane anchor protein (SdhC) is thought to be involved in transport of electrons across the membrane. In many gram-negative bacteria ubiquinone is used as an electron acceptor for Sdh. These succinate/ ubiquinone oxidoreductases contain one heme B and two conserved HIS residues.

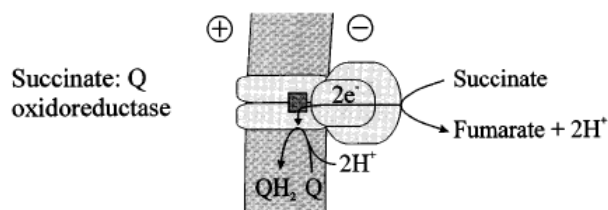


Figure 24: The oxidation of succinate to fumarate by the catalytic activities of the ubiquinone dependent succinate dehydrogenase²².

No putative ribosomal binding site could be identified for ORF 3. The conserved domain of SdhA has been identified between 2091 bp – 3690 bp; which is thought to play a role in energy production and conservation (Appendix 5). A phylogenetic tree depicting the closest related bacterial SdhA genes is shown in Figure 25. No equivalents for the SdhB, SdhC or SdhD have been identified adjacent to ORF 3.

²² Schirawski and Unden, 1998

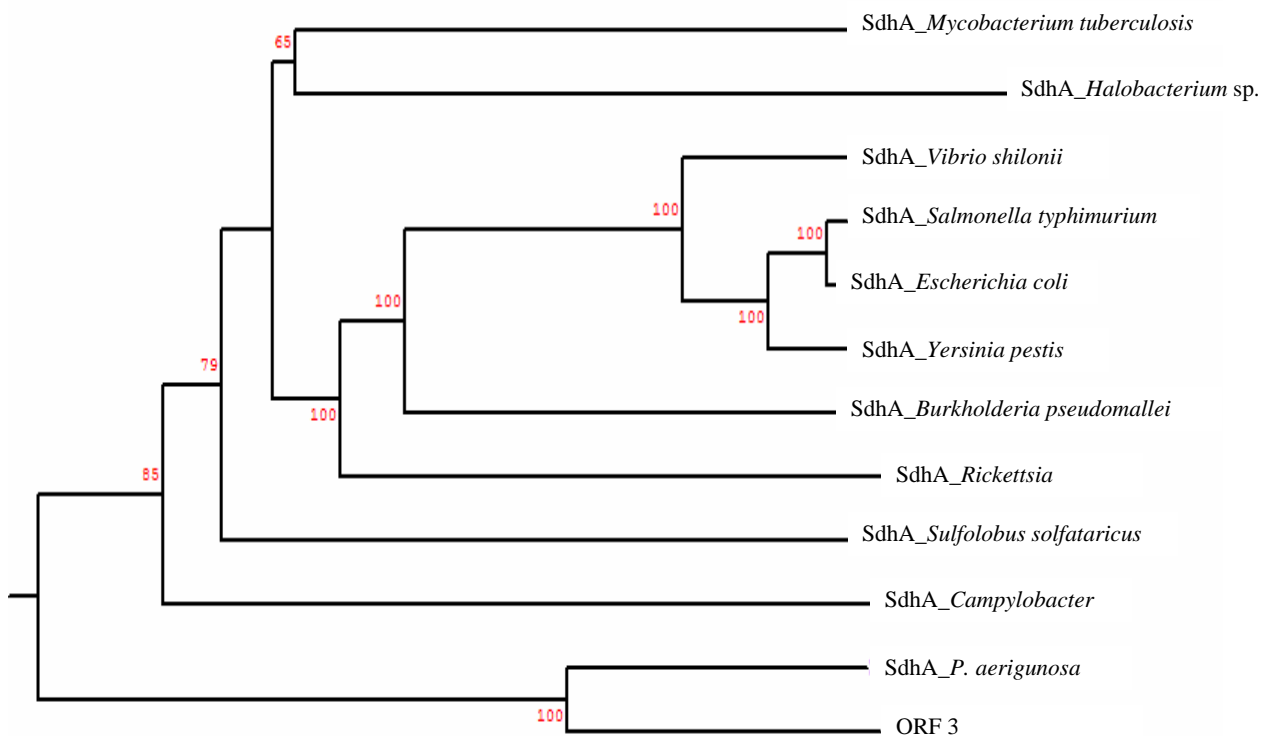


Figure 25: A phylogenetic tree of the closest homologous SdhA amino acid sequences from several organisms. Accession numbers: SdhA_*Halobacter* sp. NRC-1 (615aa; NP_280171); SdhA_*Sulfolobus solfataricus* P2 (566aa; AAK42509); SdhA_*Mycobacterium tuberculosis* H37Ra (590aa; YP_001284705); SdhA_*E. coli* 0157:H7 strain Sakai (588aa; NP_308775); SdhA_*Salmonella typhimurium* LT2 (588aa; NP_459719); SdhA_*Yersinia pestis* KIM (588aa; NP_670368); SdhA_*Vibrio shilonii* AK1 (588aa; ZP_01869369); SdhA_*Burkholderia pseudomallei* K96243 (591aa; CAH39192); SdhA_*Rickettsia prowazekii* (596aa; AAA18327); SdhA_*Campylobacter jejuni* (576aa; YP_178507).

The next open reading frame (ORF 4) of 75aa is located between 3814 bp – 4047 bp and has high similarity to a ferredoxin, 4Fe-4S from *Ralstonia picketti*. The phylogenetic tree in Figure 26 depicts the closest homologies to ORF 4. A putative ribosomal binding site was identified at 3808 bp (GGAGAA). Another ferredoxin-like protein of 107aa from *Halorhodospira halophila* SL1 was also identified (ORF 7); it is located between 6328 bp – 6651 bp and a putative ribosomal binding site was identified at 6320 bp (GGGG). ORF 4 and ORF 7 share very little homology at an amino acid level (11%).

Ferredoxin (Fd) is a natural electron carrier and ferredoxin are grouped according to the structure of its iron-sulfur clusters. These clusters are arranged as 2Fe-2S (typically associated with plants), 3Fe-4S

and 4Fe-4S cluster binding domains (Green *et al.*, 2003, Krishna *et al.*, 2006). These proteins fold into a two layer $\alpha + \beta$ structures. Bacterial ferredoxins consists of four β –strands and two α –helices with one or two [4Fe-4S] (some members also bind [3Fe-4s]) clusters bound in the core. Helical ferredoxins bind one 4Fe-4S and 3Fe-4S cluster and adopt an all helical fold.

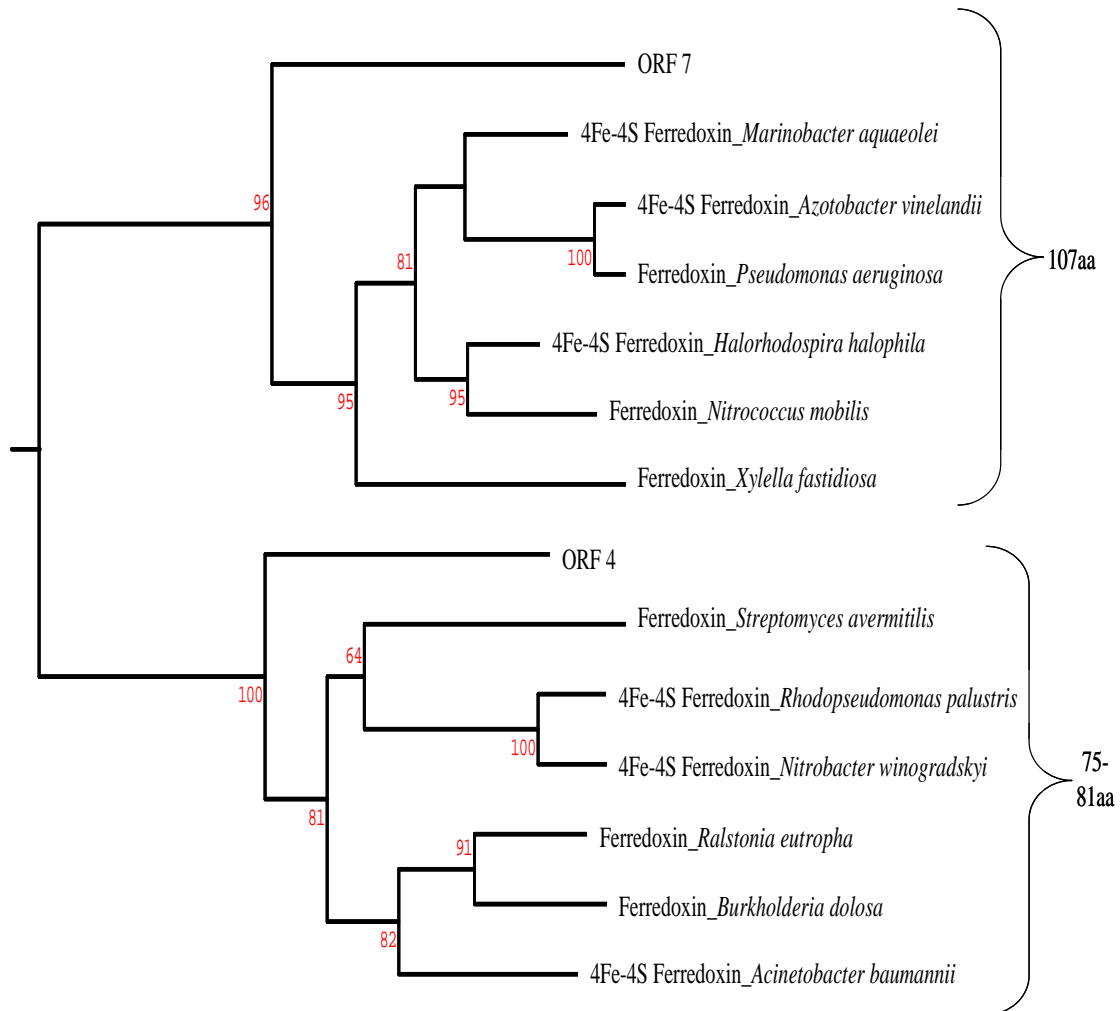


Figure 26: The phylogenetic relationship between ORF 4 and ORF 7 and their related ferredoxin genes. Accession numbers of Ferredoxins are as follows: *Marinobacter aquaeolei* VT8 (107aa; YP_959355); *Azotobacter vinelandii* AvOP (107aa; ZP_0041769); *Pseudomonas aeruginosa* C3719 (107aa; ZP_00967995); *Halorhodospira halophila* SL1 (107aa; YP_001003224); *Nitrococcus mobilis* Nb-231 (107aa; ZP_01125493); *Xylella fastidiosa* 9a5c (107aa; NP_299878); *Streptomyces avermitilis* MA-4680 (75aa; NP_826033); *Rhodopseudomonas palustris* HaA2 (77aa; YP_485100); *Nitrobacter winogradskyi* Nb-255 (77aa; YP_317295); *Ralstonia eutropha* H16 (81aa; YP_841755); *Burkholderia dolosa* AUO158 (81aa; EAY68600); *Acinetobacter baumannii* ATCC 17978 (81aa; YP_001084748).

ORF 5 is transcribed in the opposite direction to the rest of the reading frames. It is located between 5400 bp – 4186 bp it is related to a transposase-like protein ISAtc1 from *At. caldus*. ORF 13 is located

between 14572 bp and 15786 bp and it is also related the transposase-like protein ISAtc1 from *At. caldus* and has 100% homology to ORF 5. Ribosomal binding sites for ORF 5 and 13 were identified 8 bp from the transcriptional start sites (AGGAG). The direct repeats (DR) often associated with IS elements were found for both ORF 5 and 13. The DR of 7 nucleotides (TAACTAT) for ORF 5 was identified at 72 bp upstream of the transcriptional start site for ORF 5 and the DR of 8 nucleotides (AAAATATA) was found 74 bp upstream of the transcriptional start site of ORF 13. A transposase conserved domain was identified for both ORF's 5 & 13 as shown in Appendix 5. A variety of transposases from different organisms related to ORF 5 and 13 are shown in Figure 27. Insertion sequences were discussed in detail in chapter 1. IS elements can lead to recombination between plasmid and chromosomal DNA. It can also play a role in phenotypic switching or activation of metabolic pathways or colony morphology.

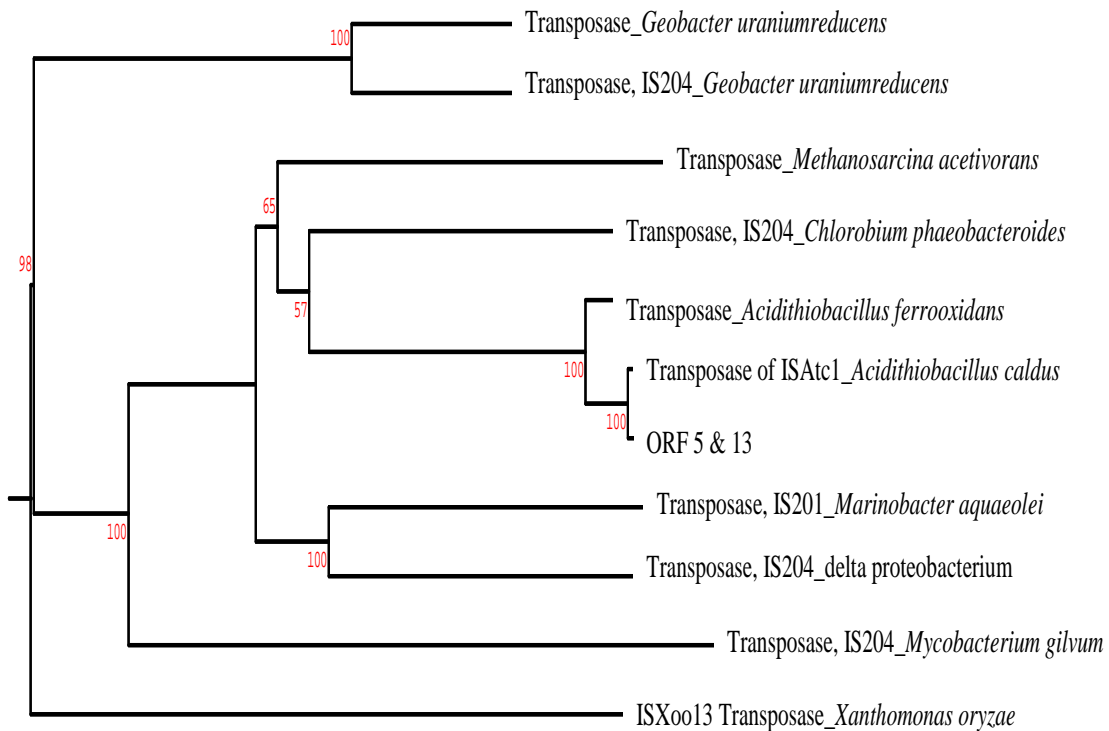


Figure 27: Various transposases have been compared to ORF 5 and 13. Accession numbers are as follows: *Geobacter uraniumreducens* Rf4 (411aa; ZP_01142339); IS204/IS1001/IS1096/IS1165_*Geobacter uraniumredecens* Rf4 (413aa; YP_001229348); *Methanosarcina acetivorans*; *Chlorobium phaeobacterioides* DSM 266 (406aa; YP_910744); *Acidithiobacillus ferrooxidans* (404aa; AAB07489); ISAtc1_*Acidithiobacillus caldus* (404aa; NP_835386); IS201/IS1001/IS1165_*Marinobacter aquaeolei* VT8 (417aa; YP_956889); IS204/IS1001/IS1096/IS1165_delta proteobacterium MLMS-1 (492aa; ZP_01288332); IS204_*Mycobacterium gilvum* PYR-GCK (432aa; YP001131597); ISXoo13 *Xanthomonas oryzae* pv. *Oryzae* MAFF 311018.

A putative open reading frame (ORF 6) was identified between 5510 bp – 5959 bp; which seems to be truncated. It has 38% identity to a PBS lyase HEAT-like repeat of *Nitrobacter winogradskyi* Nb-255 (57/148); 36% identity to a putative phycobiliprotein of *Pseudomonas aeruginosa* UCBPP-PA14 (56/154). The phycobiliproteins are involved in photosynthetic reactions by capturing light energy and transferring it to chlorophyll a. Based on their spectral properties they can be divided into 3 groups namely, phycoerythrin, phycocyanin and allophycocyanin. ORF 6 also displays a 36% identity to a FOG HEAT repeat of *Pseudomonas aeruginosa* 2192 (56/154). A conserved FOG HEAT repeat domain was found between 5693 bp – 5959 bp (Appendix 3). The presence of the conserved domain suggests that it could be a putative FOG HEAT repeat-like protein. The HEAT (Huntington-elongation-A-subunit-TOR) repeat family of tandem repeats consists of a conserved superhelical HEAT motif which is a α -helical hairpin formed by two α -helices (Park *et al*, 2005). The HEAT repeat sequences are variable in length, amino acid sequence and three dimensional structures (Andrade *et al.*, 2001). HEAT repeats are closely related to Armidillo (ARM) repeats; both these motifs are associated with eukaryotic nucleoplasmic transport, vascular transport and cytoskeletal organization. Therefore the low level of sequence identity makes it uncertain as to what the function of ORF 6 is likely to be.

ORF 8 is located between 6800 bp – 10261 bp and has 45% homology to a NADPH-dependent glutamate synthase beta chain and related oxidoreductases from *Magnetospirillum magnetotacticum* MS-1 (509/1121). It also has 45% homology to a FAD/NAD (P)-binding domain protein of *Alkalilimonoccal ehrlichei* MLHE-1 (536/1166). A possible ribosomal binding site was identified at 6767 bp (GGGG). A GltD conserved domain was found between 7473 bp and 8514 bp. As mentioned in chapter one, glutamate synthase (GOGAT) forms part of the nitrogen assimilation pathway. They are known to differ with respect to their molecular weight, subunit composition and specificity for electron donors. In Figure 28 the relatedness between ORF 8 and NADH-GOGAT, NADPH-GOGAT and Fd-GOGAT genes are shown. The product for which ORF 8 codes is smaller in size than the typical β -subunit of bacterial glutamate synthases.

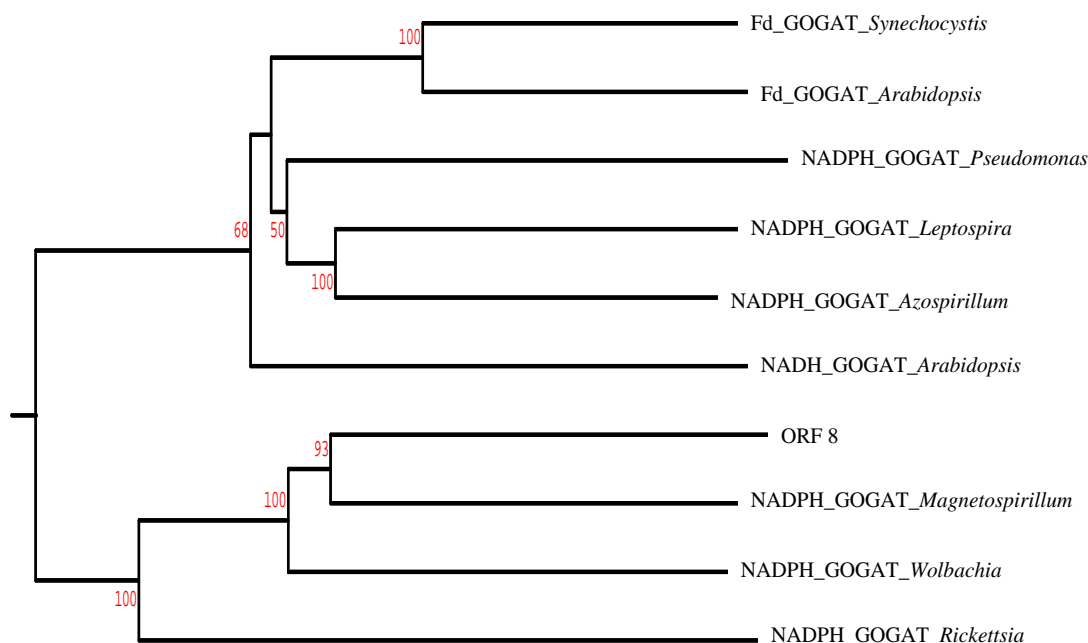


Figure 28: The phylogenetic relationship between ORF 8 and Fd-GOGAT; NADH-GOGAT and NADPH-GOGAT from other organisms. The accession numbers are as follows: Fd_GOGAT_*Synechocystis* sp. PCC 6803 (1556aa; BAA11379); Fd_GOGAT_*Arabidopsis thaliana* (1648aa; Q9ZN27); NADPH_GOGAT_*Pseudomonas entomophila* L48 (1481aa; YP_606114); NADPH_GOGAT_*Leptospira interrogans* serovar Lai Str. 56601 (1498aa; NP_711137); NADPH_GOGAT_*Azospirillum brasilense* (1515aa; Q05755); NADH_GOGAT_*Arabidopsis thaliana* (2208aa; NP200158); NADPH_GOGAT_*Magnetospirillum magnetotactim* MS-1 (1154aa; ZP00056229); NADPH_GOGAT_*Wolbachia endosymbiont* strain TRS of *Brugia malagi* (1105aa; YP_197885) and NADPH_GOGAT_*Rickettsia felis* URRWXCAl2 (1075aa; AAY60856).

ORF 9 is located between 12162 bp – 10777 bp and displays 24% homology to an outer membrane porin of *Shewanella* sp. ANA-3 and 21% homology to a putative signal peptide of *Acinetobacter* sp. ADP1. A conserved OprD domain was found between 11521 bp and 12080 bp (Appendix 5). This domain belongs to the outer membrane OprD porin family. OprD proteins facilitate the uptake of basic amino acids and small basic peptides such as arginine and gluconate (Epp *et al.*, 2001 and Tamber *et al.*, 2006).

ORF 10 was identified between 13070 bp – 12289 bp and has 54% identity to ABC-type nitrate/sulfonate/bicarbonate transport system of *Magnetospirillum magneticum* AMB-1. The diverse

Adenosine triphosphate (ATP)-binding –cassette (ABC) family of membrane transport proteins obtains their energy from the hydrolysis of ATP during the translocation of solutes across a biological membrane. This family is involved with the uptake of nutrients, signal transduction, protein secretion, antibiotic resistance, bacterial pathogenesis and several other processes. ABC transporters consist of four components: two membrane-integral domains crossing the membrane six times and two ATP-hydrolyzing domains as illustrated in Figure 29. (Schneider and Hunke, 1998).

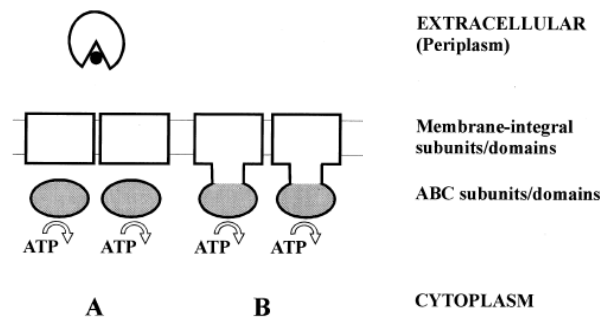


Figure 29: The organization of the ABC transporters²³. Bacterial ABC transporters are built up from individual subunits. A: Binding protein-dependent bacterial import systems; In some cases the membrane-integral domains or the ABC domains can be fused such as shown in B typical of bacterial export systems. ATP hydrolysis is coupled to the translocation of solutes across biological membranes.

The ATP-hydrolyzing domains are characterized by conserved sequences in their primary structure: Walker A and Walker B. The Walker A consist of eight nucleotide sequences namely GXXGXGKS/T (where X can vary) and Walker B constitutes a hydrophobic nucleotide binding fold (hhhhD, h: hydrophobic). The highly conserved ABC signature motif (linker peptide) is located between Walker A and B and is unique to the ABC transport family (LSGGQQ/R/KOR) (Schneider and Hunke, 1998). The membrane-integral domains contain a unique extracellular periplasmic protein component that act as a scavenger for substrate molecules and in some cases as a chemoreceptor (Schneider and Hunke, 1998).

The consensus structure of ABC domain is predicted to be folded in an alternating series of six α -helical sub-domains and five β -strands. Walker A consists of the α -helical domains and is preceded by a glycine-rich loop and Walker B consists of the β -strands. Walker A and B are linked by a helical domain or loop which is a peptide fragment of approximately 100 residues that folds into α -helical

²³ Schneider and Hunke, 1998

formation. It is thought that the Walker A motif is crucial for the binding of the β - and γ -phosphates of nucleotides while Walker B plays a role in the electron pairing of the Mg^{2+} -ion that accompanies the nucleotide.

A putative ribosomal binding site (AAGG) was identified at 13084 bp (Appendix 5). Several conserved domains have been identified: (a) ATP binding site; (b) ABC transporter signature domain; (c) Walker A/P loop; (d) Walker B; (e) D-loop; (f) Q-loop/lid; (g) H-loop/Switch region as shown in Figure 30 (Appendix 5). The secondary structure is shown in Figure 31.

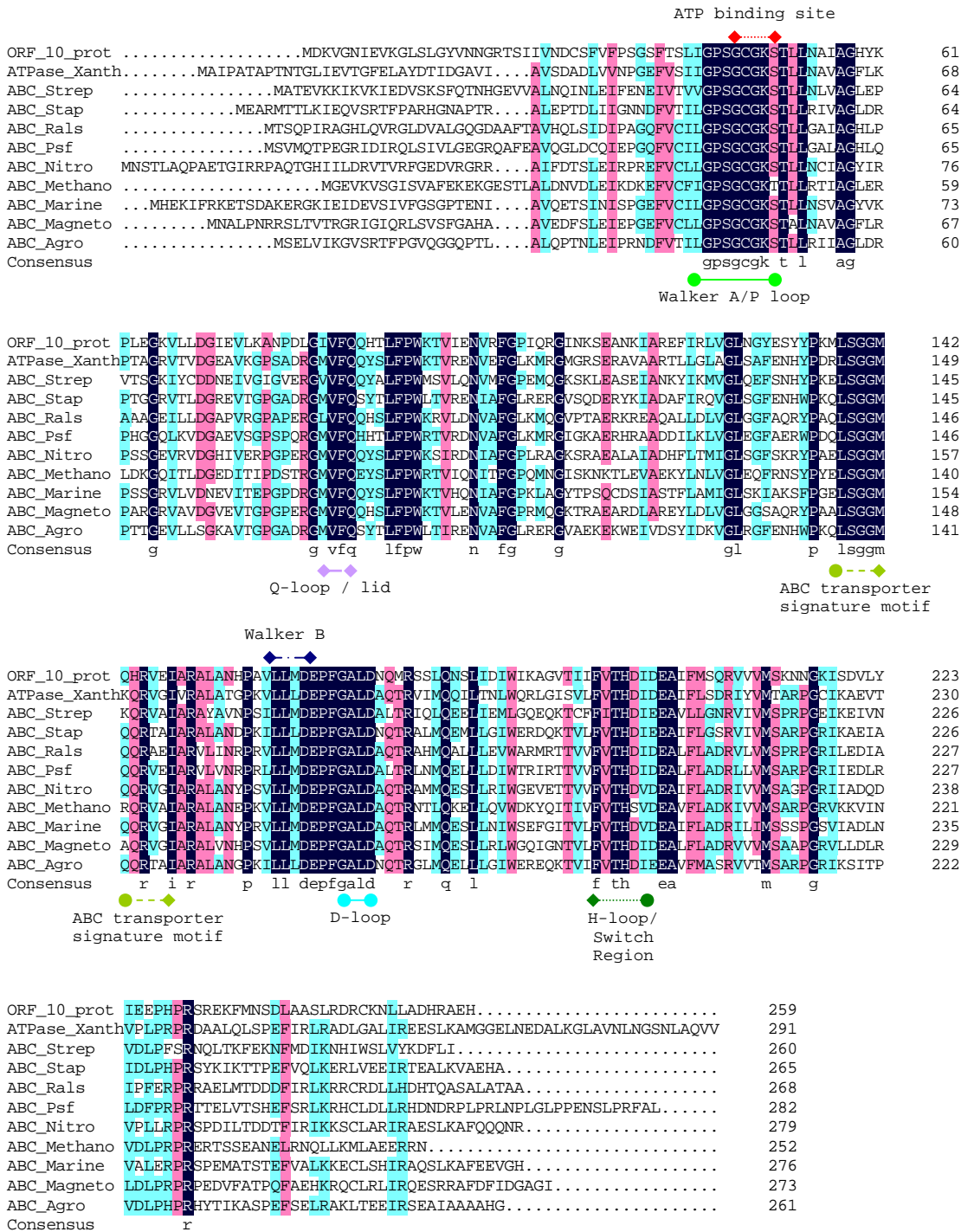


Figure 30: Multiple nucleotide sequence alignment of ORF 10 and various close homologs. The various conserved domains are indicated with pointed lines. ATPase_Xanth: ATPase_Xanthobacter autotrophicus Py2 (291aa; ZP_01200282); ABC_Strep: ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component Streptococcus suis 05ZYH33 (260aa; YP_001197585); ABC_Stap: ABC-type nitrate/sulfonate/bicarbonate transport ATPase component Stamppia

aggregate IAM 12614 (284aa; ZP_01550242); ABC_Rals: ABC-type transporter, ATPase component: Nit family_ *Ralstonia eutropha* H16 (268aa; YP_841758); ABC_Psf: ABC transporter-like_ *Pseudomonas fluorescens* PfO-1 (282aa; YP_348888); ABC_Nitro: ABC-type nitrate/sulfonate/bicarbonate transport system ATPase component_ *Nitrobacter winogradskyi* Nb-225; ABC_Methano: ABC nitrate/ sulfonate/ bicarbonate transporter, ATPase subunit_ *Methanococcoides burtonii* DSM 6242 (252aa; YP_566756); ABC_Marine: ABC-type nitrate/sulfonate/bicarbonate transport system ATPase component_Marine gamma proteobacterium HTcc2143 (276aa; ZP_01615336); ABC_Magneto: ABC-type nitrate/sulfonate/bicarbonate transport system_ *Magnetospirillum magneticum* AMB-1 (273aa; YP_421221); ABC_Agro: ABC transporter, nucleotide binding/ATPase protein_ *Agrobacterium tumefaciens* str. C58 (261aa; NP_530867).

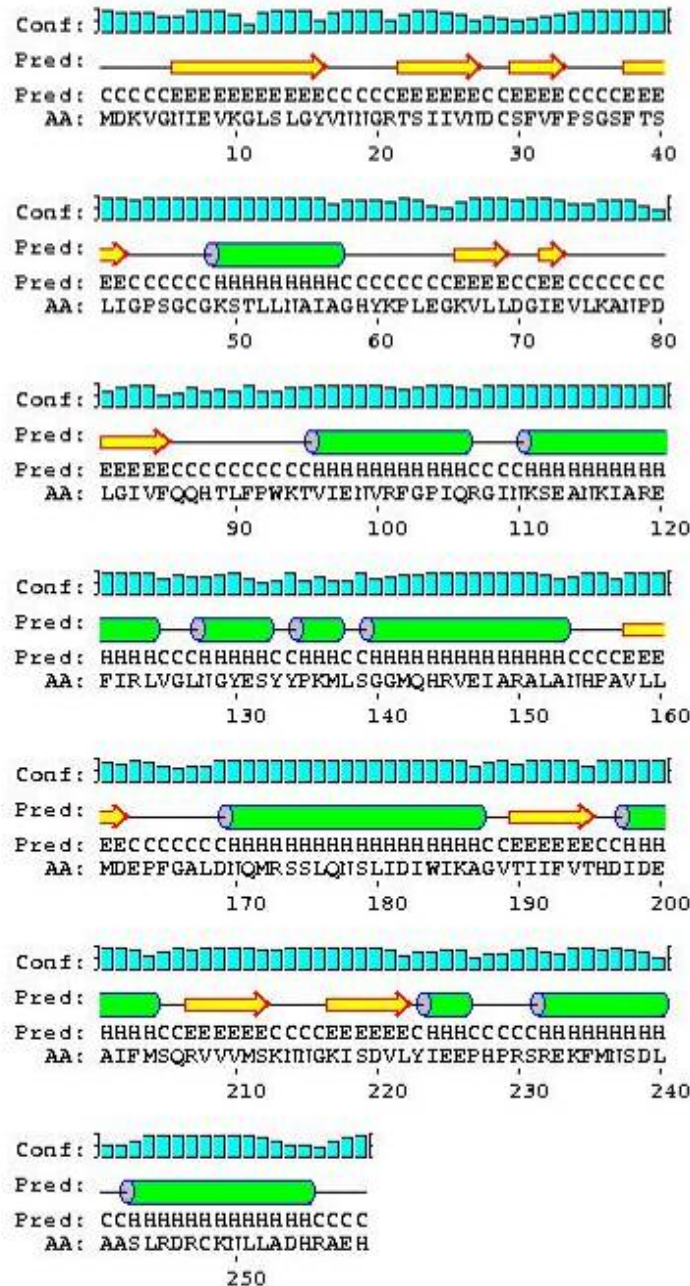


Figure 31: The PsiPRED web-based secondary structure prediction for ORF 10. The green bar represents the predicted α -helix, the yellow arrow represents the predicted β -strand.

ORF 11 is located between 13771 bp – 13085 bp and it has 48% identity (111/227) to a binding dependent transport system inner membrane component of *Nitrobacter winogradskyi* Nb-255. It also displays 46% identity (102/218) to a protein of the ABC-type nitrate/sulfonate/bicarbonate transport system of *Magnetospirillum magneticum* AMB-1 (294aa). A conserved TauC domain which is a

permease component of the ABC-type nitrate/bicarbonate transport system and a possible ribosomal binding (GAAG) site was identified at 13793 bp. The putative transmembrane domains were identified by computer analysis which is illustrated in Figure 32.

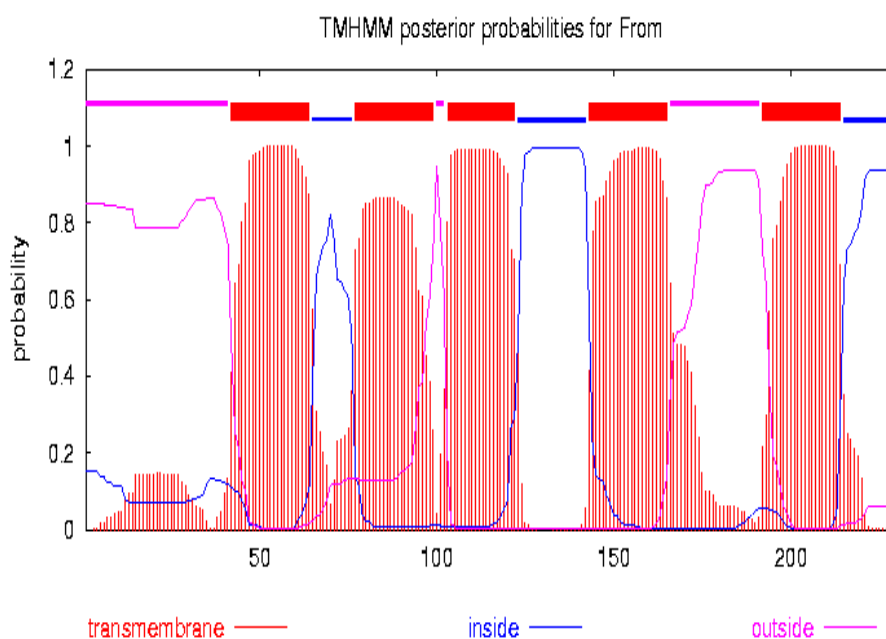


Figure 32: The putative trans-membrane spanning domains and polypeptide loops as predicted for ORF 11 by the web-based TMHMM program.

ORF 12ⁱ is located between 14499 bp – 13993 bp; it is interrupted by an insertion sequence (ORF 13). This ORF has 40% identity (69/170) to an ABC transporter sulfonate/bicarbonate/nitrate transport system substrate binding protein of *Nitrobacter winogradskyi* Nb-255 (474aa). No putative conserved domain or ribosomal binding site could be identified for this ORF. The rest of this ORF 12ⁱⁱ seems to be located between 16646 bp – 15987 bp. A possible ribosomal binding (GGAG) site was 16656 bp and a conserved TauA periplasmic component was identified as shown in Appendix 5.

The last two ORF's are closely related to proteins of which the function is not yet known. ORF 15 is located between 17568 bp – 17173 bp and has 77% identity (101/131) to a hypothetical protein pRSB101_23 from a plasmid (fully sequenced) of an uncultured bacterium. Two possible ribosomal binding sites were identified at 17569 bp (AAGGAA) or 17583 bp (AGGGA). A conserved predicted nucleic acid binding domain (containing a PIN domain) was also identified (Appendix 5). ORF 16 has 65% identity (54/83) to a hypothetical pRSB101_24 from an uncultured bacterium and is located

between 17816 bp – 17565 bp. A probable ribosomal binding (GGAGGG) site was identified at 17821 bp. No conserved domain was identified for ORF 16 (Appendix 5).

4. DISCUSSION

This study has focused on the accessory genes carried on pTcM1 in an effort to ascertain the gene carriage of pTcM1 and its possible role in *At. caldus* and the biomining environment. The BLAST searches against the NCBI-website revealed 15 putative open reading frames coding for genes not associated with plasmid maintenance functions such as replication, mobilization and stability systems. The spatial orientation and intergenic distances of ORF's 2-8 and 9-16 suggests that two divergently located operons (Operon 1 and Operon 2) might be present on this region of pTcM1 (Figure 19). The BLAST results suggest that these open reading frames encode proteins that are involved with bacterial metabolism. Putative operon 1 includes: ORF 2 (transcriptional regulator), ORF 3 (succinate dehydrogenase), ORF 4 (ferredoxin 1), ORF 6 (putative HEAT repeat), ORF 7 (ferredoxin 2) and ORF 8 (NADPH-dependent glutamate synthase, large subunit) (Table 2). Bioinformatic analysis of each of these open reading frames makes it possible to predict possible ribosomal binding sites and conserved domains based on amino acid sequence comparisons.

The role of the transcriptional regulator (ORF 2) is not clear however literature suggests that it might have a repressing effect on the expression of genes in response to environmental stimuli (Hillerich and Westpheling, 2006). The secondary structure of ORF 2 predicted with the PsiPRED web-based program; indicates that the seven α -helical and eight β -sheet structures (as shown in Figure 22), found in the conserved PhnF domain of ORF 2, confirms that the regulator belongs to the HutC subfamily. However, it is not clear what the role of the regulator on pTcM1 might be, it could be that this gene might control the regulation of operon 1 and it would be interesting to see if this is the case.

The succinate dehydrogenase operon consists of four enzymatic units namely SdhA, SdhB, SdhC and SdhD. Only the SdhA domain of the succinate dehydrogenase (ORF 3) could be identified on pTcM1. The iron sulfur protein SdhB is related to the ferredoxin proteins but no significant homology between SdhB amino acid sequences and that of ferredoxin 1 (ORF 4) could be found. Succinate dehydrogenase (*sdhA* gene) is an integral part of the TCA cycle employed by chemoorganoheterotrophic bacteria to produce fumarate from succinate. It's role on the plasmid could be to aid *At. caldus* in its energy production. Although *At. caldus* obtains its energy from the oxidation of sulfur compounds, they can

also make use of certain sugars in a form of mixotrophic metabolism. The succinate dehydrogenase could aid in utilizing other energy sources such as glucose. At this point the role of the succinate dehydrogenase is merely speculative.

The presence of two ferredoxin-like genes (ORF 4 and ORF 7) indicates that these accessory genes might be involved with electron transport. Very little homology was found between these two open reading frames. In Figure 26 it is shown that ORF 4 (77aa) and ORF 7 (107aa) are clustered in two different phylogenetic groups. BLAST results indicate that both these ferredoxins contain [4Fe-4S] clusters; which are normally associated with bacteria.

The insertion sequence elements (ORF 5 & 13) seem to have interrupted ORF 6 and ORF 12. BLAST results of these insertion sequences showed a 98% identity to a transposase-like protein of the ISAtc1 from *At. caldus* and these elements share a 100% homology to each other. One of the problems encountered in this study was the presence of insertion sequences. It seems that ORF 6, 12 and 14 have been disrupted/truncated making further analysis difficult.

ORF 6 has been interrupted by the insertion sequence jump (ORF 5). The fact that this gene is truncated makes it difficult to determine the role of this putative protein on pTcM1. It is likely that this putative phycobiliprotein is the result of horizontal gene transfer from a photosynthetic bacterium or that the low level of relatedness reflects the conservation of a domain within the putative protein rather than the whole protein.

Literature has shown that the GOGAT gene (ORF 8) can be coupled with a ferredoxin, NADH or NADPH as the electron carrier. In bacteria only the ferredoxin and NADPH electron carriers have been associated with the GOGAT gene. BLAST results have shown that this large subunit of the GOGAT is NADPH dependent. However the close proximity of the [4Fe-4S] ferredoxin directly upstream of this gene could indicate that this GOGAT might be associated with the ferredoxin. Kameya and his co-workers have (Kameya, *et al*, 2007) found a GOGAT in Cyanobacteria that is similar to the “plant-GOGAT” and coupled to a ferredoxin as electron carrier. The Fd-GOGAT has been found to have a conserved an Fd-loop which does not seem to be present in ORF 8. The lack of the Fd-loop, BLAST result and the phylogenetic arrangement (Figure 28) suggests that if the putative protein of ORF 8 might be coupled to NADPH as an electron carrier. Domains for NADPH, flavin adenine dinucleotide and [4Fe-4S] binding have been identified. However, the second subunit namely a glutamine binding

domain (α -subunit and the Cys-Asp-His catalytic triad located within the N-terminal of NADH-, and Fd-GOGAT proteins) of a functional GOGAT was not found. This could indicate that ORF 8 is not a GOGAT or that the other subunit is on the chromosome or another plasmid.

Another cluster of genes showing homology to ATP-binding cassette (ABC)-type membrane transporters have been identified. Putative operon 2 includes the following open reading frames: ORF 9 (outer membrane porin), ORF 10 (ABC type nitrate/sulfonate/bicarbonate transporter), ORF 11 (binding-dependent transport protein), ORF 12ⁱ and 12ⁱⁱ (ABC sulfonate/bicarbonate transporter), ORF 14 (Hypothetical protein pRSB101_23) and ORF 15 (Hypothetical protein pRSB101_24). This cluster of genes seem to be disrupted by another insertion sequence (ORF 13). These ABC-type systems are commonly found in bacteria and play a role in transport of nutrients, ions and metals. The chemolithotrophic *At. caldus* has less energy available than a typical heterotroph, therefore an additional energy and transport system could aid *At. caldus* to survive in these environments.

Similar types of ORF's involved in metabolism and energy functions have been identified on other plasmids of biomining bacteria. A putative ferredoxin [3Fe-4S, 4Fe-4s], an FNR regulator-type protein and NADH oxidoreductase subunit was identified on plasmid pTF5 (Dominy *et al.*, 1997). A glutaredoxin and a MerR-like regulator protein was found on the transposon Tn5467 from plasmid pTF-FC2 from *At. ferrooxidans* (Clennell *et al.*, 1995). A large antibiotic multiresistance plasmid pRSB101 (47.8 kb) was isolated from a wastewater treatment plant which harbors an ATP-binding-cassette (ABC)-type ATPase and permease, transcriptional regulators and several insertion sequences and transposases. At this point we can merely speculate what the contribution of these accessory genes will make to metabolic fitness of the host in biomining environments. The prevalence of similar types of accessory genes could be an indication of the movement of these elements and if they form part of the horizontal gene pool of these bacterial communities.

The second part of the project involves a study on the characterization and distribution of accessory genes carried on this plasmid. A question we would like to address in this study is to determine what benefit would be gained by maintaining such a large plasmid in *At. caldus* by focusing on the accessory genes.

CHAPTER THREE: DISTRIBUTION OF PUTATIVE GENES

1. INTRODUCTION

In this chapter I will focus on the distribution of the 15 putative genes found on the 18 kb region of pTcM1. In the past few years similar large plasmids have been isolated that contain metabolic and transport accessory regions. A 47.8 kb antibiotic multi-resistance plasmid pRSB101 was isolated in a waste water treatment plant. It carries a mobile genetic region which includes a multidrug transporter, that consists of an ATP-binding-cassette (ABC)-type ATPase and permease and an efflux membrane fusion protein, as well as an outermembrane protein precursor (*oprM*), several transposases and invertase/ recombinase-like genes and hypothetical proteins similar to ORF 14 and 15 (Szczepanowski *et al.*, 2004).

A number of degradative plasmids with accessory metabolic and transport functions have been discovered recently such as plasmid pA81 (98.1 kb), isolated from *Achromobacter xylosoxidans* (Jencova *et al.*, 2008); pB8 (Schlüter *et al.*, 2005), pKJK5 (Bahl *et al.*, 2007), pCT14 (Bramucci *et al.*, 2006). The modular organization of these large plasmids seems to have a general theme which includes both metabolic and transport functions.

Goldschmidt and his colleagues have shown that similar copies of the insertion elements found on pTcM1 (ORF's 5 & 13) have been identified in three of the six strains of *At. caldus* (Goldschmidt, 2004). Three strains of *At. caldus* (#6, "f" and MNG) were isolated in South Africa, these strains all contain copies of the insertion element ISAtc1. It is likely that these strains might have come in contact with the same element and this might be the reason why the other *At. caldus* strains KU, C-SH12 and BC-13 do not seem to have this element. It is also possible that the insertion elements could be type of adaptation to assist the plasmid in integrating plasmid genes into the chromosome. Similar types of IS elements were also identified in genomic DNA of five *At. ferrooxidans* strains (Tamara *et al.*, 2005). Holmes and his colleagues (Holmes *et al.*, 2001) showed that the copies of the insertion sequence, ISAfel (also known as IST1) or very closely related elements, were found in several strains of *At. ferrooxidans* and *At. thiooxidans* on both plasmid and chromosomal DNA as shown in Figure 33.

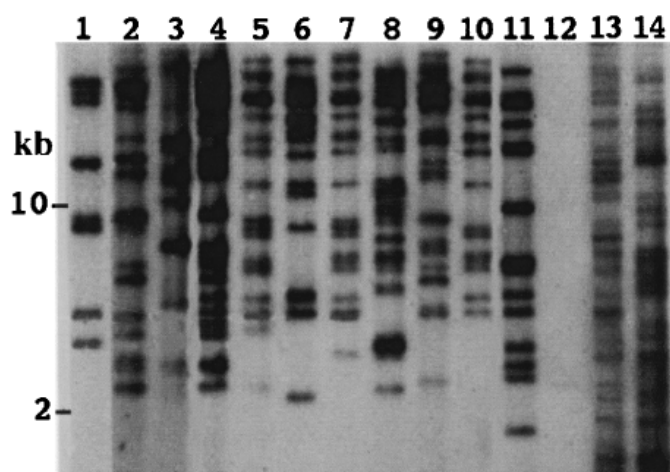


Figure 33: Southern hybridization comparison of genomic DNA from *At. ferrooxidans* (Lanes 1 -12) and *At. thiooxidans* (13-14) strains digested with *Bam*HI with a fragment from ISAfe1²⁴

Our insight into the frequency, range and evolution of plasmids is fragmented. Although most of the plasmids isolated from biomining environments are cryptic the presence of these accessory genes could play a role in the bacterium's metabolism and aid in maintaining these genes in the bacterial population by incorporating them into the chromosome.

2. MATERIALS AND METHODS

2.1 MEDIA AND GROWTH CONDITIONS

Acidithiobacilli strains were grown at 30°C and 37°C in Basal salts (1x) solution with additional 0.5 ml potassium tetrathionate (K₂O₆S₄) (Appendix 1) at a pH of 2.5 or 0.5 ml sodium thiosulphate (Na₂S₂O₃) (Appendix 1) at a pH of 4.7.

2.2 BACTERIAL STRAINS, PLASMIDS AND PROBES

The bacterial strains and plasmid constructs used in this study for total DNA extraction are summarized in Appendix 2.

²⁴ Holmes *et al.*, 2001

2.3 TOTAL DNA ISOLATION FROM ACIDITHIOBACILLI

Six strains of *Acidithiobacillus caldus* were used in this study, namely MNG, KU, “f”, #6, BC-13 and C-SH12. These cultures were harvested by centrifugation at 12000 rpm for 30 minutes. The cell pellet was resuspended in 1L of Basal salts (1x) solution (pH 2.5) and subjected to several high and low speed centrifugation spins to ensure that the cell extract is clear of excess growth media and sulfate precipitation. The cell pellet was resuspended and washed (3x) in TE buffer (pH 7.5) to neutralize the media. After the wash steps the cells were resuspended in 500µl TE buffer and incubated with 50µl Proteinase K and 3 µl RNase for 1hour at 37°C. The suspension was lysed with 100µl 10% SDS (0.8% w/v) overnight at 37°C. Proteins were precipitated using 0.5 volumes of 7.5 M (NH)₄OAc (pH 7.5). Total DNA was precipitated with 2.5 volumes of 100% ethanol (Current Protocols in Molecular Microbiology). The clean DNA pellet was resuspended in TE buffer and stored at 4°C.

2.4 DETECTION OF PUTATIVE GENES USING SOUTHERN HYBRIDIZATION

The dioxigenin-dUTP non-radioactive DNA labeling detection kit was used for labeling probes, hybridization and detection (Roche® Molecular Biochemicals). The probes used in this study are summarized in Appendix 3. Standard methods of Sambrook *et al.*, 1989 were used for hybridization experiments. Hybridizations were carried out at 40°C in DIG Easy Hybridization solution followed by two low-stringency washes at room temperature (2 x SSC, 0.1% SDS) and two high stringency washes at 65°C (0.1 x SSC, 0.1% SDS).

3 RESULTS

3.1 OCCURANCE OF THE PUTATIVE OPEN READING FRAMES IN ACIDITHIOBACILLI

Probes of the specific open reading frames were designed to determine whether the open reading frames are present in other *At. caldus* strains as well. We chose the following ORF's from the two putative operons (as discussed in chapter 2): Operon 1 – ORF's 2, 3, 5, 7 and 8; Operon 2 – ORF's 12, 13, 14 & 15 to best represent the accessory genes found on pTcM1. Two plasmids namely Tn 39 and Tn 2 from *At. caldus* strains C-SH12 and “f”, respectively, were available for screening of some of these genes. Tn 11 #1, #2 and #3 are clones of plasmid pTcM1, with insertions jumps at different locations on the plasmid. Most of these ORF's display a high degree of similarity to known genes.

3.1.1 TRANSCRIPTIONAL REGULATOR (ORF 2)

Total DNA from the six *At. caldus* strains available were digested with *Bam*HI. Plasmids Tn 11, Tn 39 and Tn 2 were also digested with *Bam*HI. The digested DNA was hybridized to a probe of the regulator-like gene (REG, Appendix 3). Positive hybridization signals for the predicted sizes of ± 6.2 kb were obtained for plasmid Tn 11 #1 (positive control) as well as total DNA from strains MNG, #6, C-SH12 and BC13 (Figure 34). The relatively weak bands observed for MNG and BC13 is likely due to a lower concentration of total DNA being loaded on the gel in relation to the other samples. Two signals of approximately ± 8 kb and $\pm 9-10$ kb were identified for strain “f”. The ± 8 kb signal from the total DNA extract corresponds to the signal obtained from plasmid Tn 2.

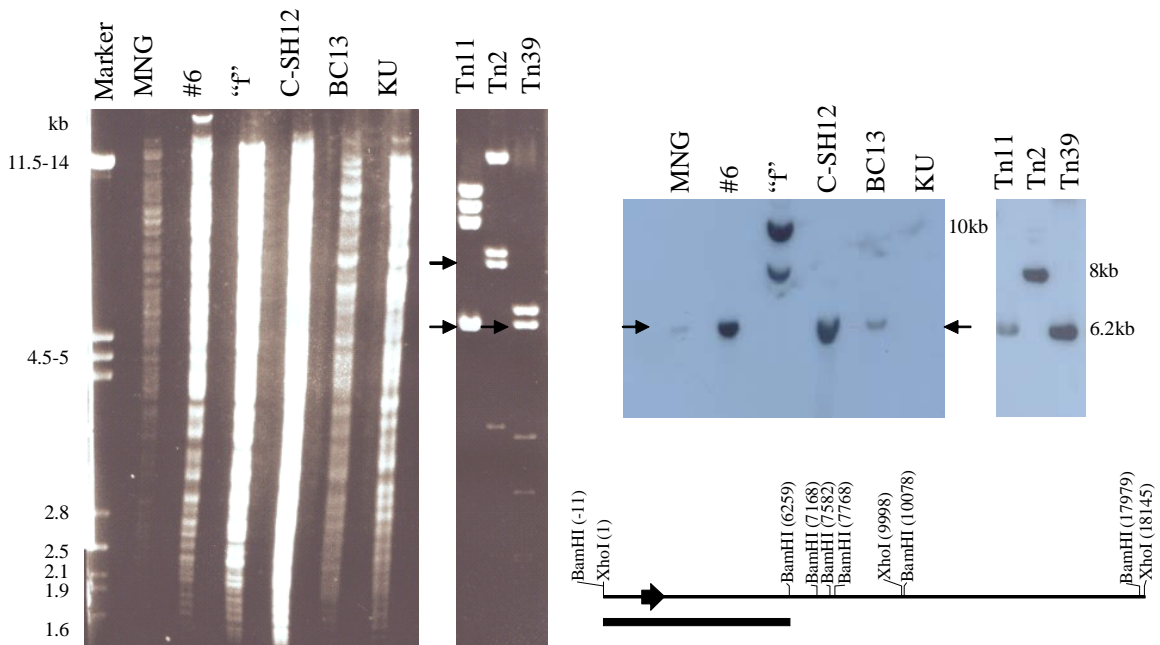


Figure 34: Southern hybridization of a DIG- labeled REG probe (300 bp fragment cut with *Xba*I and *Hind*III). Total DNA from strains MNG, #6, “f”, C-SH12, BC13, KU and plasmid DNA from Tn17, Tn 11, Tn 2 . ORF 2 is indicated by the solid arrow, the solid bar indicates the predicted fragment size of the hybridization signal for pTcM1.

3.1.2 SUCCINATE DEHYDROGENASE (ORF 3)

A 500 bp fragment cut with *SalI* and *BglII* was used as a probe for the succinate dehydrogenase open reading frame. Total DNA of the six strains as well as plasmids Tn 2 and Tn 11 #1 and #2 were digested with *XhoI*. A positive signal of ± 10 kb for the predicted size of the fragment containing the putative succinate dehydrogenase can be seen in Tn 11 #2 and Tn 2 (Figure 35). However a slightly smaller band ± 9 kb could be seen in Tn 11#1, which corresponds to the total DNA extract from MNG, #6, C-SH12 and BC13. The difference in size could possibly be the absence of the insertion sequence; which does not seem to be present in the total DNA extract from MNG (this statement is collaborated later on in this chapter – PCR results from genomic DNA of strain MNG). Two bands of ± 10 kb and ± 14 kb for total DNA from strain “f” is observed. The smaller band of ± 10 kb corresponds to the plasmid signal from Tn 2. Strain KU does not seem to carry a gene similar to the succinate dehydrogenase of pTcM1.

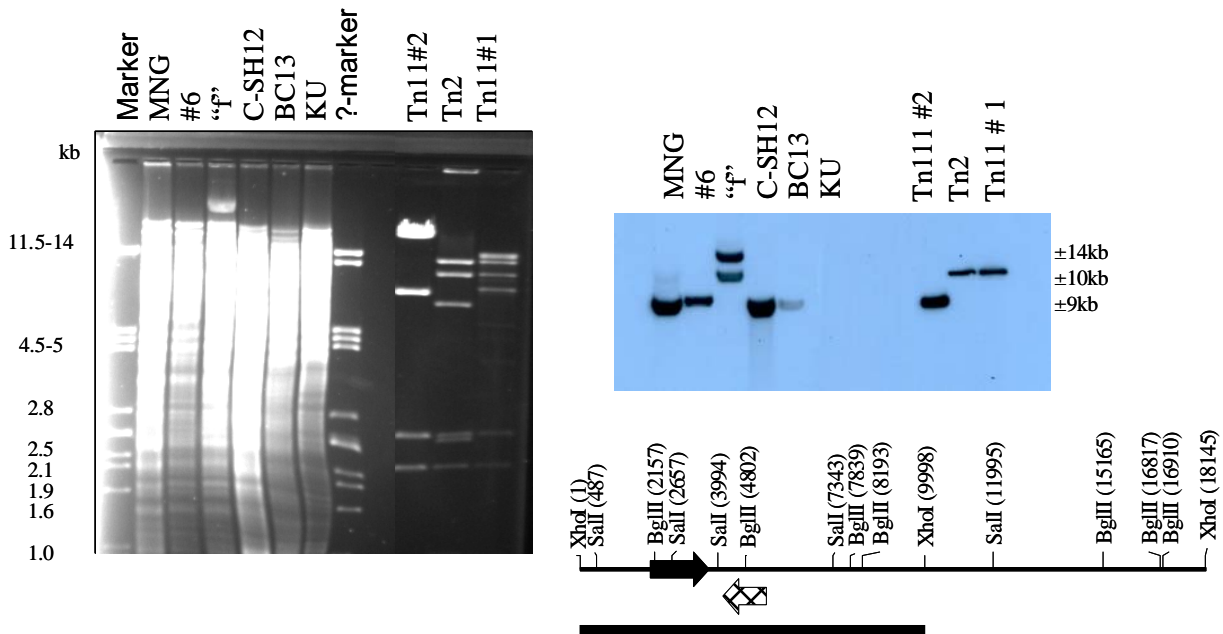


Figure 35: Southern hybridization of a DIG-labeled SUC probe (500 bp fragment cut with *BglII* and *SalI*). Total DNA of strains MNG, #6, C-SH12, BC13 and KU were cut with *XhoI*. Plasmid DNA of Tn 2 and two versions of Tn 11 were also digested with *XhoI*. ORF 3 is indicated by the solid arrow, the striped arrow indicates the position of the insertion sequence, the solid bar indicates the predicted fragment size of the hybridization signal for Tn 2.

3.1.3 FREQUENCY OF INSERTION ELEMENTS (ORF 5 AND 13)

Previous results from Goldschmidt (2004) and Holmes *et al.* (2001) have shown that insertion sequences are common in *At. caldus* as well as *At. ferrooxidans* species. Total genomic DNA was extracted as discussed in 3.2.3 and a 300 bp (cut with *Bgl*I and *Stu*I) probe from the insertion element of ORF 5 (identical to ORF 13) was used to hybridize to *Bam*HI digested DNA of the six strains of *At. caldus*. Similar to the results of Goldschmidt it was found that the strains isolated from South Africa namely “f”, #6 and MNG contain up to 15 copies of the insertion elements as shown in Figure 36.

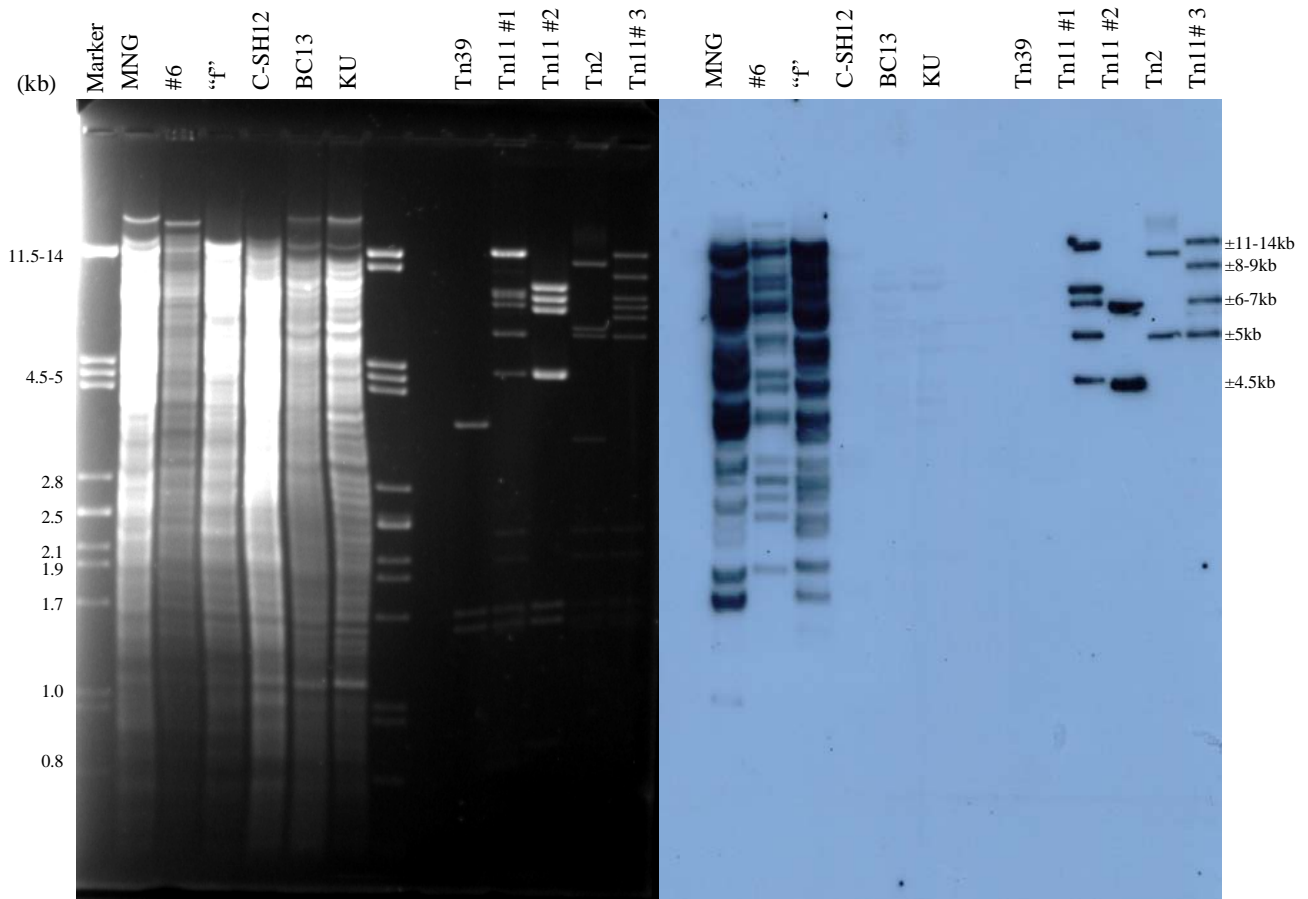


Figure 36: Southern hybridization analysis of ORF 5 and 13. Total DNA of strains MNG, #6, C-SH12, BC13 and KU were digested with *Bam*HI and probed with a clone from ORF 5. Different versions of Tn 11 # 1-3 and Tn 39 were also cut with *Bam*HI. The insert highlights faint bands that could be observed in strains C-SH12, BC13 and KU.

One of the difficulties experienced in this study was that the three versions of plasmid pTcM1, Tn 11 #1, #2, #3 gave dissimilar banding patterns, this can be seen in Figure 35. It would appear that transposition of the IS elements were presumably taking place at a higher than normal frequency when cloned in *E. coli* because the IS elements were in an unrepressed background. There seems to be multiple copies of the insertion elements present on both the chromosomal and plasmid DNA. We attempted to obtain a copy of putative operon 1 and 2 (Figure 37) that was not interrupted by the insertion element, we managed to clone an interrupted copy of operon 1 but not of operon 2.

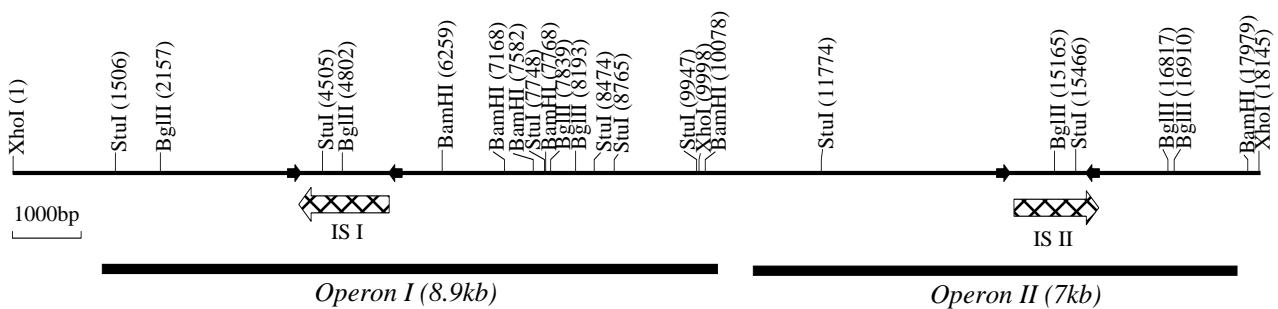


Figure 37: A physical map of the putative operons I and II (indicated by the solid bars). The insertion sequences are indicated as crossed arrows. The primer binding sites are indicated as black arrows.

Primers were designed to the flanking regions external to both insertion elements IS1 (ORF 5) and IS2 (ORF 13). The primers were designed to yield two different sized products depending on whether or not the IS elements were present in the identical location in the total genome of the different *At. caldus* strains tested. A product of 1.6 kb and 1.4 kb respectively would include IS1 and IS2 and a product of 500 bp or 300 bp would exclude IS1 and IS2; respectively. PCR amplification revealed the expected 0.5 kb products of IS1 for strain MNG and plasmids Tn 39 and Tn #1. A product of 1.6 kb could be seen for Tn 11 #2 which indicates that this region contains the insertion element whereas the smaller products do not (Figure 38). The PCR products of IS2 revealed a 1.4 kb product for strain MNG, Tn 11 #2 and no product for Tn #1. A smaller product of 0.3 kb could be seen for Tn 39 which could indicate that Tn 39 does not contain the insertion element. Faint PCR bands could be observed for the total DNA samples of MNG which indicates that there could be a chromosomal version of this region of DNA that do not contain IS2.

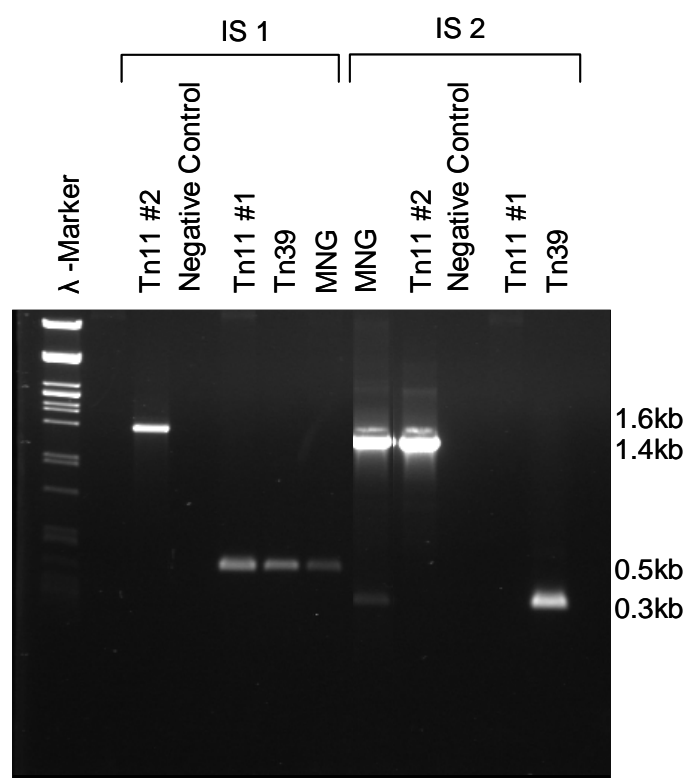


Figure 38: The flanking regions of IS1 and IS2 were PCR amplified. Total DNA of strain MNG and plasmid DNA of different preparations of the same copy of pTcM1 namely Tn 11 #1 and Tn 11 #2 and plasmid Tn 39 from C-SH12 were used as templates for the PCR reactions. The negative control does not contain any DNA.

3.1.4 FERREDOXIN (ORF 7)

We attempted to determine if the ferredoxin-like genes were present in the other strains of *At. caldus*. A PCR amplified product of ORF 7 (FER1, Appendix 3) was hybridized against total DNA of the six strains digested with *Bam*HI as shown in Figure 39. Signals of a 0.9 kb fragment were found for five strains examined; which corresponds to the plasmid copy of the ferredoxin-like open reading frame. Strain KU does not seem to contain any genes related to the probe. In the case of strain MNG and C-SH12 two faint signals of 2 kb and 4.5 kb, respectively, could also be seen. These signals could represent chromosomal copies of similar ferredoxin-like genes or partially digested DNA.

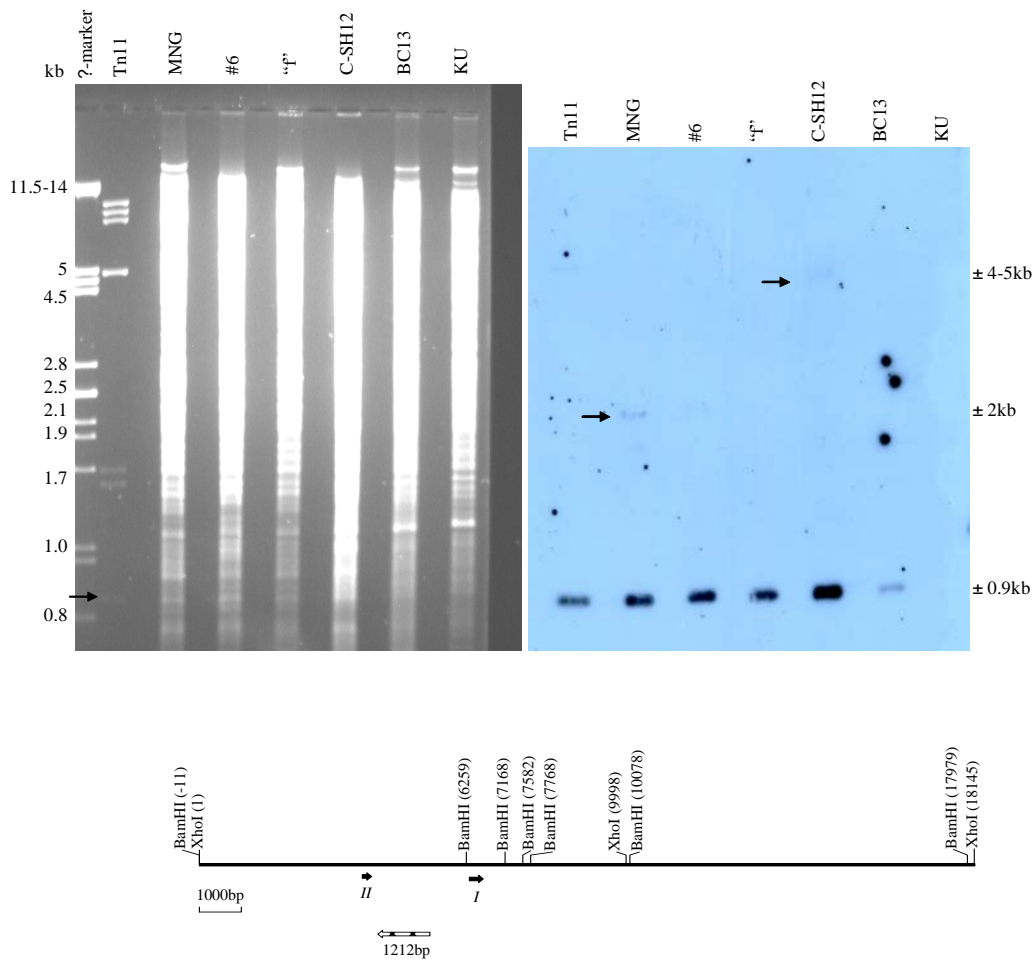


Figure 39: Southern hybridization analysis of ORF 7 (ferredoxin). Total DNA of strains MNG, #6, “f”, C-SH12, BC13 and KU and plasmid DNA of Tn 11 were digested with *Xho*I and probed with a cloned product from ORF 7. The arrows indicated faint signals that could be observed. The restriction map shows the *Bam*HI restriction sites, the insertion element (crossed arrow) and the two ferredoxin-like genes (ORF 4 & ORF 7).

3.1.5 GLUTAMATE SYNTHASE (ORF 8)

Total DNA of the six strains as well as two versions of Tn 11 (Tn 11 #1 and Tn 11 #2) and Tn 2 were digested with *Xho*I and hybridized to a labeled probe of ORF 8 (500 bp fragment cut with *Bgl*III and *Sal*I). Southern hybridization blots showed that the expected signal size of ± 10 kb for the plasmid DNA Tn 11 #2 as shown in Figure 40. The signals of Tn 11 #1 and Tn 2 are slightly smaller than that of Tn 11 #2; which could suggest that Tn 2 and Tn 11 #1 might not contain the insertion element. The signals obtained from strains MNG, #6, “f”, C-SH12 and BC13 each gave a signal in the range of 9-10

kb and all six *At. caldus* strains gave a signal of 4-5 kb as well. This could indicate that there could be a chromosomal and a plasmid copy of genes similar to that of ORF 8. The variation in the sizes of hybridization signals between different strains could be a result of different restriction enzyme sites having arisen as a result of genetic drift. The GOGAT gene might have undergone mutational changes over time.

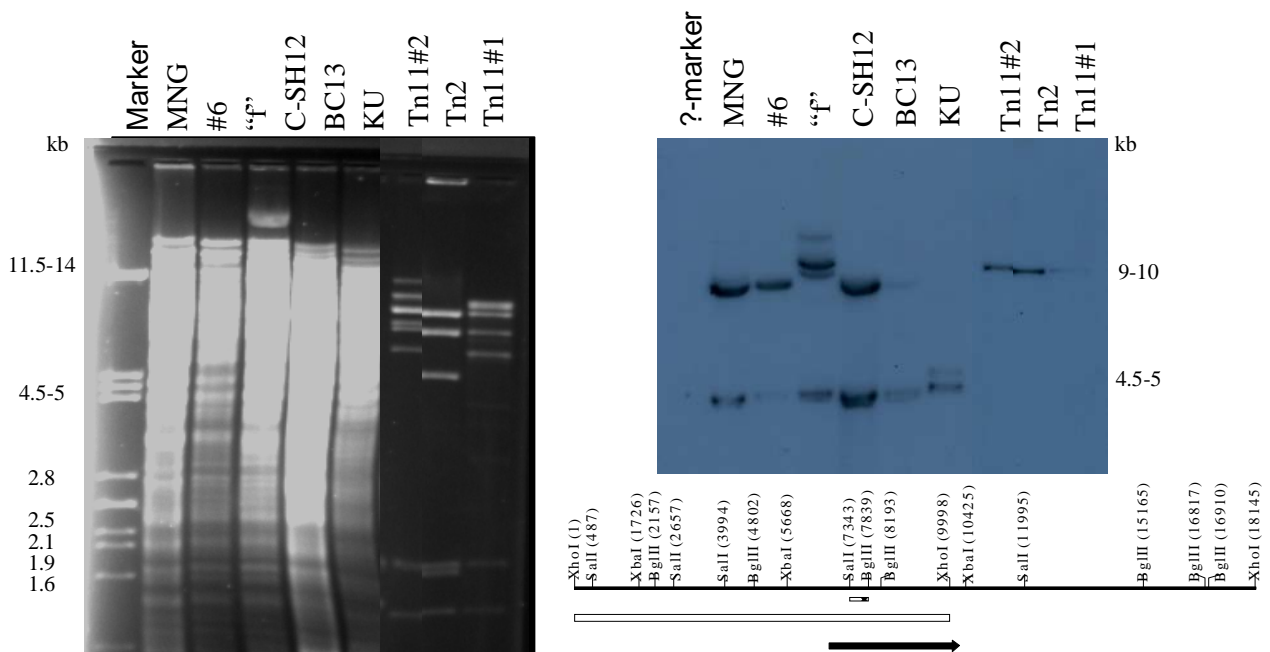


Figure 40: Southern hybridization analysis of ORF 8 (glutamate synthase-like subunit). Total DNA of the six *At. caldus* strains and two versions of Tn 11 (Tn 11 #1 and #2) and Tn 2 were digested with *XhoI* and probed with a clone from ORF 8. The restriction map of the various restrictions sites are shown. The arrow indicates ORF 8 and the open bar illustrates the predicted ± 10 kb signal for the plasmid version of ORF8. The crossed bar indicates the 500 bp *BglIII* and *Sall* fragment used as a probe for ORF 8.

3.1.6 ABC-TYPE SULFONATE/BICARBONATE-LIKE GENES (ORF 12²)

A PCR amplified DNA fragment (Appendix 3) from the ABC type open reading frame (ORF 12²) was used as a probe and hybridized to total DNA of six *At. caldus* strains digested with *XhoI*. As seen in Figure 41 strains MNG and #6 gave similar sized signals of ± 10 kb, which corresponds to the plasmid Tn 11 #1 size signal. Similar sized signals for C-SH12 and BC13 were found. Two signals for strain "f"

were observed varying in sizes of ± 15 kb and ± 10 kb. Faint bands of ± 15 kb for strains MNG and #6 similar to that of strain “f” could also be observed. No signal for strain KU was observed.

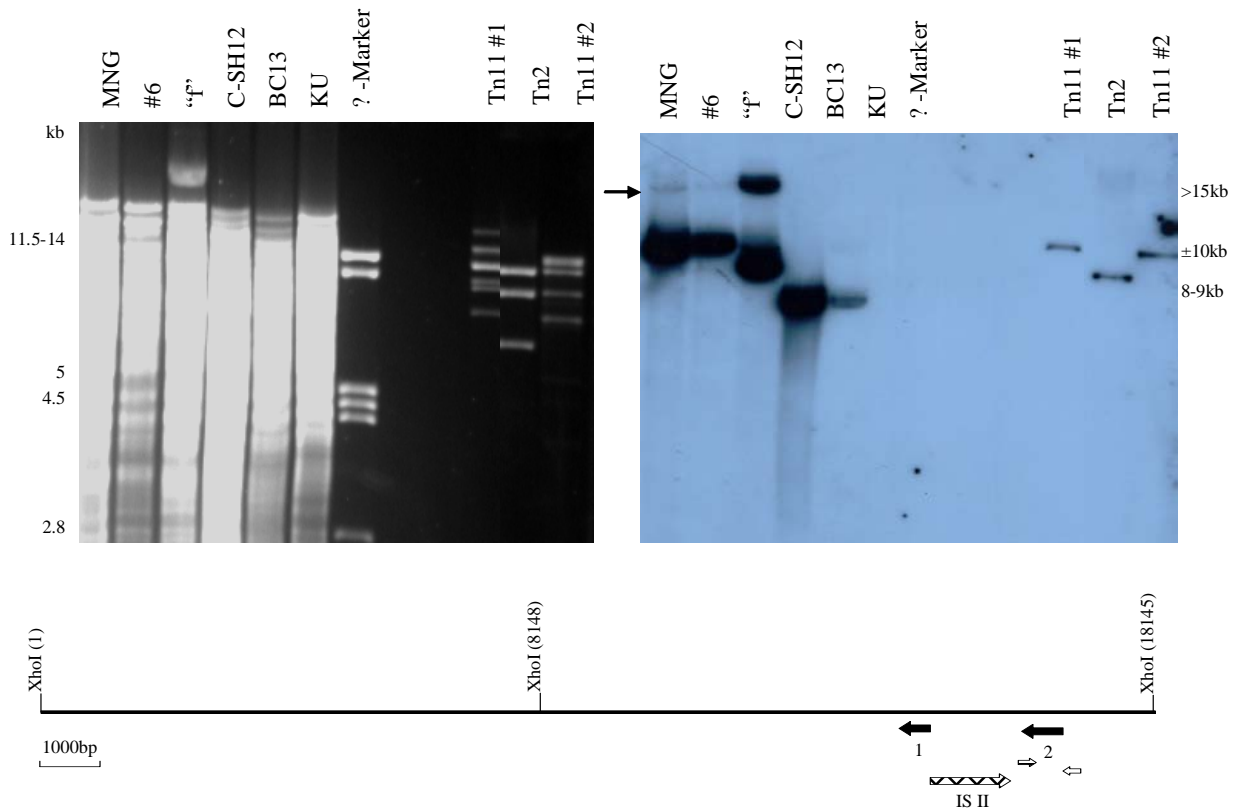


Figure 41: Plasmid DNA of Tn 11(#1 and #2) and Tn 2 as well as total DNA from the six strains of *At. caldus* were digested with *XhoI* and hybridized with a probe from ORF 12².

3.1.7 PUTATIVE pRSB101_23 AND pRSB_24 PROTEINS (ORF 14 & 15)

The multidrug resistance plasmid pRSB101 (47.8kb) was isolated from waste water treatment plants, and these have been found to act as reservoirs for antibiotic resistant bacteria and R plasmids (Szczepanowski *et al*, 2004 and Blázquez *et al.*, 1996). ORF 14 has similarity to the pRSB101_23 hypothetical protein from pRSB101. ORF 15 has similarity to another hypothetical protein from pRSB101, pRSB101_24. The presence of these two putative genes was screened for using a joint product made from a PCR product amplifying the regions flanking these two genes (pRSB, Appendix 3), as this region of the plasmid could not be subcloned into smaller fragments. Plasmids Tn 11 #1 and #2, Tn 39 and total DNA from the six *At. caldus* strains were digested with *XhoI*. Positive

hybridization signals of ± 10 kb for Tn 11 #1 and a relatively weak band of similar size could be seen for strain MNG (Figure 42). Slightly smaller signals could be observed for strain MNG, #6, and “f”; which could indicate chromosomal copies of similar type ORF’s. Signals of ± 5 kb and ± 3 kb in size were observed for C-SH12 and BC13; respectively. No signal could be observed for strain KU and Tn 11 #2. A small positive signal of 1.6 kb was also observed for plasmid Tn 39.

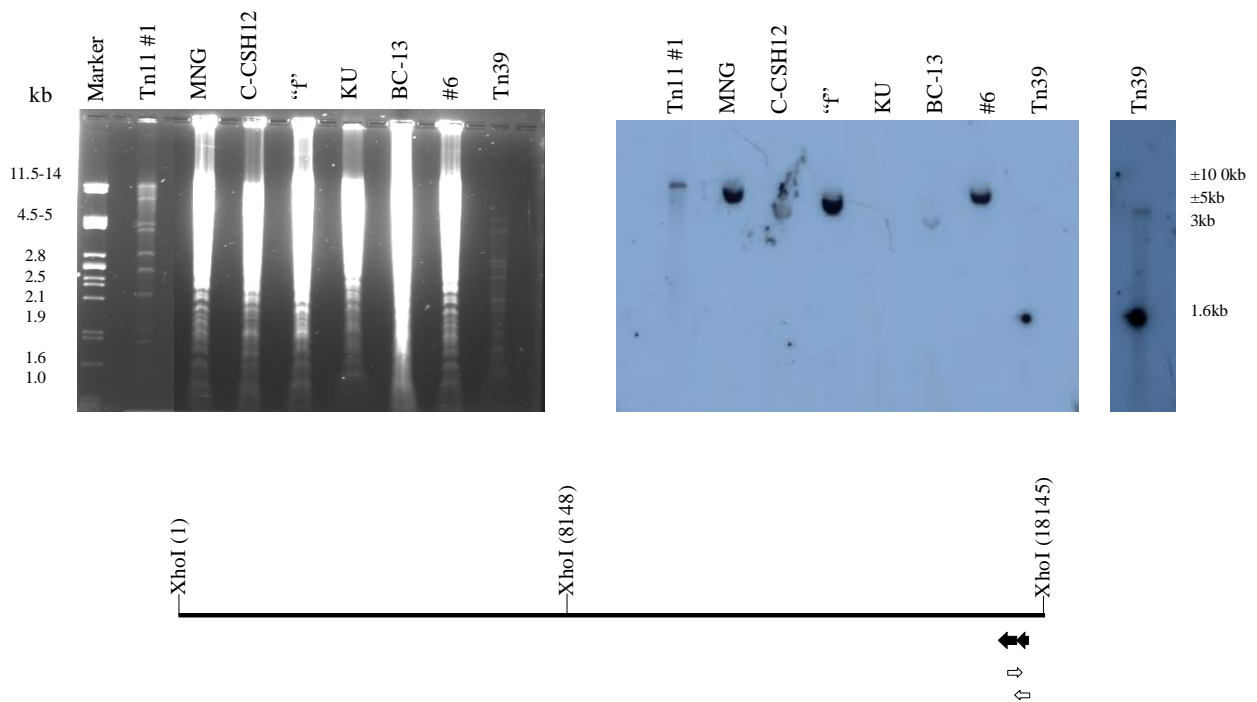


Figure 42: Southern hybridization analysis of ORF 14 and 15. Total and plasmid DNA of Tn 11 #1, Tn 39 and strains MNG, #6, “f”, C-SH12, BC13 and KU were digested with *XhoI* and hybridized to a pRSB probe. Longer exposure of plasmid Tn 39 is shown on the right hand side.

4 DISCUSSION

Positive hybridization signals with the regulator-like gene probe of ± 6.2 kb for strains MNG, #6, C-SH12, BC13 and plasmid Tn 39 were obtained (Figure 34). This indicates that these four strains carry a regulator-like gene; whether it is on a plasmid or part of the chromosome is unknown. The weaker signals for strains MNG and BC13 are likely due to a smaller concentration of DNA present on the gel. Two signals of ± 8 kb and ± 10 kb were observed for strain “f”, the ± 8 kb signal corresponds to the signal from plasmid Tn 2, which was isolated from strain “f”. Therefore it is likely that the smaller

signal could indicate a plasmid copy of the regulator gene while the other signal could be a chromosomal copy or another plasmid copy which has not been identified as yet. Strain KU does not seem to have a gene similar to the regulator-type open reading frame.

It seems that a succinate dehydrogenase-like gene is present in strains MNG, #6, C-SH12 and BC13 but not in strain KU. In Tn 11 #2 and Tn 2 the predicted signal of ± 10 kb indicates that Tn 2 could have a similar type of gene and possible operon arrangement as Tn 11 #2. The smaller signal of ± 9 kb identified for Tn 11 #1 and the four *At. caldus* strains suggest that the insertion sequence as shown in Figure 35 might not be present in these strains. The ± 10 kb signal for strain “f” corresponds to that of plasmid Tn 2 indicating that this could be a plasmid copy which contains the insertion element. The other signal of ± 14 kb could suggest that there might be a plasmid present in strain “f” with a similar copy of the succinate dehydrogenase gene that has not been identified or it could represent a chromosomal copy of the gene.

Several insertion elements have been identified in Acidithiobacilli (Goldschmidt 2004 and Holmes *et al.*, 2001). Local strains “f”, #6 and MNG seem to harbor up to 15 copies of the similar types of insertion elements as IS1 and IS2 (Figure 36). IS1 does not seem to be located at the same position on pTcM1 as indicated by the smaller 0.5 kb PCR product observed in Tn 11 #1, Tn 39 as well as MNG (Figure 38). The 1.6 kb PCR product of IS1 shown for Tn 11 #2 suggests that it could have originated during the rescuing process of pTcM1 in *E. coli*. This shows that a DNA rearrangement and transposition event took place after the plasmid was captured in *E. coli*. ORF 5 was inserted between the ferredoxin-like genes on pTcM1 during the capture or replication of pTcM1::11 in *E. coli*, southern hybridization and PCR amplification of crude plasmid from *At. caldus* strain MNG indicated that it was not present in the native DNA. When the products for IS2 were amplified, the results suggested that IS2 is present in MNG and on the plasmid Tn 11 #2 but not in Tn 11 #1. Therefore it seems that IS2 is located in the same position on pTcM1 as Tn 11 #2. The 0.3 kb PCR band observed for the total DNA samples of MNG indicated that there is a chromosomal version of this region of DNA that does not contain IS2. Plasmid Tn 39 does not seem to contain insertion elements similar to IS1 and IS2.

Similar sized signals for the ferredoxin gene were found in strain MNG, #6, “f”, C-SH12 and B-C13, which corresponds to that of the plasmid sized fragment (Figure 39). This result shows that this ferredoxin gene or a very similar type of ferredoxin gene is present in all the strains except KU. Strains MNG and C-SH12 gave two extra signals of 2 kb and 4.5 kb; respectively, which indicates that the

ferredoxin gene or a similar type ferredoxin is present on the chromosome of strain MNG and maybe on the chromosome of strain C-SH12.

A copy of the GOGAT gene seemed to be present in strains C-SH12, “f” and #6 as well as on the plasmid pTcM1 from strain MNG as each gave a signal of 10-11 kb corresponding to the plasmid signal for ORF 8 (Figure 40). No signals similar in size could be identified for strains BC13 and KU. However signals in the range of 4.5 kb were observed for all the strains examined including strains BC13 and KU which could indicate a chromosomal copy in the case of strain MNG and a chromosomal or plasmid copy or both in the other strains.

As seen in Figure 41 strains MNG and #6 gave similar sized signals of ± 10 kb for the ABC type gene which corresponds to the plasmid Tn 11 #1 and #2 size signal. Two signals for strain “f” were observed varying in sizes of more than ± 15 kb and ± 10 kb. Faint bands of ± 15 kb for strains MNG and #6 similar to that of strain “f” could be observed which suggests that there might be a chromosomal copy of the ABC type gene in strain MNG and #6. Similar sized signals of 8-9 kb for strains C-SH12 and BC13 were identified. There does not seem to be a similar type of ABC gene present in strain KU.

It was difficult to subclone and sequence pieces of DNA containing the putative pRSB101_23 and pRSB101_24 ORF's of pTcM1. Therefore a joint probe of ORF's 15 and 16 were made. Different sized bands were observed for most of the strains (Figure 42). A relatively weak band of ± 10 kb similar in size to Tn 11 could be seen for strain MNG. Very strong hybridization signals of $\pm 9-9.5$ kb was observed for strain MNG as well as for strain #6, this could indicate that these putative genes might be present on the chromosome of strains MNG and #6. Slightly smaller bands of $\pm 4.5-5$ kb were seen for strains “f” and C-SH12. The plasmid signal of Tn 39 corresponds in size to the signal from C-SH12 suggesting that a plasmid copy similar to the pRSB gene is present in this strain. Strain BC13 also gave a positive hybridization signal of ± 3 kb but no signal was seen for strain KU. Another small positive signal of 1.6 kb was observed for plasmid Tn 39 which might suggest that the arrangement of this region could be different in Tn 39.

Genes with homology to the genes present on pTcM1 seem to be present as copies on the chromosome of strain MNG as well. The plasmid might have integrated into the chromosome through homologous recombination between IS elements present on both the plasmid and the chromosome. It still remains to be seen whether or not this plasmid contributes to the metabolism or fitness of strain MNG. The

presence of the genes in the other *At. caldus* strains and on plasmids isolated from these strains shows that there is a general backbone of similar accessory genes carried on plasmids of these biomining bacteria. It still remains to be seen whether or not the accessory genes carried on this plasmid contributes to the metabolism or fitness of strain MNG or if they are just a product of horizontal gene transfer is not clear.

CHAPTER FOUR: EXPRESSION AND FUNCTIONALITY ANALYSIS OF PUTATIVE GENES

4.1 INTRODUCTION

Regulation of gene expression is crucial for life; bacteria adapt to environmental changes such as changes in nutrient sources by regulating the genes involved in metabolic pathways. Transcriptional regulators modulate the cellular physiology at the level of transcription by regulating other genes or operons and interacting with upstream regions to regulate their own transcription. The GntR family of regulators responds to various oxidized substrates related to either amino acid metabolism or metabolic pathways (Huffman and Brennan, 2002 and Vindal 2007). GntR regulators belong to the larger GalR family of regulators and it has been suggested that they play a similar role as that of the LacI (lactose repressor) of the *lac* operon (Peekhouse and Conway, 1998, Tong *et al.*, 1996, Weickert and Adhya, 1992). The *lac* operon has three structural genes β -galactosidase, β -galactosidase permease and β -galactosidase transacetylase; which are controlled by the repressor protein. The repressor protein of *E. coli* is constitutively expressed, it binds to an upstream cis-activated operator and blocks transcription of the genes involved in the utilization of lactose by distorting the conformation of the operator and inhibiting RNA polymerase to bind to its promoter and initiate transcription (Daber *et al.*, 2007). This negative regulation is relieved when the effector molecule, namely allolactose (analog of lactose), binds to the repressor and activates gene expression.

The GntR type regulators repress the genes involved in gluconate metabolism and glucose uptake (the genomic organization of GntR-type regulators is shown in Figure 43) (Frunzke *et al.*, 2008; Peekhouse and Conway, 1998). In many *Bacillus* species the GntR, GntP (gluconate permease which transports glucose into the bacterial cytoplasm) and the GntK (gluconate kinase which phosphorylates gluconate to 6-phosphogluconate) are clustered into an operon. Expression of these genes is derepressed in the presence of gluconate and subject to carbon catabolite repression by catabolite control protein CcpA and the phosphocarrier Hpr (Reizer *et al.*, 1996). In *E. coli* the *gnt* genes are also repressed by gluconate repressor GntR and activated by CRP (cAMP repressor protein). This illustrates that the expression of the *gnt* genes is controlled by the availability of gluconate and the presence of a catabolite repressor protein such as glucose. Frunzke and his coworkers (2008) showed that *Corynebacterium glutamicum* is able to adjust its uptake rates of different carbon sources (gluconate vs. glucose) to maximize the bacterium's growth rate and provide a selective advantage over its competitors.

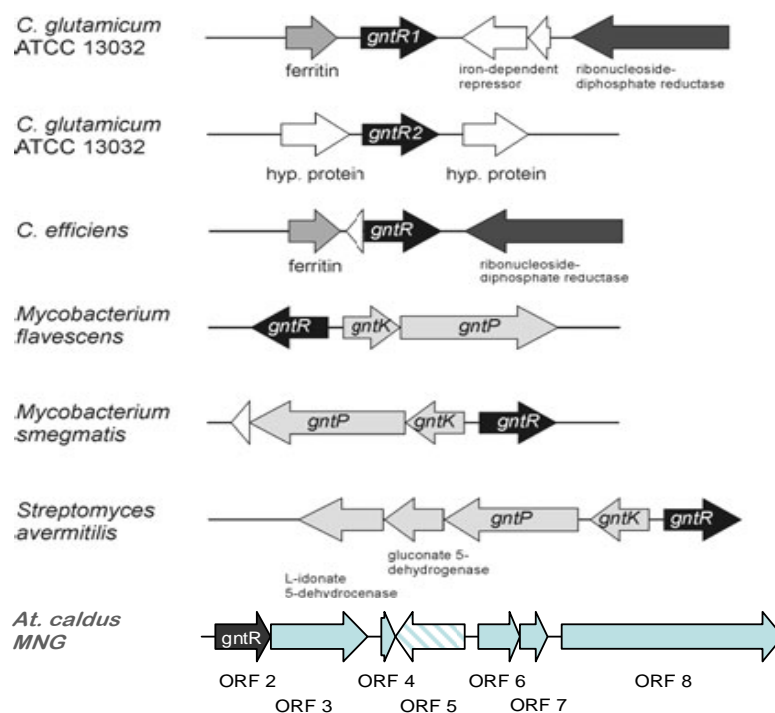


Figure 43: The genomic organization of the GntR-type regulators of *Corynebacterium*, *Streptomyces* and *Mycobacterium* species. The orientation of the genes vary. Gluconate permease (GntP), gluconate kinase (GntK) and gluconate regulator (GntR)²⁵. The putative operon 1 of *At. caldus* MNG is also shown. ORF 3: succinate dehydrogenase / fumarate reductase, ORF 4: ferredoxin 1, ORF 5: insertion element, ORF 6: hypothetical protein, ORF 7: ferredoxin 2, ORF 8: NADPH-dependent glutamate synthase.

Two divergently located operons carrying accessory genes on pTcM1 were identified (Chapter 2). The first putative operon includes ORF's 2, 3, 4, 6, 7 and 8 but the insertion element of ORF 5 seems to disrupt some of these genes (Figure 19). The second operon includes ORF's 9, 10, 11, 12, 14, 15 and 16 which are similarly interrupted by the insertion element ORF 13. The putative genes of operon 1 seem to be involved in metabolic pathways, while those of operon 2 seem to be involved in transport systems. In this chapter we determine if the genes from these operons are expressed and if they are regulated by the transcriptional regulator (ORF 2). Promoter fusion constructs were made to determine if the promoter regions of the genes are functional. β -Galactosidase activities of these constructs and the influence of the regulator *in trans* are analyzed.

²⁵ Frunzke et al., 2007, Data were taken from the bioinformatics software ERGO (Integrated Genomics). a Identity of the amino acid sequence to GntR1 (encoded by *cg2783*) of *C. glutamicum*.

4.2 MATERIALS AND METHODS

4.2.1. MEDIA AND GROWTH CONDITIONS

E. coli strains were grown at 37°C in LB (see Appendix 1) broth and agar medium supplemented with ampicillin (100µg/ml) and kanamycin (50µg/ml) as needed (Sambrook *et al.*, 1989). Acidithiobacilli strains were grown at 30°C and 37°C in Basal salts (1x) solution with additional 0.5 ml potassium tetrathionate (K₂O₆S₄) (Appendix 1) at a pH of 2.5 or 0.5 ml sodium thiosulphate (Na₂S₂O₃) (Appendix 1) at a pH of 4.7.

4.2.2. BACTERIAL STRAINS, PLASMID CONSTRUCTS AND RT-PCR PRIMERS

The bacterial strains, plasmid constructs and primers used in this study for cloning, sequencing and PCR reactions are summarized in Appendix 2 & 3.

4.2.3. TOTAL RNA EXTRACTIONS AND PURIFICATIONS

Total RNA from *At. caldus* (500 – 1000ml) strains and *E. coli* (50ml) cultures grown to mid-exponential phase and isolated as described by Trindade *et al.*, 2003. *E. coli* cultures were grown in LB broth media and *At. caldus* strains were grown in tetrathionate media (Appendix 1).

4.2.4. SLOT BLOT ANALYSIS

Total RNA (0.125ug) from *At. caldus* and *E. coli* strains were transferred to a Hybond-N⁺ nylon membrane (Amersham) using a Slot-blot Manifold® II grid from Schleicher & Schuell (Sambrook *et al.*, 1989). Dioxigenin labeled DNA probes were used to detect transcripts for the succinate dehydrogenase (ORF 3) and glutamate synthase (GOGAT, ORF 8) and hybridizations were carried out according to the manufactures recommendations. The dioxigenin-dUTP non-radioactive DNA labeling and detection kit of Roche Molecular Biochemicals was used for labeling and detection of probes.

4.2.5. PROMOTER-*lacZ* REPORTER CONSTRUCTS

PCR primers of the open reading frames for putative operon 1 and 2 were designed to amplify the putative promoter regions of the genes (Appendix 3). The primers were designed to amplify the promoter regions of the genes ± 300 -500 bp before the transcriptional start site. Linker sites containing *EcoRI* and *BamHI* restriction sites were used to insert the promoter regions into the promoterless *lacZ* reporter gene of pMC1403. The PCR products were cloned into pGEM® T-Vector and pGEM®-Easy Vector from which the PCR inserts were digested and ligated into pMC1403. The in-frame transcriptional-translational fusion products were confirmed by DNA sequencing (the position of the ribosomal binding sites relative to the nucleotide sequence is indicated in Appendix 3).

4.2.6. β -GALACTOSIDASE ACTIVITY ASSAYS

E. coli cultures were grown in 5ml LB broth medium containing the appropriate antibiotics. These overnight cultures were then diluted (50x to 100x) into fresh 5 ml LB broth medium containing the same antibiotics and were incubated for 3-4 hours (OD_{600} 0.4-0.6). *E. coli* strains grown in minimal media (Appendix 1) were grown overnight and re-inoculated into fresh media and incubated for 6-8 hours when OD_{600} readings reached the exponential phase (between 0.4 and 0.6). The β -galactosidase activity was measured in a 1000 μ l assay volume (1:9 dilutions) according to the method of Miller (Miller, J.H., 1972). The β -galactosidase assays were done in triplicate during each experiment and all experiments were repeated at least twice.

4.2.7. RT-PCR ANALYSIS OF mRNA

Total RNA was extracted as previously described (4.2.3). The Invitrogen AMV 1st Strand cDNA Synthesis Kit (Cat no. 12328-032) and the First Strand cDNA Synthesis Kit (Cat no. 04379012001) from Roche Molecular Biochemicals and were used for the synthesis of cDNA and the detection of the cDNA product during RT-PCR. The first step reverse-transcriptase reactions were carried out according to the manufactures specifications using 0.125ng of template total RNA (RNA concentration determined by a Nanodrop spectrophotometer, Genetics Department). The second step PCR reactions were carried out using 3 μ l of template cDNA product (0.1 μ g). PCR reactions was performed at 95°C for 120s; this was followed by 30 cycles alternating between the denaturing step at 95°C for 60s, the annealing step at 50-55°C for 60s and the elongation step at 72°C for 90s. The RT-PCR primers are

listed in Appendix 3. The primer pairs are illustrated in Figure 44 and summarized in Table 3. As positive control the following primer sets were used in the PCR reaction: RTGOGATFW and RTGOGATRV and RTSUCFW and RTSUCRV with a DNA template. To ensure that mRNA was not contaminated with DNA, a PCR reaction using primers designed to amplify the bacterial 16S rDNA was performed using RNA extractions as template. As negative control for the RT PCR's the internal controls were performed without AMV reverse transcriptase.

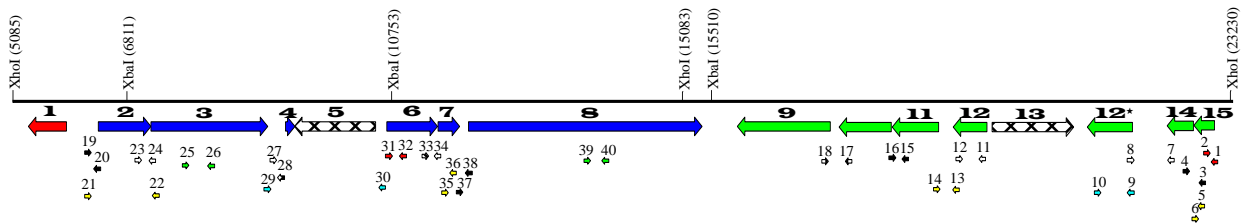


Figure 44: A physical map illustrating the primer pairs used to construct promoter-fusions, RT-PCR and cloning of the regulator gene are summarized in table 5. The Open reading frames are indicated with the arrows.

Table 3: A summary of the location and designation of the primers used.

Number	Description	Location	Number	Description	Location
1	PMF_pRSB_101_23_EcoRI	23018 bp	21	REGFW_BamHI	1069 bp
2	PMF_pRSB101_23_BamHI	17719 bp	22	REGRV_Sal	2179 bp
3	pRSBFW	17751 bp	23	PMF_SUC_EcoRI	1817 bp
4	pRSBRV	17411 bp	24	PMF_SUC_BamHI	2127 bp
5	PMF_pRSB_101_24_EcoRI	17735 bp	25	RTSUCFW	2523 bp
6	PMF_pRSB_101_24_BamHI	14543 bp	26	RTSUCRV	3002 bp
7	PMF_ABC1_EcoRI	17284bp	27	FER2FW	3829 bp
8	PMF_ABC1_BamHI	16585 bp	28	FER2RV	4042 bp
9	ABCF01	16685 bp	29	PMF_Phyco_EcoRI	3740 bp
10	ABCR01	16195 bp	30	PMF_Phyco_BamHI	5545 bp
11	ABCF2	14475 bp	31	PHYCOFW	5548 bp
12	ABCR2	14036 bp	32	PHYCORV	5857 bp
13	PMF_BIN_EcoRI	14085 bp	33	PMF_FER1_EcoRI	6090 bp
14	PMF_BIN_BamHI	13698 bp	34	PMF_FER1_BamHI	6372 bp
15	PMF_ABC3_EcoRI	13329 bp	35	FER1FW	6384 bp
16	PMF_ABC3_BamHI	13039 bp	36	FER1RV	6600 bp

17	PMF_SIG_EcoRI	12484 bp	37	PMF_GOGAT_EcoRI	6601 bp
18	PMF_SIG_BamHI	12036 bp	38	PMF_GOGAT_BamHI	6835 bp
19	PMF_REG_EcoRI	1069 bp	39	RTGOGATFW	8496 bp
20	PMF_REG_BamHI	1307 bp	40	RTGOGATRV	8866 bp

4.3. RESULTS

4.3.1 ANALYSIS OF PROMOTER ACTIVITY OF SELECTED GENES

Promoter fusion constructs were obtained for ORF's 2, 3, 7 and 8 of operon 1. Unfortunately no clones could be made of the promoter constructs for operon 2. A possible explanation for this could be that the genes of operon two might be interrupted and that the promoters might not be intact. A list of primers is given in Table 3 and the positions of these primers relative to the genes are illustrated in Figure 44.

The promoter fusion constructs were grown in both LB broth and minimal media. It was decided to include a known positive control (pLacZ3) in this assay which was known to give a strong β -galactosidase activity in previous experiments (Rawlings research group). As negative control the promoterless vector pMC1403 was used (Figure 45). These results confirm that the promoters for ORF's 2, 3, 7 and 8 are functional in *E. coli*. β -Galactosidase expression levels when fused to three of the genes (ORF's 7 and 8) were low ($\pm 50 - 70$ Miller units), but well above background. Expression of the β -galactosidase fused to the succinate gene (ORF 3) was considerably stronger than the rest (± 210 Miller units).

To determine if the regulator identified upstream of the operon serves to increase or decrease transcription on the genes downstream of the regulator gene it must be cloned and added *in trans* to the promoter constructs. Primer pair REGFW2 and REGRV2 were used to amplify the regulator gene product and introduce the restriction sites *Bam*HI and *Sal*I needed for directional cloning into pACYC184. The Reg-pACYC184 construct was provided *in trans* with the promoter fusions.

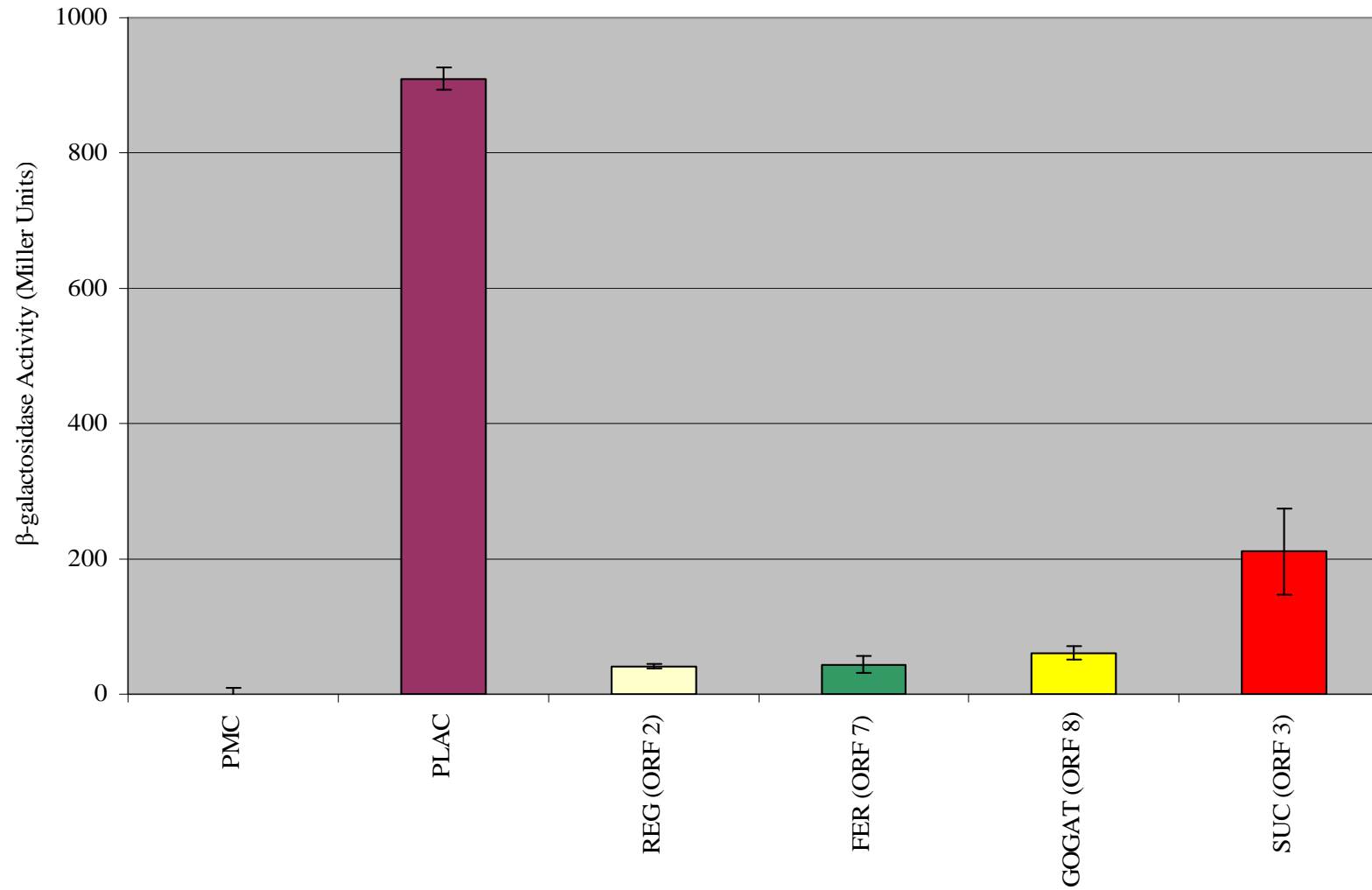


Figure 45: The β -galactosidase activity of the transcriptional regulator, ferredoxin, GOGAT and succinate dehydrogenase promoter fusion constructs grown in LB broth. At this stage there is no regulator *in trans*. As positive control construct pLacZ3 was included. As negative control the promoterless vector pMC1403 was used.

The genes identified in putative operon 1 seem to be involved with metabolism, to ensure that there would not be any metabolites present in the LB broth that could mask the regulator's activity the constructs were grown in minimal media to limit the potential influence on the regulator. The β -galactosidase assay was repeated in LB broth and minimal media (Figure 46 and 47) with the regulator *in trans* to see whether or not the β -galactosidase activity of the promoter constructs are influenced by the regulator. There is an increase in the base-line level of expression from the promoter fusion constructs when the pACYC184 constructs are placed *in trans*. The reason for this artifact is not fully clear, however the same increase was observed for all promoter fusion constructs.

The β -galactosidase activity of the constructs with the regulator *in trans* grown in LB broth showed a decrease of 49% for the succinate dehydrogenase promoter construct, 23% for the GOGAT promoter construct and 19% for the ferredoxin construct (Figure 46). While a decrease of 69% in β -galactosidase activity could be observed for the succinate dehydrogenase promoter construct with the regulator *in trans*, 26% for the ferredoxin promoter construct and 1% for the GOGAT promoter construct in minimal media was observed (Figure 47). In Figure 48 the β -galactosidase activity of the constructs with the regulator *in trans* in both minimal media and LB broth is illustrated. Even though the base-line promoter activities of the promoter fusion constructs had been increased, the decrease in β -galactosidase activity when the regulator is placed *in trans* still allows us to draw the conclusion that the regulator decreases the activity of the promoter fusion constructs, as described above, and therefore serves as a regulator.

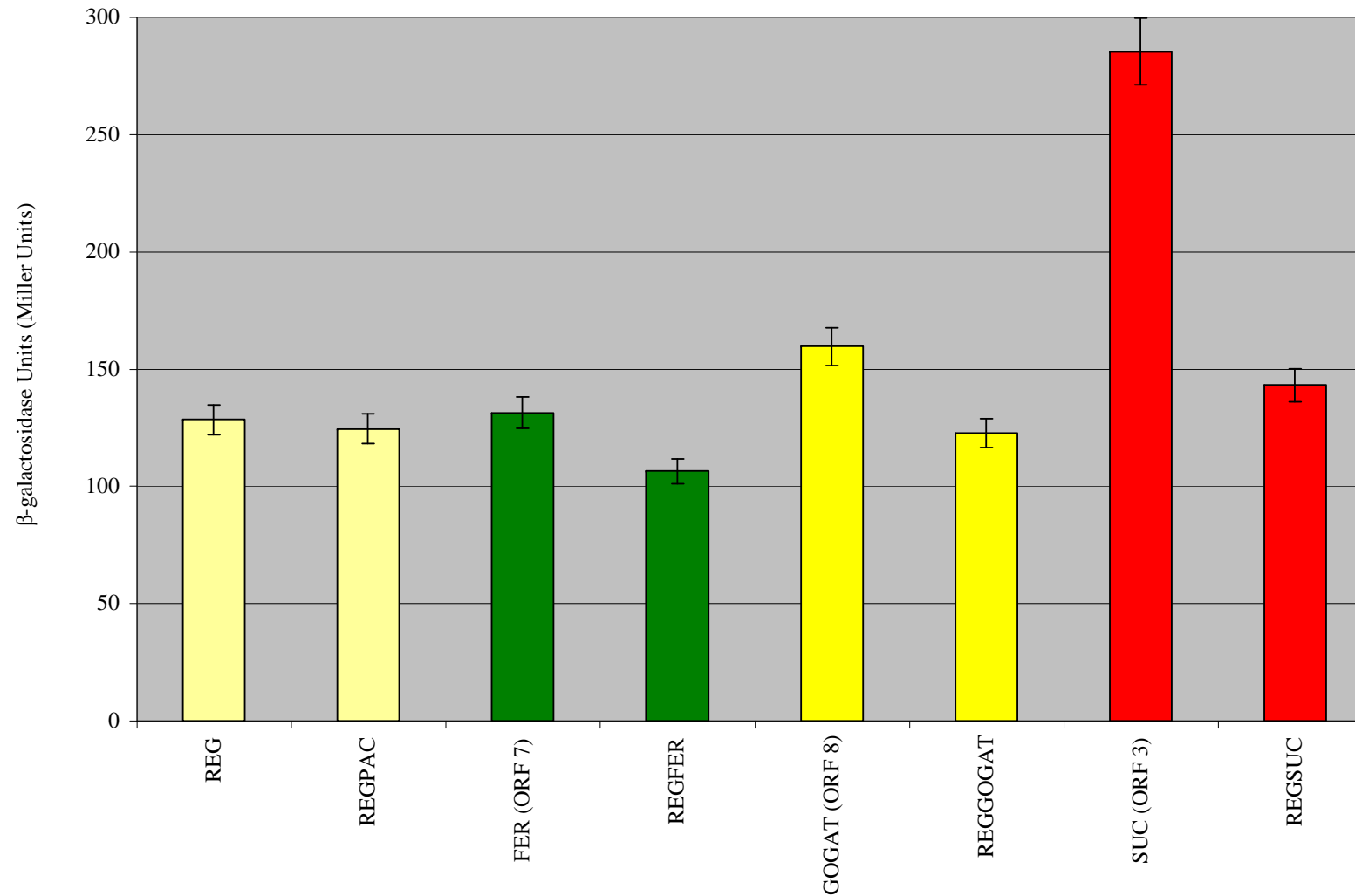


Figure 46: The β -galactosidase activity of the transcriptional regulator *in trans* with the ferredoxin (ORF 7), GOGAT (ORF 8) and succinate dehydrogenase (ORF 4) promoter fusion constructs grown in LB broth. The activity of the regulator- pACYC184 (REGPAC) construct was included.

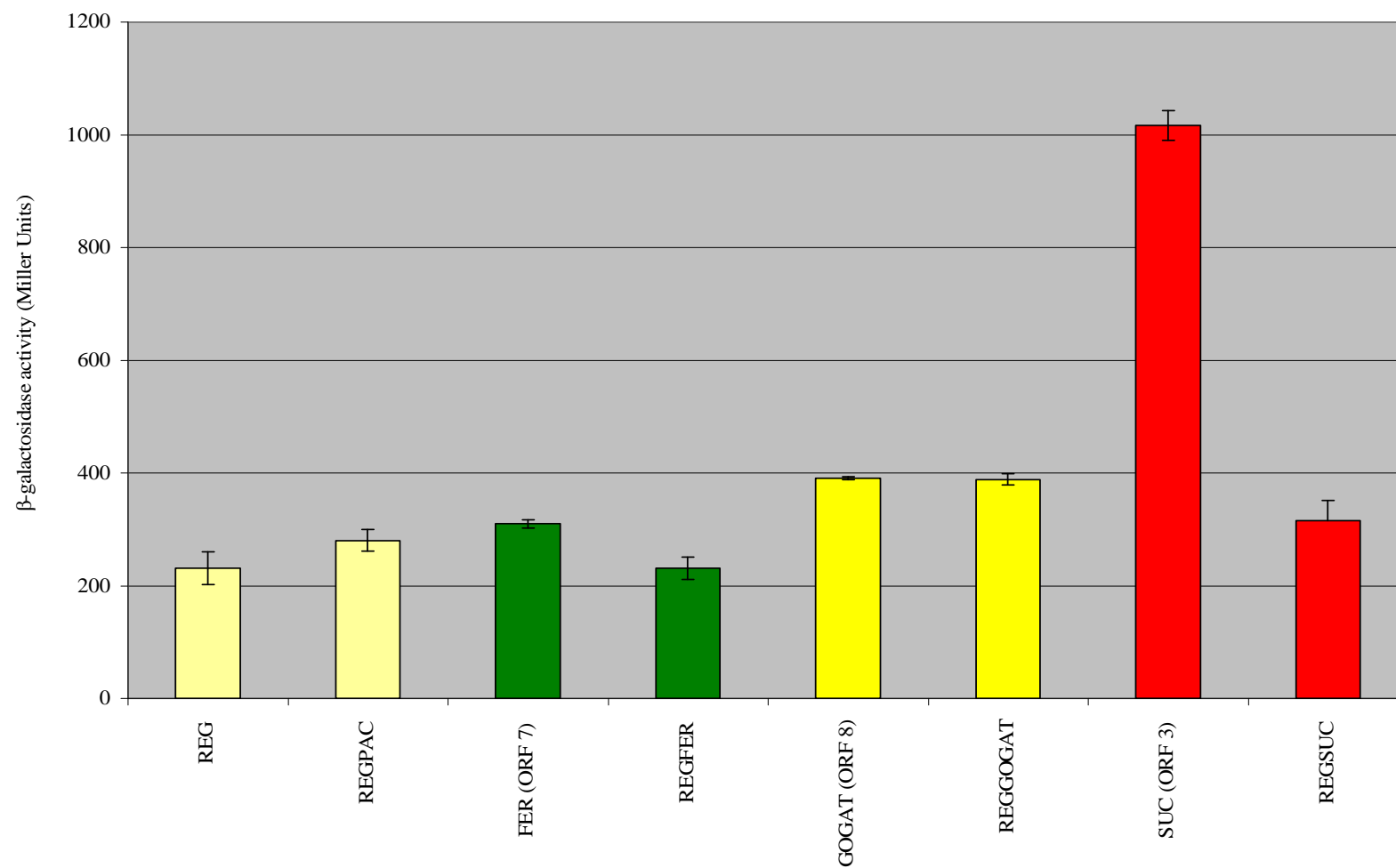


Figure 47: The β -galactosidase activity of the transcriptional regulator *in trans* with the ferredoxin (ORF 7), GOGAT (ORF 8) and succinate dehydrogenase (ORF 4) promoter fusion constructs grown in Minimal media. The activity of the regulator- pACYC184(REGPAC) construct was included.

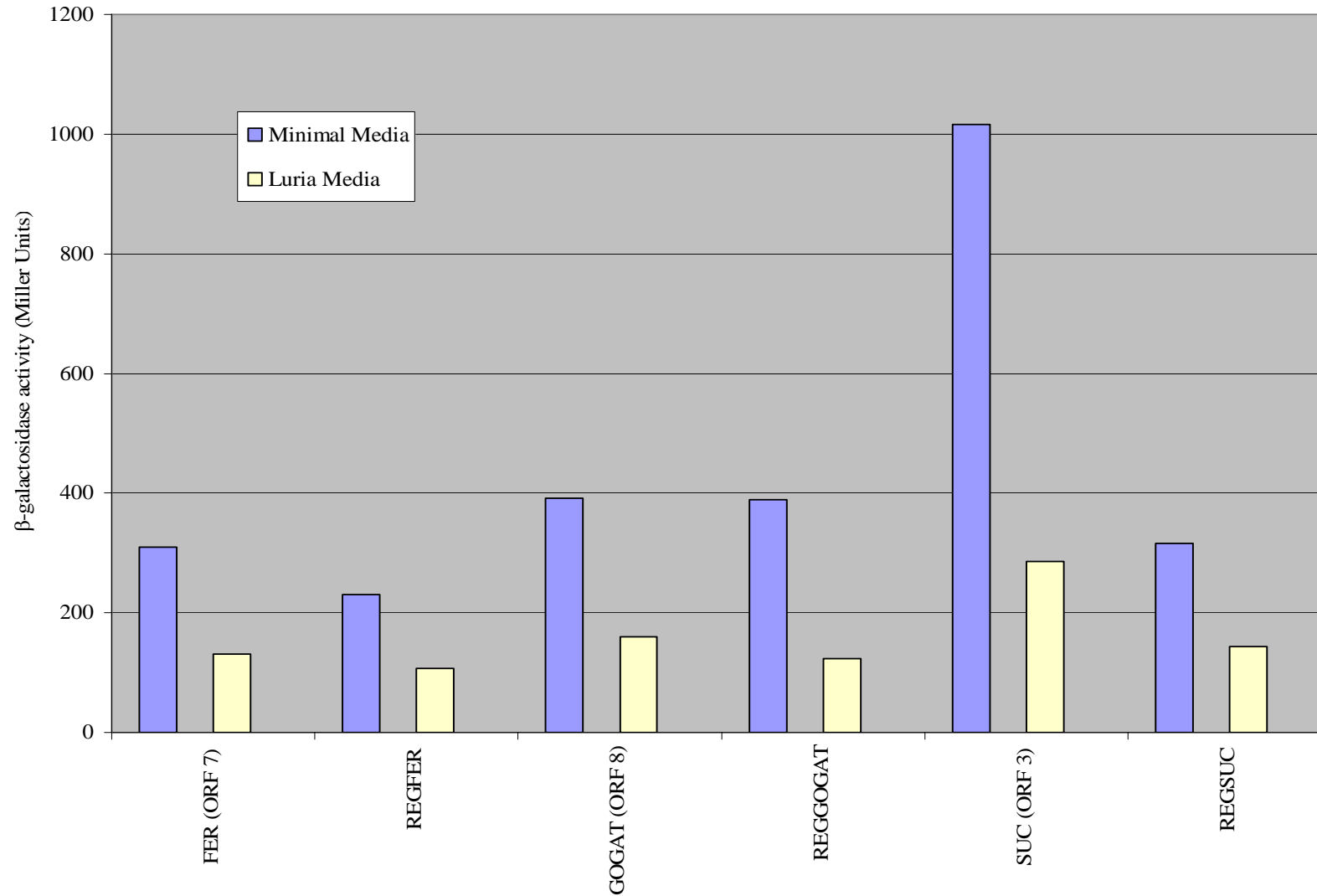


Figure 48: A comparison of the β -galactosidase activity of the transcriptional regulator *in trans* with the ferredoxin (ORF 7), GOGAT (ORF 8) and succinate dehydrogenase (ORF 4) promoter fusion constructs grown in Minimal media and LB broth.

4.3.2. DETECTION OF mRNA EXPRESSION

The study of gene expression using β -galactosidase assays could be carried out in *E. coli* only, whereas it is more relevant to study gene expression in *At. caldus*, the bacterium that was the source of the plasmid. To compare the results of the β -galactosidase activity assays with gene expression in *At. caldus*, slot blots of the genes with the highest and lowest promoter activities were used to determine if mRNA is transcribed for the genes in *At. caldus*.

Total RNA was extracted from strains KU, C-SH12, MNG, #6, “f” and BC-13 and DNA probes of the GOGAT-like and succinate dehydrogenase-like genes and they were hybridized to the blots. As negative and positive controls total RNA from the *E. coli* strains Ec-100D and Ec-100D-Tn 11 (harboring plasmid pTcM1) respectively, were used. Expression could be shown for the transcription products for strains MNG, #6, “f”, C-SH12 and BC13. (Figure 49).

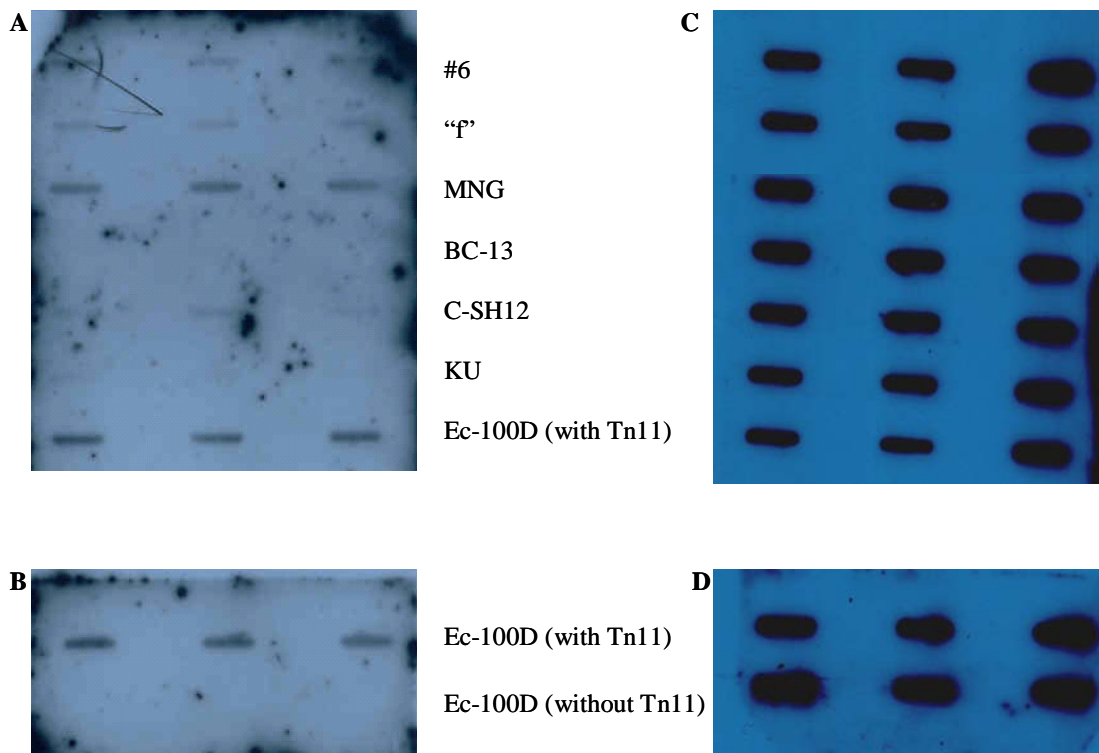


Figure 49: A: autoradiograph of total RNA from strains #6, “f”, MNG, BC13, C-SH12 and KU probed with a GOGAT-like gene. B: positive control and negative controls of strain Ec-100D with and without plasmid Tn 11. C: 16S controls of RNA samples. D: 16S controls of positive and negative controls. Maximum stringency washes were used through the experiments.

Total RNA of *At. caldus* strains were probed with the succinate dehydrogenase-like gene. Positive hybridization signals for the succinate dehydrogenase transcripts were found for strains MNG, #6, “f”, C-SH12 and BC13 (Figure 50A & B). In Figure 50 A it can be seen that the *E. coli* and *At. caldus* strain BC13 gave very strong hybridizations signals which is due to the fact that there is much more RNA present for these two strains compared to the other strains on the blot (refer to 16S control blot Figure 50 C). No transcription product was observed for strain KU.

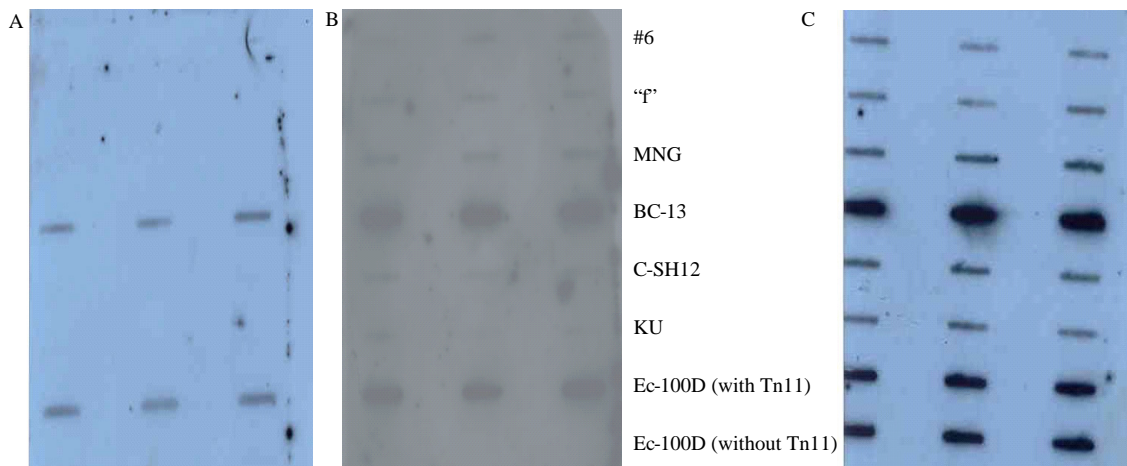


Figure 50: Slot blots autoradiographs of total RNA from *At. caldus* strains #6, “f”, MNG, BC13, C-SH12 and KU hybridized to a probe of the succinate dehydrogenase gene. A: exposure after first hour. B: longer exposure. C: 16S control of RNA samples.

4.3.3. ANALYSIS OF THE TRANSCRIPTION PRODUCTS

One of the questions that we would like to answer is whether or not the genes located within operon 1 and 2 are co-transcribed on one mRNA or whether the genes are transcribed separately. cDNA was prepared from total RNA samples as outlined in 4.2.7. An internal region of ± 400 bp from the GOGAT gene and internal region of ± 480 bp from the succinate dehydrogenase gene were used as positive controls (the location of these primer pairs are shown in Figure 44). The expected sizes of the internal controls could be observed as shown in Figure 51. A 1kb transcription product could be observed for the regulator gene.

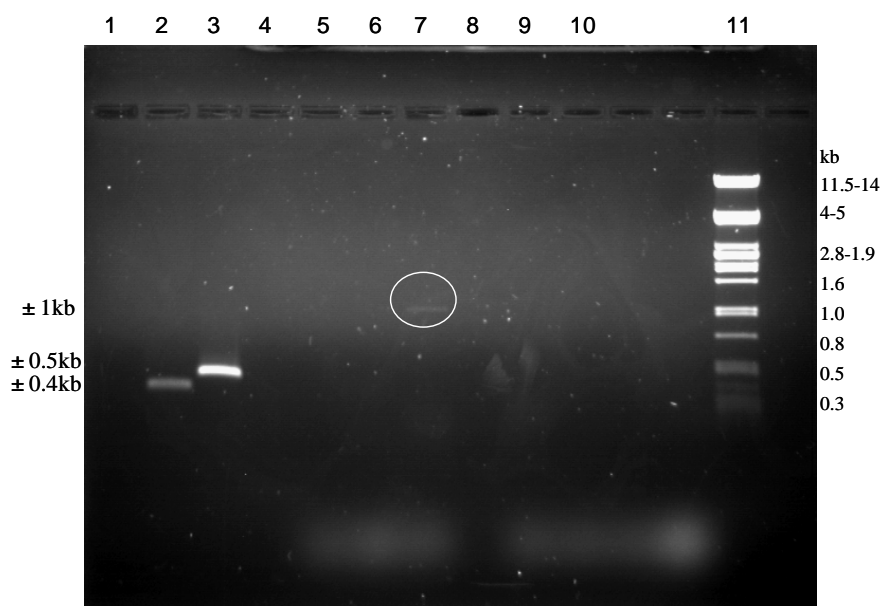


Figure 51: The reverse transcript products from *At. caldus* strain MNG. Lane 1: negative control (no AMV for internal GOGAT gene). Lane 2: positive control of the GOGAT gene (internal primer product). Lane 3: positive control of the succinate dehydrogenase gene (internal primer product). Lane 4: RT product between the GOGAT and ferredoxin #1. Lane 5: RT product between ferredoxin #1 and the putative phycobiliprotein. Lane 6: RT product between the succinate dehydrogenase and the regulator proteins. Lane 7: RT product between ORF 1 and the regulator protein. Lane 8: RT product between the putative phycobiliprotein and ORF 5. Lane 9: negative control for PCR (no DNA). Lane 10: RT product between the succinate dehydrogenase protein and ferredoxin #2. Lane 11: λ -Pst Marker.

4.4. DISCUSSION

Literature has shown that GntR protein family of transcriptional regulators usually act as repressor proteins (Hillerich and Westpheling, 2006), especially in response to environmental influences. They have been linked to the carbon cycle responding to intermediates containing carboxylate. The results support the theory that the transcriptional regulator (ORF 2) acts as a repressor protein. The β -galactosidase activity for the promoter fusion constructs of operon 1 have shown that the promoter regions are being expressed in *E. coli*. There was a significant difference observed in the β -galactosidase activity of the constructs with the regulator *in trans* in the minimal media and rich media (LB broth). We expected that the succinate dehydrogenase gene located directly downstream of the regulator gene would be the likely candidate to be influenced by the regulator. The succinate dehydrogenase displayed a very high β -galactosidase activity of over 1000 Miller Units and the effect of the added regulator *in trans* decreased its activity substantially by 701 Miller units in minimal media and 142 Miller units in LB broth (Figure 48). Similar results were observed for the regulator *in trans*

with the ferredoxin promoter constructs, a decrease in 79 Miller units in minimal media and 24.8 Miller units in LB broth was observed. There did not seem to be much of a repression effect on the GOGAT promoter constructs *in trans* with the regulator in minimal media. However there was a repression of 36.8 Miller units in the LB broth. This suggests that the GOGAT gene is suppressed in rich media but under nitrogen deprivation the repression of GOGAT gene on plasmid pTCM1 could be derepressed. This could be explained by catabolite repression or the glucose effect. In the presence of glucose and lactose, the cAMP levels will drop which results in the deactivation of the catabolite activator protein and inhibition of the *lac* operon expression. Easily catabolized carbon sources like glucose is consumed first while more complex sources are used when glucose is depleted. During this time the genes coding for the catabolism of other carbon and nitrogen sources are repressed (Wacker *et al.*, 2003). This is a common occurrence in bacteria to help conserve energy sources and adapt to environmental changes.

Slot blot hybridizations were performed with RNA probes of the GOGAT and succinate dehydrogenase genes to compare it with the β -galactosidase results. The two genes with the strongest and the weakest promoter activities were used for the hybridizations. Signals could be observed for the GOGAT mRNA products could be observed for *At. caldus* strains MNG, #6, “f” and C-SH12 but not for strains BC 13 and KU (Figure 49). This corresponds to the results obtained from the southern hybridizations in chapter 3. Strains BC13 and KU do not seem to harbor plasmid copies of the GOGAT gene. A possible explanation for the presence of the GOGAT gene on the plasmids might be that the chromosomal copy of the GOGAT gene in the *At. caldus* strains might be inactive or insufficient and an extra copy of the GOGAT is needed on plasmids. The slot blot hybridizations for the succinate dehydrogenase gene showed expression of mRNA for strains #6, “f”, MNG, C-SH12 and BC13 but not for KU (Figure 50). Due to the fact that the slot blots were not analyzed, by taking densitometer readings, to compare the level of expression exhibited by each gene we can only confirm expression of the GOGAT-like and the succinate dehydrogenase subunit RNA in *At. caldus* strains. We cannot comment on the relative levels of expression between the strains.

The final question we wished to address in this study was whether or not the genes located in putative operons 1 and 2 are being co-transcribed or transcribed individually. It seems that the genes are transcribed from their own promoters. The internal controls for the GOGAT and succinate dehydrogenase genes indicated that the cDNA was being produced but the lack of products for the other primer pairs indicate that the genes seem to be expressed from their own promoters and not as a

unit. Unfortunately all attempts to clone the promoter regions for the genes of operon 2 were unsuccessful. Sequencing of this region of the plasmid was also hampered (Chapter 2).

CHAPTER FIVE: GENERAL DISCUSSION

Local changes in DNA sequences, DNA rearrangement and DNA acquisition contribute to the overall fitness of a bacterium and determine its capacity for further evolution (Arber, 2004). Key features of plasmids are that they are able to replicate autonomously and be maintained in a cell lineage without segregational loss. Genes associated with a plasmid often spread rapidly from one genetic environment to another. Often useful genes will be incorporated into the chromosome of a host. In passing through different strains or species, plasmids are exposed to varying combinations of genes some of which are incorporated into the plasmid. Conjugative spread of the plasmid leads to new patterns of selective benefit and diversification and adaptability of a bacterium or a microbial population. Genetic elements can therefore be seen as initiators for genetic variation and important evolutionary tools (Haines, 2005).

Plasmid pTcM1 was fully sequenced by the Rawlings research group (GenBank/EMBL accession no. EU421841, Van Zyl, *et al.*, 2008^a). It was found that pTcM1 could be divided into two large regions based on sequence homology, the first region from position 1 to ± 27 kb contains putative genes involved in replication, metabolism or transport, which includes the 18 kb accessory region sequenced in this study (Figure 52). The other ± 38 kb region consists of what appears to be a “composite transposon” that carries mobilization genes, putative endonuclease and its modification enzymes and arsenic resistance genes (Figure 53).

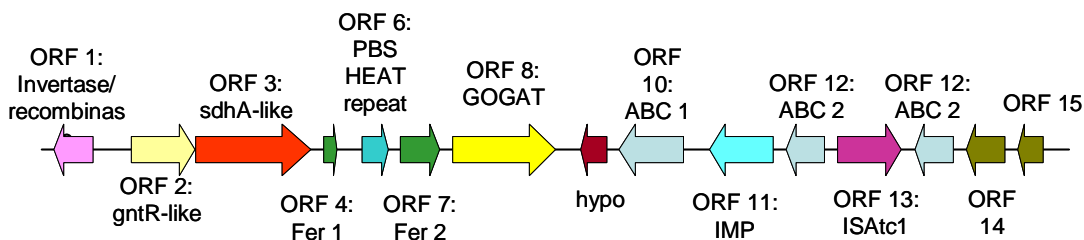


Figure 52: The 18 kb accessory region found on pTcM1.

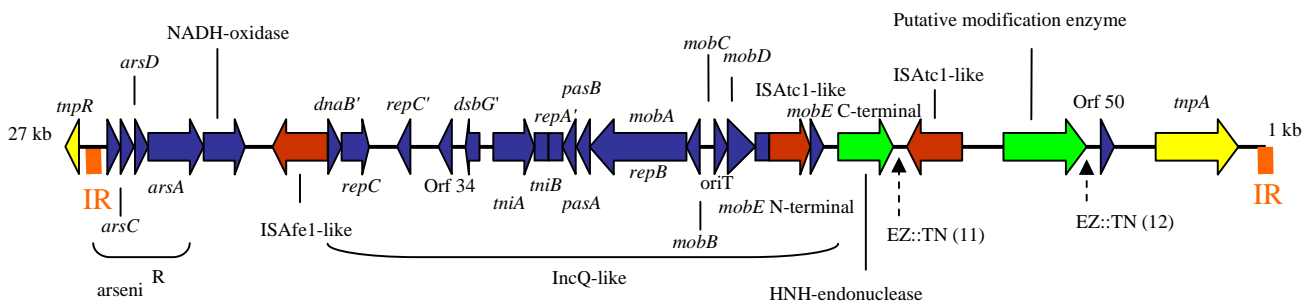


Figure 53: Schematic presentation of the composite transposon located on pTcM1²⁷.

Two other plasmids pTcF1 and pC-SH12, from *At. caldus* “F” and *At. caldus* C-SH12, respectively, were also identified by our research group, sequenced and annotated as shown in Figure 53. A common region of 26 kb was found on these two plasmids that is similar to that of pTcM1, and which includes the 18 kb accessory region and plasmid backbone genes. Another plasmid identified in *At. caldus* strain #6, which has not been cloned but investigated using restriction digests and Southern hybridization, has been shown to also contain this common 26 kb region found on the other three plasmids (Van Zyl *et al.*, 2008^a).

It is thought that plasmid pC-SH12 might be the progenitor plasmid from which plasmid pTcM1 and pTcF1 arose by acquiring the composite transposon (Figure 52 containing arsenic resistance and other genes), absent from pTcF1 (Van Zyl *et al.*, 2008^a). In chapter 3 two insertion sequences were shown to be present on the 18 kb DNA fragment as described there. However, a total of five insertion sequences were found on pTcM1 (Figure 54). An insertion element namely ISAfe-1-type element was found in the *ars* region of the plasmid, with 95.5% homology to the other four ISAtc1-like insertion elements (which includes ORF’s 5 and 13 found in the 18 kb region) (Van Zyl *et al.*, 2008^a). It was also found that two IS elements of the related plasmid pTcF1 were also ISAtc1-like and similar to ORF’s 5 and 13, but they were located in different regions, which suggests that transposition took place after the plasmids diverged.

Comparison of the plasmid backbone with those of other plasmids in the database has suggested that the general features of the backbone region are not unique to plasmids from *At. caldus* but are also found in plasmids from other environments such as activated sludge like the pRSB101-like plasmids.

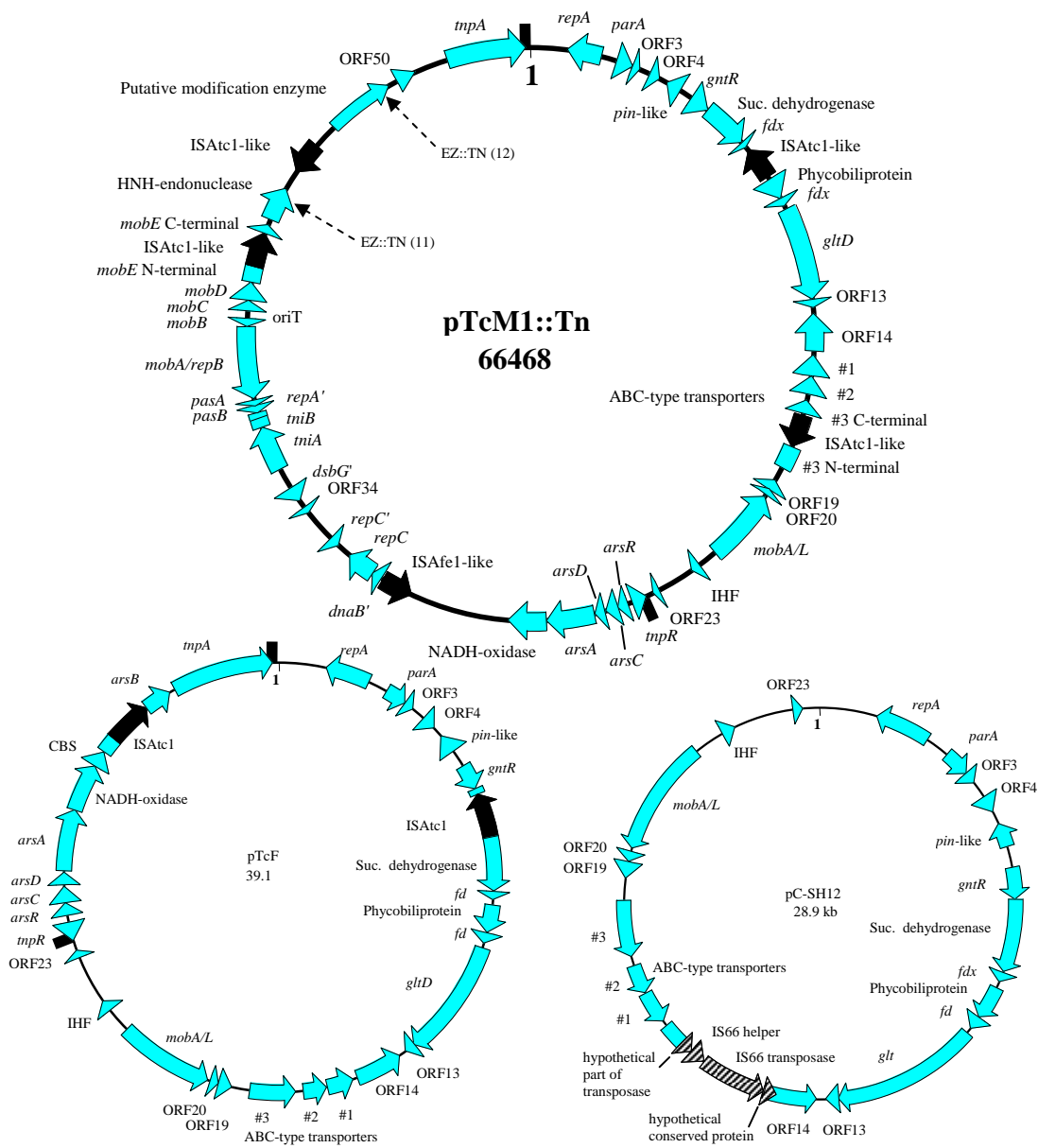


Figure 54: Genetic map of plasmids pTcM1, pTcF1 and pC-SH12²⁶. The arrows indicate the coding regions. Solid bars indicate the insertion elements.

These plasmids (pRSB101-like) have been shown to carry a strikingly similar gene complement compared to that of the *At. caldus* family of plasmids consisting of heavy metal resistance genes, genes

²⁶ Van Zyl *et al.*, 2008^a

involved in metabolic pathways as well as parts of IncQ-like plasmids. Although this class of plasmids from *At. caldus* (pTcM1, pTcF1 and pC-SH12) has not been formally classified into an incompatibility group, it is remarkably similar, in terms of the arrangement of the replication and partitioning genes as well as the complement of accessory genes it carries, to the pRSB101-like plasmids (Figure 55).

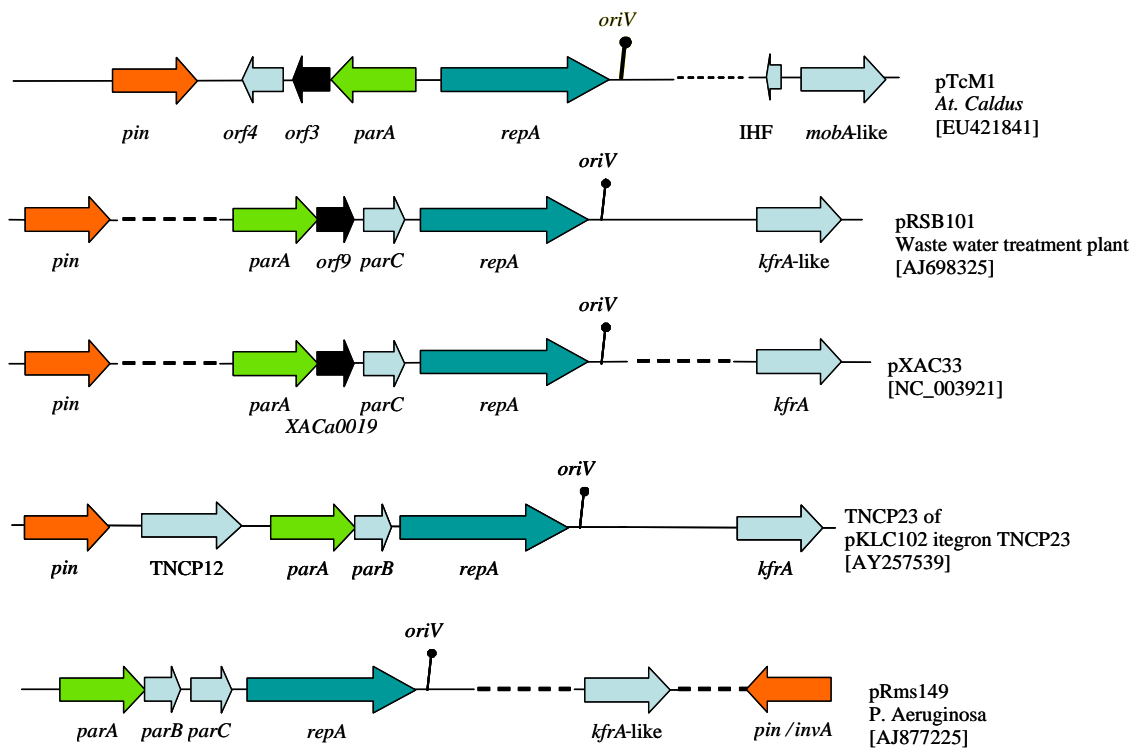


Figure 55: A comparison of the backbone region of pTcM1 to pRSB101, pXAC33, TNCP23 of pKLC102, pRms149²⁷. The dashed lines join noncontiguous regions of DNA. Putative *oriV* regions have been flagged. GenBank/EMBL accession numbers are indicated.

To determine whether or not pTcM1 forms part of the horizontal gene pool is difficult as we do not have a clear history of the biomining inoculums, which could shed light on whether the occurrence of these plasmids in strains isolated on different continents are as a result of human intervention or a natural process. Isolation of IncQ plasmids from diverse environments as varied as activated sludge from water treatment plants to acidic biomining environments demonstrates that IncQ-type plasmids have a broad-host-range and are promiscuous. One of the mobilization genes identified on pTcM1 (labeled *mobA/L*), as well as on pC-SH12 and pTcF1, is related to the IncQ-like *mobA* gene from pRSF1010. The plasmid (pTcM1) was however not mobilizable by pRP4 or pR751, both of which can

²⁷ Van zyl et al., 2008^a

mobilize IncQ-like plasmids. It may be that provided with the correct conjugative plasmid the pRSF1010 *mobA*-like gene from pTcM1 can serve to move this plasmid within the bacterial community. (Van Zyl *et al.*, 2008^a)

A second set of mobilization genes were identified on pTcM1 which are identical to the mobilization genes found on the IncQ-like pTC-F14, isolated from an *At. caldus* strain. Van Zyl and co-workers have demonstrated that plasmid pTcM1 was mobilizable using the IncP conjugative plasmid RP4 via the pTC-F14-like *mob* genes present on pTcM1 (Rawlings, 2005^b, Van Zyl *et al.*, 2003 and 2008^{a,b}). These appear to have been acquired through a transposition event and therefore may not form part of the progenitor plasmid and therefore may have no bearing on the relatedness of this plasmid (pTcM1) to other IncQ-like plasmids.

The presence of this second set of mobilizable genes together with the *mobA/L*-like gene on pTcM1, pTcF1 and pCSH12, as well as the demonstration that when provided with the correct conjugative plasmid, these plasmids are mobilizable, would suggest that they could form part of the horizontal gene pool.

The presence of the accessory genes on the plasmid in most of the *At. caldus* strains as shown by the hybridization studies in chapter 3, suggests that this region forms part of a common DNA cluster of metabolic genes in *At. caldus*. The GOGAT-like and succinate dehydrogenase-like genes could possibly provide extra biosynthetic intermediates usually associated with a TCA cycle but expected to be split into oxidative and reductive branches in an autotrophic bacterium like *At. caldus*. A gene for the catalytic α -subunit of GOGAT is not present, and therefore ORF 8 probably does not encode a functional GOGAT subunit but is possibly a gene for a product related to the β -subunit of GOGAT. Assuming that the product of the gene related to the GOGAT β -subunit fulfills a reductive role similar to this subunit (Van der Heuwel *et al.*, 2003) the electrons could be provided by the adjacent ferredoxin gene (ORF 7). Likewise, the catalytic subunit SdhA (ORF 3) could produce fumarate to be incorporated into the TCA cycle-type biosynthetic intermediates. The presence of another ferredoxin-like gene (ORF 4) adjacent to ORF 3 might suggest a link between the two. It was found that the GOGAT (ORF 8) and succinate dehydrogenase (ORF 3) genes were expressed in *At. caldus* (Chapter 4). The likelihood that the genes located between these genes might be expressed as well, is therefore good and suggests that the genes of operon 1 may all be transcribed in *At. caldus*.

The fact that the metabolism-like genes are in the same vicinity as ABC transport genes suggest they might be linked, as the same group of genes namely PBS HEAT-like repeats, succinate dehydrogenase subunit genes, ferredoxins and ABC type genes has been found to be clustered together in *Nitrobacter winogradskyi* (GenBank/EMBL accession no. CP000115.1; Protein database accession no. ABA03944.1 to ABA03939.1) and *Burkholderia multivorans* (GenBank/EMBL accession no, CP000868; Protein database accession no. ABX14981.1 to ABX14987.1) (Van Zyl et al., 2008^a). This clustering of genes related to the pTcM1 accessory genes has therefore been seen on the chromosomes of other bacteria although not on other plasmids.

If the plasmids have not spread rapidly among the *At. caldus* strains through horizontal gene transfer, then the observation that the accessory genes are still present on plasmids from different *At. caldus* strains isolated from such widely geographically separated, suggests that this particular cluster of genes has been part of these plasmids for some time. This may also indicate that the products of these genes probably benefit this bacterium although we are not sure how and to what extent.

FURTHER RESEARCH

It would be interesting to investigate the role of the GntR on regulation of the downstream genes of the putative operon 1 in more detail. If possible, to determine which effector substrates influence the repression of the downstream genes. Furthermore, the functionality of the genes from putative operon 2 have not been investigated because no promoter fusion constructs for these genes could be cloned. The amino acid sequences of ABC transporters are poorly conserved and this makes speculation on the type of substances the genes of putative operon 2 might transport difficult based on amino acid comparisons alone. A further complication would be that given the large differences in pH across the cell membranes of extreme acidophiles compared with neutrophiles like *E. coli*, it may not be possible to investigate which substances are transported by studying transport activity in a more readily retractable bacterium like *E. coli*. It is not certain if *At. caldus* transport genes would function properly in *E. coli*, however if mutants for the ABC-type transporters could be obtained from *At. caldus* this problem could be overcome. The greater challenge is to discover not only what the role of the genes present on the 18 kb region investigated in this study is, but also what the role of the entire pTcM1 plasmid is and why much of it appears to have been conserved in *At. caldus*.

LITERATURE CITED

- Actis, L., Tomasky, M. and Crosa, J. (1999).** Bacterial plasmids: replication of extrachromosomal genetic elements encoding resistance to antimicrobial compounds. *Frontiers in Bioscience*, **4**, 43-62.
- Amils, R. and Ballester, A. (1999).** Biohydrometallurgy and the environment toward the mining of the 21st century, p. 777-786. In Proceedings of the International Biohydrometallurgy Symposium IBS'99. Amils, A and Ballester A (eds.) Process Metallurgy 9A.
- An, F. and Clewell, D. (1997).** The origin of transfer (*oriT*) of the enterococcal, pheromone-responding, cytolytic plasmid pAD1 is located within the *repA* determinant. *Plasmid*, **37**, 87-94.
- Andrade, M., Petosa, C., Donoghue, S., Muller, C. and Bork, P. (2001).** Comparison of ARM and HEAT protein repeats. *Journal of Molecular Biology*, **309**, 1- 18.
- Aparicio, T., Lorenzo, P. and Perera, J. (2000).** pT3.2I, the smallest plasmid of *Thiobacillus* T3.2. *Plasmid*, **44**, 1-11.
- Arber, W. (2004).** Biological evolution: Lessons to be learned from the microbial population biology and genetics. *Research in Microbiology*, **155**, 297 -300.
- Babić, A., Lindner A., Vulić, M., Stewart, E. and Radman, M. (2008).** Direct visualisazion of horizontal gene transfer. *Science*, **319** (5869), 1533-1536.
- Bahl, M., Hansen, L., Goesmann, A. and Sorensen, S. (2007).** The multiple antibiotic resistance IncP-1 plasmid pKJK5 isolated from a soil environment is phylogenetically divergent from members of the previously established α -, β - and δ -subgroups. *Plasmid*, **58**, 31-34.
- Bartosik, D., Wlodarczyk, M. and Thomas, C. M. (1997).** Complete sequence of the replication region of *Paracoccus* (*Thiobacillus*) *versutus* pTAV1 plasmid and its correlation to several plasmids of *Agrobacterium* and *Rhizobium* species. *Plasmid*, **38**, 53-89.
- Belitsky, B., Janssen, P. and Sonenshein, A. (1995).** Sites required for GltC-dependent regulation of *Bacillus subtilis* glutamate synthase expresión. *Structural Fold Des*, **15**, 5686-5695.

- Bennette, P., Grindsted, J., Choi, C. and Richmond, M. (1978).** Characterization of Tn501, a transposon determining resistance to mercuric ions. *Molecular General Genetics*, **159**, 101-106.
- Bergamo, R. F., Novo, M. T., Veríssimo, R. V., Paulino, L. C., Stoppe, N. C., Sato, M. I., Manfio, G. P., Prado, P. I. Garcia, O. Jr. and Ottoboni, L. M. (2004).** Differentiation of *Acidithiobacillus ferrooxidans* and *At. thiooxidans* strains based on 16S-23S rDNA spacer polymorphism analysis. *Research in Microbiology* **155**, 559-567.
- Binda, C., Bossi, R.T., Wakatsuki, S., Artz, S., Coda, A. and Curti, B. (2000).** Cross-talk and ammonia channeling between active centers in the unexpected domain arrangement of glutamate synthase. *Structure*, **8**, 1299-1308.
- Blázquez, J. Navas, A., Gonzalo, P., Martinez, J. and Baquero, F. (1996)** Spread and evolution of natural plasmids harbouring transposon Tn5. *FEMS Microbiol Ecology*, **19**, 63-71.
- Boyd, J., Williams, J., Curtis, B., Kozera, C., Singh, R. and Reith, M. (2003).** Three small, cryptic plasmids from *Aromonas salmonicida* subsp. *Salmonicida* A449. *Plasmid*, **50** (2), 131-144.
- Bramucci, M., Chen, M. and Nagarajan, V. (2006).** Genetic organization of a plasmid from an industrial wastewater bioreactor. *Applied Microbial Biotechnologies*, **71**, 61-74.
- Brierley, C. L. (1997).** Mining biotechnology: Research to commercial development and beyond, p. 1-20. In Rawlings, D. (Ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.
- Brown, L.D., Dennehy, M.E. and Rawlings, D.E. (1994).** The F₁ genes of the F₁F₀ATP synthase genes from the acidophilic bacterium *Thiobacillus ferrooxidans* complement *Escherichia coli* F₁ *unc* mutants. *FEMS Microbiology Letters*, **122**, 19-26.
- Bugaytsova, Z. and Lindström, E. (2004).** Localization, purification and properties of a tetrathionate hydrolase from *Acidithiobacillus caldus*. *European Journal of Biochemistry*, **271**, 272 -280.
- Bult, C., White, O., Olsen, G., Zhou, L., Fleishman, R., Sutton, G., Blake, J., Fitzgerald, L., Clayton, L., Gocayne, J., Kerlavage, A., Doygherty, B., Tomb, J-F., Adams, M., Reich, C., Overbeek, R., Krikness, E., Weinstock, K., Merrick, J., Glodek, A., Scott, J., Geoghagen, N., Weidman, J., Fuhrman, J., Nguyen, D., Utterback, T., Kelly, J., Peterson, J., Shadow, P., Hanna,**

- M., Cotton, M., Roberts, K., Hurst, M., Kaine, B., Borodovsky, M., Klein, H-P., Frazer, C., Smith, H., Woese, C. and Venter, J. (1996).** Complete genome sequence of Methanogenic archaean, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
- Byrd, D. and Matson, S. (1997).** Nicking by transesterification: the reaction catalyzed by a relaxase. *Molecular Microbiology*, **6**, 1011-1022.
- Caspi, R., Helinski, D., Park, M. and Konieczny, F. (2000).** Interactions of DnaA proteins from distantly related bacteria with replication origin of the broad-host range plasmid RK2. *Journal of Biological Chemistry*, **275**, 18454-18461.
- Castillo, A., Taboada, H., Mendoza, A., Valderrama, B., Encarnacion, S. and Mora, J. (2000).** Role of GOGAT in carbon and nitrogen partitioning in *Rhizobium etli*. *Microbiology*, **146**, 1627-1637.
- Ceccarelli, E., Arakaki, A., Cortez, N. and Carillo, N. (2004).** Functional plasticity and catalytic efficiency in plant and bacterial ferredoxin-NADP(H) reductases. *Biochemica et Biophysica Acta*, **1698**, 155-165.
- Chakraborty, R., Deb, C., Lohia, A. and Roy, P. (1997).** Cloning and characterization of a high-copy number novel insertion sequence from chemolithotrophic *Thiobacillus ferrooxidans*. *Plasmid*, **38**, 129-134.
- Chakravarty, L., Zupancic, T., Baker, B., Kittle, J., Fry, I. and Tuovinen, O. (1995).** Characterization of the pTFI91-family replicon of *Thiobacillus ferrooxidans* plasmids. *Canadian Journal of Microbiology*, **41**, 354-365.
- Chakravarty, L. Kittle, J. and Touvinen, O. (1999).** Insertion sequence IST3091 of *Thiobacillus ferrooxidans*. *Canadian Journal of Microbiology*, **43**, 503-508.
- Chattoraj, D. (2000).** Control of plasmid DNA replication by iterons: no longer paradoxical. *Molecular Microbiology*, **37**, 476-476.
- Clennell, A., Johnston, B. and Rawlings, D.E. (1995).** Structure and function of Tn5467, a Tn 21-like transposon located on the *Thiobacillus ferrooxidans* broad-host plasmid, pTF-FC2. *Applied Environmental Microbiology*, **61**, 4223-4229.

Coram, N. and Rawlings, D. E. (2002). Molecular relationships between two groups of the genus *Leptospirillum* and the finding that *Leptospirillum ferriphilum* sp. nov. dominates South African commercial bio-oxidation tanks that operate at 40°C. *Applied and Environmental Microbiology*, **68** (2), 838-845.

Crellin, P. K. and Rood, J. I. (1997). The resolvase/invertase domain of the site-specific recombinase TnpX is functional and recognizes a target sequence that resembles the junction of the circular form of the *Clostridium perfringens* transposon Tn4451. *Journal of Bacteriology*, **179**, 5148-5156.

Crundwell, F. K. (1997). Physical chemistry of bacterial leaching, p. 177-200. *In* Rawlings, D. (Ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.

Daber, R., Stayrook, S., Rosenberg, S. and Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. *Journal of Molecular Biology*, **370**, 609-619.

Dale, J. W. and Park, S. (2004). Plasmids; gene transfer and genomic plasticity: movable genes and phase variation, p. 137-214. *In* *Molecular genetics of bacteria*, 4th Edition. John Wiley and Sons Ltd, West Sussex, England.

De groot, P., Deane, S. and Rawlings, D. (2003). A transposon-located arsenic resistance mechanism from a strain of *Acidithiobacillus caldus* isolated from commercial, arsenopyrite bio-oxidation tanks. *Hydrometallurgy*, **71**, 115-123.

Deane, S. and Rawlings, D. (1996). Cloning and sequencing of the gene for the *Thiobacillus ferrooxidans* ATCC33020 glutamate synthase (GOGAT) small subunit and complementation of an *Escherichia coli* *gltD* mutant. *Gene*, **177**, 261-263.

Del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas. (1998). Replication and control of circular bacterial plasmids. *Microbiol Molecular Biology Review*, **62**, 434-464.

DeNap, J. and Hergenrother, P. (2005). Bacterial death comes full circle: targeting plasmid replication in drug resistant bacteria. *The Royal Society of Chemistry*, **3**, 959-966.

- Dew, D. W., Lawson, E. N. and Broadhurst J. L. (1997).** The BIOX® process for bio-oxidation of gold-bearing ores or concentrates, p. 45-80. *In* Rawlings, D. (Ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.
- Dintruck H. and Knaff D. (2000).** The evolution of glutamate synthase. *Molecular Biology Rep*, **27**, 141-148.
- Dominy, C. N., Coram, N. J. and Rawlings, D. E. (1998).** Sequence analysis of plasmid pTF a 19.8 kb geographically widespread member of the *Thiobacillus ferrooxidans* pTFI91-like plasmid family. *Plasmid*, **40**, 50-57.
- Dopson, M. and Lindström, E. (1999).** Potential role of *Thiobacillus caldus* in Arsenopyrite Bioleaching. *Applied and Environmental Microbiology*, **65**, p. 36-40.
- Dopson, M., Baker-Austin, C., Koppineedi, P. R. and Bond, P. L. (2003).** Growth in sulfidic mineral environments: metal resistance mechanisms in acidiphilic micro-organisms. *Microbiology*, **149**, 1959-1970.
- Dopson, M., Lindström, E. B. and Hallberg, K. B. (2002).** ATP generation during reduced inorganic sulfur compound oxidation by *Acidithiobacillus caldus* is exclusively due to electron transport phosphorylation. *Extremophiles*, **6**, 123-129.
- Doran, K., Helinsky, D. and Konieczny, F. (1999).** Host-dependent requirement for specific DnaA boxes from plasmid RK2 replication. *Molecular Microbiology*, **33**, 490-498.
- Dorrington, R.A. and Rawlings, D.E. (1990).** Characterization of the minimum replicon of the broad-host range plasmid pTF-FC2 and similarity between pTC-FC2 and the IncQ plasmids. *Journal of Bacteriology*, **172**, 5697-5705.
- Drolet, M. and Lau, P. (1992).** Mobilization protein-DNA binding and divergent transcription at the transfer origin of the *Thiobacillus ferrooxidans* pTF1 plasmid. *Molecular Microbiology*, **6**, 1061-1071.
- Drolet, M., Zanga, P. and Lau, P. C. K. (1990).** The mobilization and origin of transfer regions of a *Thiobacillus ferrooxidans* plasmid: Relatedness to plasmids RSF1010 and pSC101. *Molecular Microbiology*, **4**, 1381-1391.

English, R. S., Lorbach, S. C., Huffman, K. M. and Shively, J. M. (1995). Isolation and characterization of the replicon of a *Thiobacillus intermedius* plasmid. *Plasmid*, **33**, 1-6.

Epp, F., Pechère, J-C., Kok, M. (2001). Raising antibodies against OprD, an outer membrane protein of *Pseudomonas aeruginosa* using translational fusions to MalE. *Journal of Microbiological Methods*, **46**, 1-8.

Espinosa, M., Cohen, S., Couturier, M., Del Solar, G., Díaz-Orejas, R., Giraldo, R., Jánniere, L., Miller, C., Osborn, M. and Thomas, C. M. (2000). Plasmid replication and copy number control, p. 1- 48. *In* The horizontal gene pool, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.

Fernandez-Tresguerres, M., Garcia de Viedma, D., Giraldo, R. and Diaz-Orejas, R. (1995). Host growth temperature and a conservation amino acid substitution in the replication protein of pPS10 influence plasmid range. *Journal of Bacteriology*, **177**, 4377-4384.

Franke, A. and Clewell, D. (1981). Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of conjugal transfer in the absence of a conjugative plasmid. *Journal of Bacteriology*, **145**, 494-502.

Friedrich, C., Rother, D., Bardischewsky, F., Quentmeier A. and Fisher, J. (2001). Oxidation of reduced Inorganic sulfur compounds by bacteria: Emergence of a common mechanism? Minireview. *Applied and Environmental Microbiology*, **67**, 2873 -2882.

Frunzke, J., Engels, V., Hasenbein, S., Gatgens, C. and Bott, M. (2008). Co-ordinated regulation of the gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators GntR1 and GntR2. *Molecular Microbiology*, **67 (2)**, 305-322.

Garcia de Viedma, D., Serrano-Lopez, A, Diaz-Orejan, R. (1995). Specific binding of the replication protein of plasmid pPS10 to direct and inverted repeats is mediated by an HTH-motif. *Nucleic Acid Research*, **23**, 5048-5054.

Gardner, M. (2003) Dissertation. An investigation into the replicon of broad host range mobilizable plasmid from the moderately thermophilic bacterium *Acidithiobacillus caldus*. University of Stellenbosch.

Gardner, M., Deane, S. and Rawlings, D. (2001). Isolation of a new broad-host-range IncQ-like plasmid, pTC-F14, from the acidophilic bacterium *Acidithiobacillus caldus* and analysis of the plasmid replicon. *Journal of Bacteriology*, **183**, 3303-3309.

Garrett, R. and Grisham, C. (1995). Nitrogen acquisition and amino acid metabolism, p.826-870. *In* Biochemistry, University of Virginia, Saunders College Publishing and Harcourt Brace College Publishers.

Geary, L. and Meister, A. (1977). On the mechanism of glutamine-dependent reductive amination of α -ketoglutarate catalyzed by glutamate synthase. *Journal of Biological Chemistry*, **252**, 16662-16669.

Gerdes, K., Ayora, S., Canosa, I., Ceglowski, P., Díaz-Orejas, R., Franch, T., Gulyaev, A. P., Jensen, R., Bugge, Kobayashi, I., Macpherson, C., Summers, D., Thomas, C. M. and Zielenkiewicz, U. (2000). Plasmid maintenance systems, p. 49-86. *In* The horizontal gene pool, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.

Giraldo, R., J. M. Andreu and R. Díaz-Orejas. (1998). Protein domains and conformational changes in the activation of RepA, a DNA replication initiator. *EMBOJ*, **17**, 4511-4526.

Goldschmidt, G. (2005) Dissertation. Cloning, sequencing and partial characterization of the accessory gene region of plasmid pTC-F14 isolated from the biomining bacterium *Acidithiobacillus caldus* "F". University of Stellenbosch.

Gorelik, M., Lunin, V., Skarina, T. and Savchenko, A. (2006). Structural characterization of the GntR/HutC family signaling domain. *Protein science* **15**, 1506-1511.

Green, A., Munro, A., Cheesman, M., Reid, G., Von Wachenfeldt, C., Chapman, S. (2003). Expression, purification and characterization of *Bacillus subtilis* ferredoxin: a potential electron transfer donor to cytochrome P450 BioI. *Journal of Inorganic Biochemistry*, **93**, 92-99.

Haines, A., Jones, K., Cheung, M. and Thomas C. (2005). The IncP-6 plasmid Rms149 consists of a small mobilizable backbone with multiple large insertions. *Journal of Bacteriology*, **187 (14)**, 4728-4738.

- Hallberg, K. and Johnson, D. (2003).** Novel acidophiles isolated from moderately acidic mine drainage waters. *Hydrometallurgy*, **71**, 139-148.
- Hallberg, K. and Lindström, E. (1994).** Characterization of *Thiobacillus caldus* sp. nov., moderately thermophilic acidophile. *Microbiology*, **140**, 3451-3456.
- Hallberg, K., Dopson, M. and Linström E. (1996).** Reduced sulfur compound oxidation by *Thiobacillus caldus*. *Journal of Bacteriology*, **178**, 6-11.
- Haydon, D.J. and Guest, J.R. (1991).** A new family of bacterial regulatory proteins. *FEMS Microbiology Letters*, **63 (2-3)**, 291-295
- Hemmiliä, I. and Mäntsälä, P. (1978).** Purification and properties of glutamate synthase and glutamate dehydrogenase from *Bacillus magaterium*. *Biochemistry Journal*, **173**, 45-52.
- Hillerich, B. and Westpheling (2006).** A new GntR Family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. *Journal of Bacteriology*, **188 (21)**, 7477-7487.
- Holmes, D., Lobus, J., Bopp, L. and Welch, G. (1984).** Cloning of *Thiobacillus ferrooxidans* plasmid in *Escherchia coli*. *Journal of Bacteriology*, **157**, 324-326.
- Holmes, D.S., Zhao, H-L., Levican, G., Ratouchniak, J., Bonnefoy, V., Varela, P. and Jedlicki, E. (2001).** ISAFE1, an ISL3 family insertion sequence from *Acidithiobacillus ferrooxidans* ATCC19859. *Journal of Bacteriology*, **183, (14)**, 4323-4329.
- Huffman, J. and Brennan, R. (2002).** Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structured Biology*, **12**, p. 98-106.
- Jencova, V., Strnad, H., Chodora, Z., Ulbrich, P., Vlcek, C., Hickey, W. and Paces, V. (2008).** Nucleotide sequence and characterization of the (halo) aromatic acid catabolic plasmid pA81 from *Achromobacter xylosoxidans* A8. *Research in Microbiology*, **159**, 118-127.
- Jin, S., Yan, W. and Wang, Z. (1992).** Transfer of IncP plasmids to extremely Acidophilic *Thiobacillus thiooxidans*. *Applied and Environmental Microbiology*, **58**, No. 1, 429- 430.

Johnson, D. (1998). Biodiversity and ecology of acidophilic microorganisms. *FEMS Microbiology Ecology*, **27**, 307 -317.

Johnson, D. (2006). Biohydrometallurgy and the environment: Intimate and important interplay. *Hydrometallurgy*, **83**, 153-166.

Johnson, D. B. and Roberto, F. F. (1997). Heterotrophic acidophiles and their roles in bioleaching of sulfide minerals, p. 259-280. *In* Rawlings, D. (Ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.

Kameya, M., Ikeda, T., Nakamura, M., Arai, H., Ishii, M. and Igarashi, Y. A novel Ferredoxin-dependent glutamate synthase from the hydrogen-oxidizing chemoautotrophic bacterium *Hydrogenobacter thermophilus* TK-6. *Journal of Bacteriology*, **189** (7), 2805-2812.

Kelly, D. P. and Wood, A. P., (2000). Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. *International Journal of Systematic and Evolutionary Microbiology*, **50**, 511-516.

Klenk, H-P., Clayton R., Tomb, J-F., Whaite, O., Nelson, K., Ketchum, K., Dodson, R., Gwinn, M., Hickey, E., Peterson, J., Richardson, D., Kerlavage, A., Graham, D., Krypides, N., Fleischmann, R., Quaxkenbush, J., Lee, N., Sutton, G., Gill, S., Kirkness, E., Dougherty, B., McKenny, K., Adams, M., Loftus, B., Peterson, S., Reich, C., McNeil, L., Badger, J., Glodek, A., Glodek, A., Zhou, L., Overbeek, R., Gocayane, J., Weidman, J., McDonald, L., Utterback, T., Cotton, M., Spriggs, T., Artiach, P., Kaine, B., Sykes, S., Sadow, P., D'Andrea, K., Bowman, C., Fujii, C., Garland, S., Mason, T., Olsen, G., Fraser, C., Smith, H., Woese, C. and Venter, J. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364-370.

Kornberg A. and T. Baker (1992). DNA replication, 2nd Edition. Freeman W.H. and Co. (Eds), New York. NY: Library of congress Cataloging-in-publication data.

Kotze, A., Tuffin, I., Deane, S. and Rawlings, D. (2006). Cloning and characterization of the chromosomal arsenic resistance genes from *Acidithiobacillus caldus* and enhanced arsenic resistance on conjugal transfer of ars genes located on transposon TnAtcArs. *Microbiology*, **152**, 3551-3560.

Krahulec, J., Kretova, M. and Grones, J. (2003). Characterization of plasmids purified from *Acetobacter pasteurianus* 2374. *Biochem Biophys Res Commun*, **310**, 94 - 97.

Krishna, S., Sadreyev, R. and Grishin, N (2006). A tale of two ferredoxins: sequence similarity and structural differences. *BMC Structural Biology*, **6:8** [online.] <http://www.biomedcentral.com/1472.6807/6/8>.

Lau, P., Drolet, M., Zanga, P. and Holmes, D. (1993). Plasmid technology for biomining Applications, p. 635-644. In A. E. Torma, M. L. Ampel and Brierley, C. (Eds.), *Biohydrometallurgical technologies*, vol. 2. TMS Press, Warrendale, Pa.

Lilley, A., Young, P. and Bailey, M. (2000). Bacterial population genetics: do plasmids maintain bacterial diversity and adaptation, p. 287-300. In *The horizontal gene pool*, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.

Madonna, M., Fuchs, R. and Brenchley, J. (1985). Fine structure analysis of *Salmonella typhimurium* glutamate synthase gene. *Journal of Bacteriology*, **161**, 353-360.

Maestro, B., Sanz, J., Diaz-Orejas, R. And Fernandez-Tresguerres, M. (2003). Modulation of pPS10 host-range by plasmid-encoded RepA initiator protein. *Journal of Bacteriology*, **185**, 1367-1375.

Martin, P., Dugan, P. and Touvinen, O. (1981). Plasmid DNA in acidophilic, chemolithotrophic thiobacilli. *Canadian Journal of Microbiology*, **27**, 850-853.

Martin, P., Dugan, P. and Touvinen, O. (1983). Uranium resistance of *Thiobacillus ferrooxidans*. *European Journal of Applied Microbial Biotechnology*, **18**, 392-395.

McGlynn, P. and Lloyd, R. (2002). Recombinational repair and restart of damaged replicative forks. *Nature Reviews Molecular Cell Biology*, **3**, 859-870.

Merlin, C., Mahillon, J., Nešvera, J. and Toussaint A. (2000). Gene recruiters and transporters: the modular structure of bacterial mobile elements, p. 363-402. In *The horizontal gene pool*, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.

Michal, G. (Edited 1999). Biochemical pathways. An atlas of biochemistry and molecular biology. Published by John Wiley & Sons, Inc., p. 46-47.

Mukhopadhyay, S. and Chattoraj, D. (2000). Replication-induced transcription of an autorepressed gene: the replication initiator gene of plasmid P1. *PNAS*, **97**, 7142-7147.

Muth, G., Furr, M., Hartmann, V. and Nohlleben, W. (1995). *Streptomyces ghanarsis* plasmid pSG5: nucleotide sequence analysis of the self-transmissible minimal replicon and characterization of the replication mode. *Plasmid*, **33**, 113-126.

Naas, T., Blot, M., Fitch, W. and Arber, W. (1994). Insertion sequence-related genetic variations in resting *Escherichia coli* K-12. *Genetics*, **136**, 721 – 730.

Naas, T., Blot, M., Fitch, W. and Arber, W. (1995). Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. *Molecular Biological Evolution*, **12**, 198 – 207.

Nagtani, H., Shimizu, M. and Valentine, R. (1971). The mechanism of ammonia assimilation in nitrogen fixing bacteria. *Archives of Microbiology*, **79**, 164-175.

Norris, P. R. (1997). Thermophiles and bioleaching, p. 247-258. In Rawlings, D. (Ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.

Norris, P. R. and Johnson, D. B. (1998). Acidophilic microorganisms, p. 133-249. In Horikoshi, K and Grant, W.D (Edited), *Extremophiles: Microbial life in Extreme environments*. Wiley series in ecological and Applied microbiology.

Oliver, G., Gosset, G., Sanchez-Pescador, R., Lazoya, E., Ku, L., Florez, N., Becerril, B. and Valle, F. (1987). Determination of the nucleotide sequence of the glutamate synthase structural genes for *Escherichia coli* K-12. *Gene*, **60**, 1-11.

Olsen, G.J., Brierley, J.A., and Brierley, C. L. (2003). Bioleaching review part B: progress in bioleaching: Applications of microbial process by the minerals industries. *Applied Microbial Biotechnology*, **63**, 249-257.

Oppen, J., Sarnovsky, R., Craig, N. and Rawlings, D. (1998). A Tn-7 like transposon is present in the *glmUS* region of the obligately chemolithotrophic bacterium *Thiobacillus ferrooxidans*. *Journal of Bacteriology*, **180** (11), p. 3007 – 3012.

Osborn, M., Bron, S., Firth, N., Holsappel, S., Huddleston, A., Kiewiet, R., Meijer, W., Seegers, J., Skurray, R., Terpstra, P., Thomas, C. M. Tietze, E. and Turner S. L. (2000). The evolution of bacterial plasmids, p. 301-362. In C.M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood Academic Publishers, Amsterdam.

Palenda, R., Vanoni, M., Perego, M., Piubelli, L., Galizzi, A., Curti B. and Zanetti, G. (1993). Glutamate synthase genes of the Diazotroph *Azospirillum brasilense*. Cloning, sequencing and analysis of functional domains. *Journal of Biological Chemistry*, **268** (5), 3099-3106.

Parini, C., Guglielmetti, S., Mora, D. and Ricci, G. (2004). Complete sequence and structural organization of pFL5 and pFL7, two cryptic plasmids from *Bacillus licheniformis*. *Plasmid*, **51** (3), 192 - 202.

Park, K. Mukhopadhyay, S. and Chatteraj, K. (1998). Requirements for regulation of origin opening of plasmid P1. *Journal of Biological Chemistry*, **273**, 24906-24911.

Park, J.H., Aravind, L. Wolff, E., Kaevel J., Kim, Y. and Park, M. (2005). Molecular cloning, expression and structural prediction of deoxyhypusine hydrolase: A HEAT repeat containing metalloenzyme. *PNAS*, **103**, 51-56.

Peekhouse, N. and Conway, T. (1998). Positive and negative transcriptional regulation of the *E. coli* gluconate regulon gene *gntT* by GntR and the cyclic AMP (c-AMP) –cAMP receptor protein complex. *Journal of Bacteriology*, **180**, 1777-1785.

R6K γ ori / KAN2 Insertion Kit. EZ:TNm Transposon tools. Cat. No. EZ1011RK. www.epicentre.com

Ravasio, S., Dossena, L., Martin-Figueroa, E., Florencio, F.J., Mattevi, A. and Morandi, P. (2002). Properties of the recombinant ferredoxin-dependent glutamate synthase of *Synechocystis* sp. PCC 6803. Comparison with the *Azospirillum brasilense* NADPH-dependent enzyme and its isolated α -subunit. *Biochemistry*, **41**, 8120-8133.

- Rawlings, D. (1995).** Restriction enzyme analysis of 16S rRNA genes for the rapid identification of *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans*, *Leptosprillum ferrooxidans* strains in leaching environments, p. 1-17. In C.A Jarez, T. Vargas, H. Toledo and J.V. Wiertz (ed), Biohydrometallurgical processing, vol II. University of Chile Press, Santiago, Chile.
- Rawlings, D. (1997).** Mesophilic, autotrophic bioleaching bacteria: description, physiology and role, p. 229-246. In Rawlings, D. (Ed.), Biomining: theory, microbes and industrial processes. Springer-Verlag Berlin Heidelberg New York Biotechnology Intelligence unit.
- Rawlings, D. (2001).** The molecular genetics of *Thiobacillus ferrooxidans* and other mesophilic, acidophilic, chemolithotrophic, iron-or sulfur-oxidizing bacteria. *Hydrometallurgy*, **59**, 187 -201.
- Rawlings, D. (2002).** Heavy metal mining using microbes. *Annual Rev. Microbiology*, **56**, 65-91.
- Rawlings, D. (2005)^a.** Characteristics and adaptability of iron-and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microbial Cell Factories*, **4**: **13**. [Online] <http://www.microbialcellfactories.com/content/4/1/13>.
- Rawlings, D. (2005)^b.** The evolution of pTF-FC2 and pTC-F14, two related plasmids of the IncQ-family. *Plasmid*, **53**, 137-147.
- Rawlings, D., David, D. and Du Plessis, C. (2003).** Biomineralization of metal-containing ores and concentrates. *Trends in Biotechnology*, **21** (1), 38- 44.
- Rawlings, D., Pretorius I-M and Woods, D.R. (1984).** Expression of *Thiobacillus ferrooxidans* origin of replication in *Escherichia coli*. *Journal of Bacteriology*, **158**, 737 -738.
- Rawlings, D., Gawith, C., Peterson, A. and Woods, D. (1983).** Characterization of plasmids and potential genetic markers in *Thiobacillus ferrooxidans*, p. 555-570. In G. Rossi, and A. E. Torma (Eds.). Recent progress in Biohydrometallurgy. Associazione Mineraria Sarda, Iglesias, Italy.
- Rawlings, D. and Kusano, T. (1994).** Molecular genetics of *Thiobacillus ferrooxidans*. *Microbiological reviews*, **58**, 39-55.
- Rawlings, D.E., Coram, N., Gardner, M. and Deane, S. (1999).** *Thiobacillus caldus* and *Leptospirillum ferrooxidans* are widely distributed in continous flow bio-oxidation tanks used to treat a variety of metal containing ores and concentrates, p. 777-786. In R. Amils and A. Ballester (ed),

Biohydrometallurgy and the environment: toward mining in the 21st century. Elsevier, Amsterdam, The Netherlands.

Reizer, J., Bergsteidt, U., Galinier, A., Küster, E., Saier, M.H., Jr. and Hillen, W. et al. (1996). Catabolite repression resistance of gnt operon expression in *Bacillus subtilis* conferred by mutation of His-15, the site of phosphoenolpyruvate-dependent phosphorylation of the phosphocarrier protein HPr. *Journal of Bacteriology*, **178**, 5480-5486.

Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. (2002). Subdivision of the Helix-turn-helix GntR family of Bacterial regulators in the FadR, HutC, MocR and YtrA subfamilies. *The journal of biological chemistry*, **277** (15), 12507-12515.

Roberto, F. (2003). Nucleotide sequence of a small cryptic plasmid from *Acidithiobacillus ferrooxidans* strain A-6. *Hydrometallurgy*, **71**, 133-137.

Rohrer, J. and Rawlings, D.E. (1992). Sequence analysis and characterization of the mobilization region of a broad-host-range plasmid pTC-FC2, isolated from *Thiobacillus ferrooxidans*. *Journal of Bacteriology*, **174**, 6230-6237.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: a Laboratory manual, 2nd edition. Cold Spring Harbour, NY: Cold Spring Harbor Laboratory.

Sanchez, H., Hevia, E., Caseres, B. and Venegas, A. (1986). Studies on native strains of *Thiobacillus ferrooxidans*. *Biotechnologies in Applied Biochemistry*, **8**, 300-308.

Saum, S., Sydow, J., Palm, P., Pfeiffer, F., Oesterhelt, D. and Muller, V. (2006). Biochemical and molecular characterization of the biosynthesis of glutamine and glutamate, two major compatible solutes in the moderately halophilic bacterium *Halobacillus halophilus*. *Journal of bacteriology*, **188** (19), 6808-6815.

Scherzinger, E., V. Haring, R. Lurz and S. Otto. (1991). Plasmid RSF1010 DNA replication *in vitro* promoted by purified RSF1010 RepA, RepB and RepC proteins. *Nucl. Acids Res.*, **19**, 1203-1211.

Schirawski, J. and Uden, G. (1998). Menaquinone-dependent succinate dehydrogenase of bacteria catalyzes reversed electron transport driven by the proton potential. *European Journal of Biochemistry*, **257**, 210-215.

Schlüter, A., Szczepanowski, R., Kurz, N., Schneiker, S., Krahn, I., and Pühler, A., (2007). Erythromycin resistance-conferring plasmid pRSB105, isolated from a sewage treatment plant, harbors a new macrolide resistance determinant, an integron-containing *Tn402*-like element, and a large region of unknown function. *Applied and Environmental Microbiology*, **73**, 1952-1960.

Schlüter, A., Heuer, H., Szczepanowski, R., Poler, S., Schneiker., Pühler, A. and Top, E. (2005). Plasmid pB8 is closely related to the prototype IncP-1 β plasmid R751 but transfers poorly to *E. coli* and carries a new transposon encoding a small multidrug resistance efflux protein. *Plasmid*, **54**, 135-148.

Schneider, D. and Lenski, R (2004). Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology*, **155**, 319- 327.

Schneider, E. and Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems : functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiology Reviews*, **22**, 1-20.

Scippers, A. and Wolfgang, S. (1999). Bacterial leaching of metal sulfides proceeds by two indirect mechanisms via thiosulfate or via polysulfides and sulfur. *Applied and Environmental Microbiology*, **65**, 319-32.

Semenza, M., Viera, M., Curutchet, G. and Donati, E (2002). The role of *At. caldus* in the bioleaching of metal sulfides. *Latin American Applied Research*, **32**, 303 –306.

Sezonov, G., Possoz, C., Friedemann, A., Pernodet, J.L. and Guarineure, M. (2000). KorSA from *Streptomyces* integrative element pSAM2 is a central transcriptional repressor: target genes and binding sites. *Journal of Bacteriology*, **182**, 1243-1250

Shiratori, T., Inoue, C., Numata, M. and Kusano, T. (1991). Characterizing and cloning of plasmids from the iron oxidizing bacterium *Thiobacillus ferrooxidans*. *Current Microbiology*, **23**, 321-326.

Smalla, K., Osborn, M and Wellington, E. M. H. (2000). Isolation and characterization of plasmids from bacteria, p. 207-248. *In* The horizontal gene pool, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.

Snyders, L. and Champness, W. (Edited 2003). Plasmids, p. 137-164; Gene transfer, 165-188 and Genomic plasticity: movable genes and phase variation, p. 189-205. *In* Molecular genetics of bacteria,

(2nd Ed). Department of Microbiology and molecular genetics, Michigan State University, East Lansing, Michigan.

Sun, S. Wang, H. and Oster, G. (2004). Asymmetry in F₁-ATPase and its application for the rotational cycle. *Biophysical Journal*, **86**, 1373-1384.

Suzuki, A. and Knaff, D. (2005). Glutamate synthase: structural, mechanistic and regulatory properties and role in the amino acid metabolism. *Photosynthesis Research*, **83**, 191-217.

Szczepenowski, R., Krahn, I., Linke, B., Goesmann, A., Pühler, A., and Schlüter A. (2004). Antibiotic multiresistance plasmid pRSB101 isolated from a waste water treatment plant is related to plasmids residing in phytopathogenic bacteria and carries eight different resistance determinants including a multidrug transport system. *Microbiology*, **150**, 3613-3630.

Tamber, S., Ochs, M. and Hancock, R. (2006). Role of the novel OprD Family of porins in nutrient uptake in *Pseudomonas auruginisa*. *Journal of Bacteriology*, **188** (1), 45-54.

Tanous, C., Chambellon, E. and Yvon, M. (2007). Sequence analysis of the mobilizable lactococcal plasmid pGdh442 encoding glutamate dehydrogenase activity. *Microbiology*, **153**, 1664-1675.

Tayson, G. ChAmpman, J., Hugenholtz, P., Allen, E., Ram, R., Richardson, P., Solovyev, V., Rubin, E., Rokhsan D. and Banfield, J. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428** (6978), 37-43.

Tayson, G., Lo I., Baker B., Allen E., Hugenholtz, P. and Bunfield, J. (2005). Genome-directed isolations of the key nitrogen fixer *Leptospirillum ferrodiazotrphum* sp. nov. from an acidophilic microbial community. *Applied and Environmental Microbiology*, **71** (10), 6319-6324.

Thomas, C. H. (2004). Evolution and population genetics of bacterial plasmids, p. 509-526. In *Plasmid Biology*, edited by Funnel, B. E. and Philips, G. J. ASM Press, Washington, DC.

Tong, S., Proco A., Isturiz, T. and Conway, T. (1996). Cloning and molecular generic characterization of the *Esherichia coli* gntR, gntK and gntU genes of GntI, the main system for gluconate metabolism. *Journal of Bacteriology*, **178**, 3260-3269.

- Top, E. M., Moëgne-Loccoz, Y., Pembroke, T. and Thomas C. M. (2000).** Phenotypic traits conferred by plasmids, p. 249-286. *In* The horizontal gene pool, edited by Thomas, C.M. School of Biological Sciences, University of Birmingham, UK. Harwood Academic Publishers.
- Tuffin, I., De groot,, P., Deane, S. and Rawlings. D. (2005).** An unusual Tn 21-like transposons containing an ara operon is present in highly arsenic-resistant strains of the biomining bacterium *Acidithiobacillus caldus*. *Microbiology*, **151**, 3027- 3039.
- Valenti, P., Polidoro, M., Buonfiglio, V., Visca, P. and Orsi, N. (1990).** Plasmid DNA profiles in *Thiobacillus ferrooxidans*. *Journal of General Applied Microbiology*, **36**, 351-355.
- Van den Heuvel, R., Svergun, D., Petroukhov, M., Coda A., Curti, B., Ravasio, S., Vanoni, M. and Mattevi, A. (2003)** The active conformation of glutamate synthase and its binding to ferredoxin. *Journal of Molecular Biology*, **330**, 113-128.
- Van den Heuvel, R., Curti, B., Bañón, M. and Mattevi, A. (2004).** Glutamate synthase: a fascinating pathway from L-glutamine to L-glutamate. *Cellular and Molecular Life Sciences*, **61**, 669-681.
- Van Zyl, L. Deane, S. and Rawlings, D. (2003).** Analysis of the mobilization region of the broad-host range IncQ-like plasmid pTC-F14 and its ability to interact with a related plasmid, pTF-FC2. *Journal of Bacteriology*, **185**, 6104-6111.
- Van Zyl, L., Deane, S., Louw, L. and Rawlings D. (2008)^a.** Presence of a family of plasmids (29 to 65 Kilobases) with a 26-kilobase common region in different strains of the sulfur oxidizing bacterium *Acidithiobacillus caldus*. *Applied and Environmental Microbiology*, **74**, 4300-4308.
- Van Zyl, L., Van Munster, J. and Rawlings, D. (2008)^b.** Construction of *arsB* and *tetH* mutants of the sulfur oxidation bacterium *Acidithiobacillus caldus* by marker exchange. *Applied and Environmental Microbiology*, **74**, 5686-5694.
- Vanoni, M. and Curti, B. (1990).** Glutamate synthase: a complex iron-sulfur flavoprotein. *Cellular and Molecular Life Science*, **55**, 617-638.
- Vindal, V., Suma, K. and Ranjan, A. (2007).** GntR family of regulators in *Mycobacterium smegmatis*: a sequence and structure based characterization. *BMC Genomics*, **8**: 289 [online.] <http://biomedcentral.com/1471-2164/8/289>.

- Wacker, I., Ludwig, H., Reif, I., Blencke, H-M., Detch, C., and Stulke, J. (2003).** The regulatory link between carbon and nitrogen metabolism in *Bacillus subtilis*: regulation of the *gltAB* operon by the catabolite control protein CcpA. *Micorbiology*, **149**, 3001-3009.
- Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996).** Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. Strain PCC6893. Sequence determination of the entire genome and assignment of potential protein coding regions. *DNA Res*, **3**, 109-136.
- Wieckert, M. and Adhya, S. (1992).** A Family of bacterial regulators homologous to Gal and Lac repressors. *The Journal of Biological Chemistry*, **267 (22)**, 15869-15874.
- Yan, D. (2007).** Protection of the glutamate pool concentration in enteric bacteria. *PNAS*, **104 (22)**, 9475-9480.
- Yates, J., Cunningham, R. and Holmes, D. (1988).** IST2: an insertion sequence from *Thiobacillus ferrooxidans*. *Proc Nat Acid Sci USA*, **85**, 7284-7287.
- Zechner, E. L., de la Cruz, F., Eisenbrandt, R., Grahn, A. M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C. M., Wilkens, B. M. and Zatyksa, M. (2000).** Conjugative-DNA transfer processes, p. 87-176. *In* The horizontal gene pool, edited by Thomas, C.M. School of Biological Sciences, University of Bermingham, UK. Harwood Academic Publishers.

APPENDIX 1

1. 10X Stock Basal Salts (500ml):

(NH ₄)SO ₄	15g
KCL	0.5g
K ₂ PO ₄	2.5g
MgSO ₄ ·7H ₂ O	2.5g
Ca(NO ₃) ₂	0.0711g
Na ₂ SO ₄	7.26g

- pH to 2.5 with conc. H₂SO₄ and autoclave

2. 1000x Stock Trace Elements (100ml):

ZnSO ₄ ·7H ₂ O	1g
CuSO ₄ ·5H ₂ O	0.1g
MnSO ₄ ·4H ₂ O	0.1g
CoSO ₄ ·7H ₂ O	0.1g
Cr(SO ₄) ₃ ·15H ₂ O	0.05g
Na ₂ B ₄ O ₇ ·10H ₂ O	0.05g
NaMoO ₄ ·2H ₂ O	0.05g
NaVO ₃ (optional)	0.01g

- Add 53µl conc. H₂SO₄ and autoclave

3. Low pH media for growth of *At. caldus*

- Make up 100ml or 500ml of 1X Basal Salts
- Add 0.5M stock potassium tetrathionate (K₂O₆S₄) stock solution (1.5g/10ml water, filter sterilize and store up to 3 weeks at 4°C)
- Ensure that the pH is still 2.5 (adjust pH with conc. H₂SO₄)
- Add 50µl of the 1000X trace elements to each 100ml's of media

4. High pH media for growth of *At. caldus*

- Make up 100ml or 500ml of 1x Basal Salts
- Add 0.5M stock sodium thiosulphate (Na₂S₂O₃) solution (2g/10ml water, filter sterilize and store at room temperature)
- Ensure pH is 4.7 (adjust pH with 10N NaOH)

5. Luria-Bertani (LB) Broth:

Tryptone
Yeast Extract
Sodium chloride
Distilled water

- Dispense into test tubes and autoclave

APPENDIX 2

Bacterial strains and cloning vectors used in this study.

Acidithiobacillus strains	Description	Source
<i>At. caldus</i> strain #6	Fairview mine, Barberton, South Africa	Rawlings et al., 1999b
<i>At. caldus</i> strain “f”	Nickel pilot plant, Billiton, South Africa	Rawlings et al., 1999
<i>At. caldus</i> strain BC-13 (ATCC 51756)	Birch coppice, Warwickshire, UK	Hallberg and Lindström, 1994
<i>At. caldus</i> strain C-SH12 (DSM 9466)	Continuous bioreactor, Brisbane, Australia	Goebel and Stackebrandt, 1994
<i>At. caldus</i> strain KU	Kingsbury coal spoil, UK	Hallberg and Lindström, 1994
<i>At. caldus</i> strain MNG	Arsenopyrite pilot plant, UCT, South Africa	Stellenbosch University laboratory collection (Murray Gardner)
<i>E. coli</i> EC-100D		
<i>E. coli</i> DH-5 α	Φ 80dlacZ Δ M15, endA1, recA1, gyrA96, thi-1, hsdR17 (rk ⁻ , mk ⁺), relA1, supE44, deoR, Δ (lacZYA-argF) U169	Promega
Plasmids		
<i>pUCbm21</i>	Amp ^r , 2725bp cloning vector derivative of pUC19	Boehringer-Mannheim
<i>pBluescript 11 Sk</i>	Amp ^r , ColE1 replicon, LacZ, cloning vector	Fermentas
<i>pBAD/MYC-HIS A, B, C</i>	4100bp pBR322-derived expression vector with His tags	Invitrogen
<i>pACYC184</i>	Tc ^r , Cm ^r , p15A replicon, cloning vector	Chang and Cohen, 1978
<i>pMC1403</i>	Amp ^r , ColE1 replicon	Casadaban et al., 1983
<i>pGEM®-T vector</i>	Amp ^r , T-tailed PCR product cloning vector	Promega
<i>pGEM® Easy vector</i>	Amp ^r , T-tailed PCR product cloning vector with <i>Eco</i> RI site in multiple cloning site	Promega

APPENDIX 3

A list of probes and primers used in this study.

Probes		
IS	A 300bp fragment digested with <i>Bgl</i> III_ <i>Stu</i> I from subclone pTcM3.9XBXB (Appendix 4)	This study
SUC	A 500bp fragment digested with <i>Bgl</i> II and <i>Sal</i> I from subclone pTcM3.9XBXB (Appendix 4)	This study
REG	A 300bp fragment digested with <i>Xba</i> I and <i>Hind</i> III from subclone pTcM1.7XBXH (Appendix 4)	This study
GOGAT	A 496bp fragment digested with <i>Bgl</i> II and <i>Sal</i> I from subclone pTcM4.7XBXB (Appendix 4)	This study
Primers used for probes	Sequence (5'-3')	
ABCF1	CGCAAAGACAATTAACATAAGC	This study
ABCR1	GGTAGAAATAACTTTCCCTTCT	This study
ABCF2	CCTCACATTAACCCAGCAAT	This study
ABCR2	CATGAGCAAACCTTTTCAGCG	This study
FER1F	CGTTTGTCCAGTTGACTGTTTC	This study
FER1R	CTTCAGGCAACTCATCTTTT	This study
FER2F	GTAAGAGTTAAAGTTCCAGTCT	This study
FER2R	TAATCAGGTATGGTATTTCTAC	This study
pRSBF1	GGAAATCGGTTTTGCACCATG	This study
pRSBR1	TCGGACTCCGTAAAACCATAC	This study
PhycoF1	CTTCCAGGCGAGCATAGAT	This study
PhycoF2	TCACTAACGGCAATTGTTGCT	This study
IS2F	CGGTGGTTGCTGCATCATAATT	This study
IS2R	GTGGACAATGATCCCAAAAGAGGT	This study
IS1F	TTATAGTCGACTGCCCAACAGAT	This study
IS1R	TAGGCTCTAGATAATCCAATAGTTT	This study
pRSB23F (EcoRI)	TTATAGAATTCAGGAAGGACCGTCACT	This study
pRSB23R (BamHI)	TAATGGATCCACCAATAAATCCCTCT	This study
pRSB24FW (EcoRI)	TATAGAATTCTGGGTATTCGGTACGGA	This study

pRSB24RV(BamH)	ATAATGGATCCGTTGGTATCCACCGC	This study
GOGATFW (EcoRI)	TTATGAATTCTTGGCCGTTGATAAAA	This study
GOGATRV (BamHI)	TATGGATCCTTTGTAGAAAAGTCC	This study
REGFW (EcoRI)	TAATGAATTCGAATCCCGACATCTTT	This study
REGRV (BamHI)	TATAGGATCCACTGTTCTGCTCC	This study
FDFW (EcoRI)	TCTAGAATTCTTGTAAAAGTATTAGCAA	This study
FDRV (BamHI)	TATAGGATCCATATATTTGCAGTT	This study
SUCFW (EcoRI)	TATAGAATTCTTATTTATGTTGCTGGAA	This study
SUCRV (BamHI)	TATAGGATCCATAGGACCTGCTGT	This study
ABC1RV (BamHI)	ATAGGATCCGTACACTATCGGTGA	This study
ABC1FW (EcoRI)	TAATGAATTCACGGTGTGATTGCGTA	This study
ABC3FW (EcoRI)	ATATGAATTCTTCCGTATGCAATTCC	This study
ABC3RV (BamHI)	AATTGGATCCTTTAACCTCAATATTTCC	This study
BINFW (EcoRI)	TAATGAATTCCAAGATTACCTCGTTC	This study
BINRV (BamHI)	TAATGGATCCAAGGAATGCTTTGAG	This study
SIGFW (EcoRI)	TAATGAATTCCTCATGATATCGATGAAG	This study
SIGRV (BamHI)	AATTGGATCCGCTTTTCAACGCATT	This study
PhycoFW (EcoRI)	TAATGAATTCCTGAGGTGAAGTCTTATG	This study
PhycoRV (BamHI)	ATATGGATCCCCCAATTGCTTTTCGC	This study
REGFW2 (BamHI)	ATAGGATCCGAATCCCGACATCTTT	This study
REGRV2 (Sall)	TTAAGTCGACTATCAGCAAAACAGC	This study
RTSUCFW	AAAAGGTTACCAGAGTTTTAAC	This study
RTSUCRV	CGCTCATTGTGTGTAATATAG	This study
RTGOGATFW	AAAAGGTTCTGCAGCGTTAC	This study
RTGOGATRV	TTCCATTCGCCGGAAGA	This study

Holmes *et al.*, 2001

APPENDIX 4

Plasmid constructs used in this study:

Plasmid Constructs	Description	Source
pTcM7KSXB	Amp ^r , a 1.2kb <i>Ksp/XbaI</i> fragment from pTcM1 and cloned into pBluescript SK/KS	This study
pTcM7XBXH	Amp ^r , a 7.7kb <i>XbaI/XhoI</i> fragment from pTcM1 and cloned into pBluescript SK/KS	This study
pTcM7ECVXB	Amp ^r , a 0.48kb <i>EcoRV/XbaI</i> fragment from pTcM1 and cloned into pUCBM21	This study
pTcM7ECVKP	Amp ^r , a 1.0kb <i>EcoRV/KpnI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM7PS	Amp ^r , a 1.1kb <i>PstI/PstI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM7ECIKP	Amp ^r , a 2.1kb <i>EcoRI/KpnI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM7SALKP	Amp ^r , a 1.5kb <i>SalI/KpnI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM7SALXB	Amp ^r , a 1.5kb <i>SalI/XbaI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM7KSKP	Amp ^r , a 1.8kb <i>KspI/KpnI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM7ECIECV	Amp ^r , a 2.1kb <i>EcoRI/EcoRV</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM7ECVI	Amp ^r , a 1.2kb <i>EcoRV/EcoRV</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM7ECVII	Amp ^r , a 2.3kb <i>EcoRV/EcoRV</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM1.7XBXH	Amp ^r , a 1.7kb <i>XbaI-XhoI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM1.7KS	Amp ^r , a 0.46kb <i>KspI/KspI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM1.7PSXH	Amp ^r , a 1.7kb <i>PstI-XhoI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM1.7NOTXB	Amp ^r , a 1.3kb <i>NotI/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM1.7NOTXH	Amp ^r , 0.40kb <i>NotI-XhoI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM4.7HIII XB	Amp ^r , a 0.38bp <i>HindIII/XbaI</i> fragment from	This study

	pTcm1 cloned into pBluescript SK/KS	
pTcM4.7XBXB	Amp ^r , a 4.7bp <i>XbaI/XbaI</i> fragment from pTcM1 into pUCBM21	This study
pTcM4.7HIII	Amp ^r , a 0.60kb <i>HindIII/HindIII</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM4.7HIIIXB	Amp ^r , a 3.7kb <i>HindIII/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM4.7KSKSI	Amp ^r , a 0.85kb <i>KspI/KspI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM4.7KSKSII	Amp ^r , a 1.4kb <i>KspI/KspI</i> fragment from pTcM1 into pBluescript SK/KS	This study
pTcM4.7PSI	Amp ^r , a 1.0kb <i>PstI/PstI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM4.7KSSAC	Amp ^r , a 0.87kb <i>KspI/SacI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM4.7PSBAM	Amp ^r , a 0.76kb <i>PstI/BamHI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9NCXB	Amp ^r , a 0.87kb <i>NcoI/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9SAL	Amp ^r , a 1.3kb <i>SalI/SalI</i> fragment from pTcM1 cloned into pBluescript SK/KS	This study
pTcM3.9PSXB	Amp ^r , a 1.1kb <i>PstI/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9XBXB	Amp ^r , a 3.9kb <i>XbaI/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9NCSAL	Amp ^r , a 0.79kb <i>NcoI/SalI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9PSSAL	Amp ^r , a 0.55kb <i>PstI/SalI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9SPXB	Amp ^r , a 1.0kb <i>SphI/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9ECVXB	Amp ^r , a 0.54kb <i>EcoRV/XbaI</i> fragment from pTcM1 cloned into pUCBM21	This study
pTcM3.9ECVSAL	Amp ^r , a 1.1kb <i>EcoRV/SalI</i> fragment from pTcM1 cloned into pUCBM21	This study
pACYCREG	Cm ^r , transcriptional regulator (ORF 2) cloned into expression vector pACYC184	This study
pGOGATPMF	Amp ^r , promoter fusion construct of Approximately 300bp upstream of ORF 8	This study

	cloned into promoterless vector pMC1403	
pSUCPMF	Amp ^r , promoter fusion construct of Approximately 300bp upstream of ORF 3 cloned into promoterless vector pMC1403	This study
pFERPMF	Amp ^r , promoter fusion construct of Approximately 300bp upstream of ORF 7 cloned into promoterless vector pMC1403	This study

APPENDIX 5

The nucleotide sequence of the 18.15kb region analyzed during this study. Amino acid residues are shown for each open reading frame. The frames are indicated with +/-1,2,3. Ribosomal binding sites (RBS) are marked with arrows. The transcription initiation and termination sites are shown in boldface and underlined. Highly conserved amino acid residues are highlighted. Direct repeat (DR) of Insertion sequences are also highlighted.

```

1   CTCGAGGACA CGGACCAGCA CTCGGCTGGG ACGACGACCG ATGCGCCCCC
   GAGCTCCTGT GCCTGGTCGT GAGCCGACCC TGCTGCTGGC TACGCGGGGG

51  AGCTGGTCGT GAGCTCCCAC TCTCCCCAGA GGTTCTGCTC CAGAGCCAGG
   TCGACCAGCA CTCGAGGGTG AGAGGGGTCT CCAAGACGAG GTCTCGGTCC

101 CGGTAGAAGC GGAATATCCC CTCCTCCGGC AGTACCTTCT CCCACTCCCG
   GCCATCTTCG CCTTATAGGG GAGGAGGCCG TCATGGAAGA GGGTGAGGGC

151 TACCCACCAG CGCTCCAGCT CGTAACGGGG TGGGGACGCC ACGGGATTCA
   ATGGGTGGTC GCGAGGTCGA GCATTGCCCC ACCCCTGCGG TGCCCTAAGT

201 CTCCCCAGGA GTCCGGGACA AGGCAACCAT CAGGAGTCCA CTGCCCTTGC
   GAGGGGTCTT CAGGCCCTGT TCCGTTGGTA GTCCTCAGGT GACGGGAACG
-3          *   S   D   V   A   R   A

251 GGCCTGCAAG GCCGCATAGA GGGAGGTCTT GCTCACCTTG ATCCTGGCCG
   CCGGACGTTC CGGCGTATCT CCCTCCAGAA CGAGTGGAAC TAGGACCGGC
-3   A   Q   L   A   A   Y   L   S   T   K   S   V   K   I   R   A

301 CGGCTTCCCG CACGTTCAGT CCGTTGGCGA GGTACTCCCG CGCCCTTTGC
   GCCGAAGGGC GTGCAAGTCA GGCAACCGCT CCATGAGGGC GCGGGAAACG
-3   A   E   R   V   N   L   G   N   A   L   Y   E   R   A   R   Q

351 AGCTTCTCGG GCGTCATCAC GGGCTTGCGT CCCCCTTTCC GCCCCGCGC
   TCGAAGAGCC CGCAGTAGTG CCCGAACGCA GGGGGAAAGG CGGGGGCGCG
-3   L   K   E   P   T   M   V   P   K   R   G   G   K   R   G   R   A

401 GGCCGCCGCC GTCAGGCCAG CCCTGGTACG CTCCCGGATC AGATCCCGCT
   CCGGCGGCGG CAGTCCGGTC GGGACCATGC GAGGGCCTAG TCTAGGGCGA
-3   A   A   A   T   L   G   A   R   T   R   E   R   I   L   D   R

451 CAAACTGACC CAGTGCCCCA AAGACGTGGA AGATGAGTCG ACCCCCAGGG
   GTTTGACTGG GTCACGGGGT TTCTGCACCT TCTACTCAGC TGGGGGTCCC
-3   F   Q   G   L   A   G   F   V   H   F   I   L   R   G   G   P

501 GTGGTGGTAT CGATGTGTTC GGTGAGGGAC CGGAAACCAA CACCCTGGGC
   CACCACCATA GCTACACAAG CCACTCCCTG GCCTTTGGTT GTGGGACCCG

```

-3 T T T D I H E T L S R F G V G Q A

551 TTCCAGTGTG GCAATGGTCT CGATCAGATG CGGTAACGAA CGTCCAAGCC
AAGGTCACAC CGTTACCAGA GCTAGTCTAC GCCATTGCTT GCAGGTTCCG
-3 E L T A I T E I L H P L S **R** G L

601 GATCCAGTCG CCAGACGACC AGCACGTCGC CATCGCGCAG ATAGGCCAGT
CTAGGTCAGC GGTCTGCTGG TCGTGCAGCG GTAGCGCGTC TATCCGGTCA
-3 D L R W V V L V D G D R L Y A L

651 GCCGCAGCGA GACCCGGGCG GTCCGATTTG GCGCCAGAGG CCGTATCCTC
CGGCGTCGCT CTGGGCCCGC CAGGCTAAAC CGCGGTCTCC GGCATAGGAG
-3 A A A L G P R D S K A G S A T D E

701 GAAAATCCGC GTGCAGCCTG CTTTCTTCAG GCGTCCGGCC TGCAAGGCGG
CTTTTAGGCG CACGTCGGAC GAAAGAAGTC CCGCAGCCGG ACGTTCCGCC
-3 F I R T C G A K K L A D A Q L A

751 TATCCTGGTC CGCGGTTCGAT ACCCGCGCGT AGCCGATCAG CGCCACGGCA
ATAGGACCAG GCGCCAGCTA TGGGCGCGCA TCGGCTAGTC GCG**GTG**CCGT
-3 D Q D A T **S** V **R** A Y G I L A V

ORF 1: start site

801 ACTCCGGTTG TCCGTCATAT CGTCCGACTA TCTTATCTTC CGGGAACTCA
T**GAGC**CCAAC AGGCAGTATA GCAGGCTGAT AGAATAGAAG GCCCTTGAGT
←
RBS

851 AGTTACAACA TATTTTTCGGA CAGTTGTCCG GGTTGACCGA CAAACGGCCG
TCAATGTTGT ATAAAAGCCT GTCAACAGGC CCAACTGGCT GTTTGCCGGC

901 TTAGATGGAC AACGCTAATG CACCGAACAG CATGTAAACG GATAATGGCG
AATCTACCTG TTGCGATTAC GTGGCTTGTC GTACATTTGC CTATTACCGC

951 ACGCGCAAGC GATACCGACG CTTGTCTATC TCTGGCCCAA CTCCCTGCCA
TGCGCGTTTC CTATGGCTGC GAACAGATAG AGACCGGGTT GAGGGACGGT

1001 GCAAATTGGG GACAATGGGC AGACGGAGGT TGGAGTCGGT CACCTTGAAG
CGTTTAAACC CTGTTACCCG TCTGCCTCCA ACCTCAGCCA GTGGAACCTC

1051 CGCCGCCGCT GTCGCCAGAG AATCCCGACA TCTTTCCCAA GGTGGATGAG
GCGGCGGCGA CAGCGGTCTC TTAGGGCTGT AGAAAGGGTT CCACCTACTC

1101 GCGATCATTG CCTGAAACGA CGAACAGGCG GAGTGCACTA TTTTAGATGG
CGTAGTAAC GGACTTTGCT GCTTGTCCGC CTCACGTGAT AAAATCTACC

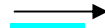
1151 TTATAGCACT ACTTTTGTGC TCTGTGCACT ATACTTGACAC ACTTGTGCGC
AATATCGTGA TGAAAACACG AGACACGTGA TATGAACGTG TGAACACGCG

1201 ATCGTCGATG TCTTAGCTTA CTGTTTTTTTA TCGATTTGCT AAAAAACCAA

TAGCAGCTAC AGAATCGAAT GACAAAAAAT AGCTAAACGA TTTTTTGGTT

1251 GCTCTGGCAC ACTTGTTGCT ACATGTATGC ACGTGTTTCA ATGGAGCAGA
CGAGACCGTG TGAACAACGA TGTACATACG TGCACAAAGT TACCTCGTCT

RBS



ORF 2: start site

PhnF



+2 V H S T T S K R A A I L A M Q
1301 ACA**GTG**CACA GTACAACGTC AAAACGCGCA GCTATTTTAG CTATGCAGCC
TGTCACGTGT CATGTTGCAG TTTTGCGCGT CGATAAAATC GATACGTCCG

+2 L Y L Q I K D D I R A N I V E G
1351 TTTATATCTA CAAATAAAGG ACGACATCAG AGCTAATATA GTCGAAGGAA
AAATATAGAT GTTTATTTCC TGCTGTAGTC TCGATTATAT CAGCTTCCTT

+2 T Y K A D D K L P S E H E M V L M
1401 CTTATAAAGC TGATGACAAG CTTCCATCAG AGCATGAGAT GGTACTCATG
GAATATTTTC ACTACTGTTC GAAGGTAGTC TCGTACTCTA CCATGAGTAC

+2 Y N A S R T T I R Q S I K D L Q
1451 TACAATGCAA GCAGAACAAC AATAAGGCAA TCTATAAAAG ACCTACAAAA
ATGTTACGTT CGTCTTGTTG TTATTCCGTT AGATATTTTC TGGATGTTTT

+2 E G L V Y T I P G K G T F I S R
1501 TGAAGGCCTA GTTTATACCA TTCCTGGCAA AGGAACTTTT ATTTCTCGTC
ACTTCCGGAT CAAATATGGT AAGGACCGTT TCCTTGAAAA TAAAGAGCAG

+2 P K V S Q T L L S L Q G F G E A M
1551 CTAAGTATC ACAGACTTTA CTATCACTGC AGGGCTTTGG CGAGGCAATG
GATTTTCATAG TGTCTGAAAT GATAGTGACG TCCCAGAAAC GCTCCGTTAC

+2 S P N G H E T Y S T I I S Q R Y
1601 TCTCCTAACG GGCACGAGAC CTATTCAACA ATAATTAGCC AGAGATACGG
AGAGGATTGC CCGTGCTCTG GATAAGTTGT TATTAATCGG TCTCTATGCC

+2 I A D K L V Q N K L R L T Q S Q
1651 TATCGCGGAT AAGTTAGTTC AAAATAAACT CAGATTAACA CAATCACAGA
ATAGCGCCTA TTCAATCAAG TTTTATTTGA GTCTAATTGT GTTAGTGTCT

+2 I F E L Q R I R Y L D R E P I S L
1701 TTTTTGAACT TCAACGGATT AGGTATCTAG ACCGAGAGCC CATATCTTTA
AAAACTTGA AGTTGCCTAA TCCATAGATC TGGCTCTCGG GTATAGAAAT

+2 D V T Y L P S D V G S K L E N E
1751 GACGTGACGT ATTTACCATC AGATGTCGGC TCCAAATTAG AAAACGAGGA
CTGCACTGCA TAAATGGTAG TCTACAGCCG AGGTTTAATC TTTTGCTCCT

D-b
domain
HTH
Motif

E-b/o
Domain

+2 **L T G K D L F M L L E S R L G I**
 1801 TTTGACTGGA AAAGACTTAT TTATGTTGCT GGAATCGCGG CTCGGCATAG
 AAAGTACCT TTTCTGAATA AATAACAACGA CCTTAGCGCC GAGCCGTATC

+2 **A L G N A E V Q L E A I L A G D N**
 1851 CTTTGGGGAA TGCAGAAGTA CAACTTGAAG CAATTTTAGC AGGGGATAAT
 GAAACCCCTT ACGTCTTCAT GTTGAACTTC GTTAAAATCG TCCCCTATTA

+2 **L A H L L N I E Q G S P I L R M E**
 1901 CTGGCACATT TATTGAACAT CGAACAAAGC TCTCCAATAC TTCGTATGGA
 GACCGTGTAATAAATACTTGTA GCTTGTTCCG AGAGGTTATG AAGCATACT

+2 **R L T Y T S L G V P L D F E Y L**
 1951 ACGTTTAAACA TATACATCCT TAGGCGTCCC ATTAGACTTT GAGTATTTAT
 TGCAAATTGT ATATGTAGGA ATCCGCAGGG TAATCTGAAA CTCATAAATA

PhnF end

+2 **Y Y R G D Y L K Y S L L L S R H G**
 2001 ACTACAGAGG GGACTACCTG AAGTATTCAC TGCTACTCTC TAGGCATGGT
 TGATGTCTCC CCTGATGGAC TTCATAAGTG ACGATGAGAG ATCCGTACCA

+2 G *

UTRA
Domain

ORF 3: start site

+1 M D K N F S E T T V D **I I V**
 2051 GGCTAACAAAT **GGACAAGAAT** TTTTCGGAAA CGACTGTAGA TATAATCGTT
 CCGATTGTTA CCTGTTCTTA AAAAGCCTTT GCTGACATCT ATATTAGCAA

+1 **V G G G T A G P M A A I A A K E**
 2101 GTTGGCGGCG GGACAGCAGG TCCTATGGCT GCCATTGCCG CAAAGGAAGC
 CAACCGCCGC CCTGTCTGTC AGGATACCGA CGGTAACGGC GTTTCCTTCG

+1 **N R D L A V L L I E K A N V K R**
 2151 CAACAGAGAT CTTGCTGTTT TGCTGATAGA AAAAGCTAAC GTAAAACGTA
 GTTGTCTCTA GAACGACAAA ACGACTATCT TTTTCGATTG CATTTTGCAT

+1 **S G A I S M G M D G L N N A V I P**
 2201 GTGGTGCTAT CTCTATGGGA ATGGACGGCC TTAATAATGC AGTGATACCA
 CACCACGATA GAGATACCCT TACCTGCCCG AATTATTACG TCACTATGGT

+1 **G H A S P E Q Y V K E I T I A N**
 2251 GGGCACGCCT CTCCAGAACA ATATGTCAAA GAGATTACCA TTGCTAACGA
 CCCGTGCGGA GAGGTCTTGT TATACAGTTT CTCTAATGGT AACGATTGCT

+1 **G I V D Q S A I L T Y A S K S F**
 2301 CGGTATAGTC GATCAATCAG CTATTCTGAC ATATGCTTCA AAAAGTTTTG
 GCCATATCAG CTAGTTAGTC GATAAGACTG TATACGAAGT TTTTCAAAC

+1 **D M I T R L D R M G V K F E K D E**

SdhA
Domain

2351 ATATGATTAC TCGCTTAGAT AGAATGGGAG TAAAGTTTGA AAAAGATGAA
 TATACTAATG AGCGAATCTA TCTTACCCTC ATTTCAAAC TTTTCTACTT
 +1 T G D F N V K K V H H L G N Y V
 2401 ACAGGTGATT TTAACGTAAA GAAAGTTCAC CACCTTGGGA ATTATGTTCT
 TGTCCACTAA AATTGCATTT CTTTCAAGTG GTGGAACCCT TAATACAAGA
 +1 P M P E G H D M K T I L Y R H L
 2451 TCCAATGCCA GAAGGGCATG ATATGAAAAC GATACTTTAT AGGCATCTTA
 AGGTTACGGT CTTCCCGTAC TATACTTTTG CTATGAAATA TCCGTAGAAT
 +1 K K H R V E I T N R K K V T R V L
 2501 AAAAGCATCG AGTTGAAATA ACAAACAGAA AAAAGGTTAC CAGAGTTTTA
 TTTTCGTAGC TCAACTTTAT TGTGTGTCTT TTTTCCAATG GTCTCAAAT
 +1 T S N D R T V S G V I A V D T R
 2551 ACATCTAACG ACCGTACCGT CAGTGGTGTA ATTGCGGTTG ATACTCGTAC
 TGTAGATTGC TGGCATGGCA GTCACCACAT TAACGCCAAC TATGAGCATG
 +1 S E I T L I H A K A V I L C T G
 2601 GTCTGAAATC AACTGATCC ATGCGAAAGC AGTAATCTTA TGCACAGGAG
 CAGACTTTAG TGTGACTAGG TACGCTTTTCG TCATTAGAAT ACGTGTCTC
 +1 A A G R L G L P A S G Y L F G T Y
 2651 CTGCCGGTTCG ACTTGGTCTT CCAGCATCAG GTTATCTTTT CGGAACATAC
 GACGGCCAGC TGAACCAGAA GGTCGTAGTC CAATAGAAAA GCCTTGTATG
 +1 E N P T N A G D G H A M A L H A
 2701 GAGAATCCTA CTAATGCTGG AGACGGGCAC GCGATGGCGT TGCATGCCGG
 CTCTTAGGAT GATTACGACC TCTGCCCGTG CGCTACCGCA ACGTACGGCC
 +1 A K L A N L E C F Q I N P L I K
 2751 GGCAAACTA GCGAATTTAG AATGTTTTCA GATAAACCCA TTAATTAAG
 CCGTTTTGAT CGCTTAAATC TTACAAAAGT CTATTTGGGT AATTAATTTT
 +1 D Y N G P A C A Y V T G P F G G Y
 2801 ACTATAATGG CCCAGCATGC GCTTACGTGA CCGGTCCATT TGGAGGATAT
 TGATATTACC GGGTCGTACG CGAATGCACT GGCCAGGTAA ACCTCCTATA
 +1 T A N S R G E R F I E C D Y W S
 2851 ACAGCTAATT CTCGTGGGGA GAGATTTATT GAGTGTGATT ACTGGAGCGG
 TGTCGATTAA GAGCACCCCT CTCTAAATAA CTCACACTAA TGACCTCGCC
 +1 Q M M M E F F K E L Q G G N G P
 2901 TCAAATGATG ATGGAATTTT TCAAGGAAC TCAAGGTGGA AACGGACCAG
 AGTTTACTAC TACCTTAAAA AGTTCCTTGA GGTTCACCT TTGCCTGGTC
 +1 V Y L K L N H L A E E T I Q T I E

SdhA
Domain

2951 TTTACTTGAA ACTAAATCAT TTGGCTGAAG AGACGATACA AACGATAGAA
 AAATGAACTT TGATTTAGTA AACCGACTTC TCTGCTATGT TTGCTATCTT
 +1 T I L H T N E R P S R G R F H S
 3001 ACTATATTAC ACACAAATGA GCGCCCGAGT CGGGGACGTT TTCACTCCGG
 TGATATAATG TGTGTTTACT CGCGGGCTCA GCCCCTGCAA AAGTGAGGCC
 +1 R G T D Y R Q Q M I E M H I S E
 3051 TCGTGGTACA GACTACCGCC AACAGATGAT AGAAATGCAC ATCTCAGAAA
 AGCACCATGT CTGATGGCGG TTGTCTACTA TCTTTACGTG TAGAGTCTTT
 +1 I G F C S G H S A S G V Y I N S N
 3101 TTGGATTTTG CAGTGGACAC AGTGCATCTG GTGTATATAT AAACAGTAAC
 AACCTAAAAC GTCACCTGTG TCACGTAGAC CACATATATA TTTGTCATTG
 +1 A E T S V K G L Y A A G D V G S
 3151 GCTGAAACGT CCGTAAAAGG GCTTTACGCA GCCGGCGATG TCGGAAGTGT
 CGACTTTGCA GGCATTTTCC CGAAATGCGT CGGCCGCTAC AGCCTTCACA
 +1 P H N Y M L G A F V Y G E I A G
 3201 ACCGCATAAT TATATGCTTG GGGCATTGT ATATGGAGAA ATAGCAGGGA
 TGGCGTATTA ATATACGAAC CCCGTAAACA TATACCTCTT TATCGTCCCT
 +1 R N A A I Y C A N T A H R S P D Q
 3251 GGAATGCTGC TATATATTGT GCAAATACAG CCCATAGAAG TCCTGACCAA
 CCTTACGACG ATATATAACA CGTTTATGTC GGGTATCTTC AGGACTGGTT
 +1 E K V R A E I N R I A R L T N R
 3301 GAGAAAGTGC GCGCTGAGAT TAATAGGATT GCCAGGCTCA CCAATCGGAA
 CTCTTTTACG CGCGACTCTA ATTATCCTAA CGGTCCGAGT GGTTAGCCTT
 +1 D G I T P Y Q M E Y K I R R I V
 3351 CGATGGTATT ACACCGTATC AAATGGAGTA TAAGATTAGA AGGATTGTTA
 GCTACCATAA TGTGGCATAG TTTACCTCAT ATTCTAATCT TCCTAACAAT
 +1 N D Y L Q P P K I G T K L E I A E
 3401 ATGATTATCT TCAACCTCCA AAAATTGGCA CAAAACCTGA GATTGCCGAA
 TACTAATAGA AGTTGGAGGT TTTTAACCGT GTTTTGAACT CTAACGGCTT
 +1 H R L A E I R D D L K Y L Y A D
 3451 CATCGTTTAG CAGAGATTAG GGATGACTTG AAATATCTTT ATGCTGATGA
 GTAGCAAATC GTCTCTAATC CCTACTGAAC TTTATAGAAA TACGACTACT
 +1 S H N L L R A L E V C D I L D C
 3501 TTCCACAAT TTAGTGC GCGC CCCTAGAAGT GTGTGATATT TTAGACTGCG
 AAGGGTGTTA AATGACGCGC GGGATCTTCA CACACTATAA AATCTGACGC
 +1 A D M A A S A S I F R T E S R W G
 3551 CAGATATGGC TGCTTCCGCT TCAATATTCA GAACGGAAAG CCGATGGGGA

SdhA Domain

GTCTATACCG ACGAAGGCGA AGTTATAAGT CTTGCCTTTC GGCTACCCCT
 +1 L Y H Y R V D Y P E R N D D S W
 3601 CTTTACCATT ACCGTGTAGA TTATCCTGAA CGAAATGATG ATTCCTGGTT
 GAAATGGTAA TGGCACATCT AATAGGACTT GCTTTACTAC TAAGGACCAA
 +1 C H V H L S K D D N G K L K C I
 3651 TTGTCATGTC CACCTGAGTA AAGATGACAA CGGAAAATA AAATGTATAA
 AACAGTACAG GTGGACTCAT TTCTACTGTT GCCTTTTGAT TTTACATATT
 +1 K V P V E P Y I C P P D E T E V K
 3701 AAGTGCCAGT AGAACCATAT ATATGTCCTC CAGACGAAAC TGAGGTGAAG
 TTCACGGTCA TCTTGGTATA TATACAGGAG GTCTGCTTTG ACTCCACTTC
 +1 S Y E R L R I A A K A S V C *
 3751 TCTTATGAAC GGCTACGAAT AGCAGCGAAA GCGTCTGTAT GTTAGTATTA
 AGAATACTTG CCGATGCTTA TCGTCGCTTT CGCAGACATA CAATCATAAT

SdhA Domain

+1 RBS → M A T S T V R V K V P V
 3801 TCGAGAAATA TAAATGGCTA CCTCAACAGT AAGAGTTAAA GTTCCAGTCT
 AGCCTCTTAT ATTTACCGAT GGAGTTGTCA TTCTCAATTT CAAGGTCAGA
 +1 S V D A S R C I A S K G C T V C V
 3851 CAGTTGATGC TAGCAGGTGC ATAGCATCTA AGGGTTGCAC GGTGTGCGTT
 GTCAACTACG ATCGTCCACG TATCGTAGAT TCCCAACGTG CCACACGCAA
 +1 D V C P L D V L V I D K V K G V
 3901 GACGTATGCC CACTTGATGT ATTAGTTATC GACAAAGTTA AAGGCGTTGC
 CTGCATACGG GTGAACTACA TAATCAATAG CTGTTTCAAT TTCCGCAACG
 +1 L M Q Y D E C W Y C T P C E V D
 3951 ACTAATGCAG TACGATGAAT GTTGGTACTG TACGCCTTGT GAAGTCGACT
 TGATTACGTC ATGCTACTTA CAACCATGAC ATGCGGAACA CTTCAGCTGA
 +1 C P T D A V S V E I P Y L I K *
 4001 GCCAACAGA TGCAGTGTCA GTAGAAATAC CATACTGAT TAAATAACTG
 CGGGTTGTCT ACGTCACAGT CATCTTTATG GTATGGACTA ATTTATTGAC
 4051 TATATTTGAA AAGGATACTG AAGATGATTT CAAGGACTAC AGACAGTACC
 ATATAAACTT TTCCTATGAC TTCTACTAAA GTTCCTGATG TCTGTCATGG
 4101 TCTGACCCTG ACGCCCGCAG GCTTCTTATT ATAGATGCGC TGGACATGGA
 ARACTGGGAC TGCGGGCGTC CGAAGAATAA TATCTACGCG ACCTGTACCT
 4151 TGATATGGAT GCAATTGATA TGGTTCCTCG TCATTTCAAG TGGGTAGCCT
 ACTATACCTA CGTTAACTAT ACCAAGGAGC AGTAAAGTTC ACCCATCGGA
 -1 DR * T P L R

4201 GAGATCCAGC TTGCCAGGA TCAGGTAGGC CATGTTGATA AAGTTCTTGT
CTCTAGGTCG AACGGGTCCT AGTCCATCCG GTACAACATAT TTCAAGAACA
-1 L D L K G L I L Y A M N I F N K

4251 GCGTGCGGTA ACCCCGAGCC TTGGCCTTGG CGGATTGAAT GAGGCTGTTG
CGCACGCCAT TGGGGCTCGG AACCGGAACC GCCTAACTTA CTCCGACAAC
-1 T R Y G R A K A K A S Q I L S N

4301 AAACCTTCCA GAATTCCATT GGTGATCTGG CTCTCGAACC ATCGGAGCAC
TTTGGAAGGT CTTAAGGTAA CCACTAGACC GAGAGCTTGG TAGCCTCGTG
-1 F G E L I G N T I Q S E F W R L V

4351 GCCATCCCAG TGATTCATGA TGGTGTAGGC GACCCTGACG ATAGGCGGCA
CGGTAGGGTC ACTAAGTACT ACCACATCCG CTGGGACTGC TATCCGCCGT
-1 G D W H N M I T Y A V R V I P P

4401 GATCGCTGGT TCTGGCGTTT TCCAACCAGG CTTTCAAGAG GGTAGCCCCC
CTAGCGACCA AGACCGCAAA AGGTTGGTCC GAAAGTTCTC CCATCGGGGG
-1 D S T R A N E L W A K L L T A G

4451 TGGTGGCGAT TCTTGATCGT GAAGATGTCC TGAAAGGTCA GGCGGAAGT
ACCACCGCTA AGAACTAGCA CTTCTACAGG ACTTTCCAGT CCGCCTTGAC
-1 Q H R N K I T F I D Q F T L R F Q

4501 GTAGGCCTGC GCCGTCTTGA GGTCTGGTC TTTGAGCAAT TCCTGCAGCT
CATCCGGACG CGGCAGAACT CCAAGACCAG AAACTCGTTA AGGACGTCGA
-1 Y A Q A T K L N Q D K L L E Q L

4551 TTTCTTTCTG CTTCAATTTG AGGTTGCAAT CGTTCTTGAG CCAAAGCCAG
AAAGAAAGAC GAAGTAAAC TCCAACGTTA GCAAGAAGT GGTTCGGTC
-1 E K Q K M K L N C D N K L W L W

4601 CGGGTCTTTT TGAGATTTGG CTGGGTGAGG ACTTCCCCCT TGCGCACGTC
GCCAGAAAA ACTCTAAACC GACCCACTCC TGAAGGGGGA ACGCGTGCAG
-1 R T K K L N P Q T L V E G K R V D

4651 GTCTACGGCC TCGTTGACGA GCTTCATGAG GTGAAAACGA TCGAAAGTGA
CAGATGCCCG AGCAACTGCT CGAAGTACTC CACTTTTGCT AGCTTTCACT
-1 D V A E N V L K M L H F R D F T

4701 TCTCCGCATT GGGCAGGTGC TCGGCAGCCC CTTTCTGGAA GGCCGGCGAG
AGAGGCGTAA CCCGTCCACG AGCCGTCTGGG GAAAGACCTT CCGGCCGCTC
-1 E A N P L H E A A G K Q F A P S

4751 AGGTCCATGC TCACATCGGT GATCGCTTCC GCGTACCAC CATGGGCCTG
TCCAGGTACG AGTGTAGCCA CTAGCGAAGG CGCGATGGTG GTACCCGGAC
-1 L D M S V D T I A E A S G G H A Q

Trans-
posase

4801 TAGATCTGCG GAGAATTTCT CAAAGGTCTT GGCATCCTTG CCGGGAGTAG
 ATCTAGACGC CTCTTAAAGA GTTTCAGAA CCGTAGGAAC GGCCCTCATC
 -1 L D A S F K E F T K A D K G P T

4851 CGAACAAAGAG TCGCCGGGCA TTCAGATCCA CGAAGAGCGT GATGTAGTCA
 GCTTGTTCTC AGCGGCCCGT AAGTCTAGGT GCTTCTCGCA CTACATCAGT
 -1 F L L R R A N L D V F L T I Y D

4901 TGTCCATGCC GACTGCTGGT TTCATCGACG CCGACGGCAT GGACATTGGC
 ACAGGTACGG CTGACGACCA AAGTAGCTGC GGCTGCCGTA CCTGTAACCG
 -1 H G H R S S T E D V G V A H V N A

4951 CATATCCACC GCAGCACGAG CTTCCGGCAC ATAATGGTCA ATCACTCGCC
 GTATAGGTGG CGTCGTGCTC GAAGCCCGTG TATTACCAGT TAGTGAGCGG
 -1 M D V A A R A E P V Y H D I V R

5001 ACAGGAGCTT GTCGGTCTCA CCGACCATGC GAGCTGCCGT CAATACCGGC
 TGTCCCTCGAA CAGCCAGAGT GGCTGGTACG CTCGACGGCA GTTATGGCCG
 -1 L L K D T E G V M R A A T L V P

5051 ATCTCCCGCA CCAGGGTCAT GATCAGCGCT TCAAAGAGCA GGGTGAAACG
 TAGAGGGCGT GGTCCCAGTA CTAGTCGCGA AGTTTCTCGT CCCACTTTGC
 -1 M E R V L T M I L A E F L L T F R

5101 CGAGCCTTCC CGCGCCCAGG GAACAGATAT CTGATGCACT CCATGCTCCT
 GCTCGGAAGG GCGCGGGTCC CTTGTCTATA GACTACGTGA GGTACGAGGA
 -1 S G E R A W P V S I Q H V G H E

5151 GGCACCTCAC ACGAGGTACA CGGGCATGGA GATAGGCTTC ATGCTGGAAG
 CCGTGAAGTG TGCTCCATGT GCCCGTACCT CTATCCGAAG TACGACCTTC
 -1 C K V R P V R A H L Y A E H Q F

5201 AAATCCATGT GCCGCCAGGT ATGGTCACGG GTATCATGTA CCGGACACTC
 TTTAGGTACA CGGCGGTCCA TACCAGTGCC CATAGTACAT GGCCTGTGAG
 -1 F D M H R W T H D R T D H V P C E

5251 CTCACCACAG ACGGAGCAAG CAAAGCGACT ACCTTTGGGA AAGTTGATGT
 GAGTGGTGTC TGCCTCGTTC GTTTCGCTGA TGGAAACCCT TTCAACTACA
 -1 E G C V S C A F R S G K P F N I

5301 GCAGATCCAG GCGTTTCTCC TCCACCGTGA AAGTCACATG ATCCACCAAC
 CGTCTAGGTC CGCAAAGAGG AGGTGGCACT TTCAGTGTAC TAGGTGGTTG
 -1 L D L R K E E V T F T V H D V L

5351 CACGGCGGAA CCAATCCTAA CGCGAGAGAA AACAGCTCTT CAGGGACCAT
 GTGCCGCCTT GGTTAGGATT GCGCTCTCTT TTGTCGAGAA GTCCCTG**GTA**
 -1 W P P V L G L A L S F L E E P V M

} Trans-
posase

ORF 5: start site

5401 CAGCTAACTC CTATTCAGGG CACGACCGCT CCACAGCATT CTACCCCCTA
GTCGATTGAG GATAAGTCCC GTGCTGGCGA GGTGTCGTAA GATGGGGGAT

←
RBS

5451 CCCACTCAAT CTGACGAAGA GCCATTGATA TAATATTATC TGGTCTCTCT
GGGTGAGTTA GACTGCTTCT CGGTAACTAT ATTATAATAG ACCAGAGAGA

DR

ORF 6: Truncated

+2 P K V R E E A A K A I G D
5501 GACTCCTCTC CGAAAGTCCG TGAAGAAGCA GCGAAAGCAA TTGGGGACTT
CTGAGGAGAG GCTTTCAGGC ACTTCTTCGT CGCTTTCGTT AACCCCTGAA

+2 P G E H M I E P L L L L L D D P
5551 TCCAGGCGAG CATATGATAG AACCATATT ATTGCTTCTT GATGATCCTG
AGTCCGCTC GTATACTATC TTGGTAATAA TAACGAAGAA CTA TAGGAC

+2 D E A V R K A A S N T L A D C N Q
5601 ATGAGGCCGT TAGAAAGGCA GCATCTAATA CACTGGCTGA TTGTAACCAG
TACTCCGGCA ATCTTCCGT CGTAGATTAT GTGACCGACT AACATTGGTC

+2 V K L L D Y L D E Y L K D K S S F
5651 GTAAACTAT TGGATTATCT AGATGAGTAT CTGAAAGATA AAAGTTCATT
CATTTTGATA ACCTAATAGA TCTACTCATA GACTTTCTAT TTTCAAGTAA

+2 V R S S I L R A V K P L K N N R
5701 CGTGCGTTCA TCCATTCTGC GAGCAGTAAA GCCTTTAAAA AACACAGAT
GCACGCAAGT AGGTAAGACG CTCGTCATTT CGGAAATTTT TTGTTGTCTA

+2 S A E Y A L K G L L D S S E A V R
5751 CAGCTGAGTA TGCCTAAAG GGTCTTTTAG ATTCATCAGA AGCAGTTAGA
GTCGACTCAT ACGCGATTTT CCAGAAAATC TAAGTAGTCT TCGTCAATCT

+2 I E S I G I L G Y L K N P D Y I
5801 ATTGAAAGCA TTGGTATTCT TGGATATCTA AAAAATCCTG ATTATATTTT
TAACTTTCGT AACCATAGA ACCTATAGAT TTTT TAGGAC TAATATAAAG

+2 S L A T I A V S D N S P N V R R
5851 GTCCTTAGCA ACAATTGCCG TTAGTGATAA CTCGCCAAAT GTTAGAAGGG
CAGAAATCGT TGTTAACGGC AATCACTATT GAGCGGTTTA CAATCTTCCC

+2 V A M G A L G Y S S S E V A C S A
5901 TTGCCATGGG CGCGCTTGGG TATTCGTCTT CTGAGGTAGC GTGCTCTGCG
AACGGTACCC GCGCGAACCT ATAAGCAGAA GACTCCATCG CACGAGACGC

+2 L V D
5951 CTTGTAGATG GCCTTGAGG TATATCCTGG CAAGTGCGCG CCGAGTCCGC

FOG
HEAT
repeat

GAACATCTAC CGGAACCTCT ATATAGGACC GTTCACGCGC GGCTCAGGCG
6001 CGCAACAATC GGTAAGATTA AAGCTCCGAG CCATGTTTGC AATAAGCTTA
CGCTTGTTAG CCATTCTAAT TTCGAGGCTC GGTACAAACG TTATTCTGAAT
6051 TAGAATTACT CGAAAATGAT GAATATTGGC AGGTAATTGT AAAAGTATTA
ATCTTAATGA GCTTTTACTA CTTATAACCG TCCATTAACA TTTTCATAAT
6101 GCAACTATAG GAGCGATCAG ATGTATATCA TCTGTTAATG TAGTTAGCGC
CGTTGATATC CTCGCTAGTC TACATATAGT AGACAATTAC ATCAATCGCG
6151 ATTTCTCTTG CATCAAGAGC CGAACGTACG CAAAGAAGCG GCTGGCTGCT
TAAAGAGAAC GTAGTTCTCG GCTTGCATGC GTTTCTTCGC CGACCGACGA
6201 TAGGTGAACT TGGTGACCAA TCTGCTTTAG CAGCATTAGA GGCTGCATTA
ATCCACTTGA ACCACTGGTT AGACGAAATC GTCGTAATCT CCGACGTAAT
6251 CATGACAGGG ATCCGGATGT TAGAAAGATC GCAGCATTCT CAATTGCTAA
GTACTGTCCC TAGGCCTACA ATCTTTCTAG CGTCGTAAGC GTTAACGATT

ORF 7: start site

+1
6301 GTTAGATCGT TCTTTC **GGGG** TATAGTTATG ACTTATGTAG TGACTGATAA
CAATCTAGCA AGAAAGCCCC ATATCAATAC TGAATACATC ACTGACTATT
M T Y **V V T D**

+1
6351 **C V N C K Y M D C V D V C P V D**
CTGTGTTAAC TGCAAATATA TGGATTGCGT TGACGTTTGT CCAGTTGACT
GACACAATTG ACGTTTATAT ACCTAACGCA ACTGCAAACA GGTCAACTGA

+1
6401 **C F H E G K N F L V I D P S V C I**
GTTTCCACGA AGGAAAAAAT TTCTTAGTTA TAGATCCAAG TGTTTGCATA
CAAAGGTGCT TCCTTTTTTTA AAGAATCAAT ATCTAGGTTT ACAAACGTAT

+1
6451 **D C G V C E P E C P A S A I Y K**
GATTGCGGTG TTTGCGAGCC AGAGTGTCCG GCGTCAGCTA TATATAAAGA
CTAACGCCAC AAACGCTCGG TCTCACAGGC CGCAGTCGAT ATATATTTCT

+1
6501 **S D L P D E** F V A Y L D I N K K
CTCTGATCTT CCTGATGAAT TTGTTGCATA TCTTGATATT AACAAAAAAC
GAGACTAGAA GGACTACTTA AACAAACGTAT AGAACTATAA TTGTTTTTTT

+1
6551 L S S S W P L I K Y K K D E L P E
TTAGCAGTTC TTGGCCGTTG ATAAAATACA AAAAAGATGA GTTGCCTGAA
AATCGTCAAG AACCGGCAAC TATTTTATGT TTTTCTACT CAACGGACTT

+1
6601 A H K W D G I P N K R S Y I E A
GCACATAAAT GGGATGGAAT ACCAAATAAG AGAAGCTATA TTGAAGCTTA
CGTGTATTTA CCCTACCTTA TGGTTTATTC TCTTCGATAT AACTTCGAAT

COG1146
Ferredoxin
motif

+1 M N C R C R S F C *
 6651 GATGAATTGT CGATGTCGGT CATTCTGTTA GCTAAATATA GCATGATTAT
 CTACTTAACA GCTACAGCCA GTAAGACAAT CGATTTATAT CGTACTAATA

6701 TTAATATATT TTCTTTATCC ATAAATGACC GAAATCAAAT AGACGATTGA
 AATTATATAA AAGAAATAGG TATTTACTGG CTTTAGTTTA TCTGCTAACT

ORF 8: Start site

+2 RBS M K F T D F C
 6751 TCGGGTTAAA TATACGGGGT AGATATTAAT GAAGTTTACT GACTTTTGTGTA
 AGCCCAATTT ATATGCCCCA TCTATAATTA CTTCAAATGA CTGAAAACAT

+2 M N D Y L S T D F S T K I D G L F
 6801 TGAATGATTA TCTATCTACG GACTTTTCTA CAAAGATAGA TGGACTATTT
 ACTTACTAAT AGATAGATGC CTGAAAAGAT GTTTCCTATCT ACCTGATAAA

+2 L L F L E E K K E G L A K K L V
 6851 CTTTTATTTT TTGAAGAAAA AAAAGAAGGG CTTGCCAAAA AACTGGTGAC
 GAAAATAAAG AACTTCTTTT TTTTCTTCCC GAACGGTTTT TTGACCACTG

+2 F R E S N S P I E K E K S S F L
 6901 GTTCCGAGAA AGCAATAGCC CTATAGAAAA GGAGAAAAGC TCCTTTCTTA
 CAAGGCTCTT TCGTTATCGG GATATCTTTT CCTCTTTTTCG AGGAAAGAAT

+2 I S L A I H L E K F I A L F F D I
 6951 TATCTCTTGC TATACATCTT GAGAAGTTTA TTGCTCTATT CTTGATATA
 ATAGAGAACG ATATGTAGAA CTCTTCAAAT AACGAGATAA GAAGCTATAT

+2 S P H V N S L Q E R T L I H D P
 7001 AGCCCTCACG TTAATTCCCT GCAGGAGCGT ACGCTCATCC ATGACCCGGT
 TCGGGAGTGC AATTAAGGGA CGTCCTCGCA TGCGAGTAGG TACTGGGCCA

+2 M A F K K H F V L R R A R R Y R
 7051 CATGGCCTTC AAGAAACACT TCGTGCTGCG CCGTGCGCGG CGCTATCGTG
 GTACCGGAAG TTCTTTGTGA AGCACGACGC GGCACGCGCC GCGATAGCAC

+2 G E F P E S F A E L S A W L D E E
 7101 GCGAGTTTCC CGAGTCTTTT GCCGAGCTTT CCGCCTGGCT TGACGAGGAA
 CGCTCAAAGG GCTCAGAAAA CGGCTCGAAA GCGGACCGA ACTGCTCCTT

+2 I Q Q A H L D P A D P E L A I A
 7151 ATCCAGCAGG CGCACCTGGA TCCCGCCGAT CCGGAGCTCG CCATCGCCCG
 TAGGTCGTCC GCGTGGACCT AGGGCGGCTA GGCTCGAGC GGTAGCGGGC

+2 L G E N W L A D E A R H A E A I
 7201 TCTGGGGGAA AACTGGCTGG CCGACGAGGC CCGGCATGCC GAGGCCATCC
 AGACCCCTT TTGACCGACC GGCTGCTCCG GGCCGTACGG CTCCGGTAGG

+2 L K L T R W C A L A L T D P E G Q
 7251 TCAAGTTGAC CCGCTGGTGT GCCCTCGCCC TTACCGATCC CGAGGGGCAG
 AGTTCAACTG GGCACCACA CGGGAGCGGG AATGGCTAGG GCTCCCCGTC

+2 H R V A D W P S F R L P Q R V D H
 7301 CACAGGGTGG CGGACTGGCC GAGTTTTTCGT TTGCCACAGC GGGTCGACCA
 GTGTCCCACC GCCTGACCGG CTCAAAGCA AACGGTGTCG CCCAGCTGGT

+2 E A L V P L Q R Y T A R P D Q P
 7351 CGAGGCACTG GTCCCCTTGC AGCGCTATAC CGCCCGTCCC GATCAGCCCT
 GCTCCGTGAC CAGGGGAACG TCGCGATATG GCGGGCAGGG CTAGTCGGGA

+2 W M G D P S Q L R R R D G F H L T
 7401 GGATGGGCGA TCCATCCCAG CTGCGCCGTC GCGATGGTTT TCACCTGACG
 CCTACCCGCT AGGTAGGGTC GACGCGGCAG CGCTACCAA AGTGGACTGC

+2 D P R M D A R **G V Q S E V H Y C**
 7451 GACCCGCGCA TGGACGCCCC CGGCGTCCAG AGCGAAGTGC ACTATTGCCT
 CTGGGCGCGT ACCTGCGGGC GCCGCAGGTC TCGCTTCACG TGATAACGGA

+2 **Y C H D H D G D F C S R G F P E**
 7501 GTATTGCCAC GACCACGATG GCGACTTCTG CTCCCGGGGT TTCCCGGAAA
 CATAACGGTG CTGGTGCTAC CGCTGAAGAC GAGGGCCCCA AAGGGCCTTT

+2 **K K G V P E L G F K V D P L D V V**
 7551 AGAAAGGCGT TCCGGAGCTC GGCTTCAAGG TGGATCCGCT GGACGTGGTC
 TCTTTCCGCA AGGCCTCGAG CCGAAGTTCC ACCTAGGCGA CCTGCACCAG

+2 **L T G C P L E E K I S E M Q I L**
 7601 CTCACGGGCT GTCCCCTGGA AGAGAAAATC TCGGAGATGC AGATCCTGCG
 GAGTGCCCCG CAGGGGACCT TCTCTTTT TAG AGCCTCTACG TCTAGGACGC

+2 **R D G F N I A A L A M I M V D N**
 7651 CCGCGACGGT TTCAATATCG CCGCCCTGGC CATGATCATG GTGGACAACC
 GGCCTGCCA AAGTTATAGC GGCGGGACCG GTACTAGTAC CACCTGTTGG

+2 **P M L P A T G H R I C N D C M K A**
 7701 CGATGTTGCC GGCCACGGGA CATCGCATCT GTAACGACTG CATGAAGGCC
 GCTACAACGG CCGGTGCCCT GTAGCGTAGA CATTGCTGAC GTACTTCCGG

+2 **C V Y Q K Q D P V N I P Q V E T**
 7751 TCGTCTACC AAAAGCAGGA TCCGGTGAAC ATTCCCAGG TAGAGACCCG
 ACGCAGATGG TTTTCGTCTT AGGCCACTTG TAAGGGGTCC ATCTCTGGGC

+2 **V L S D V L S L P Y G V E I Y D**
 7801 GGTGCTCAGC GATGTCTTGT CCCTGCCCTA TGGCGTCGAG ATCTACGACC
 CCACGAGTCG CTACAGAACA GGGACGGGAT ACCGCAGCTC TAGATGCTGG

gltD
conserved
domain

```

+2 L L C R W N P L R Q T Q Y L A R D
7851 TCCTCTGTCG CTGGAATCCC CTGCGCCAGA CCCAATATCT GGCACGGGAC
AGGAGACAGC GACCTTAGGG GACGCGGTCT GGGTTATAGA CCGTGCCCTG

+2 Y N G R R V L V V G M G P A G F
7901 TACAACGGTC GGCGGGTGCT GGTGGTGGGC ATGGGTCCGG CAGGCTTTAC
ATGTTGCCAG CCGCCCACGA CCACCACCCG TACCCAGGCC GTCCGAAATG

+2 M A H H L L Q E G C A V V G I D
7951 CATGGCCCAC CACCTCCTGC AGGAAGGCTG CGCCGTGGTG GGCATCGACG
GTACCGGGTG GTGGAGGACG TCCTTCCGAC GCGGCACCAC CCGTAGCTGC

+2 G L K I E P L A E E W L R Q P V R
8001 GACTCAAGAT CGAACCGCTG GCCGAGGAGT GGCTGCGGCA GCCGTTGCGT
CTGAGTTCTA GCTTGCGGAC CGGCTCCTCA CCGACGCCGT CCGCCACGCA

+2 S W A A L Q E D L D E R L L L G
8051 TCCTGGGCGG CCCTACAGGA GGATCTGGAC GAGCGCCTGC TGCTGGGTTT
AGGACCCGCC GGGATGTCCT CCTAGACCTG CTCGCGGACG ACGACCCAAA

+2 G G V A E Y G I T V R W D K N F
8101 CGGTGGCGTT GCGGAATACG GTATCACCGT GCGCTGGGAC AAGAACTTTC
GCCACCGCAA CGCCTTATGC CATAGTGGCA CGCGACCCTG TTCTTGAAAG

+2 L K L I Y L S L A R R P L F Q I F
8151 TCAAGCTCAT CTATCTGAGC CTTGCCCGGC GGCCCTTTT CCAGATCTTC
AGTTCGAGTA GATAGACTCG GAACGGGCCG CCGGGGAAAA GGTCTAGAAG

+2 G G V R L G G T L T L E D A W S
8201 GCGGCGTGC GCCTGGGTGG TACCCTGACC CTGGAGGATG CTTGGTCCAG
CCGCCGCACG CGGACCCACC ATGGGACTGG GACCTCCTAC GAACCAGGTC

+2 G F D H V C L A T G A G L P R V
8251 CGGTTTTGAT CACGTCTGCC TGGCAACGGG TGCCGGTCTG CCCCGGGTGC
GCCAAAGCTA GTGCAGACGG ACCGTTGCC ACGCCAGAC GGGGCCACG

+2 L S L P G S L A R G M R Q A S D F
8301 TGTCCTTGCC CGGGAGTCTC GCCCGCGGCA TGCGTCAGGC CAGCGATTTT
ACAGGAACGG GCCCTCAGAG CGGGCGCCGT ACGCAGTCCG GTCGCTAAAA

+2 L M A L Q L T G A G K Q S S L A
8351 CTCATGGCCC TGCAACTCAC CGGGGCAGGG AAACAGAGCA GTCTTGCCAA
GAGTACCGGG ACGTTGAGTG GCCCCGTCCC TTTGTCTCGT CAGAACGGTT

+2 L Q V R L P A V V I G G G L T A
8401 TCTCCAGGTA CGGTTGCCCG CCGTGGTCAT CCGTGGTGGT CTGACGGCCG
AGAGGTCCAT GCCAACGGGC GGCACCAGTA GCCACCACCA GACTGCCGGC

+2 V D T A T E V Q A Y Y V R Q V E K

```

gltD
conserved
domain

8451	TGGATACGGC AACGGAAGTG CAGGCCTACT ACGTGCGCCA GGTCGAAAAG ACCTATGCCG TTGCCTTCAC GTCCGGATGA TGCACGCGGT CCAGCTTTTC	} gltD conserved domain
+2	V L Q R Y E A M A A A R G E A A	
8501	GTTCTGCAGC GTTACGAGGC CATGGCGGCG GCGCGGGGAG AAGCGGCGGT CAAGACGTCG CAATGCTCCG GTACCGCCGC CGCGCCCCTC TTCGCCGCCA	
+2	R A G L S E E D A T I L D E F L	
8551	ACGGGCTGGA CTCAGCGAGG AAGATGCCAC CATCCTCGAT GAATTTCTGG TGCCCGACCT GAGTCGCTCC TTCTACGGTG GTAGGAGCTA CTTAAAGACC	
+2	A H G R A V R A E R E R A K A A G	
8601	CACACGGTCG GGCCGTGCGT GCCGAGCGCG AGCGCGCCAA GGCTGCCGGG GTGTGCCAGC CCGGCACGCA CGGCTCGCGC TCGCGCGGTT CCGACGGCCC	
+2	E Q P N F I P L L H A W G G V T	
8651	GAGCAGCCCA ACTTCATTCC CCTGCTCCAC GCCTGGGGTG GGGTGACGCT CTCGTCGGGT TGAAGTAAGG GGACGAGGTG CGGACCCAC CCCACTGCGA	
+2	A Y R K G M Q Q S P A Y T R N H	
8701	GGCTTACCGC AAGGGTATGC AACAAATCCCC TGCCTACACC CGCAATCACG CCGAATGGCG TTCCCATACG TTGTTAGGGG ACGGATGTGG GCGTTAGTGC	
+2	E E L I K A L Q E G I F Y Q E G L	
8751	AGGAACTCAT CAAGGCCTTG CAGGAGGGGA TCTTCTACCA AGAGGGCCTG TCCTTGAGTA GTTCCGGAAC GTCCTCCCCT AGAAGATGGT TCTCCCGAC	
+2	D P L R A E L D G Y G H I A A M	
8801	GACCCCCTGC GGGCCGAGCT GGATGGCTAC GGACACATTG CGGCCATGGT CTGGGGGACG CCCGGCTCGA CCTACCGATG CCTGTGTAAC GCCGGTACCA	
+2	F R R M E E V E G R W R P S E V	
8851	CTTCCGGCGA ATGGAAGAGG TGGAGGGGCG CTGGCGTCCG AGTGAAGTGG GAAGGCCGCT TACCTTCTCC ACCTCCCCGC GACCGCAGGC TCACTTCACC	
+2	E L H L P A R A V F V A A G T V P	
8901	AGCTGCACCT GCCCGCCCGC GCCGTTTTCG TCGCCGAGG GACGGTTCCC TCGACGTGGA CGGGCGGGCG CGGCAAAGC AGCGGCGTCC CTGCCAAGGG	
+2	N T I Y E R E Y P G T L T L D G	
8951	AATACCATCT ACGAGCGCGA GTATCCCGGG ACCTTGACGC TGGATGGCGA TTATGGTAGA TGCTCGCGCT CATAGGGCCC TGGAACTGCG ACCTACCGCT	
+2	H F Q T Y V G H P H S L Q A V Q	
9001	TCATTTCCAG ACCTATGTGG GACACCCCA TAGTCTGCAG GCAGTACAGG AGTAAAGGTC TGGATACACC CTGTGGGGGT ATCAGACGTC CGTCATGTCC	
+2	V A A D C K S P D F G P F T S Y R	
9051	TGGCGGCGGA CTGCAAGTCG CCCGATTTTG GCCCCTTTAC CTCCTACCGC	

ACCGCCGCCT GACGTTTCAGC GGGCTAAAAC CGGGGAAATG GAGGATGGCG
 +2 D E R D H R V S F V G D A H P V
 9101 GATGAACGGG ATCATCGCGT GAGCTTCGTG GGCGACGCGC ACCCCGTCTT
 CTA CTTGCC TAGTAGCGCA CTCGAAGCAC CCGCTGCGCG TGGGGCAGAA
 +2 H G S V V K A I A S A K R S Y P
 9151 TCACGGCAGC GTGGTCAAGG CCATCGCCTC GGCCAAGCGG AGCTACCCCG
 AGTGCCGTCG CACCAGTTCC GGTAGCGGAG CCGGTTTCGCC TCGATGGGGC
 +2 E L M E H L R Q L P M Q D I P L E
 9201 AGCTCATGGA GCATCTACGA CAGCTGCCCA TGCAGGACAT TCCCTTGGAA
 TCGAGTACCT CGTAGATGCT GTCGACGGGT ACGTCCTGTA AGGGAACCTT
 +2 S F Q R E V A L R L G M R I R R
 9251 TCCTTTTCAGC GGGAGGTGGC GCTGCGTTTG GGTATGCGCA TCCGCCGCAT
 AGGAAAGTCG CCTCCACCG CGACGCAAAC CCATACGCGT AGGCGGCGTA
 +2 D R S N P A V C E L W V E A P A
 9301 CGACCGCAGC AACCCGGCAG TCTGCGAGCT CTGGGTGGAG GCTCCGGCAG
 GCTGGCGTCG TTGGGCCGTC AGACGCTCGA GACCCACCTC CGAGGCCGTC
 +2 A A R N F R P G Q F F R L Q S Y E
 9351 CCGCCCGAAA CTTCCGCCCT GGACAGTTCT TTCGTCTCCA GAGTTACGAG
 GCGGGGCTTT GAAGGCGGGA CCTGTCAAGA AAGCAGAGGT CTCAATGCTC
 +2 A M S P V V Q G T R L Q I P V L
 9401 GCGATGAGTC CGGTGGTGCA GGGGACACGC CTGCAAATAC CGGTGCTCAC
 CGCTACTCAG GCCACCACGT CCCCTGTGCG GACGTTTATG GCCACGAGTG
 +2 V S G A G V E D D C I R L M V L
 9451 GGTGAGCGGC GCCGGCGTCG AGGACGACTG TATCCGCCTC ATGGTATTGC
 CCACTCGCCG CGGCCGCAGC TCCTGCTGAC ATAGGCGGAG TACCATAACG
 +2 Q W G T A P R L V G R L R E G D P
 9501 AGTGGGGAAC GGCACCGCGT CTGGTGGGCC GCCTGCGGGA AGGGGACCCA
 TCACCCCTTG CCGTGGCGCA GACCACCCGG CGGACGCCCT TCCCCTGGGT
 +2 V V L M G P T G A P T D I P K D
 9551 GTGGTACTCA TGGGTCCCAC GGGCGCACCG ACGGATATCC CCAAGGACAA
 CACCATGAGT ACCCAGGGTG CCCGCGTGGC TGCCATATAGG GGTTCTCTGT
 +2 T I L V V A G R W G A A V M M D
 9601 GACCATTCTG GTGGTCGCCG GACGCTGGGG TGCTGCGGTG ATGATGGACA
 CTGGTAAGAC CACCAGCGGC CTGCGACCCC ACGACGCCAC TACTACCTGT
 +2 I G P A L R A A G N R V L Y V A A
 9651 TTGGTCCAGC GCTGCGGGCG GCGGGTAACC GCGTTCTCTA TGTGGCGGCT
 AACCAAGTCG CGACGCCCGC CGCCATTGG CGCAAGAGAT ACACCGCCGA

+2 L G E A R E L D H Q E L L E A A
 9701 CTGGGGGAAG CGCGGGA ACT GGACCATCAG GAGCTCCTGG AAGCCGCCGC
 GACCCCTTC GCGCCCTTGA CCTGGTAGTC CTCGAGGACC TTCGGCGGCG

+2 D Q I V W C V A K G P Q I T A R
 9751 GGATCAGATC GTCTGGTGTG TCGCCAAGGG TCCGCAGATT ACGGCGCGGC
 CCTAGTCTAG CAGACCACAC AGCGGTTCCC AGGCGTCTAA TGCCGCGCCG

+2 R P Q D I S V T A S D M V Q L L L
 9801 GCCCACAGGA CATCAGCGTC ACTGCCAGCG ACATGGTGCA GCTACTGCTG
 CGGGTGTCTT GTAGTCGCAG TGACGGTCCG TGTACCACGT CGATGACGAC

+2 D Y S A G Q L G K A D G N I A L
 9851 GACTATTCCG CCGGCCAGTT GGGAAAAGCG GACGGCAACA TTGCCCTCAA
 CTGATAAGGC GGCCGGTCAA CCCTTTTTCG CTGCCGTTGT AACGGGAGTT

+2 E V D R L M V M G S T G L L K A
 9901 AGAGGTGGAT CGGCTCATGG TGATGGGTTC CACCGGACTG CTCAAGGCCT
 TCTCCACCTA GCCGAGTACC ACTACCCAAG GTGGCCTGAC GAGTTCCGGA

+2 F Q A A L R G S L E S C F R A D L
 9951 TTCAGGCCGC TCTACGGGGA AGCCTCGAAA GCTGTTTCCG AGCCGATCTC
 AAGTCCGGCG AGATGCCCTT TCGGAGCTTT CGACAAAGGC TCGGCTAGAG

+2 E A V A T V G S P M Q C M L K G
 10001 GAGGCCGTGG CCACCGTTGG CAGTCCCATG CAATGCATGT TGAAGGGAGT
 CTCCGGCACC GGTGGCAACC GTCAGGGTAC GTTACGTACA ACTTCCCTCA

+2 C A Q C L Q W Q V D P Q T G Q R
 10051 CTGCGCCCAG TGTCTGCAGT GGCAGGTGGA TCCGCAGACG GGGCAGCGCA
 GACGCGGGTC ACAGACGTCA CCGTCCACCT AGGCGTCTGC CCCGTGCGGT

+2 T Q A V F S C A G Q D Q P L A W V
 10101 CCCAGGCCGT TTTCAGTTGC GCCGGGCAAG ACCAGCCCCT GGCCTGGGTG
 GGGTCCGGCA AAAGTCAACG CGGCCCGTTC TGGTCGGGGA CCGGACCCAC

+2 D L D H L A A R Q S Q N R L L E
 10151 GATCTGGATC ATCTGGCCGC GCGGCAGAGC CAAAACCGGC TGTTGGAGCG
 CTAGACCTAG TAGACCGGCG CGCCGTCTCG GTTTTGGCCG ACAACCTCGC

+2 L S S Q W L D H V F K Q G G R N
 10201 CCTGTCCAGT CAGTGGTTGG ATCACGTTTT TAAGCAAGGA GGACGAAATG
 GGACAGGTCA GTCACCAACC TAGTGCAAAA ATTCGTTCTT CCTGCTTTAC

+2 V I F *
 10251 TTATTTTTTG ATACCATAAA GAGAGTGTCA AACTTCATAA AACTTCATA
 AATAAAAAAC TATGGTATTT CTCTCACAGT TTGAAGTATT TTTGAAGTAT

10301 CGATAAAGAG AGAAATAGAT GCCAGAGTTG TGGAAATGCCA CTTATGTTTCG
GCTATTTCTC TCTTTATCTA CGGTCTCAAC ACCTTACGGT GAATACAAGC

10351 ATAAAAATCC TGTTAAAAGC GGAACATATT GTAGCTACTG CCACAACGGA
TATTTTTAGG ACAATTTTTCG CCTTGTATAA CATCGATGAC GGTGTTGCCT

10401 ACAGATTTTG TTAATATCTT TATATCTAGA AAAGAGATGA CGGAGAAAAT
TGTCTAAAAC AATTATAGAA ATATAGATCT TTTCTCTACT GCCTCTTTTA

10451 CGATAATATC CTTAAAAAAA GAAATACCCC GTATTATATC AGACTATATA
GCTATTATAG GAATTTTTTTT CTTTATGGGG CATAATATAG TCTGATATAT

10501 TTAAAATCCG GATAGCGACA CTAAAACGAT GGATTTTTTAA AACGTCGCCA
AATTTTAGGC CTATCGCTGT GATTTTGCTA CCTAAAATTT TTGCAGCGGT

10551 AAAAAGTCGT GAACGATTTT ATCTGATCTG TGAATCTGCA GAAAAATCGG
TTTTTCAGCA CTTGCTAAAG TAGACTAGAC ACTTAGACGT CTTTTTAGCC

10601 TCGCGGTGTG GACAGCAGTT ATCTGCCTAG TACCTTCGCC TCTACCGCCG
AGCGCCACAC CTGTCGTCAA TAGACGGATC ATGGAAGCGG AGATGGCGGC

10651 CATGCAGTCG TCTTTGCCGC TCATCCAGTA TCCCTTCGAA CGGTTTGTGT
GTACGTCAGC AGAAACGGCG AGTAGGTCAT AGGGAAGCTT GCCAAACACA

10701 CGAGCTGCTA TCAGTGCGTC TAATTCCATA GCATATGGCC ATGCGGTATT
GCTCGACGAT AGTCACGCAG ATTAAGGTAT CGTATACCGG TACGCCATAA

10751 AAAAGTGAAA GTTATCTAAC GTCTTCTCAG AAAGCGTAGG TCAACATGAC
TTTTCACTTT CAATAGATTG CAGAAGAGT CTTTCGCATCC AGTTGTACTG
-1 * F A Y T L M V

10801 GCGATTATAG ATAAAGTAGC CTGTGGAAAG ATCGCCATGT GCCACCTCTA
CGCTAATATC TATTTTCATCG GACACCTTTC TAGCGGTACA CGGTGGAGAT
-1 R N Y I F Y G T S L D G H A V E

10851 CACGATCACG GATGGAAAGA CCTTTGATCC ATCCTTGAGG GAAGTAGGTC
GTGCTAGTGC CTACCTTTCT GGAAACTAGG TAGGAACTCC CTTTCATCCAG
-1 R D R I S L G K I W G Q P F Y T

10901 AGATCACCGT AGATATCATT GGCCTAGCCA GCAGTATGCA GGTGGTATTG
TCTAGTGGCA TCTATAGTAA CCGCATCGGT CGTCATACGT CCACCATAAC
-1 L D G Y I D N A Y G A T H L H Y Q

10951 GGCATAGGCG GCCATAAACA GAAACTGATT GCTGAGGAAG TCTTGGGTCC
CCGTATCCGC CGGTATTTGT CTTTGACTAA CGACTCCTTC AGAACCCAGG
-1 A Y A A M F L F Q N S L F D Q T

11001 ACTTGACCTT CCAGCCATTT CCCGCCCCA TTTCGACCAA GCCTCGGATC
TGAACTGGAA GGTCGGTAAA GGGCCGGGGT AAAGCTGGTT CGGAGCCTAG

```

-1 K V K W G N G P G M E V L G R I
11051 ATCGAAGTGG TATACAGGGG ATCGGAGGCA TAGCCAGCGG TGTAGGGCGA
TAGCTTCACC ATATGTCCCC TAGCCTCCGT ATCGGTTCGCC ACATCCCGCT
-1 M S T T Y L P D S A Y G A T Y P S
11101 AACGATGGCA CCGCCGCCCA CAGCCCCCTG ATGATAGATG ATCTCATTGT
TTGCTACCGT GCGGCGGGT GTCGGGGGAC TACTATCTAC TAGAGTAACA
-1 V I A G G G V A G Q H Y I I E N
11151 AGCTGGCCGC CAGCTGGCCA TTCCAAGGA GAGAACTGCC GCCATTATAG
TCGACCGGCG GTCGACCGGT AAGGGTTCCT CTCTTGACGG CGGTAATATC
-1 S A A L Q G N G L L S S G G N Y
11201 TCAATACCGC CAATCACACC CCACGCAGTG CTGTTGACGC CGTTGCCTTC
AGTTATGGCG GTTAGTGTGG GGTGCGTCAC GACAACTGCG GCAACGGAAG
-1 D I G G I V G W A T S N V G N G E
11251 AGGAATACCT CCAACGATGG CGCCAGATCC AAGGCTCGCA TTGAGGAGAC
TCCTTATGGA GGTGCTACC GCGGTCTAGG TTCCGAGCGT AACTCCTCTG
-1 P I G G V I A G S G L S A N L L
11301 TGTTACCCTC CCATTGACGA TTAACTTGCG CGCCAAAGAA GGGATCAAAA
ACAATGGGAG GGTAAGTCT AATTGAACGC GCGGTTTCTT CCCTAGTTTT
-1 N G E W Q R N V Q A G F F P D F
11351 CCACTTCCCG TTTTGAGGGT GTRACTGCGCA TCTCCATATA CCGTGTGGGC
GGTGAAGGGC AAAACTCCCA CATGACGCGT AGAGGTATAT GGCACACCCG
-1 G S G T K L T Y Q A D G Y V T H A
11401 AAAGCTGTAG AAATTATAGT ACCAAGCCTG GGCATTGATG CCATAACGCT
TTTCGACATC TTTAATATCA TGGTTCGGAC CCGTAACTAC GGTATTGCGA
-1 F S Y F N Y Y W A Q A N I G Y R
11451 TGTAGCTGGT GCCAAAGGCG AGGATGCCAT TAGATGTCCG ATCAGTATCT
ACATCGACCA CGGTTTCCGC TCCTACGGTA ATCTACAGCC TAGTCATAGA
-1 Y S T G F A L I G N S T P D T D
11501 GGCAGTCCTC CAACGCCACC ATAAATGGGA TCGCTATCAT AATTAGGCTT
CCGTCAGGAG GTTGCGGTGG TATTTACCCT AGCGATAGTA TTAATCCGAA
-1 P L G G V G G Y I P D S D Y N P K
11551 GTAGTAGAGG TTATCTCGAT AGTAATCACC CGACGTACGG CTCTTCCAAC
CATCATCTCC AATAGAGCTA TCATTAGTGG GCTGCATGCC GAGAAGGTTG
-1 Y Y L N D R Y Y D G S T R S K W
11601 GGAAGATACG CATAACATAG ATATTCCAAC CGCGGTACGG AGAAACCTGC
CCTTCTATGC GTATGGTATC TATAAGGTTG GCGCCATGCC TCTTTGGACG
-1 F I R M G Y I N W G R Y P S V Q

```

} OprD Domain

11651 GCAAAAACAC CTTGGTAGGT AGCAGGTATA ACGCGGGAAT CGGCGCCATT
CGTTTTTGTG GAACCATCCA TCGTCCATAT TGCGCCCTTA GCCGCGGTAA
-1 A F V G Q Y T A P I V R S D A G N

11701 GACCCAGGGG GTGTTGACGA TCTGGTTGCC AGCACGAATG AGCACGATCT
CTGGGTCCCC CACAACCTGCT AGACCAACGG TCGTGCTTAC TCGTGCTAGA
-1 V W P T N V I Q N G A R I L V I

11751 GTGGGATCTC ATACTGCAGA TAGGCCTGCC CCAGAGCGTT GATGGAGTCC
CACCTAGAG TATGACGTCT ATCCGGACGG GGTCTCGCAA CTACCTCAGG
-1 P I E Y Q L Y A Q G L A N I S D

11801 TGGGTGCCCA TAAGGGTTAC ATCTGGGTTG ACGCGATTAG CGCCAAACT
ACCCACGGGT ATTCCCAATG TAGACCCAAC TGCGCTAATC GCGGTTTTGA
-1 Q T G M L T V D P N V R N A G L S

11851 GTTGCCCGTG TAGAAACCAA CATCGGCGCT GAAGCCC CGG AGGTAGCCCG
CAACCGGCAC ATCTTTGGTT GTAGCCGCGA CTTCCGGGCGC TCCATCGGGC
-1 N A T Y F G V D A S F G A L Y G

11901 TATGAGCGTT GATGAAGCCA CCTAGGGAGA AGGCACTAAG ATTCGTAGCA
ATACTCGCAA CTACTTCGGT GGATCCCTCT TCCGTGATTC TAAGCATCGT
-1 H A N I F G G L S F A S L N T A

11951 TTTTTGCCAA AGAAACGATC GAAATAATAG GCTCGGATCT GGCCGTGAC
AAAAACGGTT TCTTTGCTAG CTTTATTATC CGAGCCTAGA CCGGCAGCTG
-1 N K G F F R D F Y Y A R I Q G D V

12001 GGTCCCATCG GTGATCACGC TTTTCAACGC ATTGGCGCTA CTGTCAGCGT
CCAGGGTAGC CACTAGTGCG AAAAGTTGCG TAACCGCGAT GACAGTCGCA
-1 T G D T I V S K L A N A S S D A

12051 ACGCTGTTGC ATTAATAACA ATCATGGTTA GAAACAGAGA AGAGCACACA
TGCGACAACG TAATTATTGT TAGTACCAAT CTTTGTCTCT TCTCGTGTGT
-1 A T A N I V I M T L F L S S C V

12101 TACGTCAACC TGTTCATGAG CATATCCTCA TTCAGTGTCA ATATCTGTAC
ATGCAGTTGG ACAAGTACTC GTATAGGAGT AAGTCACAGT TATAGACATG
-1 Y T L R N M L M D E N L T L I Q V

12151 ATATATAGGC AAATAGATAT CATTATTCCA CCCAGCCTAC AATGAACTAT
TATATATCCG **TT**TATCTATA GTAATAAGG**T** GGGTCGGATG TTAATTGATA
-1 Y I P L RBS

ORF 9: Start site

12201 AATATATCTA AATAAAAATA AAATTACATT TATATTATCA CATGGCTACA
TTATATAGAT TTATTTTTAT TTTAATGTAA ATATAATAGT GTACCGATGT

OprD
Domain

12251 TATTATTCTA TGTTATCATA TTAATTATCC TCGCGTTAGA TTAATGCTCT
ATAATAAGAT ACAATAGTAT AATTAATAGG AGCGCAATCT AATTACGAGA
-2 * H E

12301 GCACGATGGT CTGCAAGCAA GTTTTTGCAC CGGTCTCTTA AGCTTGCTGC
CGTGCTACCA GACGTTTCGTT CAAAAACGTG GCCAGAGAAT TCGAACGACG
-2 A R H D A L L N K C R D R L S A A

12351 CAAATCGCTA TTCATAAACT TCTCTCTGGA TCTTGGATGT GGCTCTTCTA
GTTTAGCGAT AAGTATTTGA AGAGAGACCT AGAACCTACA CCGAGAAGAT
-2 L D S N M F K E R S R P H P E **E**

12401 TATAAAGAAC ATCAGATATC TTGCCATTAT TCTTTGACAT AACCACCACT
ATATTTCTTG TAGTCTATAG AACGTAATA AGAAACTGTA TTGTTGGTGA
-2 **Y L V D S I K G N N K S M V V V**

12451 CTCTGAGACA TAAATATTGC TTCATCGATA TCATGAGTAA CAAAATTAT
GAGACTCTGT ATTTATAACG AAGTAGCTAT AGTACTCATT GTTTTAAATA
-2 **R Q S M F I A E D I^g D^g H^{ag} T^g V^g F^g I^g I**

12501 TGTCACCCCA GCTTTGATCC AAATATCAAT CAGACTATTT TGTAATGACG
ACAGTGGGGT CGAAACTAGG TTTATAGTTA GTCTGATAAA ACATTACTGC
-2 **T V G A K I W I D I L S N Q L S**

12551 ATCTCATCTG ATTGTCGAGA GCCCCAAAAG GCTCATCCAT CAAAAGCACC
TAGAGTAGAC TAACAGCTCT CGGGGTTTTTC CGAGTAGGTA GTTTTCGTGG
-2 **R M Q N D^e L^e A^e G^e F P E^{ad}D^{ad} M^d L^d L^d V^d**

12601 GCAGGGTGGT TTGCGAGGGC GCGAGCAATT TCGACTCTAT GTTGCATGCC
CGTCCCACCA AACGCTCCCG CGCTCGTTAA AGCTGAGATA CAACGTACGG
-2 **A P H N A L A R A I E^b V^b R^b H^b Q^b M^b G^b**

12651 GCCAGACAAC ATTTTTGGAT AATATGACTC ATATCCGTTT AATCCAACCA
CGGTCTGTTG TAAAAACCTA TTATACTGAG TATAGGCAAA TTAGGTTGGT
-2 **G^b S^b L^b M K P Y Y S E Y G N L G V**

12701 GTCTAATGAA TTCTCGCGCT ATTTTATTTG CCTCAGACTT ATTAATCCCT
CAGATTACTT AAGAGCGCGA TAAAATAAAC GGAGTCTGAA TAATTAGGGA
-2 **R I F E R A I K N A E S K N I G**

12751 CTCTGTATTG GACCAAACG TACATTTTCT ATCACTGTCT TCCATGGGAA
GAGACATAAC CTGGTTTTGC ATGTAAGA TAGTGACAGA AGGTACCCTT
-2 **R Q I P G F R V N E I V T K W P F**

12801 TAATGTGTGT TGTGAAATA CTATGCCAAG ATCAGGGTTT GCTTTTAGTA
ATTACACACA ACAACTTTAT GATACGGTTC TAGTCCCAA CGAAAATCAT
-2 **L T H Q Q^{af} F^f V^f I^f G L D P N A K L V**

ABC-NrtD-
SSuB
Transporter
domain

12851 CTTCTATCCC ATCAAGCAAT ACCTTGCCTT CTAATGGTTT ATAATGACCA
 GAAGATAGGG TAGTTCGTTA TGGAACGGAA GATTACCAA TATTACTGGT
 -2 E I G D L L V K G E L P K Y H G

12901 GCTATCGCAT TCAACAAGGT ACTTTTCCCA CAACCTGAAG GGCCAATAAG
 CGATAGCGTA AGTTGTTCCA TGAAAAGGGT GTTGGACTTC CCGGTTATTC
 -2 A I A N L L T^a S^{ac} K^{ac} G^c C^{ac} G^{ac} S^c P^c G^c I L

12951 GGAAGTGAAT GATCCAGACG GAAAAACGAA GGAGCAATCA TTAACAATGA
 CCTTCACTTA CTAGGTCTGC CTTTTTGCTT CCTCGTTAGT AATTGTTACT
 -2 S T F S G S P F V F S C D N V I

13001 TTGATGTGCG CCCATTATTA ACGTATCCCA AACTGAGTCC TTTAACCTCA
 AACTACACGC GGGTAATAAT TGCATAGGGT TTGACTCAGG AAATTGGAGT
 -2 S T R G N N V Y G L S L G K V E

13051 ATATTTCCAA CTTTATCCAT TTTAGTTTAA AGCCTTATTC CGCCCATGGA
 TATAAAGGTT GGAATAGGTA AAATCAAATT TCGGAATAAG GCGGGTACCT
 -2 I N G V K D M RBS

ABC-NrtD-
SSuB
Transporter
domain

ORF 10: start site

^a ATP binding site; ^b ABC transporter signature; ^c Walker A/P loop;
^d Walker B; ^e D-Loop; ^f Q-loop/Lid; ^g H-loop/Switch region

-3 * E A W P

13101 ACAACTATTT GTTTCATAAC ATAGATCAGT TTGGTGCTTA CAGCGCTAAG
 TGTGATAAAA CAAAGTATTG TATCTAGTCA AACCACGAAT GTCGCGATTC
 -3 V V I Q K M V Y I L K T S V A S L

13151 TATCCGATT GTAACCATGC CAACAATTAT ATTGCCATAT TGAATTAGTT
 ATAAGGCTAA CATTGGTACG GTTGTTAATA TAACGGTATA ACTTAATCAA
 -3 I G I T V M G V I I N G Y Q I L

13201 CATAAGACTC CCAAGTAAAA TATCCTATGC CAAATCTTCC GCTAATCATT
 GTATTCTGAG GGTTCATTTT ATAGGATACG GTTTAGAAGG CGATTAGTAA
 -3 Y S E W T F Y G I G F R G S I M

13251 TCCCCAGCTA ATAATGCAAA CCACGCACCG CCCATACCAA GGGCCATTCC
 AGGGGTCGAT TATTACGTTT GGTGCGTGGC GGGTATGGTT CCCGTAAGG
 -3 E G A L L A F W A G G M G L A M G

13301 AACCGTTATG TTCGGAATTG CATA CGGAAT CACTACGTTA AACAAAACGC
 TTGGCAATAC AAGCCTTAAC GTATGCCTTA GTGATGCAAT TTGTTTTGCG
 -3 V T I N P I A Y P I V V N F L V

TauC
permease
component

13351 TAGTTTTTTT TGCGCCCAA GAACGTGCGG CTTTTATTAA TTCGTACGGC
 ATCAAAAAA ACGCGGGTTT CTTGCACGCC GAAAATAATT AAGCATGCCG
 -3 T K K A G L S R A A K I L E Y P

13401 ACACGTCTTA CACCATCCCA TGTGTTGAGC GCAATCGGGA ATACTGCTCC
 TGTGCAGAAT GTGGTAGGGT ACACAACCTCG CGTTAGCCCT TATGACGAGG
 -3 V R R V G D W T N L A I P F V A G

13451 AAGAAACGTA ATAAAGATGA TGCTTTCTTC CGTTGATGGC CACATTAGAA
 TTCTTTGCAT TATTTCTACT ACGAAAGAAG GCAACTACCG GTGTAATCTT
 -3 L F T I F I I S E E T S P W M L

13501 TAGCCATTGG TACCCAGGCT ATTGCCGGTA TTGGCCTAAG AATCTCTAAA
 ATCGGTAACC ATGGGTCCGA TAACGGCCAT AACCGGATTC TTAGAGATTT
 -3 A M P V W A I A P I P R L I E L

13551 TATGGAGAAA TAATAGCCCT TAGAATCGTG CTATAGCCTA CTGAGACACC
 ATACCTCTTT ATTATCGGGA ATCTTAGCAC GATATCGGAT GACTCTGTGG
 -3 Y P S I I A R L I T S Y G V S V G

13601 AACCGCTATC CCAACAATTG TTCCAATGGT AAATGCTACA CTAATCCTAA
 TTGGCGATAG GGTTGTTAAC AAGGTTACCA TTTACGATGT GATTAGGATT
 -3 V A I G V I T G I T F A V S I R

13651 TCATGCTCCA AAGTATGTCT TTCCAGTATG CTTGGAAGA TAATGAAGCA
 AGTACGAGGT TTCATACAGA AAGGTCATAC GGAACCTTCT ATTACTTCGT
 -3 M S W L I D K W Y A K S S L S A

13701 AGGAATGCTT TGAGAACGTG CATAACAGTG GGGATATTGG AAAAGTCTAA
 TCCTTACGAA ACTCTTGAC GATTGTGAC CCCTATAACC TTTTCAGATT
 -3 L F A K L V H M V T P I N S F D L

13751 AACAAATTTA ACGTGCGAAG CTGTCAAGAC TTCGCCATAA CGCTAACCCCT
 TTGTTTTAAAT TGCACGCTTC GACAGTTCTG AAGCGGTATT GCGATTGGGA
 -3 V F K V H S A T L

TauC
permease
component

RBS

ORF 11: start site

13801 GCAAAAACGG ATAACGTAGA CAAGATTATA GATCCCAATC GTGACTGTAC
 CGTTTTTGCC TATTGCATCT GTTCTAATAT CTAGGGTTAG CACTGACATG

13851 TAATAACAGC AGATATGGAT ATGGCCGATA GGTATTCTGT ACCCTTTGGC
 ATTATTGTCT TCTATACCTA TACCGGCTAT CCATAAGACA TGGGAAACCG

13901 TCGGATTCCC TTTTCGCAGCA GCTGCTGGCA ATTTTCTTTC AGCGACCTGG
 ACGCTAAGGG AAAGCGTCGT CGACGACCGT TAAAAGAAAG TCGCTGGACC

13951 CGATATTCTT CAACGTTGTT CATTTTTGTC ACCATACCAA ATTTATACCT

GCTATAAGAA GTTGCAACAA GTAAAAACAG TGGTATGGTT TAAAATATGGA
-1 * V

14001 TGGAGATAAT TTTGTTATAG GTGTAGACTT GTGTTCCATG AGCAAACCTTT
ACCTCTATTA AAACAATATC CACATCTGAA CACAAGGTAC TCGTTTGAAA
-1 S I I K N Y T Y V Q T G H A F K

14051 TCAGCGCCTT GTTTTGTTCAT GAACGAGGTA ATCTTGTTTC CGATCATTAC
AGTCGCGGAA CAAAACAGTA CTTGCTCCAT TAGAACAAAG GCTAGTAATG
-1 E A G Q K T M F S T I K N G I M V

14101 ATAATATGCA TATTTCGAAA AGAGTTTCAA TCCAGAATCG TGGTCATAAA
TATTATACGT ATAAAGCTTT TCTCAAAGTT AGGTCTTAGC ACCAGTATTT
-1 Y Y A Y K S F L K L G S D H D Y

14151 CATAATTGC ATCTATTTTT TTGCGCGAAG CTATCGCAGA TTTAGCTGCC
GTATGTAACG TAGATAAAAA AACGCGCTTC GATAGCGTCT AAATCGACGG
-1 Y M A D I K K R S A I A S K A A

IR

14201 AATAGCATCT GTTTAACAGT CTTATATTTT CTTATTCCCT TTCCGGCATA
TTATCGTAGA CAAATTGTCA GAATATAAAA GAATAAGGGA AAGGCCGTAT
-1 L L M Q K V T K Y K R I G K G A Y

14251 CCAAATCTCT GGCGGCAGGT TGACATTTTT TTCTGGGTTG TAAGGTGTGG
GGTTTAGAGA CCGCCGTCCA ACTGTAAAAA AAGACCCAAC ATTCCACACC
-1 W I E P P L N V N K E P N Y P T

14301 CGGTGGTTGC TGCATCATAA TTCAATCCAA GCTGCTTATA TGCTAGCTTC
GCCACCAACG ACGTAGTATT AAGTTAGGTT CGACGAATAT ACGATCGAAG
-1 T T A A D Y N L G L Q K Y A L K

14351 TGAAAGGATG GATTTACCCA TTGTTCAAAA TTTAACGTAG GAATATGTGC
ACTTTCCTAC CTAAATGGGT AACAAAGTTTT AAATTGCATC CTTATACACG
-1 Q F S P N V W Q E F N L T P I H A

14401 ATATTTTCGAC AGTATACTAT GGTCAATTTT TAGCGTGTCA ACCCATTTTT
TATAAAGCTG TCATATGATA CCAGTATAAA ATCGCACAGT TGGGTAAAAA
-1 Y K S L I S H D Y K L T D V W K

DR

14451 CTGTTATTGC TGGGTTTAAAT GTGAGGTAGC CACCATGACT AAAATATAGG
GACAATAACG ACCCAAATTA CACTCCATCG GTGGTACTGA TTTTATATCC
-1 T I A P N L T L Y G G H S F Y L

ORF 12: Truncated

14501 CTCTTCGTCA GATTGAGTGG GTAGGGGGTA GAATGCTGTG GAGCGGTCGT
GAGAAGCAGT CTAACCTACC CATCCCCCAT CTTACGACAC CTCGCCAGCA

ORF 13: start site

```

+1           RBS           M V P E E L F S L
14551 GCCCTGAATA GGAGTTAGCT GATGGTCCCT GAAGAGCTGT TTTCTCTCGC
      CGGGACTTAT CCTCAATCGA CTACCAGGGA CTTCTCGACA AAAGAGAGCG

+1   L G L V P P W L V D H V T F T V
14601 GTTAGGATTG GTTCCGCCGT GGTGGTGGGA TCATGTGACT TTCACGGTGG
      CAATCCTAAC CAAGGCGGCA CCAACCACCT AGTACACTGA AAGTGCCACC

+1 E E K R L D L H I N F P K G S R F
14651 AGGAGAAACG CCTGGATCTG CACATCAACT TTCCCAAAGG TAGTCGCTTT
      TCCTCTTTGC GGACCTAGAC GTGTAGTTGA AAGGGTTTCC ATCAGCGAAA

+1   A C S V C G E E C P V H D T R D
14701 GCTTGCTCCG TCTGTGGTGA GGAGTGTCCG GTACATGATA CCCGTGACCA
      CGAACGAGGC AGACACCACT CCTCACAGGC CATGTACTAT GGGCACTGGT

+1   T W R H M D F F Q H E A Y L H A
14751 TACCTGGCGG CACATGGATT TCTTCCAGCA TGAAGCCTAT CTCCATGCCC
      ATGGACCGCC GTGTACCTAA AGAAGGTCGT ACTTCGGATA GAGGTACGGG

+1   R V P R V K C Q E H G V H Q I S V
14801 GTGTACCTCG TGTGAAGTGC CAGGAGCATG GAGTGCATCA GATATCTGTT
      CACATGGAGC ACACTTCACG GTCCTCGTAC CTCACGTAGT CTATAGACAA

+1   P W A R E G S R F T L L F E A L
14851 CCCTGGGCGC GGAAGGCTC GCGTTTCACC CTGCTCTTTG AAGCGCTGAT
      GGGACCCGCG CCCTTCCGAG CGCAAAGTGG GACGAGAAAC TTCGCGACTA

+1   M T L V R E M P V L T A A R M V
14901 CATGACCCTG GTGCGGGAGA TGCCGGTATT GACGGCAGCT CGCATGGTCG
      GTACTGGGAC CACGCCCTCT ACGGCCATAA CTGCCGTCGA GCGTACCAGC

+1   G E T D K L L W R V I D H Y V P E
14951 GTGAGACCGA CAAGCTCCTG TGGCGAGTGA TTGACCATTA TGTGCCCGAA
      CACTCTGGCT GTTCGAGGAC ACCGCTCACT AACTGGTAAT ACACGGGCTT

+1   A R A A V D M A N V H A V G V D E
15001 GCTCGTGCTG CGGTGGATAT GGCCAATGTC CATGCCGTCG GCGTCGATGA
      CGAGCACGAC GCCACCTATA CCGGTTACAG GTACGGCAGC CGCAGCTACT

+1   T S S R H G H D Y I T L F V D L
15051 AACCAGCAGT CGGCATGGAC ATGACTACAT CACGCTCTTC GTGGATCTGA
      TTGGTCGTCA GCCGTACCTG TACTGATGTA GTGCGAGAAG CACCTAGACT

+1   N A R R L L F A T P G K D A K T F
15101 ATGCCCGGCG ACTCTTGTTT GCTACTCCCG GCAAGGATGC CAAGACCTTT
      TACGGGCCGC TGAGAACAAG CGATGAGGGC CGTTCCTACG GTTCTGGAAA

```

Trans-
posase

```

+1  E K F S A D L Q A H G G S A E A
15151 GAGAAATTCT CCGCAGATCT ACAGGCCCAT GGTGGTAGCG CGGAAGCGAT
      CTCTTTAAGA GCGTCTAGA TGTCCGGGTA CCACCATCGC GCCTTCGCTA

+1  T D V S M D L S P A F Q K G A A
15201 CACCGATGTG AGCATGGACC TCTCGCCGGC CTTCCAGAAA GGGGCTGCCG
      GTGGCTACAC TCGTACCTGG AGAGCGGCCG GAAGGTCTTT CCCCAGCGGC

+1  E H L P N A E I T F D R F H L M K
15251 AGCACCTGCC CAATGCGGAG ATCACTTTTCG ATCGTTTTCA CCTCATGAAG
      TCGTGGACGG GTTACGCCTC TAGTGAAAGC TAGCAAAGT GGAGTACTTC

+1  L V N E A V D D V R K G E V L T
15301 CTCGTCAACG AGGCCGTAGA CGACGTGCGC AAGGGGGAAG TCCTCACCCA
      GAGCAGTTGC TCCGGCATCT GCTGCACGCG TTCCCCTTC AGGAGTGGGT

+1  P N L K K T R W L W L K N D C N
15351 GCCAAATCTC AAAAAGACCC GCTGGCTTTG GCTCAAGAAC GATTGCAACC
      CGGTTTAGAG TTTTCTGGG CGACCGAAAC CGAGTTCTTG CTAACGTTGG

+1  L K M K Q K E K L Q E L L K D Q N
15401 TCAAAATGAA GCAGAAAGAA AAGCTGCAGG AATTGCTCAA AGACCAGAAC
      AGTTTTACTT CGTCTTTCTT TTCGACGTCC TTAACGAGTT TCTGGTCTTG

+1  L K T A Q A Y Q F R L T F Q D I
15451 CTCAAGACGG CGCAGGCCTA CCAGTTCCGC CTGACCTTTC AGGACATCTT
      GAGTTCTGCC GCGTCCGGAT GGTCAAGGCG GACTGGAAAG TCCTGTAGAA

+1  T I K N R H Q G A T L L K A W L
15501 CACGATCAAG AATCGCCACC AGGGGGCTAC CCTCTTGAAA GCCTGGTTGG
      GTGCTAGTTC TTAGCGGTGG TCCCCGATG GGAGAACTTT CGGACCAACC

+1  E N A R T S D L P P I V R V A Y T
15551 AAAACGCCAG AACCAGCGAT CTGCCGCTA TCGTCAGGGT CGCCTACACC
      TTTTGCGGTC TTGGTCGCTA GACGGCGGAT AGCAGTCCCA GCGGATGTGG

+1  I M N H W D G V L R W F E S Q I
15601 ATCATGAATC ACTGGGATGG CGTGCTCCGA TGGTTCGAGA GCCAGATCAC
      TAGTACTTAG TGACCCTACC GCACGAGGCT ACCAAGCTCT CGGTCTAGTG

+1  N G I L E G F N S L I Q S A K A
15651 CAATGGAATT CTGGAAGGTT TCAACAGCCT CATTCAATCC GCCAAGGCCA
      GTTACCTTAA GACCTTCCAA AGTTGTCGGA GTAAGTTAGG CGGTTCCGGT

+1  K A R G Y R T H K N F I N M A Y L
15701 AGGCTCGGGG TTACCGCACG CACAAGAACT TTATCAACAT GGCCTACCTG
      TCCGAGCCCC AATGGCGTGC GTGTTCTTGA AATAGTTGTA CCGGATGGAC

```

Trans-
posase

+1 I L G K L D L R L P T *

15751 ATCCTGGGCA AGCTGGATCT CAGGCTACCC ACTTGAAATG ACGAGGAACC
TAGGACCCGT TCGACCTAGA GTCCGATGGG TGAAC~~TTT~~TAC TGCTCCTTGG

DR

15801 AAAATATAAA TATAATACCT CTTTTGGGAT CATTGTCCAC TTTGCCATCA
TTTTATATTT ATATTATGGA GAAAACCCTA GTAACAGGTG AAACGGTAGT

15851 TTTTGCTTGC AACATATGGG TTCTTCGTAA TCCATGCATC AGCAGCTATT
AAAACGAACG TTGTATACCC AAGAAGCATT AGGTACGTAG TCGTCGATAA

15901 ACTGCTTATC AAAAGCTACA CTATCTGTGG ATACTTCTCC GCAAAGATTT
TGACGAATAG TTTTCGATGT GATAGACACC TATGAAGAGG CGTTTCTAAA

15951 GTAACCACTG CTCCGTGTAA ATATGGGACC CGTGTATCAA TACCGTTGTA
CATTGGTGAC GAGGCACATT TATACCCTGG GCACATAGTT ATGGCAACAT
-2 * Y R Q

16001 AATCATTCTT CCAACACCCT TTGTA~~CT~~CCA TATACTCAGA AATTGGACAA
TTAGTAAGAA GGTGTGGGA AACATGAGGT ATATGAGTCT TTAACCTGTT
-2 D N K W C G K Y E M Y E S I P C

16051 AAATCTGCGT GGGCTGCGAT CTTATCTGCT GTGATTGCTG CTATTCCAGC
TTTAGACGCA CCCGACGCTA GAATAGACGA CACTAACGAC GATAAGGTGCG
-2 F D A H A A I K D A T I A A I G A

16101 CATCGGGCTC TGATTAATGA TATGCAAGGA AGACATCGGA ATGTGCTTGT
GTAGCCCGAG ACTAATTACT ATACGTTCCCT TCTGTAGCCT TACACGAACA
-2 M P S Q N I I H L S S M P I H K

16151 CTTGTGCCAT CTCATATAGC ATTCCCCAAG ATGAACTTCC TATTGGGGTA
GAACACGGTA GAGTATATCG TAAGGGGTTT TACTTGAAGG ATAACCCCAT
-2 Q A M E Y L M G W S S S G I P T

16201 GAAATAACTT TCCCTTCTAA TTCTCCAACA TTTGTAATTT TTGACTTAAT
CTTTATTGAA AGGGAAGATT AAGAGGTTGT AAACATTAAA AACTGAATTA
-2 S I V K G E L E G V N T I K S K I

16251 CGGAACTACT ATGCCGTTTC CTGCCCCGTC CAGATTGTAT CCAGTCATTG
GCCTTGATGA TACGGCAAAG GACGGGGCAG GTCTAACATA GGTCAGTAAC
-2 P V V I G N G A G D L N Y G T M

16301 TTATAAGGTA CGAGCGCTCA TTCTTTGTTT GCTGAAATTT GGCTCCATTA
AATATTCCAT GCTCGCGAGT AAGAAACAAA CGACTTTAAA CCGAGGTAAT
-2 I L Y S R E N K T Q Q F K A G N

16351 ACAACTAATG GATAATCACC CATAGTACCA AAATCTAGTT TGTTGGCTAT
TGTTGATTAC CTATTAGTGG GTATCATGGT TTTAGATCAA ACAACCGATA

TauA
periplasmic
component

-2 V V L P Y D G M T G F D L K N A I

16401 CATCATGTTT GTTATTGGTG CACCAGAACT GTAGTCGCGC CAAATAATAT
 GTAGTACAAA CAATAACCAC GTGGTCTTGA CATCAGCGCG GTTTATTATA
 -2 M M N T I P A G S S Y D R W I I

16451 TATATTTTAC GTTAGCATAT TTGCCGGAAT GAGGAAGGTA TTTTTGTAAT
 ATATAAAATG CAATCGTATA AACGGCCTTA CTCCTTCCAT AAAAACATTA
 -2 Y K V N A Y K G S H P L Y K Q L

16501 AGATGCAGGC CCTTAATTAC CGTCCCAGCA GGATAGGTAT CGGTGCACAT
 TCTACGTCGG GGAATTAATG GCAGGGTTCGT CCTATCCATA GCCACGTGTA
 -2 L H L G K I V T G A P Y T D T C M

16551 ACTTTGATAG CCGATGCTTA TGTTAATTGT CTTTGC GTAC ACTATCGGTG
 TGAAACTATC GGCTACGAAT ACAATTAACA GAAACGCATG TGATAGCCAC
 -2 S Q Y G I S I N I T K A Y V I P

16601 AGACAGCCAC AAAGCAAGCA AAAGTAAAGA CTGGCAACCG TTTTCATATTC
 TCTGTCCGGT TTTTCGTTTCGT TTTTCATTTCT GACCGTTGGC AAAGTATAAG
 -2 V A V F C A F T F V P L R K M

ORF 14: start site

16651 TGCTCCAAGC GATTGATACA TGTCCGAACA TGTCATGCAA CACCTGTGCC
 ACGAGGTTTCG CTAACTATGT ACAGGCTTGT ACAGTACGTT GTGGACACGG
 ←
 RBS

16701 AGGAAAACCT CTTTATAAAA AGGGCAGTGA GTCCTTGTA AATGCGTAAT
 TCCTTTTGGG GGAATATTTT TCCCGTCACT CAGGAACATT TTACGCATTA

16751 AATAGTGCGT GGTGATAGAT GGATGTGCAG TGTTGCACTG GTATGGGGAG
 TTATCACGCA CCACTATCTA CCTACACGTC ACAACGTGAC CATACCCCTC

16801 GTCGGCCTCT GTTGCCAGAT CTCTCATTCC TCACGTTGTT CTTGTGCTTC
 CAGCCGGAGA CAACGGTCTA GAGAGTAAGG AGTGCAACAA GAACACGAAG

16851 AGGCGTGGCC ACTTGACTCC AGTTAACCTG ATTTACATCC AGCAGCACCG
 TCCGCACCGG TGAAGTGAAG TCAATTGGAC TAAATGTAGG TCGTCGTGGC

16901 ATGGGGACTA GATCTGACGT GGCAATGGGT TGGCTCCACG TCTTTACTGG
 TACCCCTGAT CTAGACTGCA CCGTTACCCA ACCGAGGTGC AGAAATGACC

16951 CTTCAAGGCC ACCATCTAGG TATCCAATTC TTACCTCGTT ATATACAGTC
 GAAGTTCCGG TGGTAGATCC ATAGGTGAAG AATGGAGCAA TATATGTCAG

17001 CCGTGTTCTG AAGGCTGAAC GGGTGTGAAC ATGATCAATC AAAAAATCCA
 GGCACAAGAC TTCCGACTTG CCCACACTTG TACTAGTTAG TGTTTTAGGT

} TauA
 periplasmic
 component


```

17051 CCCTTTTGA GAGGCCCGGA GTTTCGAGAA TCCGGGCCTT GGTGCTGCTG
      GGGAAAAACT CTCCGGGCCT CAAAGCTCTT AGGCCCGGAA CCACGACGAC

17101 GCCTCGGCGT TCTCGTAACC CCCTATCGCT TTGCGCCGGG CGCCATTAC
      CCGAGCCGCA AGAGCATTGG GGGATAGCGA AACGCGGCC CCGGTAAGTG

17151 GGCTTTCCAG AACAAAGGCTC CCTCACAAAA GCCGAACGGA GTGCCCTGC
      CCGAAAGGTC TTGTTCCGAG GGAGTGT TTTT CGGCTTGCCT CACGGGGACG
-1      * L L R V S H G Q

17201 GCAGCCAGAA TCCGCACCGC ACGCTTGTCG AAGGACACAA AGGTCTCTCC
      CGTCGGTCTT AGGCGTGGCG TGCGAACAGC TTCCTGTGTT TCCAGAGAGG
-1 A A L I R V A R K D F S V F T E G

17251 ACCCAGGCAG TACCCTTCGT ACGCAATCAC ACCGTCCGCA AAATCTCCAC
      TGGGTCCGTC ATGGGAAGCA TGCCTTAGTG TGGCAGGCGT TTTAGAGGTG
-1 G L C Y G E Y A I V G D A F D G

17301 CGGCCTCCAA TACCAACAAC CCCGCTTCCG CTCCCGCTCG GTCCACATCT
      GCCGGAGGTT ATGGTTGTTG GGGCGAAGGC GAGGGCGAGC CAGGTGTAGA
-1 A E L V L L G A E A G A R D V D

17351 ACGTTGCCAG CCGATAGCAG CGCCCGAATC GCACCGGCCA CATCGGACTC
      TGCAACGGTC GGCTATCGTC GCGGGCTTAG CGTGGCCGGT GTAGCCTGAG
-1 V N G A S L L A R I A G A V D S E

17401 CGTAAAACCA TACACCCGAA GGAGTACCCA CACGAACTCG CACAAACACG
      GCATTTTGGT ATGTGGGCTT CCTCATGGGT GTGCTTGAGC GTGTTTGTGC
-1 T F G Y V R L L V W V F E C L C

17451 GCAACGAGAC CGCGACCCGC TCCGCTTCCC GCAATACCGC AGCCGCCGCC
      CGTTGCTCTG GCGCTGGGCG AGGCGAAGGG CGTTATGGCG TCGGCGGCGG
-1 L S V A V R E A E R L V A A A A

17501 TGCCTTGCG CTGGATCGTC ACATACCGCA GCACGGACCA GAACGTTGGT
      ACGCGAACGC GACCTAGCAG TGTATGGCGT CGTGCCTGGT CTTGCAACCA
-1 Q A Q A P D D C V A A R V L V N T

17551 ATCCACCGCG ACCTTCATTC CTTACCCGAC CATCCCTGGG CAATCGCCTC
      TAGGTGGCGC TGGAAGTAAG GAA TGGGCTG GTAGGGA CCC GTTAGCGGAG
-1 D V A V K M ← RBS ← RBS

      ORF 15: start site

-2      * E K G S W G Q A I A E

17601 ATTCATGTCT TCGATGCTGG CCACCCTCCG GGTCTTACCC GCAAGCACAC
      TAAGTACAGA AGCTACGACC GGTGGGAGGC CCAGAATGGG CGTTCGTGTG

```

Predicted
nucleic
acid
binding
domain

-2 N M D E I S A V R R T K G A L V

17651 CAATAAATCC CTCTACTGTC CCTGCCGGGC GCCCTGCCGT GAGCAGGACC
 GTTATTTAGG GAGATGACAG GGACGGCCCG CGGGACGGCA CTCGTCCTGG
 -2 I F G E V T G A P R G A T L L V

17701 CTCCCACCGG GCAGTAGCTC CACCTCGATC TGGTCTCCGT ACCGAATACC
 GAGGGTGGCC CGTCATCGAG GTGGAGCTAG ACCAGAGGCA TGGCTTATGG
 -2 R G G P L L E V E I Q D G Y R I G

17751 CACATGGTGC AAAACCGATT TCCGGAGCGT CACTTGCCCC TTCGCCGTAA
 GTGTACCACG TTTTGGCTAA AGGCCTCGCA GTGAACGGGG AAGCGGCATT
 -2 V H H L V S K R L T V Q G K A T

17801 CCTTCAAGGT CGTCACGCTG CCCTCCATCC TCACGAAACA CCCATCTTCT
 GGAAGTTCCA GCA**GTG**CGAC **GGGAGG**TAGG AGTGCTTTGT GGGTAGAAGA
 -2 K L T T V **RBS**

ORF 16: start site

17851 ATAGTAATGG AAATACGCCT TACCGAGAAA CAAGGCGCAC TACCAATCGT
 TATCATTACC TTTATGCGGA ATGGCTCTTT GTTCCGCGTG ATGGTTAGCA

17901 CCCCCAGACC ACGCTTCGGA GTGACGGTCC TTCCTCGCGC CCGATCCGCC
 GGGGGTCTGG TGCGAAGCCT CACTGCCAGG AAGGAGCGCG GGCTAGGCGG

17951 GCCGCATTGG GCATCCCCAC CTCCATCCGG ATCCGCTCGA CCACCCGAAC
 CGGCGTAACC CGTAGGGGTG GAGGTAGGCC TAGGCGAGCT GGTGGGCTTG

18001 CAAATCCTGG TCCGCCGCCT TCGCTGCGCG AAAGAGAGCC ATCAGCTGCG
 GTTTAGGACC AGGCGGCGGA AGCGACGCGC TTTCTCTCGG TAGTCGACGC

18051 CGTCCCCAAC GACCGCACCG GGCTTCCCAT CCGCCGACCA CCAGCGAAAC
 GCAGGGGTTG CTGGCGTGGC CCGAAGGGTA GCGGCTGGT GGTGCTTTG

18101 CCCTCCGCCG TTTCTGGAT AGTGAACCCT CGGAACAACC TGTCTCGAG
 GGGAGGCGGC AAAGGACCTA TCACTTGGGA GCCTTGTTGG ACAGGAGCTC