Molecular analysis and characterisation of plasmid pTF5 from *Thiobacillus ferrooxidans* ATCC33020

Clifford Noel Dominy

In fulfilment of the requirements for the degree of
Doctor of Philosophy in the Faculty of Science,
University of Cape Town.

CAPE TOWN

FEBRUARY 1997

To my wife Patricia,
for her patience, love and understanding
over the years.

and to my parents,
Ashton and Denise Dominy
for their encouragement and support.

CERTIFICATION OF SUPERVISOR

In terms of paragraph 9 of "General regulations for the degree of Ph.D." I as supervisor of the candidate Clifford Noel Dominy, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

Professor D.E. Rawlings
Department of Microbiology
University of Cape Town

Acknowledgements

I am greatly indebted to my supervisor, Professor Doug Rawlings for his enthusiastic guidance and support over the duration of this project. His confidence and optimism have served as a source of inspiration over the years.

Many thanks go to members of the "Thiobacillus Research Unit", both past and present, who freely gave advice (and reagents) in times of crises. Specifically I would like to thank Dave Berger for his initial supervision of the project and Shelly Deane for invaluable help with the pulsed field gel electrophoresis.

Thanks also go to Ann Jaffray, Di de Villiers, Anne-Marie Clennel, Stan Botman, Charlie Hendrikse, Nikki Campbell and Genevieve Wilson whose administrative and technical assistance eased my life considerably within the department. A special thanks must go to Di James for advice and assistance with, what must have been, every commercial sequencing kit in existence between 1991 and 1996. I am indebted to Nicolette Coram who finished off the "last few kb" of pTF5 sequencing as part of her B.Sc (Hons) studies.

I also feel obliged to thank "The Louts", in particular Will Bourn, Jono Lane, Joe Santangelo, Gordon D. Brown and Ant Smith, for many an evening spent in "The Laboratory" critically evaluating various aspects of this work.

TABLE OF CONTENTS

ABSTRACT

Plasmid pTF5, an endogenous plasmid of *Thiobacillus ferrooxidans* was isolated and cloned from a *T. ferrooxidans* ATCC33020 pHC79 cosmid library. The plasmid was isolated due to its ability to confer metronidazole sensitivity to an *Escherichia coli* metronidazole resistant mutant. The region responsible for this phenotype was narrowed down to a 319 bp fragment which did not appear to contain an open reading frame. A nearly identical fragment had been previously reported as being part of an origin of vegetative replication (*oriV*) of the narrow host-range plasmid, pTFI91 from *T. ferrooxidans* TFI91 (Chakravarty et al. 1995). Plasmid pTF5 differed from pTFI91 in at least two aspects; its size, (pTF5 was 19.8 kbp versus 9.8 kbp for pTFI91), and in the number of iterons present in its *oriV*. Plasmid pTF5 was completely sequenced and found to contain four open reading frames (ORFs) which encoded putative iron-sulphur containing proteins. Two of the ORFs with amino acid homology to ferredoxin-like (*fdxA*) and *fnr*-like regulatory genes respectively, appeared to be divergently transcribed from the other two ORFs, one of which had amino acid sequence similarities to a prismane-like protein (*psmA*) and the other to a NADH oxidoreductase protein (*redA*). *In vitro* transcription-translation analysis showed that only the *fdxA* gene was recognised by a heterologous *E. coli* system, with the *psmA* and *redA* genes requiring *E. coli*-derived promoters in order to be expressed. Evidence of a "transpositional scar" was detected between the *fnr*-like and *fdxA* genes, due to the presence of two incomplete ORFs with sequence identity to a transketolase gene and transposase gene from *E. coli* and *T. ferrooxidans* respectively. Plasmid pTF5 was present in *T. ferrooxidans* at a copy number of 2 to 4 per genome. As was the case with other *T. ferrooxidans* pTFI91-like plasmids, pTF5 could not replicate in *E. coli* (Rawlings and Woods 1985, Valenti et al. 1990, Chakravarty et al. 1995) and was not detected in the *Thiobacillus thiooxidans* or *Leptospirillum ferrooxidans* strains tested. Pulsed field gel electrophoresis showed that the approximately 6 kbp of sequence of pTF5, which

included the iron-sulphur protein-encoding ORFs, was also present on the chromosome of strain ATCC33020. These ORFs were not essential to the survival of the host as a number of *T. ferrooxidans* strains did not contain either the plasmid or the chromosomally located ORFs. Some of the *T. ferrooxidans* strains tested appeared not to contain plasmids, although pTF5-like plasmid sequence was detected and had apparently become integrated into the hosts chromosome. The iron-sulphur protein-encoding ORFs of pTF5 had organisational similarity to simple electron transport pathways which in other bacteria are responsible for the degradation of aromatic compounds. Due to the possibility of a redox active role by products of the pTF5 ORFs, resistance to a number of heavy metal compounds was tested in *E. coli* strains which contained constructs which expressed the proteins *in vitro*. The presence of these constructs in the heterologous host did not alter the tolerance of the *E. coli* strain to either uranium or silver.

ABBREVIATIONS

A	adenosine
aa	amino acids
Amp	ampicillin
ATCC	American Type Culture Collection
bp	base pairs
BSA	bovine serum albumin
C	cytosine
C-	carboxy terminal (of protein)
cccDNA	covalently closed circular DNA
CoA	coenzyme A
Cm	chloramphenicol
^{o}C	degree's Centigrade
d	day(s)
Da	daltons
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
Dnase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DTT	1,4-dithio-L-threitol
EDTA	ethylene-diaminetetra-acetic acid
Em	erythromycin
EtBr	ethidium bromide
g	grams
G	guanine
h	hour(s)
IHF	integration host factor
IPTG	isopropyl-B-D-thiogalactopyranoside
IS	insertion sequence element
kbp	kilobase pairs
kDa	kilodaltons
LB	Luria broth
M	molar
min	minutes
Mr	relative molecular mass
N-	amino terminal (of protein)
NADH	nicotinamide adenine dinucleotide (reduced)
nt	nucleotide

OD_{600}	optical density at 600 nanometres
ORF	open reading frame
oriV	origin of vegetative replication
PAGE	polyacrylamide gel electrophoresis
r	resistant (superscript)
RBS	ribosome binding site
RNA	ribonucleic acid
s	sensitive (superscript)
s	seconds
SDS	sodium dodecyl sulphate
sp	species
T	thymidine
TBE	tris-borate EDTA buffer
Tn	transposon
Tris	Tris(hydroxymethyl)aminomethane
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
α	alpha
β	beta
Δ	delta
λ	lambda
μ	micro
σ	sigma

Chapter 1

Literature Review

Literature Review

The primary aim of this investigation was the isolation and characterisation of genes from *Thiobacillus ferrooxidans* ATCC33020 which encode redox-active proteins. *T. ferrooxidans* has an unusual physiology in that it is a Gram negative autotroph which grows under very acidic conditions (pH = 2). Consequently many of the genetic techniques used in neutrophilic, heterotrophic bacteria cannot be employed in molecular genetic studies of *T. ferrooxidans*. An example of this would be the difficulty of generating electron transport mutants in this micro-organism. These enzymes play a central metabolic role in supplying energy to the growing cell and any mutation introduced into these pathways would have a lethal effect on the cell. This sort of investigation therefore employed the use of a more indirect method to isolate these genes. The antimicrobial agent metronidazole was used to select for redox-active phenotypes in a metronidazole resistant *E. coli* mutant. This approach had been used successfully in isolating redox active genes from other bacteria (Santangelo et al. 1991). This dissertation reports the isolation and characterisation of a number of genes whose products have homology to redox-active proteins, which were unexpectedly discovered on an endogenous *T. ferrooxidans* plasmid.

A major focus of this review will be the plasmids of *T. ferrooxidans*. A separate section will review the progress which has been made in the use of plasmids in the development of genetic systems for the thiobacilli. The discovery of electron-transport protein-encoding genes on plasmid pTF5 has led to the inclusion of relevant aspects of this area of *Thiobacillus* research in the review. Three fundamental classes of electron transferring reactions will be dealt with - the enzymes involved in i) ferrous iron respiration, ii) the sulphur oxidation pathway, and finally iii) enzymes involved in the redox reactions of selected heavy metal elements and compounds present in *T. ferrooxidans* ecological niche.

The genus *Thiobacillus* encompasses a widely varying group of sulphur-oxidising bacteria. It includes both acidophilic and neutrophilic organisms, some of which are capable of both autotrophic and heterotrophic metabolism. The most studied species within this genus is *T. ferrooxidans*. Members of this species are obligate autotrophs capable of growth at acidic pH's of 1.5 - 2.5 (Holt et al. 1994). Some bacteria, such as *T. ferrooxidans* m1, capable of growing at neutral pH have been included in this group, but taxonomic studies based on 16S rDNA sequencing have indicated that these bacteria should be reclassified as a distinct genus separate from *Thiobacillus* (Lane et al. 1992).

T. ferrooxidans is capable of oxidising both ferrous iron as well as sulphur compounds and is of economic importance in the bioleaching of pyrite ores. Recently the role of *T. ferrooxidans* in some biooxidation processes has been reassessed and the bacterium might play a less significant role than was once believed (Rawlings, 1995, Espejo et al. 1995). An industrial drawback in using this bacterium in commercial bioleaching, is its slow growth rate (Rawlings and Silver, 1995). Consequently efforts have been made to study the genes which are responsible for nutrition, such as the fixation of carbon dioxide (Kusano et al. 1991) and nitrogen (Pretorius et al. 1986, Berger et al. 1990, and Kilkenny et al. 1994). Equally important is an understanding of genes encoding proteins involved in the generation of energy such as those of the respiratory chain (Kusano et al. 1992b, Casimiro et al. 1995; Hall et al. 1996, Elbehti and Lemesle-Meunier, 1996) and ATP synthase apparatus (Brown et al. 1994).

1.1 Plasmid studies in the genus *Thiobacillus*

As with most other facets of genetic investigation, in the genus *Thiobacillus,* the majority of endogenous plasmids reported have been found in *T. ferrooxidans*. These plasmids have consequently been analysed in the greatest detail. Thus far most plasmids isolated from members of the genus have been cryptic showing no obvious phenotype such as antibiotic or heavy metal resistance (Rawlings and Kusano, 1994). An exception is plasmid pTF-FC2 discussed in section 1.1.2. There has been circumstantial evidence that

some plasmids may play a role in resistance to heavy metals commonly found in *Thiobacillus* habitats (Holmes et al. 1983).

1.1.1 Plasmids of various *Thiobacillus* species other than *Thiobacillus ferrooxidans*

Several plasmids have been detected in the recently described *Thiobacillus* T3.2 (de Siloniz et al. 1993, Lorenzo et al. 1995). T3.2 is an obligate autotroph capable of oxidising iron, sulphur and related compounds at mesophilic temperatures. It possesses a single long flagellum which renders it highly motile, thus distinguishing it from the more common *T. ferrooxidans* species.

Using pulsed field gel electrophoresis (PFGE) Lorenzo et al. (1995) detected three extrachromosomal elements present in T3.2. The largest ran slightly ahead of the chromosomal DNA and could possibly represent a very large plasmid. The two smaller bands could be distinguished from one another by restriction analysis. The larger of the two was cloned using a number of different restriction endonucleases. It was found to be 15.3 kbp in size and was designated pT3.21 (Lorenzo et al, 1995). Little has been reported on this plasmid except that its physical map is notably different from any other *Thiobacillus* plasmid reported thus far.

Facultatively heterotrophic thiobacilli have also been shown to contain plasmids (Mao et al. 1980). Two plasmids were described in *Thiobacillus acidophilus*. They were designated pAF1 and pAF2 and had estimated molecular weights of 18.36 and 11.95 $x10^6$ Da respectively. At least three bands were also detected in what was reported to be a heterotophic *T. ferrooxidans* strain KG-4 (Mao et al. 1980). Due to heterogeneic colony morphology a single colony was isolated (named TFG-1) and tested for the presence of plasmids. The three plasmid bands were estimated to be 40.06, 21.32 and 2.53 $x10^6$ Da in size and were named pTF1, pTF2 and pTF3 respectively. Subsequent analysis has shed some doubt on the purity of these strains and it is possible that some of the plasmids

reported in these early studies may have originated from a contaminating *Acidophilium* species (Harrison, 1982).

Extrachromosomal DNA has also been observed in the facultatively heterotrophic *Thiobacillus versutus*, formerly known as *Thiobacillus* A2 (Bednarska et al. 1983). The two plasmids reported from this organism had estimated molecular weights of 57 amd 3.5 $x10^6$ Da and were called pTA1 and pTA2 respectively (Bednarska et al. 1983). pTA2 (redesignated pTAV2) was subjected to further analysis and found to be a linear molecule (Wlodarczyk and Nowicka, 1988). At that time the presence of linear plasmids had only been reported in *Streptomyces* (Hayakawa et al. 1979). Evidence for the linearity of pTAV2 was based on electrophoretic characteristics of the plasmid during different lytic procedures. This together with electron microscope measurements showed pTAV2 to be 3.7 kbp in size (Wlodarczyk and Nowicka, 1988). Of interest was the observation that pTAV2 when heated and then reannealed migrated through agarose gels at a size of only half its original 3.7 kbp. The authors speculated that this could be due to internal homology of the single stranded DNA resulting in the formation of a hairpin loop (Wlodarczyk and Nowicka, 1988).

The same group analysed the larger pTA1 (now pTAV1) plasmid from the same bacterium (Bartosik et al. 1995). They estimated that pTAV1 was 107 kbp and like most other *Thiobacillus* plasmids, this plasmid remains cryptic. The cloning of a selectable marker (1.3 kbp kanamycin resistance gene) into the plasmid was used to identify the minimal replicon of pTAV1. Several mini-replicons, which replicated in *T. versutus* were isolated in this way (pTAV1 could not replicate in *E. coli*). The smallest named pTAV202 contained 4 kbp of pTAV1 DNA.

In vitro analysis of the proteins from this region showed that two polypeptides were produced. They were approximately 26 and 45 kDa in size and are thought to be involved in replicative functions (Bartosik et al. 1995). The mini-replicon region could not be stably maintained in *T. versutus* without selection. A 14 kbp region from pTAV1 was

found to confer stability to the mini-replicons, but no further information has been reported on this region.

Davidson and Summers (1983) screened a number of *Thiobacillus* species for plasmids in an attempt to establish a transformation system for introducing recombinant DNA into the organisms. Plasmid DNA was not observed in *T. neapolitanus* or *T. perometabolis*. *T. intermedius* had a plasmid which migrated between the 24.6 and 38 $x10^6$ Da markers. *T. acidophilus* ATCC 27807, used in the study produced at least four plasmid bands on agarose electrophoresis. They were estimated to be greater than 100, 54, 18, 12 $x10^6$ Da in size. This agreed with the independent size estimates of 12 and 18 $x10^6$ Da observed in isolates of this species by Mao et al. (1980).

1.1.2 The plasmids of obligately autotrophic *Thiobacillus ferrooxidans*

The first plasmids isolated from chemolithotrophic strains of *Thiobacillus ferrooxidans* were detected in strains found in uranium mines (Martin et. al. 1981). The results of the study serve to illustrate the great diversity of plasmids found in this species, and have been reproduced in Table 1.1. The authors noted that *T. ferrooxidans* strain TF35, when grown on either iron or tetrathionate contained the same size plasmid (13 MDa). Furthermore when TF35 was cultured continuously on potassium tetrathionate, it did not lose either the plasmid or its ability to oxidise iron. In combination with the fact that some of the strains analysed did not contain any plasmids, it was considered unlikely that the plasmids encoded essential functions, such as the ability to oxidise iron (Martin et.al. 1981).

Further restriction analysis of the plasmids from TF-35 (pTF35O) and TF-29 (pTF291 and pTF292) were undertaken (Dolan and Tuovinen, 1986). The results showed that one of the plasmids (pTF292) from TF-29 was similar to pTF35O. A physical map of the 20

kbp pTF292 plasmid was constructed, and two *Bam*HI fragments (2.9 and 4.1 kbp) were cloned into the pUC8 cloning vector.

Table 1.1 : Plasmid profiles among obligately autotrophic strains of *T. ferrooxidans* demonstrate that the species harbour a diverse range of plasmids (Martin et al. 1981)

Strain	Designation	Size x10^6 kDa
TFI-14	pTFI-140	36
TFI-16	pTFI-160	46
TFI-18	pTFI-180	50
TFI-22	pTFI-220	12
TFI-24	pTFI-240	9.5
	pTFI-241	23
TFI-29	pTFI-290	6.5
	pTFI-291	19
	pTFI-292	31
TFI-30	pTFI-300	4.9
	pTFI-301	7
	pTFI-302	9
	pTFI-303	22
	pTFI-304	36
TFI-32	pTFI-320	5.1
	pTFI-321	8.1
	pTFI-322	13
	pTFI-323	19
TFI-34	pTFI-340	14
	pTFI-341	19
	pTFI-342	33
TFI-35	pTFI-350	13
TFI-41	pTFI-410	6.7
	pTFI-411	10

In an attempt to identify suitable plasmids for use as cloning vectors, Rawlings et. al. (1983) detected plasmids in three of six *T. ferrooxidans* strains screened. Strain FC contained plasmids of 4.5, 12.4 and 27.6 kbp, strain WLR had two plasmids (11.6 and 27.5 kbp respectively) and TF35 contained a single 19.2 kbp plasmid designated pTF35 (Rawlings et. al. 1983). Plasmid pTF35 contained a single *Hin*dIII restriction site and this

was used to clone pTF35 into the *E. coli* vector pBR322 (Rawlings and Woods, 1985). Further analysis of pTF35 showed that it could not replicate in *E. coli,* nor could it be mobilised by IncN type plasmids such as RP4 (Rawlings et al. 1986b).

To investigate whether plasmids from *T. ferrooxidans* ATCC33020 contained elements conferring resistance to toxic metals, Holmes et al. (1983) grew the bacterium in the presence of uranium. The authors intended to investigate whether the presence of plasmids within the strain contributed to uranium resistance. They identified at least three plasmids within *T. ferrooxidans* 33020, with sizes of 6.7, 16 and greater than 23 kbp. The 6.7 kbp plasmid (pTF1) was cloned into an *E. coli* vector and named pTF100. A physical restriction enzyme map was obtained for the plasmid and hybridisation experiments revealed that it had very little homology with the other larger plasmids within the cell. Holmes and co-workers proposed a nomenclature for these plasmids based on the initials of the species followed by a number. Hence the larger plasmids became pTF2 and pTF3 (Holmes et al. 1984). No causal relationship could be established however, between the disappearance of the plasmids and an increased sensitivity to uranium.

Plasmid pTF1 described above was originally cloned into a ColE1 based vector (Holmes et al. 1984). A similar if not identical plasmid was cloned independently by Rawlings and Woods (1985) and named 33020-1. They established that this plasmid could be mobilised between *E. coli* by IncP plasmids. This work was continued by Lau and co-workers and is described later in this section.

Italian researchers have speculated that the genes for mercury and silver resistance may be located on a 19 MDa plasmid found in a number of *T. ferrooxidans* strains isolated near Rome (Visca et.al. 1986). Two plasmids, sized 6 and 24 MDa, were present in all the strains used in this study. In addition to this, strain TF-MS harboured plasmids of 11 and 19 MDa, strain TF-MF 9.5 and 19 MDa plasmids, strain TF-MB a 9.5 MDa plasmid and strain TF-NM1 a 19 MDa plasmid. All of the plasmid-containing *T. ferrooxidans* strains were tested for ampicillin, tetracycline and chloramphenicol resistance, as well as for

heavy metal resistance (arsenite, arsenate, mercury and silver). All of the bacteria showed very similar resistance levels to the three antibiotics and the arsenic-containing compounds. However, strains harbouring the 19 MDa plasmid all had increased levels of mercury and silver resistance. The authors admitted that this was only circumstantial evidence and that further analysis would be required to test their hypothesis (Visca et. al. 1988). Further research by this group into plasmids found in Italian isolates of *T. ferrooxidans* showed that other very different plasmids were also present. Of particular relevance to this thesis (see Chapter 2) was the discovery of a 20 kbp plasmid which was present in strains TFOB, TFOF, TFOS, TFVS and TFMSR. Some of the strains (TFOS, TFOB, TFOF) also contained a second smaller 9.6 kbp plasmid which also appeared to be widespread (Valenti et. al. 1989). The 20 kbp plasmid, named pTFO, was mapped and was found to be nearly identical to plasmid pTF-35 cloned by Rawlings et. al. (1983). Plasmid pTFO seems to have a world-wide distribution, as it was also detected in strains TFNM1 and TFNM3 from New Mexico, and TFP4 from a remote Italian site (Valenti et. al. 1989, and Valenti et. al. 1990).

Further characterisation was carried out on pTFO and a more detailed physical map allowed the cloning of a 2.3 kbp fragment of the plasmid (Polidoro et al. 1993). Total protein extracts from *E. coli* containing this clone as well as total protein from TFMSR were subjected to SDS-PAGE analysis. A highly expressed 38 kDa protein present in both the *E. coli* and *T. ferrooxidans* extracts was found (Polidoro et al. 1993). Attempts to cure plasmid pTFO from TFMSR were unsuccessful. Further characterisation of the plasmid showed that pTFO was unable to replicate in *E. coli*, a property also observed with pTF35 (Rawlings et.al. 1983).

Sanchez and co-workers detected a 9.8 kbp plasmid in a number of *T. ferrooxidans* strains isolated from copper mine waters in Chile (Sanchez et al. 1986). Of particular interest was the wide distribution of the plasmid as the strains were isolated from mines up to 450 kilometre's apart. In a study in Japan over a hundred independent strains of *T. ferrooxidans* were isolated from six different Japanese mines (Shiratori et al. 1991).

Seventy three percent of the strains isolated contained plasmids in the size range of over 2 kbp. Southern hybridisation analysis showed varying degree's of homology between these plasmids. It is evident that not only are some plasmids widely distributed amongst *T. ferrooxidans* strains (Sanchez et al. 1986, Valenti et al. 1990, Pramila et al. 1996), but also some share common DNA sequences (Shiratori et al. 1991). Comparative analysis of a number of plasmids isolated from geographically distinct locations was undertaken (Chakravarty et. al. 1995). Using restriction analysis the authors proposed that there was at least one distinct family of plasmids found within the *T. ferrooxidans* species. This distinct family was classified on the structural features of its replicon. At least part of the replicon was located on a 2.2 kbp *Sac*I restriction fragment which cross-hybridised to a number of plasmids of different sizes. Southern hybridisation analysis indicated that the DNA flanking the replicon had significant, but less well conserved homology (Chakravarty et. al. 1995). These features appeared to be present in the another two members of the TFI91 plasmid family. The other plasmids were pTF35I a 15 kbp plasmid from strain TF35, and pDSM583 a 20.2 kbp plasmid from strain DSM 583 (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH).

The 2.2 kbp *Sac*I fragment from a 9.8 kbp plasmid designated pTFI91, was cloned and sequenced. The sequence had features of a plasmid origin of replication as well as an insertion sequence named IS3091. The origin of replication was divided into three regions, based on the structural features contained within each segment. Region I had nine tandemly arranged direct repeats of up to 33 bp in length. Region II contained a DnaA binding site as well as an integration host factor (IHF) binding site followed by a series of shorter 10 bp direct repeats. This whole region was characterised by the low GC mol % of its sequence. Region III has an unusually high GC mol % and did not contain any specific secondary structure (Chakravarty et al. 1995).

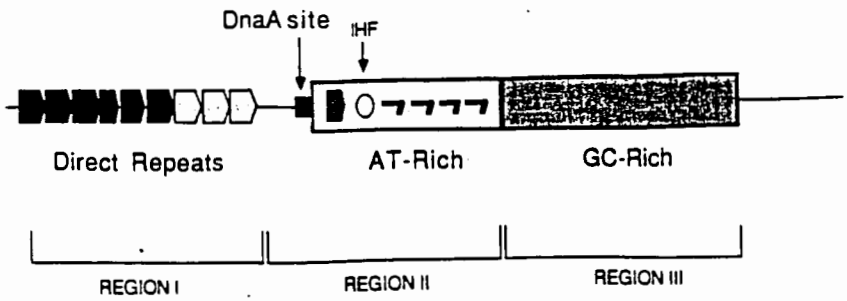


Figure 1.1 : The origin of vegetative replication (*oriV*) of plasmid pTFI91 (Chakravarty et al. 1995). Region I shows the nine iterons of the *oriV*, region II contained the DnaA and IHF host-protein binding sites, whilst region III displayed an unusually high G+C mol%.

The two most extensively studied plasmids from *T. ferrooxidans* are pTF1 and pTF-FC2. Plasmid pTF1 from *T. ferrooxidans* ATCC33020 (Holmes et al. 1984) was analysed further with a view to developing it as a vector for *T. ferrooxidans* (Lau et al. 1989). The mobilisation region of pTF1 was isolated on a 2.8 kbp DNA fragment (Figure 1.2). The region contained potentially 6 ORFs coding for polypeptides ranging from 75 to 328 aa. Two of the ORFs showed similarity to the mobilisation genes from RSF1010, and were called *mobL* and *mobS* respectively (Drolet and Lau, 1992). Due to ORFs 3, 4, and 5 overlapping *mobL*, it is unlikely that they could encode functional products. A 38 bp putative palindromic sequence was identified as a possible origin of plasmid transfer (*oriT*) (Drolet et al. 1990). The site of nicking was determined for the pTF1 *oriT* and found to be within the intergenic region between the two divergently transcribed *mob* genes. The nick site was adjacent to an inverted repeat sequence, a phenomenon which also occurs in RSF1010 (Drolet et al. 1990). Gel retardation assays showed that the 28 kDa MobL protein bound to a 42-mer single stranded oligonucleotide, which represented the nicked T-DNA of plasmid pTF1 (Drolet and Lau, 1992). It was suggested that the

MobS gene, which did not appear to bind to the DNA, aided mobilization by directly or indirectly binding to MobL during transfer.

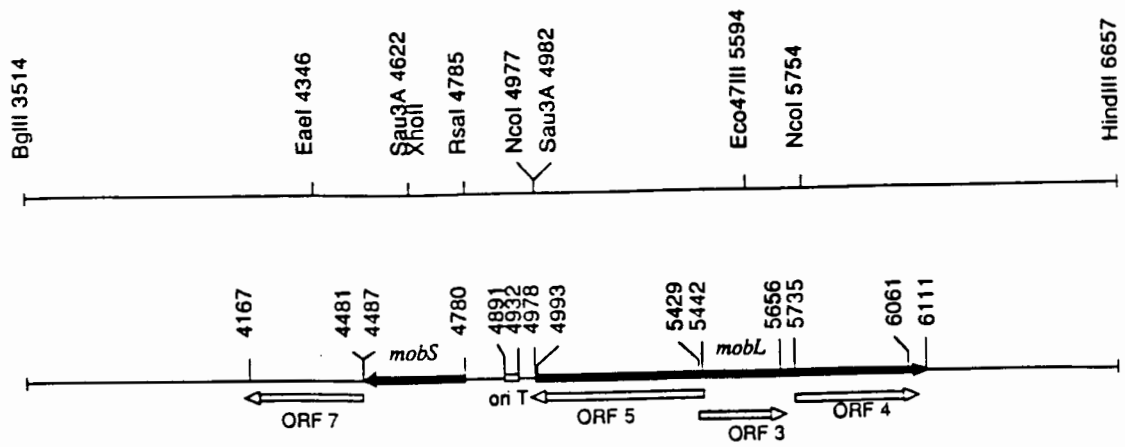


Figure 1.2 : A physical map of the mob region of pTF1 showing the *oriT, mobS* and *mobL* genes (Drolet and Lau, 1992). Proteins corresponding to ORFs 3, 4, 5 and 7 have not been reported.

A common feature amongst the plasmids reviewed thus far is that they are all cryptic and replicate exclusively in their host organisms. One exception is plasmid pTF-FC2 from *T. ferrooxidans* FC1, a strain isolated from an inoculum that was being adapted for use in a South African bioleaching reactor. Plasmid pTF-FC2 was sequenced and found to consist of three distinct regions, a region responsible for replication, a second region involved in mobilization and a third region which contained a Tn*21*-like transposon (Rawlings and Kusano, 1994). The region responsible for replication was subjected to further analysis (Dorrington and Rawlings, 1989,1990). The copy number of the plasmid was estimated to be twelve to fifteen per chromosome. A series of 3x22 bp inverted repeats responsible for plasmid incompatibility and copy number control were reported. Immediately adjacent to these repeats was the *oriV* which was 75% homologous to the origin of vegetative replication of the IncQ plasmids (Dorrington and Rawlings, 1989).

In vitro transcription-translation analysis of the proteins involved in the replication revealed the presence of a 32 kDa RepA-like protein and a 33 kDa RepC-like protein. In IncQ plasmids, the RepA protein acts as a DNA-dependant ATPase whilst the RepC product is responsible for initiating replication (Sherzinger et al. 1984). The pTF-FC2 RepA and RepC proteins shared 43% and 60% amino acid identity with the IncQ RepA and RepC homologues respectively. Upstream of this operon lie three open reading frames designated ORFs 3,4,5 whose functions were thought to be involved in plasmid stability (Dorrington and Rawlings, 1990). Upstream of these ORFs lies a gene with 23% aa identity to the IncQ RepB primase protein. *In vitro* transcription-translation analysis showed that the *repB* gene produced a 40 kDa protein which was shown to function as a pTF-FC2 *oriV* specific primase (Dorrington et al. 1991).

The second region of pTF-FC2 was found to be involved in the mobilization of the plasmid (Rohrer and Rawlings, 1992). The region responsible for mobilization was contained within a 3.5 kbp fragment (Figure 1.3). In contrast with the IncQ-like replication region, this region had similarity to IncP-type plasmids. The proteins involved in mobilization designated MobA - MobE had between 24% and 33% aa identity to the homologous Tra proteins in the RP4 plasmid. Clones containing the *mob* fragment were able to be mobilised from donor to recipient *E. coli* cells by RP4 (Rohrer and Rawlings, 1992).

The third region of pTF-FC2 was found to contain a Tn*21*-like transposon element designated Tn*5467* (Clennel et al. 1995). The transposon is bordered by 38 bp inverted repeats which were identical to those from transposon Tn*21*. Of interest was the finding that the borders flanked open reading frames for a redox-active glutaredoxin-like protein, as well as a protein with sequence similarity to the regulatory protein (MerR) encoded by the mercury resistance operon. In addition to this, an open reading frame was identified

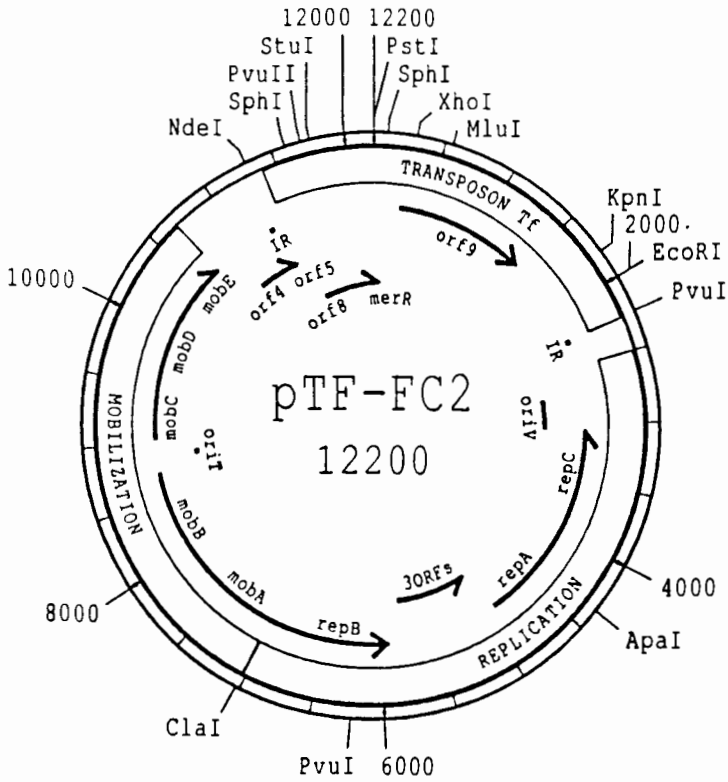


Figure 1.3 : Physical map of plasmid pTF-FC2 showing the mobilization, replication and the transposon - like regions (Rawlings et al. 1993).

which had homology to multi-drug resistance membrane transport proteins. The glutaredoxin-like open reading frame, could complement an *E. coli* thioredoxin mutant for several thioredoxin dependant functions. Furthermore, sequences homologous to the Tn*21* resolvase and transposase was discovered, but these open reading frames were truncated. It was found that the transposon was inactive in pTF-FC2 unless complemented *in trans* with the genes for the Tn*21* transposase and resolvase (Clennel et al. 1995). Since this region represents a glutaredoxin-complementing phenotype which can be selected for in the appropriate *E. coli* mutant, pTF-FC2 can no longer be considered to be a cryptic plasmid. However, the role of these three ORFs within the host organism, remains to be determined.

Plasmid pTF-FC2 was an ideal candidate for use as a shuttle vector between heterotrophic bacteria and the autotrophic *T. ferrooxidans* (Rawlings et al. 1984a). The IncP plasmid, RP4 can replicate in *T. novellus* (Kulpa et al. 1983) and pTF-FC2 can replicate in both *T. novellus* as well as *E. coli*. This could allow for the development of a two-step transformation process, whereby recombinant DNA cloned in pTF-FC2 could be transferred from *E. coli* to *T. ferrooxidans* via a *T. novellus* intermediate (Rawlings et al. 1984b). Attempts to use such a two-step conjugation system for introducing DNA into *T. ferrooxidans* have been unsuccessful.

1.2 Insertion sequences from *T. ferrooxidans*

To date the only insertion elements described for the *Thiobacillus* genus have come from *Thiobacillus ferrooxidans*. Two families of repeated DNA sequences have been reported from strain ATCC19859 (Yates and Holmes, 1987). Both families were approximately 1 kbp in size and were present in up to 25 copies in the genome. Six copies of Family 1 were found on a 45 kbp plasmid in strain ATCC19859. This resulted in an estimated 15 fold higher repeated sequence density on the plasmid than on the chromosome (Yates and Holmes, 1987).

Family 2 was the first to be sequenced (Yates et al. 1988). It was a 1408 bp insertion sequence element and was therefore renamed IST2. It had imperfectly conserved 25 bp terminal inverted repeats, which were flanked by 9 bp target site duplications (Yates et al. 1988). IST2 contained three open reading frames, the largest of which was 888 bp in length and encoded a probable transposase gene. Two copies of the IST2 elements were sequenced and were shown to be identical, although they had inserted into different regions of the chromosome. IST2 appeared not to exhibit target site specificity (Yates et al. 1988).

IST2 was shown to be active within the *T. ferrooxidans* cell (Cadiz et al. 1994) and it was suggested that it might be involved in genomic rearrangements in the *T. ferrooxidans* chromosome. Differential colony morphologies were observed among *T. ferrooxidans* growing on solid media (Schrader and Holmes, 1988), which led investigators to speculate that IST2 may be involved in a phenotypic switching event. This was based on observations that, depending on the media used to propagate the cells, at least two different restriction banding patterns were observed following Southern hybridisation using IST2 as a probe. This implied that IST2 was responsible for the phenotypic switch which might possibly induce a different metabolic pathway involved in oxidation within the cell. However it was observed that when the cells were transferred onto the initial media, not all of the restriction bands reverted to the initial pattern, suggesting that the switching event does not proceed in a controlled fashion. The "phenotypic switching " event deserves further investigation before the precise role of insertion elements may be determined in this process. Furthermore, although the results of this investigation demonstrated clear evidence of transposition, recombinational activity could not be excluded (Cadiz et al. 1994).

The nucleotide sequence of the family 1 repeated sequence was determined and renamed IST1 due to its identification as a 1300 bp insertion element (Zhao and Holmes, 1993). It consisted of two major open reading frames, with 26 bp imperfect terminal repeats and a 5 bp target site duplication sequence. There was little homology between IST1 and IST2, and very little similarity with other insertion elements. IST1 was also found to undergo positional changes within the *Thiobacillus* genome. Mobility of IST1 was observed amongst cells which had been adapted to growth in toxic concentrations of Cu^{2+} (Holmes and Ul Haq, 1989). Unlike IST2, IST1 inserted reversibly into a specific *Bam*HI genomic fragment (Zhao and Holmes, 1993). This insertion resulted in a concomitant loss of iron oxidising ability in a manner similar to IST2. The region of *T. ferrooxidans* chromosome which flanks the insertion site was sequenced. A 411 bp partial open reading frame was found, but it had no significant homology to any other genes, and its role (if any) in iron

oxidation remains to be determined (Zhao and Holmes, 1993). It was proposed, that insertion elements aid the bacterium in rapidly adapting to new environments.

1.3 Development of genetic tools for the genus *Thiobacillus*.

The development of reliable genetic systems in the thiobacilli will greatly facilitate the understanding of basic metabolism within this genus. A number of factors need to be addressed in order to develop these systems. The generation and propagation of mutants in the genus is an essential preliminary step in the understanding of genetic regulation of these bacteria. Reliable methods of introducing recombinant DNA into the cells are required, along with the expression of selectable markers which can reliably distinguish transformed thiobacilli from untransformed cells. This section serves to outline the progress made in developing these tools in both facultatively autotrophic thiobacilli as well as in the autotrophic *T. ferrooxidans* and *T. thiooxidans*.

1.3.1 Transformation and conjugation in the facultative thiobacilli

The transformation of facultatively autotrophic members of the *Thiobacillus* genus is somewhat easier due to their neutrophilic and heterotrophic growth. *Thiobacillus thioparus* has been shown to be naturally transformable. Amino acid auxotrophs were transformed to prototrophy by an unknown competency system using mechanically sheared chromosomal DNA (Yankofsky et al. 1983). Conjugation between bacteria requires cell to cell contact in order for DNA transfer to occur. In heterotrophic bacteria conjugation has been readily accomplished due to the presence of a large number of auxotophic markers being available in these systems (Ippen-Ihler, 1989). The broad host-range plasmid RP1 has been conjugated into *Thiobacillus neapolitanus* at frequencies comparable to those of *E. coli* and *P. aeruginosa* (Kulpa et al. 1983).

Thiobacillus novellus has also been shown to be capable of conjugation with heterotrophic bacteria (Davidson and Summers, 1983). A number of plasmids, notably

RP4 and its derivatives, were conjugable between *E. coli* and *T. novellus* at neutral pH. Furthermore most of the antibiotic selectable markers used in the study were expressed in *T. novellus*. Similar experiments between *E. coli* and *T. intermedius, T. perometabolis, T. neapolitanus* and *T. acidophilus* failed to produce viable transconjugants. However, most of the above thiobacilli (*T. acidophilus* being the exception) could be conjugated with *T. novellus* under acidophilic conditions. This observation led to the development of the two step conjugation process between *E. coli, T. novellus* and the obligately autotrophic *T. ferrooxidans* discussed in section 1.3.2.

Isoleucine and lysine auxotrophic mutants of *T. novellus* and *T. versutus* have been generated by transpositional mutagenesis techniques using plasmid incompatibility (Davidson et al. 1985). In this experiment Davidson and co-workers introduced a transposon into either *T. novellus* or *T. versutus* by introducing it into the cell on a plasmid. A second plasmid was present in the bacterium which was of the same incompatibility group as the plasmid-containing the transposon. By selecting for the second plasmid and the marker on the transposon, the transposon was forced to integrate into the genome of the host cell. Transposon Tn*501* carried on plasmid RP4 was used to generate *T. novellus* mutants deficient in the oxidation of reduced sulphur compounds such as elemental sulphur, thiosulphate and tetrathionate as well as the ability to fix carbon dioxide. Transposon Tn*1721* insertions into the *T. versutus* genome, provided mutants with sensitivity to arsenate and gentamycin. Auxotrophs for isoleucine-valine, arginine, phenylalanine, valine and panthothenate were also isolated (Davidson et al. 1985).

1.3.2 Conjugation and electroporation in *T. ferrooxidans* and *T. thiooxidans*

Initial attempts to conjugate foreign DNA into *T. ferrooxidans* in a two step process were made, using *T. novellus,* as an intermediate host (Davidson and Summers, 1983). Recombinant DNA was transferred from genetically well characterised systems such as

E. coli to *T. novellus* in a heterotrophic (and well defined) environment. Once the plasmid DNA was resident in the *T. novellus*, conjugation conditions could be altered to suit the autotrophic acidic environment of *T. ferrooxidans*. The plasmid DNA would then be conjugated for a second time from the *T. novellus* donor to the new recipient, *T. ferrooxidans*. This approach was unsuccessful due to the inability to distinguish the *T. ferrooxidans* transconjugant from the donor *T. novellus* cell.

Due to the absence of suitable selectable markers, broad host-range plasmids have been employed to bring about a direct one-step conjugation in these obligate acidophiles. Broad host range IncP plasmids RP4, R68.45, RP1::Tn*501* and pUB307 were used in conjugation experiments in both *T. thiooxidans* and *T. ferrooxidans* (Jin et al. 1992, Peng et al. 1994). In both of these cases *Escherichia coli* was used as the donor organism and the mating media was adjusted to pH levels of 4.8. Yeast extract (0.05% (w/v)) was added to support the donor *E. coli* cells. The kanamycin and streptomycin resistance markers used in the *T. ferrooxidans* experiment were sufficiently stable at pH 4.8 to select against non-transconjugants (Peng et al. 1994). In the *T. thiooxidans* study the RP4 plasmid was able to express kanamycin and tetracycline resistance in the recipient cell, whilst the ampicillin resistance marker was inactive. The RP4 plasmid could also be conjugated in the reverse direction from *T. thiooxidans* to *E. coli*, a technique which would greatly improve the genetic investigation of *T. thiooxidans* (Jin et al. 1992).

T. ferrooxidans when oxidising iron, cannot grow at pH levels over pH 2.5 due in part to the fact that rusticyanin a major component in the iron oxidising pathway is unstable above that level (Ingledew and Cobley, 1980). This requires the use of factors such as heavy metal resistance as selection markers at low pH levels. Strains of *T. ferrooxidans* have been identified with increased levels of resistance to uranium (Holmes et al. 1983), arsenic (Rawlings et al. 1984b, Kondratyeva et al. 1995), zinc (Kondratyeva et al. 1995) and mercury (Inoue et al. 1991). Mercury resistance has been the only heavy metal marker to be successfully tested in *T. ferrooxidans* (Kusano et al. 1992).

There has been one report of successful electroporation of *T. ferrooxidans* (Kusano et al. 1992). Appropriate vectors were constructed which expressed the mercury resistance operon (*mer*) cloned from *T. ferrooxidans* E-15. These vectors either consisted of endogenous cryptic plasmids from *T. ferrooxidans* or the broad host-range IncQ pKT240 plasmid. The potential shuttle vectors were mixed with plasmid-free, mercury sensitive *T. ferrooxidans* cells and were electroporated into the cells. Of the thirty *T. ferrooxidans* strains tested, only a single strain, Y4-3 produced viable progeny with four of the five plasmids used. Cell viability was estimated to be 55 to 67% and the transformation efficiency was between 120-200 CFU/μg plasmid DNA (Kusano et al. 1992). The authors admitted however, that the use of mercury as a selectable marker may be limited due to the small differences in the tolerance levels between the wild type sensitive strains and the transformed cells (Kusano et al. 1992).

So far the review has focused on plasmids and the challenges faced in developing a reliable genetic system in the thiobacilli. This is a fundamental development which must be mastered before molecular biologists can gain a better understanding of the metabolic processes which govern the growth and replication of these economically valuable bacteria.

1.4. The molecular biology of redox active proteins and their genes in the thiobacilli

The rest of the review will focus on the biochemistry and molecular biology of the redox active proteins of the genus *Thiobacillus* and related bacteria. Redox-active proteins are enzymes involved in the transport of electrons within the cell and the review will concentrate the central pathways for the oxidation of reduced iron and sulphur compounds, as well as the role of selected heavy metals in electron transport within the

cell. Due to the status of current knowledge in the field, emphasis will be focused on the redox-active proteins of *T. ferrooxidans.*

1.4.1 Ferrous iron respiratory pathways

The ability to grow by the aerobic respiration of ferrous iron occurs throughout the eubacteria and archaea (Blake et al. 1993). So far at least four different mechanisms have been identified based on the differing novel cytochromes used in each. That as many as four distinct mechanisms exist probably indicates that the ability to oxidise ferrous iron has evolved independently on more than one occasion. The best known iron oxidation system is the blue rusticyanin-based oxidation system of *Thiobacillus ferrooxidans*. A variant of rusticyanin has also been detected in strain M-1, a bacterium which cannot oxidise sulphur compounds and should be placed in a separate genus to the other thiobacilli. Based on 16S rRNA analysis both were found to be clearly different purple proteobacteria (Lane et al. 1992). A second iron-oxidation mechanism is found in a group of moderately thermophilic Gram positive bacteria represented by *Sulfobacillus thermosulfidooxidans.* These bacteria contain an unusual acid stable yellow chromophore thought to be involved in ferrous iron oxidation (Blake et al. 1993). The oxidation of ferrous iron is a general property of this group which is thought to comprise a number of separate bacterial genera. *Leptospirillum ferrooxidans* is a representative of bacteria which possess a third type of iron oxidation system. These bacteria contain a deep red cytochrome of 17.9 kDa in strain DSM2705 and 12 kDa in strain P3A (Blake et al. 1993). This indicates that there may be at least two different iron oxidation pathways in the *L. ferrooxidans* species. Certain *Archaea,* such as *Metallosphaera sedula* also oxidise ferrous iron. These extremely thermophilic *archaea* produce a novel membrane bound yellow chromophore which is distinct from the BC-1 soluble cytochrome (Blake et al. 1993). Only the well described rusticyanin based mechanism of *Thiobacillus ferrooxidans* will be reviewed, as it is the only mechanism that has been studied in detail.

1.4.2 Iron oxidation pathway of *Thiobacillus ferrooxidans*

T. ferrooxidans oxidises ferrous iron to provide a source of electrons, for its metabolic processes (Holt et al. 1994). As early as 1963, researchers isolated an iron-cytochrome *c* reductase from *T. ferrooxidans* (Blaylock and Nason, 1963). The enzyme was active and readily oxidised ferrous iron, donating the electron to a horse ferricytochrome *c*. The researchers also detected endogenous cytochromes (*a, c* and *b*) within *T. ferrooxidans* and proposed an electron transporting pathway based on the observation that all of the cytochromes were able to oxidise ferrous iron (Blaylock and Nason, 1963). Previously a cytochrome *a* molecule had been identified as an *a*1 type, in a bacterium thought to be *T. ferrooxidans* (Vernon et al. 1960). This allowed researchers to propose a simple electron transfer chain as is shown below.

Fe^{2+} ⟶ cytochrome *c* ⟶ cytochrome *a* ⟶ O_2

The blue copper protein, rusticyanin, was the next enzyme from *T. ferrooxidans* identified as playing an important part in the oxidation of iron (Cox and Boxer, 1978). These authors found that the rusticyanin protein was 16.5 kDa in size and had approximately one copper atom per protein molecule. They proposed that rusticyanin was the initial electron acceptor in the iron respiratory pathway. Optical studies on the respiratory chain components showed the presence of two *a1*-type cytochromes, a number of *c*-type cytochromes, and two *b*-type cytochromes (Ingledew and Cobley, 1980). Using this information, in conjunction with the discovery of a ferredoxin-like component in the respiratory complex, a model for the pathway was further refined as shown.

Fe^{2+} ⟶ rusticyanin ⟶ cyt *c*(2) ⟶ cyt *c* ⟶ cyt *a*1*a*1 ⟶ O_2

This early biochemistry provided a framework for the analysis of individual proteins of the respiratory pathway. The following section deals with each component of the pathway as it is understood at present.

In search of the primary ferrous ion electron acceptor

In 1986 a new component of the respiratory chain was identified (Fry et al. 1986). Using electron paramagnetic resonance spectroscopy it was shown that an enzyme from *T. ferrooxidans* was able to reduce rusticyanin, and was responsible for the direct oxidation of ferrous iron. It had the spectroscopic characteristics of a 4Fe-xS moiety and it was proposed that this group constituted the redox-active site of an iron-oxidoreductase enzyme (Fry et al. 1986).

Expression studies using *T. ferrooxidans* ATCC13661 showed that when the bacterium was transferred from sulphur to a ferrous iron medium, three proteins were induced (Mjoli and Kulpa, 1988). They were the 16.5 kD rusticyanin protein, a 32 kD protein and a 92 kD glycoprotein. The 92 kD glycoprotein was shown to be an integral membrane protein which was thought to function as a porin protein. The oxidation of iron was not induced when the glycosylation of the protein was blocked. One could speculate that the 92 kD porin protein could well be the 4Fe-xS-containing protein identified by Fry and co-workers (Mjoli and Kulpa, 1988).

In addition to the above finding, Blake et al. (1988) provided evidence for the existence of a cytochrome *c* protein which was responsible for the initial transfer of electrons from ferrous iron to rusticyanin. Fukumori et al. (1988) purified an iron-oxidising enzyme which had a molecular weight of 63 kDa and was responsible for reducing cytochrome *c*552 in the presence of ferrous iron. This was, therefore, a second candidate as the primary electron acceptor in the iron respiratory pathway. A similar protein has been described from a French isolate of *T. ferrooxidans* (Cavazza et al. 1995, and Cavazza and Bruschi, 1995).

The gene for another iron(II) oxidase was cloned by reverse genetics using the amino acid sequence of the N-terminal portion of the purified enzyme. (Kusano et al. 1992). The predicted protein of this iron oxidase *(iro)* gene showed homology to a number of bacterial ferredoxins usually referred to by the acronym HiPIP (High-redox potential iron-sulphur proteins). The *iro* gene encodes a 273 base pair open reading frame which would code for a 90 aa protein. Thirty seven aa were predicted to function as a signal sequence for membrane translocation. The remaining 53 aa protein (approximately 6.2 kDa) must, therefore, be a subunit of a homomultimeric protein to produce the 63 kDa active complex (Kusano et al. 1992). The fact that the ferredoxin-like subunit aggregated into a multimer was unusual in this class of protein, having only been observed in one other case (Meyer et al. 1990). Northern hybridisation showed that the *iro* gene is not part of an operon, producing two transcripts of 450 and 600 bp in size (Kusano et al. 1992).

Another potential candidate for the primary electron acceptor of ferrous iron was the 19 kD protein iron: rusticyanin oxidoreductase (Blake and Shute, 1994). It was noted that this protein contains a cytochrome *c* type moiety which was retained by the partially pure protein. However Yamanaka and Fukumori (1995) have criticised this finding and suggested that the protein having being purified with detergents, could have introduced one of the membrane bound cytochrome *c* molecules *c*552(m) or *c*550(m) in the preparation which may account for the oxidation of ferrous iron. Nevertheless, the original researchers remain convinced that this represents a cytochrome *c* type molecule with iron: rusticyanin oxidoreductase properties (Casimiro et al. 1995).

The *c*-type cytochromes

It was noticed that the Fe(II) oxidase enzyme could not directly reduce rusticyanin unless trace amounts of cytochrome *c*552(s) were present (Fukumori et al. 1988). This soluble cytochrome along with its membrane bound homologues *c*552(m) and *c*550(m) have all been detected in cells which are actively oxidising ferrous iron (Mansch and Sand, 1992).

The three cytochrome *c* molecules are acid stable with molecular weights of 60, 30, and 25 kDa. Of interest was the fact that these components appeared to be membrane associated, and only became soluble on acidification (Mansch and Sand, 1992). The two membrane-bound cytochromes were purified and characterised independently (Tamegai et al. 1994, and Valkova-Valchanova and Chan, 1994). Cytochromes *c*-552(m) and *c*-550(m) were 22.3 kDa and 51 kDa in size and were able to reduce the cytochrome *c* oxidase of *T. ferrooxidans* Fe-1 (Tamegai et al. 1994). In comparing the reaction rates of the three cytochromes it was noted that the order of reactivity with the oxidase was unequal, and the three could be ranked as follows

cytochrome *c*-550(m) > cytochrome *c*-552(s) >>> cytochrome *c*-552(m)

It was interesting to note the simultaneous isolation of two different membrane-bound cytochrome *c* molecules from *T. ferrooxidans* ATCC13661 (Valkova-Valchanova and Chan, 1994). These two cytochromes had molecular weights of 21 and 68 kDa respectively and suggests that different strains of bacteria classified as *T. ferrooxidans* may well have slightly different respiratory iron oxidation pathways.

The discrepancy between these two results has been addressed (Yamanaka and Fukumori, 1995). Cytochrome *c*-552 (21kDa) has a slightly different aa composition to the 22.3 kDa *c*-552(m) cytochrome, but nevertheless may well have an analogous function. A similar *c*552 cytochrome has been described from a French isolate of *T. ferrooxidans* (Cavazza et al. 1995, and Cavazza and Bruschi, 1995). However, the discrepancy in size between the 51 kDa *c*-550(m) cytochrome of Fe-1 and the 68 kDa counterpart from ATCC13661 indicates that they may be two very different proteins.

Rusticyanin

Rusticyanin is a blue copper protein isolated from a number of strains of *T. ferrooxidans* (Blake et al. 1991). The 16.5 kDa protein has homology to a number of other copper-

containing proteins such as amicyanin from *Pseudomonas* AM1 and plastocyanin from *Pseudomonas nigra var italica* (Ronk et al. 1991). However it does differ from these proteins in at least two ways. Firstly, the enzyme is active at pH2, which although unusual for this class of enzyme, is not surprising considering its periplasmic location in the cell. Secondly although it contains just one copper atom per molecule, it nevertheless exhibits a high redox potential of +680 mV (Ingledew and Cobley, 1980). Other members of the blue copper protein class range in redox potential from +184 mV to +370 mV (Ronk et al. 1991). Chou-Fasman and Garnier-Robson plots have predicted that the rusticyanin protein is composed of an 21 aa α helix followed by six β sheets to form what is known as a β barrel structure. The authors predicted that the copper ligand was bound to the rusticyanin molecule at residues Cys 138, His 143, Met 148 and Asp 73 (Ronk et al. 1991).

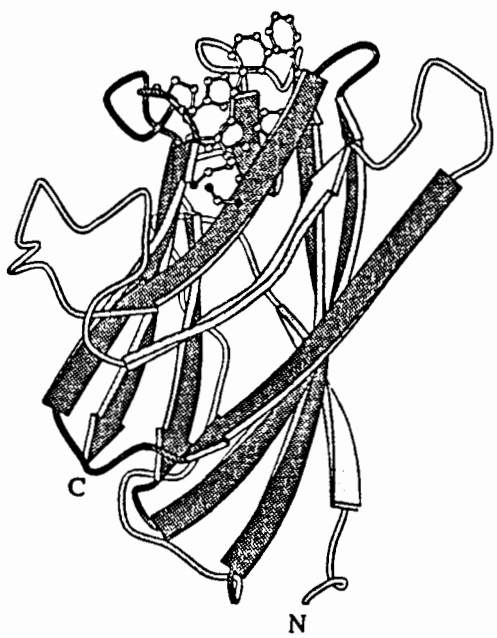


Figure 1.4 : Proposed energy-minimised model for the rusticyanin protein (Grossmann et al. 1995)

Three rusticyanin proteins have been sequenced, all of which show a high degree of homology to each other (> 90% aa identity) (Grossmann et al. 1995, Cavazza et al. 1995,

and Cavazza and Bruschi, 1995). X-ray absorption analysis of two of the variants disagreed with the β barrel model proposed by Ronk et al. (1991). The authors felt that in the light of new copper proteins being characterised, a β barrel model could not adequately explain the high redox potential of rusticyanin (Grossmann et al. 1995).

All agreed that three of the amino acids in rusticyanin which had previously been predicted to bind to the copper ligand (Cys 138, His 143 and Met 148) were correct. Based on the X-ray absorption study, the fourth amino acid binding position was predicted to be the His 85 residue. This agreed with the analysis of Nunzi et al. (1993) and Yano et al. (1991). An increased distance between the copper ligand in its oxidised (1.99 Å) and reduced (2.07 Å) forms was predicted, which would account for the unusually high redox potential of the protein. An energy-minimised model was presented based on the folding pattern of domain I of nitrite reductase as is shown in Figure 1.4 (Grossmann et al. 1995).

In order to ensure high levels of expressed rusticyanin, an artificial gene was constructed in *E. coli* with eight oligomers designed using the *E. coli* codon usage patterns (Casimiro et al. 1995). The gene was synthesised using a "one pot" PCR protocol, and the expressed synthetic rusticyanin protein was spectroscopically identical to the native *T. ferrooxidans* form. Site directed mutagenesis experiments using this artificial gene produced convincing evidence that His 85 was indeed the fourth ligand site to the copper molecule (Casimiro et al. 1995). The endogenous rusticyanin gene from *T. ferrooxidans* ATCC 23270 has recently been cloned and sequenced (Hall et al. 1996).

Cytochrome *c* oxidase

The terminal oxidase in the ferrous iron oxidation pathway is believed to be the cytochrome *c* oxidase characterised by Kai et al. (1989, 1992). The protein had an absorption peak at 595 nm and was originally classified as an *a*1 cytochrome (Vernon et

suggested that the molecular weight of the oxidase was 169 kD which would require that more than one copy of some of the polypeptides be present in the final protein complex. It is of interest that this cytochrome *c* oxidase has identical properties to an independently described molybdenum oxidase from another strain of *T. ferrooxidans* AP19-3 (Sugio et al. 1992a). Similarly other cytochrome oxidases from *Nitrobacter winogradskyi* and *Pseudomonas aeruginosa* were able to oxidise molybdenum (Fujiwara et al. 1992). It is quite possible therefore, that the *T. ferrooxidans* cytochrome *c* oxidase may be involved in other redox pathways (Yamanaka and Fukumori, 1995).

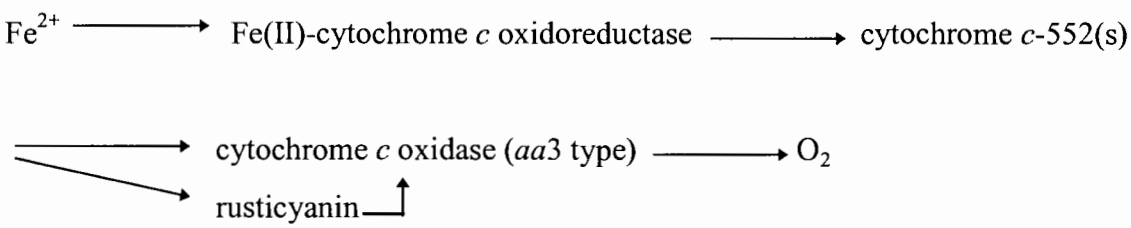
1.4.3 Towards a unified ferrous iron respiratory pathway in *T. ferrooxidans*

Periplasmic components of the ferrous iron respiratory pathway have to be acid stable proteins as they are required to function in the low pH milieu of the periplasm (Yamanaka et al. 1993, and Yamanaka and Fukumori, 1995). Possible components of the pathway which have been identified thus far are compared in Table 1.2.

Table 1.2 : Components of the iron (II) oxidation pathway of *T. ferrooxidans*. The components have been grouped into a) potential initial oxidation components b) cytochromes c) rusticyanin d) terminal oxidase.

Name	Size	Strain	Source
a)			
Iron (II) oxidase	63 kDa	Fe-1	Blake et al.(1988)
Iron : Rusticyanin oxidoreductase	19 kDa	ATCC23270	Blake and Shute, (1994)
Porin protein	92 kDa	ATCC13661	Mjoli and Kulpa, (1988)
Smaller protein	32 kDa	ATCC13661	Mjoli and Kulpa, (1988)
b)			
Cytochrome c-550(m)	51 kDa	Fe-1	Tamegai et al. (1994)
Cytochrome c-550	68 kDa	ATCC13661	Valcova-Valchanova and Chan, (1995)
Cytochrome c-552(m)	22.3 kDa	Fe-1	Tamegai et al. (1994)
Cytochrome c-552(s)	22.3 kDa	Fe-1	Tamegai et al. (1994)
Cytochrome c-552	21 kDa	ATCC13661	Valcova-Valchanova and Chan, (1995)
Cytochromes c-type	21,30,46 kDa	unknown	Elbehti et al. (1996)
c)			
Rusticyanin	16.5 kDa	Many	Cox and Boxer, (1978)
d)			
Cytochrome aa3 oxidase	169 kDa	AP 19-3	Kai et al. (1992)

On this basis Yamanaka and Fukumori (1995) proposed the following ferrous iron oxidation pathway in *T. ferrooxidans*.



It should be noted from the above pathway that assuming the iron: rusticyanin oxidoreductase described by Blake and Shute (1994) is not the same as the Fe-

cytochrome *c* oxidoreductase, the two complexes may well function in tandem. Furthermore, there is evidence that the three cytochromes *c*-552(s), *c*-552(m) and *c*-550(m) are all capable of oxidising the Fe-cytochrome oxidoreductase to varying degrees and could possibly all substitute for this function. Rusticyanin does not play a key role in this pathway, but rather serves to expand the pathway by providing an alternate means of supplying the cytochrome c oxidase with electrons. A further possible role is the protection of the iron-cytochrome oxidoreductase molecule from destruction by the cytochrome c-552(s) (Yamanaka and Fukumori, 1995).

In addition to this the role of the porin protein described by Mjoli and Kulpa (1988) has not been defined in the above pathway and should be investigated further. Future work may show that the pathway is not linear in nature and that under varying conditions different proteins may play important roles in the oxidation of iron. It must also be remembered that different strains of *T. ferrooxidans* have shown variation in some of the components of the respiratory pathway implying that slightly differing mechanisms may operate in these bacteria and this may be confusing our understanding of the pathway as a whole.

1.4.4 Sulphur oxidation pathway

A feature of *Thiobacillus ferrooxidans* is its ability to oxidise both reduced iron and sulphur compounds. This gives it an advantage in bioleaching systems where it can attack pyrite (FeS_2) ores without necessarily requiring another bacterium to oxidise either one or the other component of the ore. The oxidation of reduced sulphur provides more energy for growth than the equivalent iron mechanism (Pronk et al. 1990).

Enzymes of the sulphur oxidation pathway

Sugio et al. (1985) isolated a ferric ion-reducing enzyme (FIR) from *Thiobacillus ferrooxidans* AP 19-3. They proposed that it was involved in the oxidation of reduced sulphur compounds as is outlined in Figure 1.5. The FIR functions by cycling ferric ions back to ferrous ions and has coupled this process to the concomitant oxidation of sulphur (Sugio et al. 1985). In this way, a link was established between the sulphur and ferrous iron oxidation pathways.

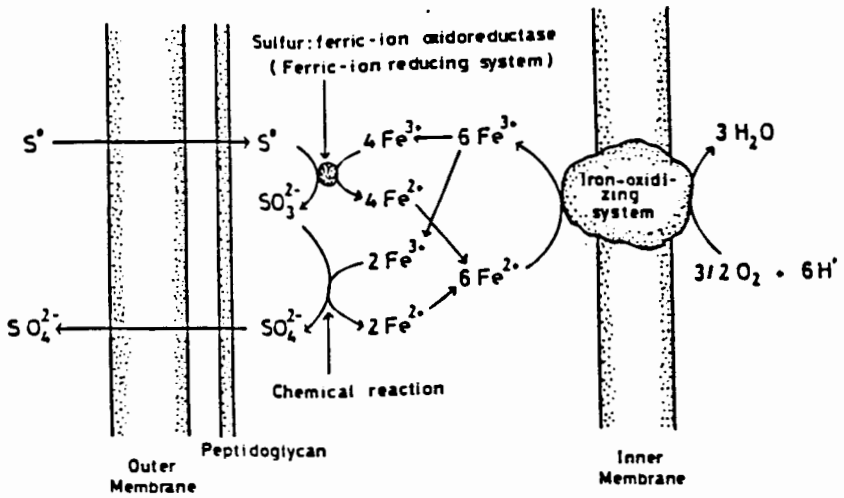


Figure 1.5 : The role of FIR in the reduction of ferric ions during sulphur oxidation recycling ferrous iron for use in the iron oxidation pathway (iron-oxidising system) (Sugio et al. 1985)

It has been reported that cupric ions in the presence of elemental sulphur could be reduced enzymatically by the hydrogen sulphide - ferric ion oxidoreductase (SFORase) enzyme (Sugio et al. 1990). Copper is not the only metal that can be reduced by this enzyme. SFORase can reduce Fe(III) or Mo(VI) with both hydrogen sulphide or elemental sulphur. This suggests that various metals such as molybdenum and copper may play an intricate role as electron acceptors in the oxidation mechanisms of both iron and sulphur. The

and sulphur. The SFORase enzyme has been found in many *T. ferrooxidans* strains tested as well as in a number of *L. ferrooxidans* strains (Sugio et al. 1992b).

Another sulphur-oxidising enzyme to be isolated from *T. ferrooxidans* was the sulphur: ferric ion oxidoreductase from strain AP19-3 (Sugio et al. 1987). The intact enzyme had a molecular weight of 46 kDa and was comprised of two identical 23 kDa subunits. As was the case for the iron oxidation enzymes, the S:Fe(III) oxidoreductase was detected in the periplasmic fraction of the cells. An interesting feature was that when *T. ferrooxidans* strain AP19-3 was grown in a sulphur medium, the iron oxidase discussed in the iron section was also expressed along with the S:Fe(III) oxidoreductase. This suggests that the iron oxidase may play a role in both oxidation systems (Sugio et al. 1988a). The reaction results in the oxidation of elemental sulphur to produce sulphite as is shown in the following chemical equation.

$$S_8 + 14\,Fe^{3+} + 12\,O_2 + 30\,OH^- \longrightarrow 8\,SO_3^{2-} + 14\,Fe^{2+}$$

T. ferrooxidans AP19-3 was also shown to possess a new type of sulphite oxidase (Sugio et al. 1988b). This enzyme was found in the plasma membrane of the cell and was responsible for reducing ferric ions back to the ferrous form. It uses the ferric iron to oxidise sulphites into more harmless sulphates. Sulphite is a potentially toxic compound and strain AP19-3 requires this oxidase to aid the natural chemical oxidation of sulphite to sulphate. The presence of the sulphite oxidase in the plasma membrane is important as sulphite cannot be allowed to enter into the cytoplasm (Sugio et al. 1988b).

A sulphite:ferric iron oxidoreductase has been purified from strain AP19-3 (Sugio et al. 1992b). The role of this enzyme appears to be under the regulation of elemental sulphur. It was inhibited by ferrous iron ions, resulting in the preferential oxidation of ferrous iron when the organism is in a position to oxidise both iron and sulphur. It had a molecular weight of 660 kDa and was comprised of two different subunits of 61 and 59 kDa respectively (Sugio et al. 1992b).

A sulphur binding (SBP) protein which binds reversibly to hydrogen sulphide has been purified from *T. ferrooxidans* (Suzuki et al. 1993). The role of this enzyme is to supply the SFORase enzyme with its sulphide substrate. It had a molecular weight of 16 kDa and was estimated to bind 2.3 H_2S molecules.

Two different types of sulphite reductases were detected in *T. ferrooxidans* AP19-3 (Sugio et al. 1995). They differed in the cofactors which are associated with them as well as in their cellular location. The NADPH dependant sulphite reductase was found in the cytoplasm and is probably involved in anabolic pathways. The NADH dependant sulphite reductase was found in the periplasm and is responsible for the reduction of sulphites to produce hydrogen sulphides (Sugio et al. 1995). The enzyme functions as a monomer and has a molecular weight of approximately 54 kDa.

T. ferrooxidans has been shown to oxidise elemental sulphur in an anaerobic environment (Pronk et al. 1992). Oxygen is the electron acceptor during aerobic growth, but under anaerobic conditions ferric ions could substitute for oxygen. This has implications in the leaching environment due to the fact that large amounts of ferric iron can be generated on the aerobic surfaces, which can then be utilised by cells at greater depths in a dump for example where oxygen is scarce (Sand et al. 1995). It was noted that in very acidic environments (pH <2) ferric ions are the preferred electron acceptor to oxygen, a situation which is reversed at higher pH levels.

1.4.5 A model for the oxidation of sulphur compounds in *Thiobacillus ferrooxidans.*

Two theories have been proposed to explain the mechanism of leaching of metal sulphides - a direct contact (sessile) mechanism and an indirect (planktonic) mechanism (Blake et al. 1994). The direct attack mechanism involved exclusively enzymatic reactions of the bioleaching bacteria. The indirect mechanism required a chemical

oxidation step in conjunction with the bacterial reactions (Sand et al. 1995). The two mechanisms are as follows.

Direct attack

$$Cu_2S + 0.5\,O_2 + 2H^+ \longrightarrow Cu^{2+} + CuS + H_2O$$
$$CuS + 0.5\,O_2 + 2H^+ \longrightarrow Cu^{2+} + S_0 + H_2O$$
$$S_0 + 1.5\,O_2 + H_2O \longrightarrow SO_4^{2-} + 2H^+$$

$$Cu_2S + 2.5\,O_2 + 2H^+ \longrightarrow 2Cu^{2+} + SO_4^{2-} + H_2O$$

Indirect attack of a metal sulphide (MS)

$$MS + 2Fe^{3+} \longrightarrow M^{2+} + S_0 + 2Fe^{2+}$$
$$2H^+ + 0.5O_2 + 2Fe^{2+} \longrightarrow 2Fe^{3+} + H_2O$$
$$S_0 + 1.5\,O_2 + H_2O \longrightarrow SO_4^{2-} + 2H^+$$

$$MS + 2O_2 \longrightarrow M^{2+} + SO_4^{2-}$$

A number of difficulties with the direct attack mechanism have been highlighted (Sand et al. 1995). Firstly, much of the research in support of the direct attack has been carried out on pure synthetic metal sulphides, rather than on natural compounds. Secondly, sulphur oxidation is completely dependent on the presence of ferric irons. Thirdly, it has been shown that even when *T. ferrooxidans* is growing in an iron-free sulphur medium, between 0.5 and 5% of its dry weight are ferric ions which have been tightly bound to the extracellular polymeric compounds which surround the outer membrane (Sand et al. 1995).

In the environment, *T. ferrooxidans* cells attach to metal sulphide compounds such as pyrite by an exopolymer layer. The bacteria then form a biofilm one to two cells thick (Sand et al. 1995). The exopolymer layer consists of polythionate granula as well as ferric irons. Adhesion to the surface is believed to occur when the net surface charge of the bacterial exopolymer layer was sufficient to overcome the repulsive electrical charge present on the sulphide mineral surface (Blake et al. 1994). This may be aided in part by a

non-uniform charge being present on the bacterial surface, thus allowing it to attach to the ore in an area of lower charge density.

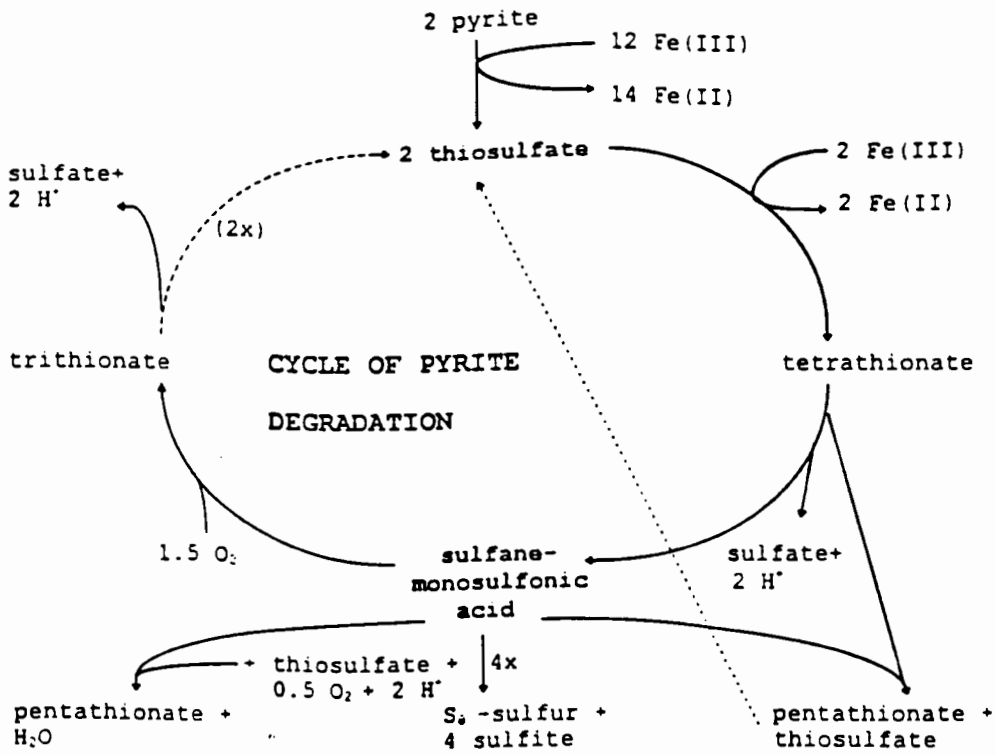


Figure 1.6 : Indirect oxidation of pyrite by *T. ferrooxidans* within the exopolymer layer of the bacterium (Schippers et al. 1996).

Figure 1.6 illustrates the proposed leaching mechanism of pyritic ores in the "reaction compartment" of the exopolymer layer (Schippers et al. 1996). The key sulphur compounds were identified as thiosulphate, tetrathionate and disulphane-monosulfonic acid. Trithionate is rapidly hydrolysed to thiosulphate whereupon it reenters the cycle (Schippers et al. 1996).

1.5 The role of heavy metals in the bioleaching environment

Metals are abundant in the bioleaching environment and a number of elements such as copper (Nielsen and Beck, 1972) and molybdenum (Sugio et al. 1992a), may be used as

electron donors and electron acceptors during iron and sulphur oxidation (section 1.4.4). This part of the review completes the section on redox mechanisms within *T. ferrooxidans*, and covers three selected elements found in the bioleaching habitat. These elements, uranium, mercury, and arsenic have been included here because they are prevalent in the ecological niche of *T. ferrooxidans* and are detoxified in biological systems by oxidation-reduction and efflux pathways.

1.5.1 Uranium

Uranium reduction from the insoluble U^{6+} to the insoluble (and less toxic) U^{4+} has been reported in a number of *Desulfovibrio* species (Lovley, 1993). Uranium toxicity has also been investigated in *T. ferrooxidans* cultures (Tuovinen and Kelly, 1974a, and Tuovinen and Kelly, 1974b). It was noted that *T. ferrooxidans* was inhibited by 2mM uranyl sulphate and this was thought to be due to an inhibition of iron oxidation (Tuovinen and Kelly, 1974a). Excessive amounts of uranium (5-100 mM) caused an immediate termination of carbon dioxide fixation in *T. ferrooxidans* cells. Tolerance to the element was brought about by successive subculturings in increased concentrations of uranium (Tuovinen and Kelly, 1974b). As pointed out earlier, resistance to uranium in a *T. ferrooxidans* culture has been correlated with the loss of a 20 kbp plasmid from *T. ferrooxidans*, although a direct causal relationship between the two phenomena was never established (Holmes et al. 1983).

The oxidation of uranuous ions takes place in a two step equation

$$\begin{array}{lllllll} 2U^{4+} & + & O_2 & + & 4H^+ & \longrightarrow 2U^{6+} & + & 2H_2O \\ 2U^{6+} & + & 4H_2O & & & \longrightarrow 2UO_2^{2+} & + & 4H^+ \\ \hline 2U^{4+} & + & O_2 & + & 2H_2O & \longrightarrow 2UO_2^{2+} & + & 4H^+ \end{array}$$

($\Delta G = -130.4$ kJ.mol^{-1} at 30°C)

The oxidation of uranous ion in the above equation has been shown to be coupled to carbon dioxide fixation (Ivarson, 1980, and Dispirito and Tuovinen, 1982a,b). Furthermore, both uranyl ions and ferric ions are able to oxidise rusticyanin implying that the two oxidation mechanisms may be intricately linked.

1.5.2 Arsenic

Arsenic is prevalent in metal sulphide ores (Summers and Silver, 1978), and the detoxification of arsenic is crucial to bioleaching bacteria. Biological systems detoxify arsenate ions by reducing them to the more toxic arsenite form and then exporting the arsenite from the cell (Silver, 1996). Arsenic resistance has been reported in a number of *T. ferrooxidans* strains (Kondratyeva et al. 1995). These authors found a number of isolates which were resistant to up to 4 g/l As^{3+}. A South African strain of *T. ferrooxidans*, FC13, was isolated which exhibited resistance to 13 g/l total arsenic, the concentration at which the compound precipitated from solution under the conditions at which the biooxidation plant from which the strain was isolated operates (Rawlings, unpublished data). The genes for arsenic resistance have recently been cloned from a *T. ferrooxidans* ATCC33020 genomic library, and expressed in *E. coli* (Butcher, personal communication).

1.5.3 Mercury

Mercury is a toxic element which is prevalent in metal ores. The ability to reduce mercury to a less toxic form would be a great advantage to autotrophic bacteria in the biomining environment. In 1982, strains of *T. ferrooxidans* were reported which were resistant to 5 μM Hg^{2+} and this was shown to be due to a mercuric reductase enzyme (Olson et al. 1982). A mercuric ion-reducing enzyme was isolated from *T. ferrooxidans* TFI-29 (Booth and Williams, 1984). It was a flavoprotein with a molecular weight of 130 kDa, and consisted of a 54 kDa and 62 kDa subunit. The mercury reductase gene has also

been successfully used as a selectable marker in the transformation of *T. ferrooxidans* (see section 1.3.2).

The mercury resistance operon has been cloned and sequenced (Inoue et al. 1989). Like most mercury resistance operons it was inducible by Hg^{2+}. Analysis of the upstream promoter / operator region of this operon showed homology at the DNA level to the promoter sequences of Tn*501* and plasmid R100, both of which encode mercury resistance (Inoue et al. 1990). Interestingly two copies of the *mer* operon regulatory gene, *merR1* and *merR2* were detected within 5 kbp of each other, The intergenic region between these two genes was sequenced and found to contain other mer operon genes, some of which had also been duplicated (*merC*). Primer extension analysis revealed that both *merR* genes were expressed in *T. ferrooxidans* (Inoue et al. 1991) This organisation of the *mer* operon, where the *merR* regulator gene is separate from the rest of the operon has only been reported in *T. ferrooxidans*.

As discussed earlier, a gene with homology to *merR* has been identified on plasmid pTF-FC2 upstream of an ORF with homology to multidrug resistance proteins (Clennel et al. 1995). This suggests the possibility of another heavy metal resistance mechanism, an observation which is receiving further investigation (Rawlings, personal communication).

1.6 Aims of this project

The original aim of the project was to isolate genes from *T. ferrooxidans* which encoded redox-active proteins. Investigation of the genes which encode these proteins was considered to be an important step in the understanding of the metabolism of *T. ferrooxidans* as a whole. With this in mind, a metronidazole activation assay was employed to isolate *T. ferrooxidans* genes whose products could reduce metronidazole in an *E. coli* mutant host. Chapter 2 describes the isolation of a cosmid which confers metronidazole sensitivity to this *E. coli* mutant. An unexpected finding was that the insert

of the cosmid originated from an endogenous *T. ferrooxidans* plasmid, designated pTF5. Surprisingly the metronidazole-activating phenotype was contained on a 319 bp DNA fragment which did not appear to contain any ORFs, but rather formed part of the *oriV* of the plasmid.

Redox-active protein-encoding genes were however present on pTF5, although they did not contribute to metronidazole sensitivity in *E. coli* F19. Subsequent aims of the project included firstly, molecular characterisation and expression of the genes and products in an *E. coli* host (Chapter 3). This was done in the hope of obtaining evidence for the expression of these genes in *T. ferrooxidans* as well as determining a possible role for the corresponding proteins in that bacterium. Secondly, the geographic diversity of the plasmid was investigated using pulsed field gel electrophoresis (Chapter 4), as pTF5-like plasmids had previously been detected by other researchers. Finally, characterisation of plasmid pTF5 was carried out, in order to determine features such as cellular copy number and its differing relationships with a number of *T. ferrooxidans* strains.

Chapter 2

The activation of metronidazole by plasmid pTF5

Contents

The activation of metronidazole by plasmid pTF5

Summary

A *Thiobacillus ferrooxidans* total genomic library was screened for redox-active protein-encoding genes which could activate metronidazole under anaerobic conditions. A family of cosmids which had a number of restriction fragments in common and which were able to confer wild type levels of metronidazole sensitivity to a resistant *E.coli* mutant was isolated. The metronidazole sensitive phenotype was isolated on a 319 base pair *Xba*I-*Sal*I restriction fragment. The fragment was sequenced and found to have extensive homology to a plasmid origin of replication isolated by Chakravarty et al. (1995) from *Thiobacillus ferrooxidans* TFI91. The plasmid, pTF5, was cloned from two members of the cosmid family and was found to be larger (19.8 kbp) than the 9.8 kbp pTFI91. A second difference was that the pTF5 *oriV* contained fewer regulatory features (iterons) than the pTFI91 sequence. The complete replicon of pTF5 was sequenced and characterised along with its flanking regions. *In vitro* transcription - translation reactions from an *E. coli* derived system, indicated that there are a number of highly expressed open reading frames on plasmid pTF5. Potential proteins were mapped onto the DNA sequence, and computer database searches revealed that they had homology to a number of plasmid proteins involved in replication, plasmid stability and recombination.

2.1 Introduction

This discussion serves to introduce the antimicrobial agent, metronidazole. It is not intended to be an exhaustive review of the drug, but rather a concise appraisal of its electrochemical properties; the resistance mechanisms which have been described for it, and how the two may be utilised by molecular biologists as a tool to screen for proteins capable of performing oxidation-reduction reactions.

2.1.1 Metronidazole : A pharmacological perspective

Metronidazole 1-(2-hydroxyethyl)2-methyl-5-nitroimidizole (Figure 2.1), has long been recognized as an effective antimicrobial agent in the treatment of obligate anaerobic infections (Johnson, 1993). Metronidazole was originally chemically synthesized from azomycin, an antibiotic isolated from *Streptomyces* (McFadzean, 1986) over thirty years ago.

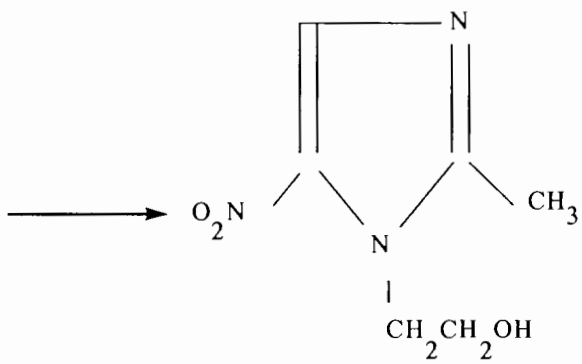


Figure 2.1 : The chemical structure of metronidazole, showing the reactive nitro group (arrow) believed to be reduced to form toxic intermediates in an anaerobic atmosphere (Johnson, 1993).

The drug is chemically inert in aerobic systems, but in the absence of oxygen can be rapidly reduced by enzymes, producing toxic intermediates (Johnson, 1993). The reduction pathway is believed to proceed via the mechanism shown in Figure 2.2 (Quon et al. 1992). The most significant bacteriocidal effect of metronidazole is the production of ion radicals during the reduction of the drug (Church et al. 1980). Metronidazole is readily taken up by most cells, the rate of uptake having been shown to be proportional to the rate of metronidazole reduction (Edwards, 1993).

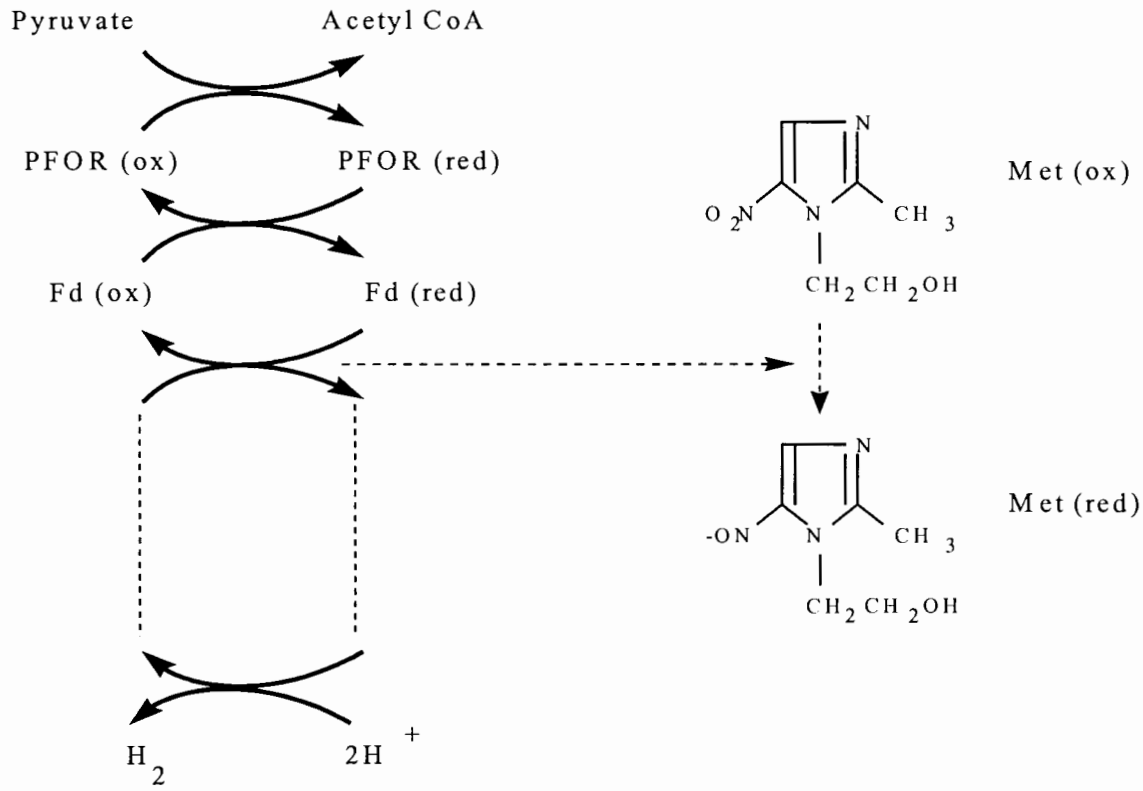


Figure 2.2 : The activation of metronidazole (Met(ox)) to its reduced toxic intermediates (Met(red)) by redox-active enzymes in the hydrogenosome of *T. vaginalis* (Quon et al. 1992). PFOR is the pyruvate-ferredoxin oxidoreductase enzyme, Fd is the ferredoxin protein.

Metronidazole appears to bind DNA preferentially in adenine-thymine (AT) rich regions (Edwards et al. 1980 and Tocher and Edwards, 1992) thus targeting the free-radical containing intermediates to vital genomic functions in the cell. Furthermore there is also evidence that once taken up, up to 30% of metronidazole binds to proteins within the cell (Muller and Lindmark, 1976).

2.1.2 Resistance mechanisms

Relative to other commercial antibiotics, resistance to metronidazole is rarely reported (Edwards, 1993). Resistance to the drug has been intensively studied in order to facilitate the design of new antimicrobial agents which may circumvent these mechanisms (Rowley et al. 1979).

Microorganisms sensitive to metronidazole, possess redox-active enzymes which are responsible for the reduction event. Examples include ferredoxins in *Clostridium acetobutylicum* (O'Brien and Morris, 1971), *Clostridium pasteurianum* (Lockerby et al. 1984), *Giardia duodenalis* (Townson et al. 1994) and *Trichomonas vaginalis* (Quon et al. 1992); and pyruvate:Fd oxidoreductases from *Bacteroides fragilis* (Narikawa, 1986). Organisms such as *C. pasteurianum* and *T. vaginalis* which are resistant to metronidazole, have been noted to have reduced expression of these enzymes (Quon et al. 1992). Other candidate enzymes implicated in this reduction reaction include hydrogenases from *C. pasteurianum* (Church and Laishley, 1995), *Trichomonas foetus* (Lindmark and Muller, 1976) and *T. vaginalis* (Kulda et al. 1993), general "nitroreductases" in *Escherichia coli* (Yeung et al. 1984), lactate and pyruvate dehydrogenases from *B. fragilis* (Narikawa et al. 1991, Britz and Wilkinson, 1979) and *Clostridium perfringens* (Sindar et al. 1982).

A family of plasmids has been identified in *B. fragilis* which encode resistance to the 5' nitroimidazole drug family, which includes metronidazole (Reysset et al. 1992, 1993). Three plasmids have been described in this family, each of which encodes a different nitroimidazole resistance (*nim*) gene. These *nim* genes are expressed from promoters located in adjacent insertion elements (Haggoud et al. 1995, Trinh et al. 1995). Two of the *nim* genes, *nim*A and *nim*B, share 70 % amino acid identity to each other, suggesting a common evolutionary lineage (Haggoud et al. 1995), whilst the third, *nim*C is genetically distinct. The mechanism by which these genes confer metronidazole resistance to the host organism has not been elucidated.

Other plasmids from *B. fragilis*, which contain metronidazole resistance conferring phenotypes have also been described (Wehnert et al. 1990, Dachs et al. 1995). A small cryptic 6 kbp plasmid pBFC1 was found to harbour a 195 bp open reading frame responsible for the sensitivity to metronidazole in *E. coli* (Wehnert et al. 1992). The 7.3 kDa MetA protein produced by this gene, was not involved in metronidazole reduction but rather acted at the level of DNA repair (Dachs et al. 1995). Enhanced DNA repair mechanisms in bacteria have been suggested to contribute to metronidazole resistance, by reversing the oxidative damage done to the chromosomal DNA of the host (Yeung et al. 1984).

2.1.3 Metronidazole as a tool in molecular biology

Metronidazole has a an electropotential of -485 mV (Peterson, 1988) and can thus oxidise enzymes with a more electronegative potential. The drug acts as an electron sink in these reactions, diverting electrons away from electron transfer pathways in order to reductively activate metronidazole. This has been exploited as a means of isolating genes encoding redox-active proteins in *Rhodobacter sphaeroides* (Wall et al. 1984) and *C. acetobutylicum* (Santangelo et al. 1991). It was also used in the isolation of nitrogen fixation mutants in *Rhodopseudomonas capsulatus* (Willison and Vignais, 1982) and *Rhodobacter capsulatus* (Schmehl et al. 1993).

In order to set up a potential metronidazole assay system in *E. coli*, the mechanisms of resistance first needed to be identified and understood in that host. Yeung et al. (1984) investigated the interaction of metronidazole with DNA repair mutants of *E. coli*. These mutants, particularly those with deficient *rec*A and *uvr*B genes could not repair DNA damage induced by metronidazole, and were thus more sensitive to the drug than wild type strains.

It had also been noted that strains of *E. coli* that could not reduce nitrates and chlorates, were less susceptible to metronidazole (Yeung et al. 1984). The implication of this was that in *E. coli* there existed a generalised "nitroreductase" enzyme (or enzymes) which was responsible for reducing metronidazole to its bacteriocidal intermediates. Yeung et al. (1984) used a kinetic model for metronidazole reduction originally proposed by Chrystal et al. (1980) in order to determine *E. coli* susceptibility to the drug (Fig 2.3)

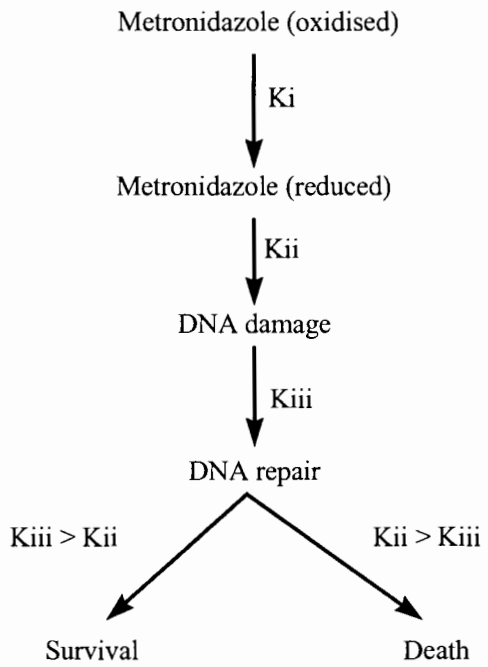


Figure 2.3 : A kinetic model for metronidazole metabolism in *E.coli*, which determines the susceptibility of a particular cell to the drug (Chrystal et al. 1980). Ki, Kii and Kiii are kinetic rate constants representing rate of reduction, DNA damage, and DNA repair respectively.

Santangelo et al. (1991) used this principle to generate an *E. coli* mutant which was both recombinase (*recA*) and "nitroreductase" deficient. The authors used Tn*PhoA* mutagenesis of *E. coli* CC118 (*recA*) to isolate mutants which were unable to reduce the indicator dye benzylviologen. A number of these mutants were then screened for

increased resistance to metronidazole. One mutant F19, showed the greatest resistance to metronidazole (25 μg/ml), when compared to the parent construct (5 μg/ml). The authors used strain F19 as a heterologous host to screen for "redox" genes from *C. acetobutylicum* which could confer wild type levels of metronidazole sensitivity to the mutant (Santangelo et al. 1991).

A *C. acetobutylicum* P262 genomic library was screened for genes which could reduce metronidazole in *E. coli* F19. Twenty five clones were isolated and were classified according to the degree of sensitivity to metronidazole conferred on the F19 mutant host (Santangelo et al. 1991). One of the most sensitive clones, was subcloned and the region conferring sensitivy to metronidazole was sequenced. An open reading frame with predicted amino acid sequence homology to a flavoprotein was identified. This was a likely candidate for reducing metronidazole in this obligate anaerobe. The authors speculated that the flavoprotein may function in *C. acetobutylicum* by acting as an electron carrier in the phosphoroclastic reaction (Santangelo et al. 1991). This process (see Figure 2.2) involves the anaerobic breakdown of pyruvate to acetyl CoA by pyruvate-ferredoxin-oxidoreductase. Excess reducing equivalents are disposed of by a hydrogenase enzyme accepting electrons from the reduced ferredoxin to produce hydrogen gas (Quon et al. 1992). It has been shown in *C. pasteurianum*, that flavodoxin can substitute for the ferredoxin enzyme in this phosphoroclastic reaction (Knight et al. 1966). This reaction had been implicated in the reduction of metronidazole (Narikawa, 1986).

A second clone from this study has also been characterised (Davison et al. 1995). Unlike the previous *C. acetobutylicum* genomic clone which conferred levels of sensitivity to metronidazole equal to *E. coli* wild type cells, pMET7C exhibited only a slight reactivity to the drug. Sequencing of this clone showed that a 972 bp open reading frame with homology to a catabolite control protein, CcpA, from *Bacillus subtilis* was present on the clone responsible for the metronidazole sensitive phenotype. The protein, RegA,

produced from this open reading frame was shown to regulate starch degradation in *C. acetobutylicum* (Davison et al. 1995). Its role (if any) in the reduction of metronidazole was not determined.

2.2 Materials and methods

Materials and methods used in more than one chapter are presented in Appendix C. In addition, descriptions of software programs used in the computer analysis section have been included in Appendix C.6.

2.2.1 Bacterial strains and cosmid clones

Bacterial strains used in this study are shown in Table 2.1. Cosmid clones conferring a metronidazole sensitive phenotype were isolated in an initial screening by D.K. Berger. The cosmids, pMET7, pMET9, pMET16, pMET17 all contained *T. ferrooxidans* insert DNA and shared a number of common restriction fragments.

Table 2.1 : Bacterial strains used in the isolation of the metronidazole activating origin of replication (*oriV*) from pTF5

Bacteria	Strains	Genotype	Reference
E. coli	CC118	*araD139Δ(ara,leu) 7697ΔlacX74 phoAΔ20,galEK,thi,rpsE,rpoB, argEam, recA1*	Manoil and Beckwith, (1985)
E. coli	F19	CC118 ::Tn*PhoA* *(pgk, yehB)* - this study	Santangelo et al. (1991)
T. ferrooxidans	ATCC33020		ATCC, Rockville MD, USA

2.2.2 Colony isolation assay

A total genomic cosmid library of *T. ferrooxidans* ATCC33020 (Ramesar R., Ph.D. thesis, 1988) was screened for the expression of a metronidazole sensitive phenotype. *E. coli* mutant F19 (Santangelo et al. 1991) was transduced with cosmid vector pHC79, containing *T. ferrooxidans* total genomic DNA inserts packaged according to the method of Hohn and Collins (1980). Transduced F19 cells were plated on two different concentrations of metronidazole (Sigma) containing selective media. This media consisted of Luria agar (Sambrook et al. 1989) supplemented with 0.5% glucose (Saarchem), 0.5% sodium nitrate (Saarchem) to act as an anaerobic electron acceptor, and 100 mg/ml ampicillin (Sigma) to select for the packaged cosmids.

The two concentrations of metronidazole used were 50 mg/l (Met-50 plates) and 85 mg/l (Met-85 plates). The lower concentration of metronidazole (50 mg/l) acted as a positive control, allowing all transductants to grow under the assay conditions used. It was presumed that the higher metronidazole concentration (85 mg/l) selectively inhibited the growth of any F19 clone which contained *T. ferrooxidans* total genomic DNA capable of activating metronidazole. To ensure that an even inoculum of cells were delivered to both concentrations of metronidazole-containing plates, clones were replica plated from Luria agar (ampicillin 100 μg/ml) master plates (Met-0 plates). Media used in this procedure was aged overnight in a reducing atmosphere. Transductants, once replica plated onto the appropriate media, were allowed to grow for 60 h at room temperature in a Forma Scientific 1024 anaerobic chamber in an atmosphere consisting of 85% CO_2, 10% N_2, 5% H_2 and the relative growth of the cosmid clones on the two different metronidazole concentrations was observed.

2.2.3 Single Cell Reduction assay

Cosmid transductants which were able to grow on the Met-0 and Met-50 media, but were inhibited by the Met-85 media, were tested for minimum inhibitory concentration levels, using a single cell reduction assay. As this assay relied on the growth of a single *E. coli* cell on increasing concentrations of metronidazole, lower concentrations of metronidazole were used than for the colony isolation assay. Standard metronidazole concentrations of 0, 5, 10, 15, 20, 25 mg/l were used in the single cell reduction assay. Selective metronidazole media was prepared in an identical manner to the colony isolation assay. Cosmids identified as conferring metronidazole sensitivity to F19 were transformed into that host using standard methods (Sambrook et al. 1989). Transformants were grown in Luria broth for 15 minutes at 37°C and then diluted $1/10^4$ in sterile water. 100 µl samples of this diluent were then plated onto the pre-reduced metronidazole media in an anaerobic environment. Transformants were grown anaerobically for 60 h at room temperature and the minimum inhibitory concentration for each clone was determined by the highest concentration of metronidazole on which isolated colonies could grow.

2.2.4 Analysis of Tn*PhoA* insertion sites

The *E. coli* F19 mutant had been produced by Tn*PhoA* insertional mutagenesis (Santangelo et al. 1991). Characterisation of the mutant had indicated that it was unable to reduce nitrates or chlorates. In order to further characterise the mutant, the site (or sites) of Tn*PhoA* integration was determined by marker rescue cloning. *E. coli* F19 total genomic DNA was isolated according to the method described in Sambrook et al. (1989). Total genomic DNA was digested with a restriction endonuclease which cut the Tn*PhoA* element at a single site on one or the other side of the kanamycin resistance marker. In the cloning of the right hand junction between Tn*PhoA* and the F19 chromosme, *Eco*RI was

used (See figure 2.8). In a similar manner, *Sal*I was used for the cloning of the left hand junction.

These digests were then incubated with a combination of *Eco*RV and *Stu*I which generate blunt ended restriction fragments and which did not cleave within the Km^r containing fragment of Tn*PhoA*. Digests were ligated using T4 ligase (Boehringer Mannheim) into pBluescript-KS cloning vector (Stratagene) which had been digested with either *Eco*RI or *Sal*I, and *Eco*RV. The ligation mix was transformed into *E. coli* JM109 a *rec*A mutant in order to prevent the possibility of homologous recombination between the host *E. coli* and the F19 genomic clone. Clones containing the ampicillin resistance marker of pBS-KS and the kanamycin resistance marker of Tn*PhoA* were selected on Luria agar supplemented with ampicillin (100 mg/l) and kanamycin (50 mg/l).

2.2.5 Nested deletions using exonucleaseIII

On occasion, nested deletions (Henikoff, 1987) were required in both sequencing strategies as well as in mapping putative *in vitro* proteins onto DNA sequence. Deletions were obtained using standard protocols (Sambrook et al. 1989), an example of which is decribed here. Two μgrams of pCD550, containing three highly expressed ORFs was restricted with *Xba*I to generate 5'-OH ends suitable for the exonuclease III digestion. The restriction digest was precipitated and resuspended in A restriction buffer (Boehringer Mannheim). Digestion with *Apa*I then produced a 3'-OH end in order to protect the vector from degradation by exonuclease III. The reaction was started by adding one unit of exonuclease III and incubating at 37°C. Samples were taken at 30 sec intervals and placed in S1 nuclease buffer at 4°C. S1 nuclease (1 unit) was added to each time interval to remove single stranded DNA not digested by Exonuclease III. Samples from each time interval were then ligated to achieve the circle closure of the deleted pCD550 clone.

2.3 Results

2.3.1 Isolation of a cosmid family sensitive to metronidazole

Preliminary screening of the *T. ferrooxidans* genomic cosmid library was carried out by D.K. Berger. Using the colony isolation assay, six cosmids were isolated which were able to activate metronidazole. This study was initiated by determining the minimum inhibitory concentrations (MIC's) for each of these cosmids using the single cell assay. The results are shown in Table 2.2

Table 2.2 : Minimum inhibitory concentrations (MIC's) of members of the *T. ferrooxidans* cosmid family which were able to activate metronidazole in *E. coli* F19.

Cosmid clone	MIC (mg/l)
pMET2	10
pMET7	10
pMET9	10
pMET16	10
pMET17	10
pHC79	20

The cosmids were subjected to restriction analysis (see Figure 2.4) which showed that all but one of them shared common restriction fragments. This indicated that there were presumably at least two families of cosmids in the *T. ferrooxidans* gene bank with the ability to reduce metronidazole.

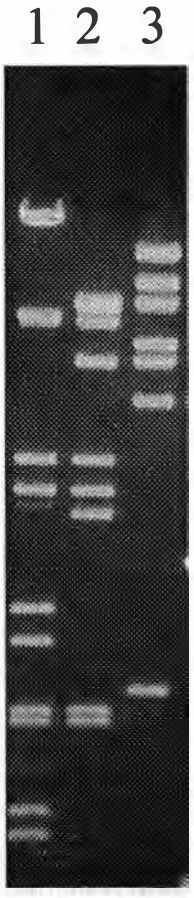


Figure 2.4 : Restriction analysis of *T. ferrooxidans* cosmids which conferred sensitivity to metronidazole shows that pMET7 and pMET17 contain similar DNA. Lane 1, contains pMET7, lane 2, pMET17, lane 3, pMET2, an unrelated cosmid clone have all been digested with *Eco*RI.

2.3.2 Characterisation of the metronidazole reducing phenotype

Cosmid pMET17 was selected for subcloning and identifying the region responsible for activating metronidazole. DNA fragments from pMET17 were subcloned into the pBS-KS cloning vector (Stratagene, San Diego) and were tested for the ability to activate the drug.

A 2.1 kbp *Sac*I (pCD300) fragment was subcloned which showed very high levels of sensitivity to metronidazole (see Figure 2.5). A *Xba*I-*Sal*I fragment within this region

(pCD313), which retained the MET^s phenotype was sequenced using T7 Sequenase. The fragment was 319 bp in length, and retained the full MET^s phenotype. The DNA sequence was compared to sequences lodged in the GenEMBL DNA databases using the BLAST algorithm, and had a high degree of identity (98%) to a plasmid origin of vegetative replication (*oriV*) from *T. ferrooxidans* TFI91 (Chakravarty et al. 1995). The 319 bp *Sal*I-*Xba*I metronidazole active fragment was located between the IHF binding domain of the *oriV* and the start site of a transposase gene of an insertion element (Figure 2.5). The pTF5 *oriV* thus had an almost identical structure to that described in pTFI91.

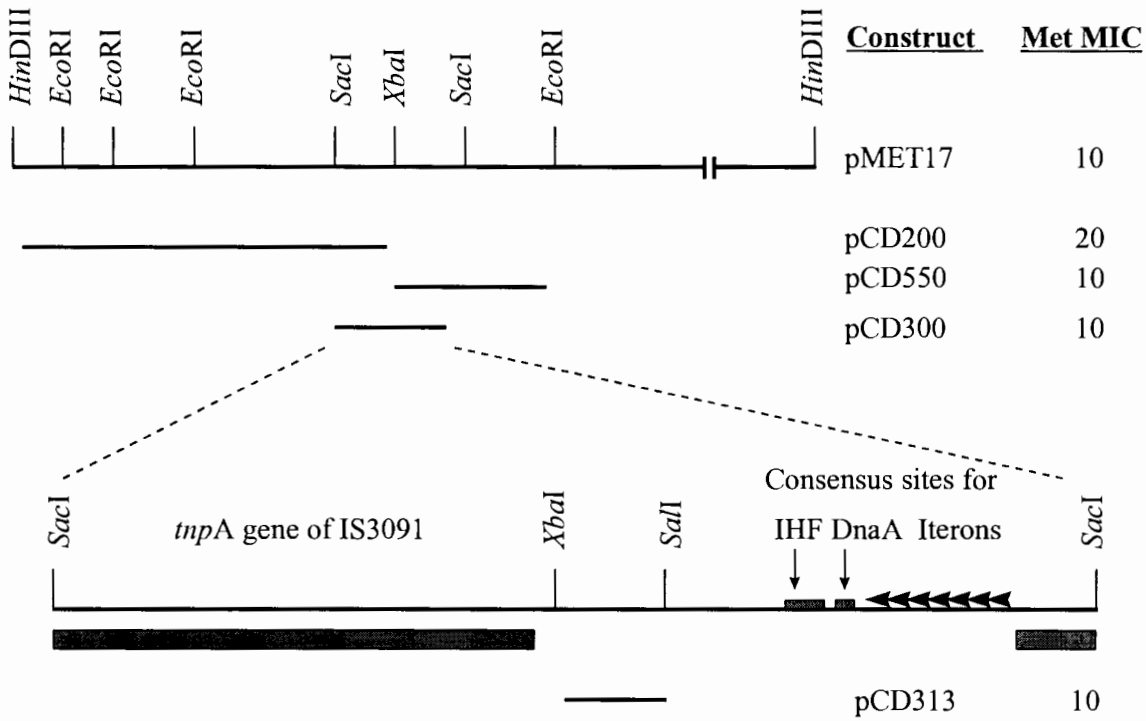


Figure 2.5 : The *oriV* region of pTF5. Subclones of pMET17 showed that the region conferring metronidazole sensitivity to *E. coli* F19, was present on a 2.2 kbp *Sac*I fragment. Clones pCD200 (*Xba*I-*Hin*dIII) and pCD550 (*Xba*I-*Eco*RI) further narrowed this region to a 1.1 kbp *Xba*I-*Sac*I fragment. This led to the cloning of pCD313, a 319 bp *Xba*I-*Sal*I fragment. pCD300 showed strong sequence identity (98%) to the *ori*V region of pTFI91 represented in Figure 1.1.

Thus it appeared as if an endogenous *T. ferrooxidans* plasmid and been incorporated into the cosmid gene library when it was constructed. Further evidence that the DNA insert in pCD313 is plasmid associated is presented in Chapter 4. The plasmid, designated pTF5, was cloned intact from two members of the cosmid family, pMET17 and pMET7 (as outlined in Figure 2.6).

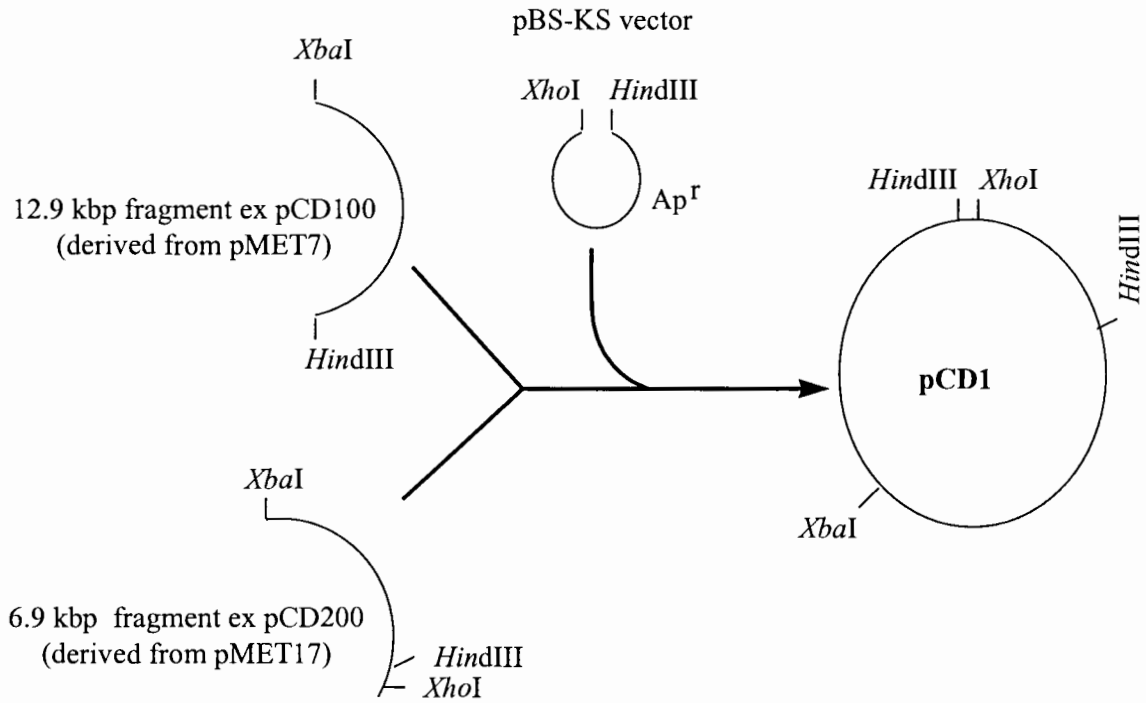


Figure 2.6 : Cloning of pTF5 into pBS-KS cloning vector. Clone pCD100 was a 12.9 kbp *Xba*I-*Hin*dIII fragment cloned into pBS-KS from pMET7. Clone pCD200 was a 6.9 kbp *Xba*I-*Hin*dIII (pBS-KS) subclone of pMET17.

To ensure that no genetic information had been either gained or lost by the cloning procedure, independently derived fragments were cloned from either pMET7 or pMET17. The restriction patterns of these constructs are shown in Figure 2.7. Lanes 9-11 confirmed that the predicted ligation reaction detailed in Figure 2.6 was correct.

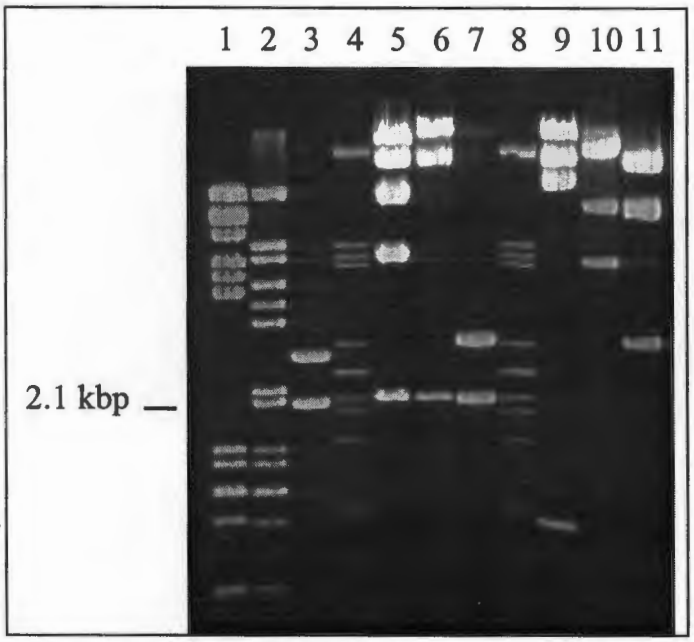


Figure 2.7 : Restriction analysis of pCD1 and its parent clones during the reconstruction of the endogenous pTF5 plasmid. Lanes 1-3 *Sph*I digests of pMET7, pMET17 and pCD300 clones respectively. Lanes 5-7 *Sac*I digests of the pMET7, pMET17 and pCD400 subclones. Lane 4, 8 contain bacteriophage λ *Pst*I molecular weight standards. Lanes 9 - 11, *Xba*I-*Hin*dIII digests of pMET7, pMET17 and the pCD1 subclone.

The 12.9 kbp *Xba*I-*Hin*dIII fragment from pMET7 (lane 9) was cloned into pBS-KS to form pCD100. Similarly the 6.9 kbp *Xba*I-*Hind*III from pMET17 (lane 10) was also cloned into pBS-KS to form pCD200. Using the restriction sites present in that vectors multiple cloning site, it was then possible to use different restriction sites to clone both fragments into a new vector. The 2.2 kbp *Sac*I restriction fragment is present in both cosmids as well as the pCD1 clone of pTF5 (lanes 5 to 7). This indicates that the *Xba*I restriction site which is internal to that fragment was correctly reconstituted during the pCD cloning event. In a similar manner the *Sph*I restriction digests (lanes 1 to 3) confirm the integrity of the *Hin*dIII cloning site. Clone pCD400 (lane 3) is a *Sph*I fragment which

contains the *Hin*dIII restriction site. It was cloned from pMET17 into pUC-BM21 and sequenced (Chapter 3).

Further evidence of the integrity of pTF5 was obtained by comparing the physical restriction maps of the highly homologous pCD1 and pDER301 and this is presented in section 2.3.6. Sequence was also obtained from the pCD1 clone using pUC Universal sequencing primers (Stratagene). The sequence agreed with the pCD400 sequence in the vicinity of the *Hin*dIII site confirming that it too was intact.

2.3.3 Characterisation of *E. coli* mutant, F19

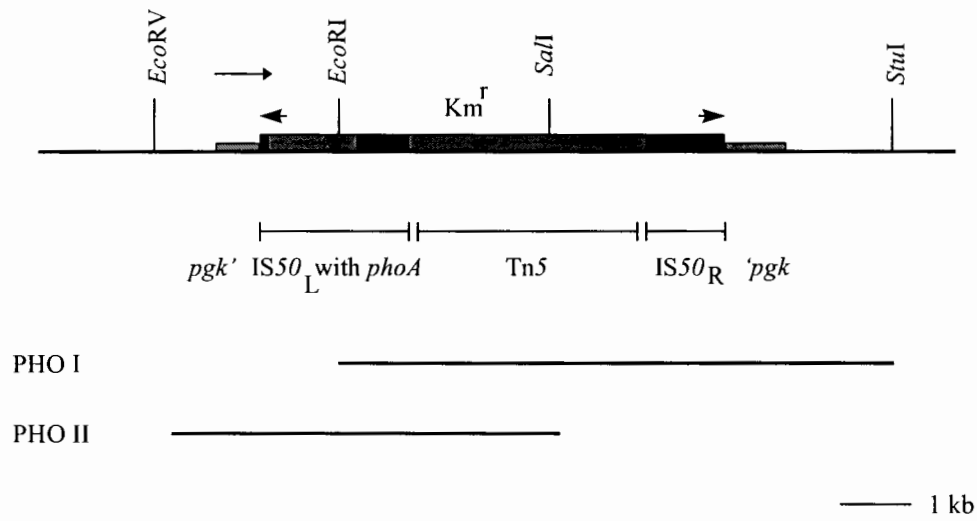


Figure 2.8 : A preliminary physical map indicating the point of integration of the Tn*PhoA* transposon into the 3-phophoglycerate kinase (*pgk*) gene on the *E. coli* F19 chromosome. Relevant subclones are shown as black bars beneath the map, arrowheads show the positions of sequencing primers used in the experiment. The arrow indicates the direction of *pgk* transcription.

The 319 bp pCD313 clone which conferred sensitivity to *E. coli* F19 did not code for a recognizable ORF. This unexpected result made it difficult to adequately explain the MET^s phenotype conferred to the F19 mutant, by this fragment. Santangelo et al. (1991)

had shown that the *E. coli* F19 mutant was unable to reduce nitrates or chlorates. Further genetic characterisation of the F19 mutant was undertaken to determine whether alternate factors could be influencing the assay.

The site of insertion of Tn*PhoA* into the *E. coli* F19 genome was identified by marker rescue cloning as described in Materials and Methods (Section 2.2.4). Four colonies were isolated, three from cloning experiments designed to isolate the right border of the Tn*PhoA* insertion, and a single colony from the experiment designed to clone the left border.

DNA primers were synthesized which were complimentary to the Tn*PhoA* inverted repeats as indicated in Figure 2.8 (previous page). Restriction analysis of these recombinants indicated that there were two types of clones from colonies containing the right border of Tn*PhoA*. DNA sequencing of a member of each of these clones revealed that the Tn*PhoA* had integrated into two different sites in the *E. coli* F19 genome. Perfect homology to the previously published *E. coli* K12 3-phosphoglycerate kinase (*pgk*) gene was obtained from BLAST analysis of the GenEMBL nucleic acid databases as well as to a hypothetical gene designated *yeh*B, the function of which is unknown. Sequence from the sole colony containing the left hand integration site confirmed the point of integration into the *pgk* gene as being the codon for V_{135}. The *pgk* and *yeh*B genes have been assigned positions on the *E. coli* map of 65.42 and 46.54 minutes respectively (*E. coli* Database Collection - http://susi.bio.uni-giessen.de/cgi-bin/ecsuch/)

No further analysis was undertaken as it was evident that F19 had at least two copies of the Tn*PhoA* transposon in its genome. Determining the mechanism of metronidazole resistance in this mutant would be complicated by the double insertion and was beyond the scope of this investigation.

2.3.4 Sequence analysis of pTF5 replicon

Both strands of plasmid pTF5 were sequenced according to standard methods (Appendix C.3). Figure 2.9 shows the nucleic acid sequence from the *Sph*I site (5641) to the *Apa*I site (14524), a region which includes the pTFI91-like *ori*V / Met^s fragment.

Figure 2.9 (overleaf) : Nucleic acid sequence of the 8.9 kbp *Sph*I-*Apa*I pTF5 fragment containing the *oriV* and ORFs with aa sequence identity to proteins involved in plasmid replication, recombination and transposition. Putative stem-loop sequences and promoter recognition sequences have been underlined. The long 33 bp iterons of the *ori*V have been numbered 1 - 7, whilst the shorter 9 bp iterons have been labelled a - d, both represented by arrows. Putative ribosome binding sites are in bold print. The numbering of the nucleic acid sequence is relative to the first base of the unique *Hin*dIII restriction site of pTF5.

*Sph*I

5641 GCATGCCCATGAAATGATATTTGGATATTCCATGGCCGTGGCTGTTGGTTTTTTGTTGACCGCAGTCAAAAACTGGACGG

5721 GTCACCAGACGCTTCGCGGATATGGATTATTATTACTAGGTTTGTTCTGGCTTATGGCACGCATAATGCCGTTCACCGAT

5801 TTTCCTGGCGCACTGATGATCATGGCGCTGTTCGACCTGTTGTTTGATGCCACTGTTTGCATCGCCTTACTGTATCCTTT

5881 GGTCCGCAGCAAGCAATGGAAACAGCTTGGCATATGGTCCATATTATTACTAATGACGACAGGTAATACATTGTTCTATA

5961 TAGGACTCATCTGGAAGCATCCTGAGCTGACGCAGATTGGGTTGTTTTTCGGCCTCTATCTGATTATCGCGCTGATTATG

6041 CTGATGGGACAGCGGGTAATGCCGTTCTTTATTGAAAGAGGCGTTGGTTATCCGGTGGTTTTGATGGTTCGGCCGTGGGT

*Bam*HI

6121 AAACGTCAGTAGCGTGATTCTTATGATGTTTTTTATAGTGGCTGTTGTATGGGATCCGATTTCACTGATTACGGCTTTGG

6201 TGGCCATCGCGTTAGTTTTTTTTGCAGGTTCTACAAATGATCGGATGGTATACTCCGGGCATATGGCGTAACCTTTGTTAT

6281 GGAGTCTTTACCTGGCTTATGGATGGATATTCATTGGATTTATTTTGACTGCGCTTGTTCATTGGATACCCATCAATCCA

*Eco*RI

6361 ATGTTAGGAATTCATGCTTTTGCATATGGTGGGGTAGGCATGATGACCATTGGCATGATGGCGAGAGTCTCTCTAGGGCA

6441 CACAGGAAGAAACATCTTTACACCCCCACAGGCCTTGAACTGGATGTTCTTTTCGTTGTTTCTTGGTGCCATTGTGCGCG

6521 TGATTATTCCAATAATATTGCCCATGCAGTACCGAATGTGGATTGAGGTGTCGCAAGGCCTGTGGATTATTGCCTTCGTC

6601 GGTTTTATCATCGTTTATGCGCCCATGCTTATAAAGCCACGGATCGATGGCCGATATGGATAATGAAATAAATAGAACAA

6681 CGTGGATTGTTATTTATGGAATCACAAGTGATGGGGTAAAATTCCGCCCTTCGTCGTGGGCGGAAATGCTTATTGAGGGG

SL2 (-10.6 kJ.mol^{-1})

6761 GTTGGGTTAACATCCTTTGGCAGAGACCACAAAATTCATTATACAAGACATATTTGGACAATATTAACGAATGGCAATGC

*Pme*I

6841 CGGCATTATAATGAATGGCTGGTTTAAACATGAGCAGCCTATAGCATTTGCTGAGATTATGCGCTTTGCCAATAATAATA

6921 GATTAACAGTCAATACGTATAACAATACCATCGCTCTTCTATAACCACTGAAGTTTAGACATGCAGCCTCCTCGCCAGCT

7001 TCTATGTGTGGCGATATGGCGTAGTAAATGTGGAGCAGGATTGAGTCAGGCGACCTTTTTGCGGGGTGGGTCATCGGTAT

* A V K K R P P D D T

7081 CGGCCCAAAAGGCGGGGGCGGGGCACTGGAATTTCCCGGACTTCGGGTCGTAGCCCTTCCAGAATGGAACCCGCAGAAAT

D A W F A P A P C Q F K G S K P D Y G K W F P V R L F

*Pst*I

7161 TGCAGATGAAGCCATTGGGCCATGAGTTCTGCGGTGACAGGAGTAGCGATTACCCAGGGCGCTACTGCAGTCATGGGGAA

Q L H L W Q A M L E A T V P T A I V W P A V A T M P F

7241 ATCTTCGGATAGGGTTTCGGCGAGGAGTTGCACCATGGACCAAGCGATGCAGCGGATTTGCATCCAGCGTTCCAAGACGG

D E S L T E A L L Q V M S W A I C R I Q M W R E L V

7321 TGCGTTTTTGCTGCCAGAGGTTATGGATGCCCCACCAGCGTTTGAGATTGTGGAAGATCGGTTCGATCTGCCAGCGGTTG

T R K Q Q W L N H I G W W R K L N H F I P E I Q W R N

7401 CTGTAGAGGACGACGATTTCGCTGGCGGTGAGATCCGTCTCGGAAGCGAGATAGAGGCGGGTGGGAGACCACCGATGATG

S Y L V V I E S A T L D T E S A L Y L R T P S W R H H

7481 CTTTTCATCGTAGATCGCACTCCATACGGCACGGACCGGATGGCCGTTCAGGAATCGTGCATTGGCGATGATGGATTGAA

K E D Y I A S W V A R V P H G N L F R A N A I I S Q L

7561 GACGGACCTTCTGCCGTTTCCCATAGAGATCCAGAGAGACGATCTCGGTGGGCAGCGCAGCCATGTCTTCCACGGTGTAC

R V K Q R K G Y L D L S V I E T P L A A M D E V T Y

7641 TTGTTTCCATATATCCGAGGACGCCCGCGTTGGGATGTCCTCGGCGGTGGGTCGCGGAGCAGTACAGTATCTTTGCGCAC

K N G Y I R P R G R Q S T R P P P D R L L V T D K R V

*Pst*I

7721 CTGGCCGATGAACTGCTGCCGGCTGCAGCGTTGCTGACGCAACAGGGGCAAGAGCAGGCGCCGCCGCATGAACCAAGCAT

Q G I F Q Q R S C R Q Q R L L P L L L R R R M F W A

7801 CGGTGAGCAACCGGACGGGTACGGGGATGTGGGGTATAAAGGCGCGCAGCAATGCTCCGGCGATTTTCAGCTTGTGGGTG

D T L L R V P V P I H P I F A R L L A G A I K L K H T

7881 TTGCCGGTACTGGGCACCAGGCGCAAGCGGATGGGAAAGGAAATGGCTTTGCCGGAACGGGTACGCAGGGTCAACGCCAG

N G T S P V L R L R I P F S I A K G S R T R L T L A L

7961 ACCCACCCAATTCTGGGCGAGGACATAACTGGGGCGGTTGTGCTTATGACTGTGGTCGAAGCGGATGGCCGCGCCTGGGG

G V W N Q A L V Y S P R N H K H S H D F R I A A G P

*Sal*I *Pst*I

8041 CCTTGGTAGAGTGCCGCAGGACCATCGTGTCGTCGACGACTAACTCGACGATCAGGGGCTTGGGCAAGGACTGCAGGAGC

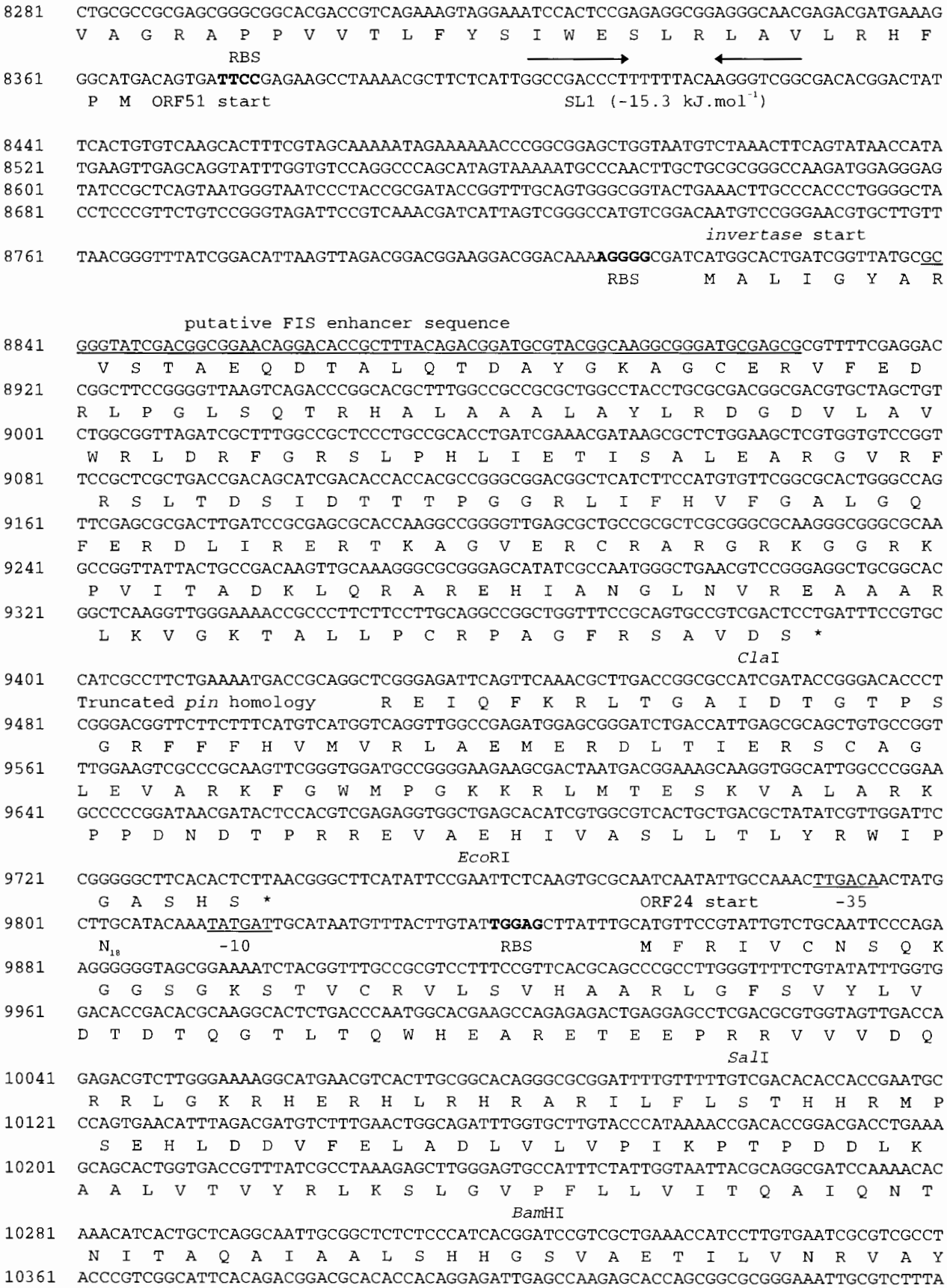
A K T S H R L V M T D D V V L E V I L P K P L S Q L L

8121 CAGCGTACCTGGGTTATGGCCAACGCCTGTGTACGCAAGGACGCACGCTCCAGGAGTTTGTAATAGGTGGTCCAGTGGCA

W R V Q T I A L A Q T R L S A R E L L K Y Y T T W H C

8201 GCGGAGGGAGATGCTGCTGAGGGCGCGGGTGACCCAGCCTTCGGGCGAGATCAGGCAACCCGCAAAGCAGCTCCACAAAA

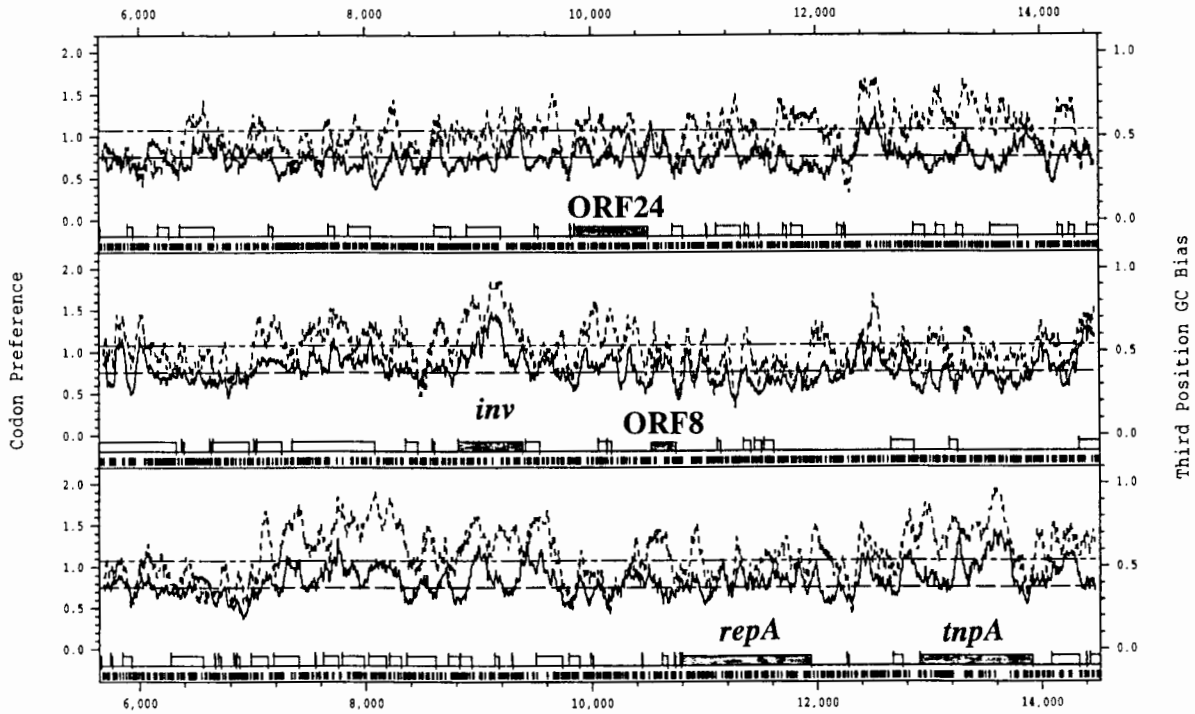
R L S I S S L A R T V W G E P S I L C G A F C S W L


RBS
P M ORF51 start
SL1 (-15.3 kJ.mol-1)
invertase start
RBS
putative FIS enhancer sequence
ClaI
Truncated pin homology
EcoRI
ORF24 start
-35
N18
-10
RBS
SalI
BamHI

10361	ACCCGTCGGCATTCACAGACGGACGCACACCACAGGAGATTGAGCCAAGAGCACCAGCGGCGCGGGAAATTGCGTCTTTA
	P S A F T D G R T P Q E I E P R A P A A R E I A S L
	*Kpn*I RBS
10441	TGGGATAACATACTATCATACTTGCATACTGGTACCGTTGAGACAAGGAGTGAAGCGCGTGGCTAAACCATCTGCTCT**AA**
	W D N I L S Y L H T G T V E T R S E A R G *
10521	**CAG**TTGACGTGATGAACGCACCGCAGGTCTCGCCTATAGGTGGCTTACCATCTCGCGGCTTTGCCGAGCCCCACCAGAAC
	ORF8 start M N A P Q V S P I G G L P S R G F A E P H Q N
10601	AAACAGCAAGCAGCACCTAAAAATGATGCGCTCGTGCAAATCCGTTGTTCTAAGGCGGCAGCTAAAGAGGTGAAGCGCGC
	K Q Q A A P K N D A L V Q I R C S K A A A K E V K R A
10681	TGCCGTTGAAGCGGAAATGACGATTAGCGAGTTCATGCTTGTATGCTTTCATGCTTATATGAAACGATGATTTTTTGTTG
	A V E A E M T I S E F M L V C F H A Y M K R *
	RBS
10761	GCATACTAGTATGCTAGTATGTGCGAGCA**AGGAT**AATAAATGGCAGACAACAATGTGACCAAGAAAAATCCCCCTACACG
	***repA* start** M A D N N V T K K N P P T R
10841	CAGATCAGTAGCGCCGGTGTCTGATACCGCTTTTGCCGGTTGGCAGCTAAGCCTTTTCCAGGGCTTCCTAGCCAACACGG
	R S V A P V S D T A F A G W Q L S L F Q G F L A N T D
10921	ACGACCAGGTCGAATCACTCTCAAACGCCGTTGACTTGTGGGACAGCATACCGCGTTACTCTATTTCACGAGCCAGAATG
	D Q V E S L S N A V D L W D S I P R Y S I S R A R M
11001	AACACCATGCGAACAGCTGATGGGTTTCTAGGCGTTGCGAGTTGTCATTCCACTATCGTGGTAGAGCATATACGGCAAGA
	N T M R T A D G F L G V A S C H S T I V V E H I R Q E
11081	ATATATCCCGCACAGGTTAAAAATTAATGATGGCCAATGGAAAAGCTATTATCCGAGCGCGCGTGAAGAGCTTGTCGAAT
	Y I P H R L K I N D G Q W K S Y Y P S A R E E L V E Y
	*Cla*I
11161	ATGCCTTGCGGAAAATCTCTGCTGAACAGGGCGCTGGCTTCTTTGATCGATCAACTTATCGCAGCGGAGCCCGATTTTCT
	A L R K I S A E Q G A G F F D R S T Y R S G A R F S
	*Eco*RV
11241	TTATATCAGCTCCGTAAAGAGTTGGAGCAACAAGGACACCAGCTTGCGTATGACCAGATCATCGAGGCGCTCGATATCCT
	L Y Q L R K E L E Q Q G H Q L A Y D Q I I E A L D I L
11321	CTCTCTGAGTAGTATCGAAATTGAATGCGCAACAGACAGTGGAGATGGGGCCTTCGCTCGCTCCACATATTTTGCTGCCT
	S L S S I E I E C A T D S G D G A F A R S T Y F A A L
11401	TAAGCGGTGTTAAACGCAAGGATTACGAAACCCATCGCGACACTCGATGGATAGCACAATTCCATCCGTTAGTGACGCAG
	S G V K R K D Y E T H R D T R W I A Q F H P L V T Q
	*Cla*I
11481	AGTATCGATCATGTGACCTATCGACAGTTTAACTATCAGAGAATGATGACATGCAGCACTCAGCTTGCCCGATGGTTGAT
	S I D H V T Y R Q F N Y Q R M M T C S T Q L A R W L I
11561	CGGACAATTGGTCTTGAAATATACCCAAGCTGCCATGCTCAACAGTTTCGAAATACGTTATAGCACCATCAAGCGTGATA
	G Q L V L K Y T Q A A M L N S F E I R Y S T I K R D S
11641	GCGCCCTATTGGCAGGGTACAAACTGGATCGTCAGGCGATAGCAGCCTTGGACCAAGCGTGGGATGAGCTCAAAAGCCTG
	A L L A G Y K L D R Q A I A A L D Q A W D E L K S L
11721	GGCGCGCTATCTACCGTAAAAAAGATCGAGCAGCGGGGAGCGCGCAGCAAACTAGAGGATGTTATTTATACGTTACATCC
	G A L S T V K K I E Q R G A R S K L E D V I Y T L H P
11801	GACGCAGGAATTTGTTGCCGAGCAGAAGGCAGCGAACCGTCGTCAGAACAATGCGAAAGATGGTATCAGCAGCGCGGTAG
	T Q E F V A E Q K A A N R R Q N N A K D G I S S A V E
11881	AGATGCAAAATCGAGTAGAGCAGCTTAACAAAAAACATCCTGTTCGTGCGACGAGACAGGGGAAATAGGTGACGGGTCAG
	M Q N R V E Q L N K K H P V R A T R Q G K *
11961	CCCTCCGGACAGGGGAAATAGGTGACGGGTCGGCCCTCCGGACAGGGGAAATAGGTGACGGGTCGGCCCTCCGGACAGGG
	1 2 3
12041	GTTAGTTTTTAACAGGGGAAATAGGTGACGGGTCAGCCCTCCGGACAGGGGAAATAGGTGACGGGTCGTCAAAAAGACAG
	4 5
12121	GGGAAATAGGTGACGGGTCGTCAAAAAGACAGGGGAAATAGGTGACGGGTCGTCAAAGAATCGCTCTCAATGACCCCCAA
	6 7
12201	TC**TGTGGATAA**CTGCTAAGTCTTTGCCTTAAATTCATACCTCGGCAGGGGAAATGGGTGACGGATGAAAATCA**TAACAAA**
	DnaA consensus

```
                                                                             PacI
12281  ATAATTTTGTTGGAAAATTTCGCCATTTTTCAGCCCTTATCCTTTTTATCCTTTAATTAACCTTCTTTTATCCTTAGGCT
       IHF consensus                                    -------->  -------->  -------->  -------->
                                                            a          b          c          d

12361  TCGCAAACTCCGCCGCCGTCTTCGTGTCGCCCCCACCGACTCGAACCCACGCCCTTCTGCTACGCGTCCGCCTGCGTGTC
12441  CCACCCGGCACCTCCCAACTCTTTAAAGAGCGCCCACTTGCCTCCGGCCAGGCGCGAACATCCCGGCGGGACCCAGCAGG
                                          SalI
12521  TGCCCCTGTTCCCGCTTCGGCTTGCCGGGTCGACCCCCTTTGGCTGCGTCCCCCAACTCCTATACGGCAGCGGCTCCCCC
12601  TAAAATTTTCCAAAAAGCGGGTGATTTTAGCTCAGGGCGCATTTTTGGTGACCGGGTGATGGAAATATACCGGAATGGAA
12681  TTTGCGTCGCTTCTAGGGCAAATACGGGCTATTTACGGCGGGGTTCTGCATCAGGCTTCTGGTCGGTTGTGCTGCGAATG
12761  ATGCTTGGGACCGGCCTTGCTTGCGTTCGTTACCGTCACCGTGCGCTGCTGACCGGCTTTCTTCGAGTGCAAAGTGAGCG
                             XbaI                                         RBS
12841  GTTTCTACTGGGGGGGCGCGCTCTAGAATGAAGTGCAACATCTCCGGTCGAAGGATGGCGGAGGTGAGTGAATGGGGACG
        EcoRV                                                       tnpA start  M  G  T
12921  AGATATCAGCAGTTGCAGTCGGAGCAGAGGAACCAGATTCAGCGAGGGCTGAACGAGGGGTTGAGTATGCGGGCCGTGGC
       R  Y  Q  Q  L  Q  S  E  Q  R  N  Q  I  Q  R  G  L  N  E  G  L  S  M  R  A  V  A
13001  CAAGCAGATAGGGAGGAGTCCCAGCACGGTCAGCCGGGAGGTACGCCGGGGTTTGGTGGGAGAAACCTACGATGCGATAC
        K  Q  I  G  R  S  P  S  T  V  S  R  E  V  R  R  G  L  V  G  E  T  Y  D  A  I  Q
13081  AAGGCCGGGAGGGACGGCGCAGAGGCGTTCGTAAGGGGGTTAGAAAGCTGGTGGGAGGCGCGCCCTTAACCAACGCGGTG
         G  R  E  G  R  R  R  G  V  R  K  G  V  R  K  L  V  G  G  A  P  L  T  N  A  V
13161  ACACACGCTATCCTGCAAAGGAAATGGTCACCAGAGCAGGTGGCCGGGAGGTTGCGGATGGACTATCCCGAGGACAAGCA
       T  H  A  I  L  Q  R  K  W  S  P  E  Q  V  A  G  R  L  R  M  D  Y  P  E  D  K  Q
13241  GTGGCGTGTCTCCCATGAGACCATTTATCAGTTCATCTATGCCCACCCGGCCGGTGAGCTGCGTAAGGCGCTGATAGCGG
        W  R  V  S  H  E  T  I  Y  Q  F  I  Y  A  H  P  A  G  E  L  R  K  A  L  I  A  A
13321  CGCTACGCCAGGGGACACGCAAAGCGCAAGCGCGCACACGCGGAAAGGACCGGCGCGGACAACTGCGGAACATGCGTTCC
         L  R  Q  G  T  R  K  A  Q  A  R  T  R  G  K  D  R  R  G  Q  L  R  N  M  R  S
13401  ATCGGGGAGCGTCCCTTGGAGGCCCAAGACCGCGAGATACCCGGCCACTGGGAAGGAGACTTCATCAAAGGGGCTTTCAA
       I  G  E  R  P  L  E  A  Q  D  R  E  I  P  G  H  W  E  G  D  F  I  K  G  A  F  N
13481  CGGCAGCGCCATTGGTACTCTGGTGGAGCGCAGCAGCCGTTTCGTGCTTCTGGTCAGGATGGAAGGCACCGATGCCGACG
        G  S  A  I  G  T  L  V  E  R  S  S  R  F  V  L  L  V  R  M  E  G  T  D  A  D  A
                                     SphI
13561  CGGCCCTGGAGGGGTTCACCAGGCGCATGCCCTTGCCCAAGTCCATCCTGCGGACCCTCACCTATGACCAGGGCAAGGAG
         A  L  E  G  F  T  R  R  M  P  L  P  K  S  I  L  R  T  L  T  Y  D  Q  G  K  E
13641  ATGGCACGGCACGAGGAGCTGGAGCGCAAGGTGGGCATCCGTATCTACTTTGCCGACCCGCATAGTCCTTGGCAGCGCCC
       M  A  R  H  E  E  L  E  R  K  V  G  I  R  I  Y  F  A  D  P  H  S  P  W  Q  R  P
13721  AACCAACGAGAACACCAATGGTCTCCTGCGCCAGTATCTTCCCCAAAGGACGGATTTATCAGGATATTCACAACGCCGCT
        T  N  E  N  T  N  G  L  L  R  Q  Y  L  P  Q  R  T  D  L  S  G  Y  S  Q  R  R  L
13801  TGACGCAGGTGGCGGAAGAGCTCAACAATCGCCCAAGAAAATCTTTGGGATTCCGAACGCCAGCAGAAGTAATAGCACAG
         T  Q  V  A  E  E  L  N  N  R  P  R  K  S  L  G  F  R  T  P  A  E  V  I  A  Q
13881  CAAATCATGCAGTTAAACAGTGGTGTTGCGCTTCAAATTTGAAACCGCCGGGTGTAAGCCGGAATCCCAGATTTTTTCGG
       Q  I  M  Q  L  N  S  G  V  A  L  Q  I  *
13961  CAGGAGGTTCTGCGCCAGAATCGCGGGGTTTGTCCGCTTCCGGACGGTTTGACGACTCTGAAACGCTTGCCCCTGCTCGA
14041  TCTGTGTAGTAACGTATCTACATTGATCCAGCATACGGGGTGCCATCATGACCATTACTACCTTATCCAGCCGCGAGTTC
14121  AACCAGGGGGCGAGCCAGGCAAAACGGGCCGCAAACAATGGACCAGTGTTCATCACCGATCGAGGCCGACCGGCCCATGT
14201  GCTGATGAGCTTTGAGGATTATCAGCGGCTCACCAAGCAGCGACGCAACATTGCCGATGCGCTGGCTATGCCGGGTATTG
14281  CCGACATCGAGTTCGAGCCGCCGCGTGTGACGATTGAAACGCGTCCGGCGGATTTCTCATGAGGTTTGTTCTCGATACCA
14361  ACGTGGTGTCTGAACTGCGCAAGGTGCGGCTTGGTAAGGCCGATGTGAACGTGACGGCATGGACGGAAAGCGTGGATGCC
14441  GCCGATCTCTTTGTGTCAGCCATCACCATCCTGGAACTGGAGCTTGGCGTTCTGTCGATGGAGCGAAAGGACGCCACCCA
         ApaI
14521  GGGGCCC
```

Forward strand



Reverse strand

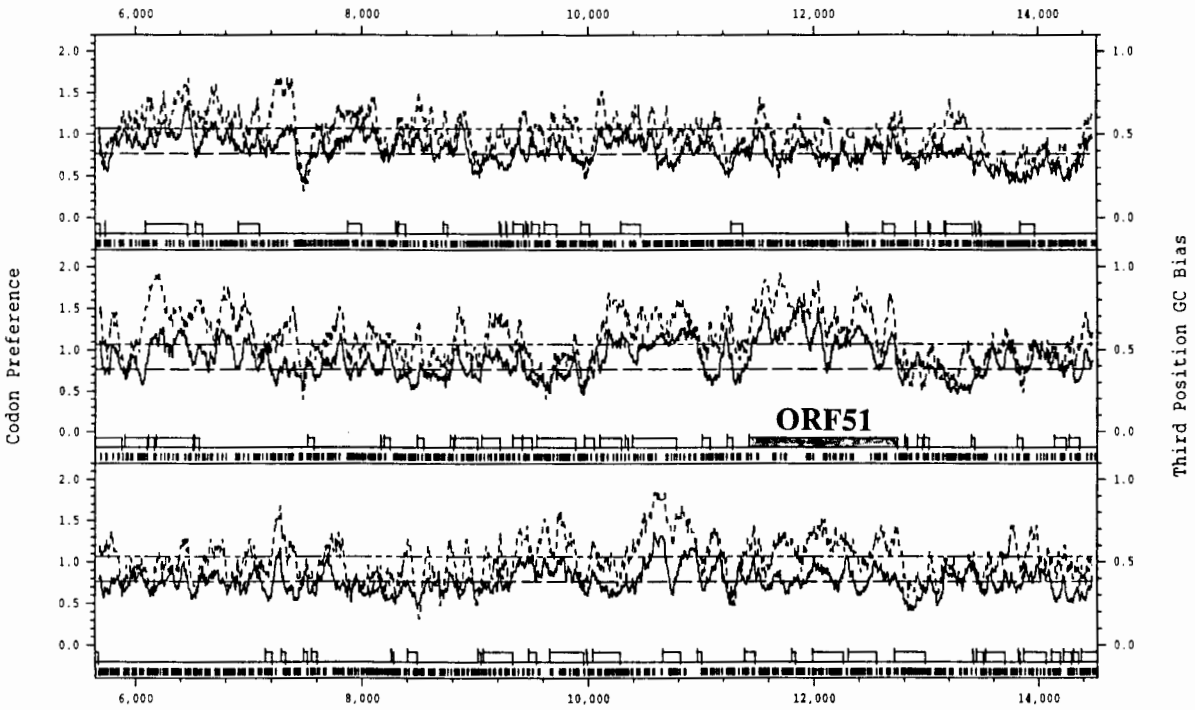


Figure 2.10a shows the predicted open reading frames of the 8.9 kbp *Sph*I-*Apa*I fragment using the codon preference table and GC bias information derived from previously sequenced *T. ferrooxidans* chromosomal genes (Rawlings et al. 1991). A physical map (Figure 2.10b) shows the appropriate restriction endonuclease sites along with a summary of the ORFs.

b.

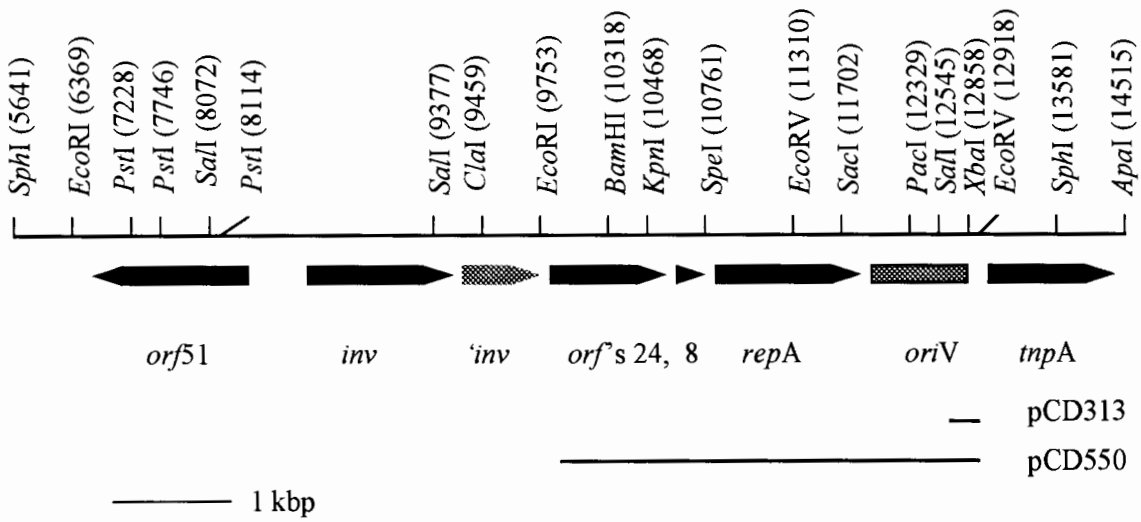


Figure 2.10a (opposite) : CODONPREFERENCE analysis of the 9 kbp *Sph*I-*Apa*I restriction fragment identifying the predicted ORFs according to codon preference tables for *T. ferrooxidans* chromosomal genes (Rawlings et al. 1991). b : Physical map of the pTF5 replicon and flanking sequence showing the open reading frames and appropriate restriction sites. The restriction sites have been numbered relative to the first base of the *Hin*dIII restriction site with which pTF5 was cloned.

ORF51 was identified as a 1317 bp open reading frame based on *T. ferrooxidans* codon preference patterns (Figure 2.10a). This ORF was found on the complimentary DNA strand to the other genes of the pTF5 replicon (Figure 2.10b) and coded for a predicted protein of 439 amino acids (Mr = 50.8 kDa) . Database searches using the BLAST program (Altschul et al. 1990) did not show significant overall homology to other genes

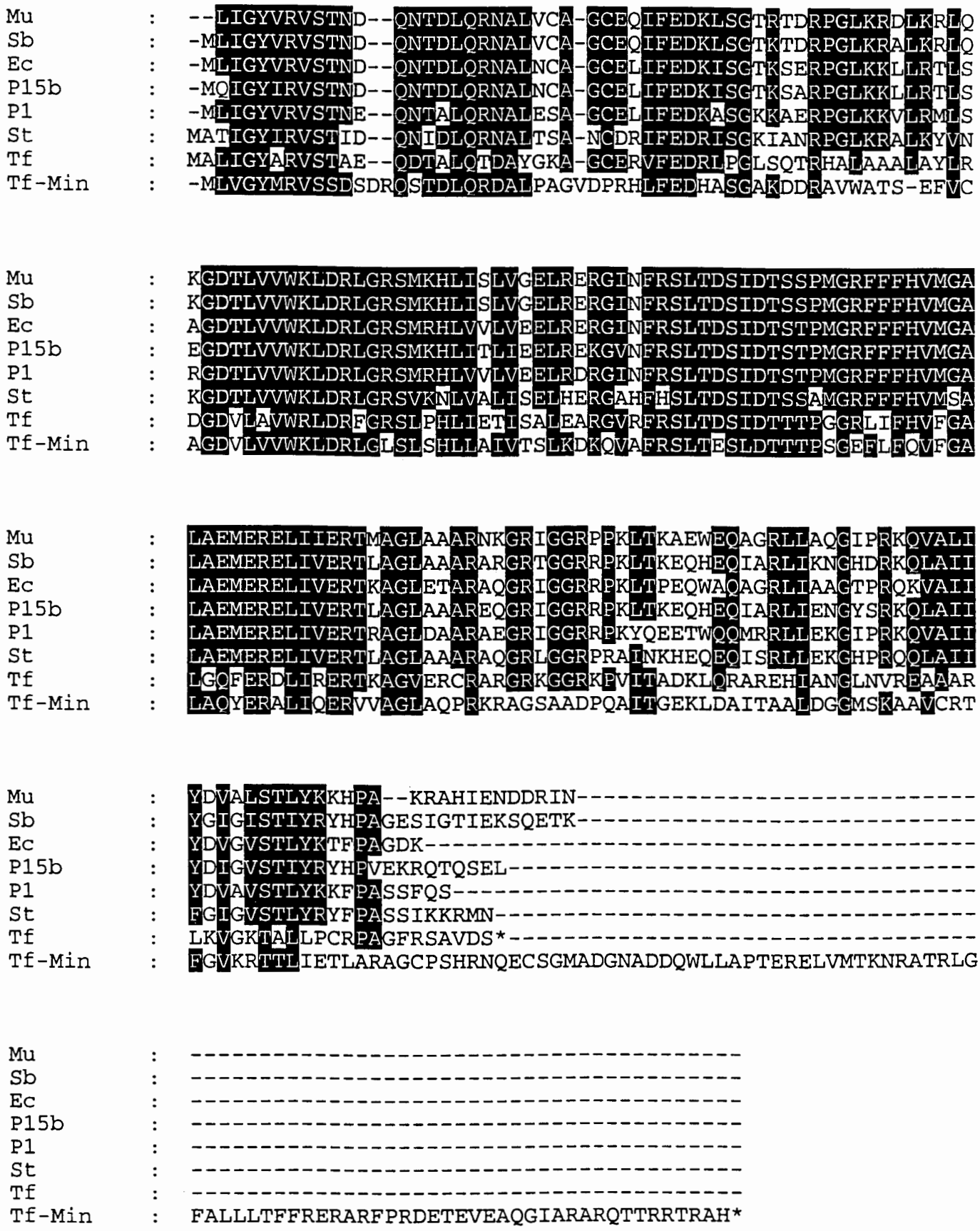
or proteins in the GenEMBL database. Analysis of the various predicted domains of ORF51 using the BLAST algorithm with BEAUTY post-processing (Worley et al. 1995) showed that ORF51 deduced amino acid sequence contained transposase-like protein motifs (Figure 2.11). These motifs were similar to those from transposases reported in *Bacillus thuringiensis* insertion elements (Rezsohazy et al. 1992). As no other IS-related features, such as target site duplications or terminal inverted repeats, were observed, this was the only evidence to the potential function of ORF51 in pTF5.

```
pTF5 ORF51     321 YSNRWQIEPIFHNLKRWWGIHNLWQQKRTVLE 352
                   YS RWQIE IF   K  + IH+    KR  LE
Bt Tnp         342 YSLRWQIEIIFKTWKSLFQIHDWQNIKRERLE 373
```

Figure 2.11 BEAUTY analysis (Worley et al. 1995) of ORF51 showed the similarity of transposase protein motif of IS231 from *B. thuringiensis* (Rezsohazy et al. 1992) to the predicted ORF51 amino acid sequence.

Structural features were noted which may play regulatory roles in the expression of ORF51. Predicted stem-loop structures, as well as direct repeated sequences were identified using the DNAMAN sequence analysis software. The first potential stem-loop structure (SL1) was found 25 bp upstream of the predicted "GGAA" ribosome binding site of ORF51 (Figure 2.9). It has a 9 bp stem structure with an 8 bp loop, and a predicted Gibbs free energy of -15.3 $kJmol^{-1}$. A second stem-loop structure (SL2) was identified 330 bp downstream of the predicted ORF51 stop codon. It has a predicted Gibbs free energy of -10.6 $kJmol^{-1}$.

Upstream of the ORF51 open reading frame was the start of a 591 bp open reading frame with homology to genes encoding invertase/ recombinase enzymes from a number of bacterial and bacteriophage systems. The invertase enzyme is responsible for inverting large segments of DNA such as the *Neisseria gonnorhea* antigen switching mechanism used in avoiding human immune responses (Hagdom et al. 1985). The invertase-like ORF of pTF5, contains a region with homology to a FIS site (factor for inversion stimulation) (Figure 2.12a) which in other invertase systems functions as a


Mu : --LIGYVRVSTND--QNTDLQRNALVCA-GCEQIFEDKLSGTRTDRPGLKRDLKRLQ
Sb : -MLIGYVRVSTND--QNTDLQRNALVCA-GCEQIFEDKLSGTKTDRPGLKRALKRLQ
Ec : -MLIGYVRVSTND--QNTDLQRNALNCA-GCELIFEDKISGTKSERPGLKKLLRTLS
P15b : -MQIGYIRVSTND--QNTDLQRNALNCA-GCELIFEDKISGTKSARPGLKKLLRTLS
P1 : -MLIGYVRVSTNE--QNTALQRNALESA-GCELIFEDKASGKKAERPGLKKVLRMLS
St : MATIGYIRVSTID--QNIDLQRNALTSA-NCDRIFEDRISGKIANRPGLKRALKYVN
Tf : MALIGYARVSTAE--QDTALQTDAYGKA-GCERVFEDRLPGLSQTRHALAAAALAYLR
Tf-Min : -MLVGYMRVSSDSDRQSTDLQRDALPAGVDPRHLFEDHASGAKDDRAVWATS-EFVC
Mu : KGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFHVMGA
Sb : KGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFHVMGA
Ec : AGDTLVVWKLDRLGRSMRHLVVLVEELRERGINFRSLTDSIDTSTPMGRFFFHVMGA
P15b : EGDTLVVWKLDRLGRSMKHLITLIEELRERKGVNFRSLTDSIDTSTPMGRFFFHVMGA
P1 : RGDTLVVWKLDRLGRSMRHLVVLVEELRDRGINFRSLTDSIDTSTPMGRFFFHVMGA
St : KGDTLVVWKLDRLGRSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFHVMSA
Tf : DGDVLAVWRLDRFGRSLPHLIETISALEARGVRFRSLTDSIDTTTPGGRLIFHVFGA
Tf-Min : AGDVLVVWKLDRLGLSLSHLLAIVTSLKDKQVAFRSLTESLDTTTPSGEFLFQVFGA
Mu : LAEMERELIIERTMAGLAAARNKGRIGGRPPKLTKAEWEQAGRLLAQGIPRKQVALI
Sb : LAEMERELIVERTLAGLAAARARGRTGGRRPKLTKEQHEQIARLIKNGHDRKQLAII
Ec : LAEMERELIVERTKAGLETARAQGRIGGRRPKLTPEQWAQAGRLIAAGTPRQKVAII
P15b : LAEMERELIVERTLAGLAAAREQGRIGGRRPKLTKEQHEQIARLIENGYSRKQLAII
P1 : LAEMERELIVERTRAGLDAARAEGRIGGRRPKYQEETWQQMRRLLEKGIPRKQVAII
St : LAEMERELIVERTLAGLAAARAQGRLGGRPRAINKHEQEQISRLLEKGHPRQQLAII
Tf : LGQFERDLIRERTKAGVERCRARGRKGGRKPVITADKLQRAREHIANGLNVREAAAR
Tf-Min : LAQYERALIQERVVAGLAQPRKRAGSAADPQAITGEKLDAITAALDGGMSKAAVCRT
Mu : YDVALSTLYKKHPA--KRAHIENDDRIN----------
Sb : YGIGISTIYRYHPAGESIGTIEKSQETK----------
Ec : YDVGVSTLYKTFPAGDK----------
P15b : YDIGVSTIYRYHPVEKRQTQSEL----------
P1 : YDVAVSTLYKKFPASSFQS----------
St : FGIGVSTLYRYFPASSIKKRMN----------
Tf : LKVGKTALLPCRPAGFRSAVDS*----------
Tf-Min : FGVKRTTLIETLARAGCPSHRNQECSGMADGNADDQWLLAPTERELVMTKNRATRLG
Mu : ----------
Sb : ----------
Ec : ----------
P15b : ----------
P1 : ----------
St : ----------
Tf : ----------
Tf-Min : FALLLTFFRERARFPRDETEVEAQGIARARQTTRRTRAH*

recombinational enhancer sequence. Homology of the predicted pTF5 *inv*-like protein to other members of this class of enzyme, is shown in Figure 2.12b.

```
FIS     CGCGTATCAACAAATGAACAAAACACTGCTTTACAACGAAA
        ||  |||||  ||    |||||  ||||  |||||||    |
pTF5    CGGGTATCGACGGCGGAACAGGACACCGCTTTACAGACGGA

FIS     CGCTCTTGAAAGCGCAGGATGTGAGCT
        ||    | ||||||  |||||  ||||
pTF5    TGCGTACGCAAGCGCGGGATGCGAGCG
```

Figure 2.12a (opposite) : Comparison of the pTF5 invertase-like ORFs putative FIS enhancer sequence to the consensus FIS sequence of the Din family of invertase/recombinases. (Glasgow et al. 1989) b (opposite): Multiple sequence alignment of the *T.ferrooxidans inv*-like gene with its nearest homologs. Shown here are *S. typhimurium* Hin (St), bacteriophage Mu Gin (Mu), *Shigella boydii* PinB (Sb), bacteriophage P1 Cin (P1), *E. coli* Pin (Ec), *E. coli* p15B Min (p15B), *T. ferrooxidans* Min (Tf-Min) and *T. ferrooxidans* Inv (Tf).

Immediately downstream of the invertase-like ORF is a second invertase-like ORF which had the initial 282 base pairs (94 aa) of the 588 bp ORF deleted. The truncated invertase-like ORF was not identical to the intact ORF at either the protein or nucleic acid level, and consequently is unlikely to be the result of a recent gene duplication event. This organisation of a complete invertase ORF immediately upstream of a truncated invertase ORF has been reported in similar systems (Galas and Chandler, 1989).

Downstream of the invertase-like ORFs are three genes which are highly expressed in *E. coli* (Section 2.3.5). The genes are found on a 3111 bp *Eco*RI-*Xba*I DNA (pCD550) fragment which also included the previously described metronidazole sensitising fragment (Figure 2.10b). The ORF which extended for 654 bp and predicted to code for a 24.3 kDa polypeptide was named ORF24. It had amino acid sequence homology to a number of proteins believed to be involved in partitioning functions in other plasmids

such as the ParA protein from the pTAR plasmid of *Agrobacterium tumefaciens* (Gallie and Kado, 1987). Interestingly it also had sequence similarity to the N-terminal domain of the MinD enzyme which is involved in septum site determination in *Synechocystis* sp.. This domain contains a conserved ATP/GTP binding site motif "KGGSGKSTV" (Figure 2.13). Other proteins homologous to the ORF24 gene product include the VirC1 proteins from *Agrobacterium rhizogenes* Ri plasmid, as well as *A. tumefaciens* Ti plasmid homolog (Gallie and Kado, 1987). The VirC1 enzyme is believed to be involved in nicking of the pRi or pTi plasmids during conjugal transfer (Hirayama et al. 1988).

```
pTF5 ORF24  4  IVCNSQKGGSGKSTVCRVLSVHAARLGFSVYLVDTDTQGTLTQWHEARETEE
               +V  S KGG+GKST   VL    A  G  V ++D D   +LT W  A E  E
pTAR ParA   4  VVVASSKGGAGKSTTAVVLGTELAHKGVPVTMLDCDPNRSLTIWANAGEVPE

ORF24     101  ADLVLVPIKP
               ADLVL+P++P
ParA      101  ADLVLIPMRP

ORF24     135  VITQAIQNTNITAQAIAALSHHGSVAETILVNRVAYPSAFTDG
               +++ AI++   T    + + +   + E  LV R AY + F  G
ParA      144  MVSPAIRSHEYTGIKASLIENGVEIIEPPLVERTAYSALFQFG
```

Figure 2.13 : Regions of aa sequence similarity between the predicted ORF24 protein and the ParA protein from pTAR plasmid of *A. tumefaciens* (Gallie and Kado, 1987) as identified by the BLAST-BEAUTY algorithm. A predicted ATP/GTP binding domain (residues 9-16 in box) was identified by the PROSITE protein motif database.

Analysis of the region upstream of ORF24 revealed the presence of a putative *E. coli* σ^{70} sigma factor promoter recognition sequence. The -35 sequence "TTGACA" had perfect consensus to this type of sigma factor recognition sequence, whilst the -10 region "TATGAT" had 5/6 bp homology to the consensus sequence. Furthermore the spacing between the two elements is 18 bp which falls within the $N_{16\text{-}18}$ range typical for σ^{70} promoters (Harley and Reynolds, 1987). No other promoter-like sequences could be

identified. It is possible that this promoter may be responsible for expressing the three ORFs immediately downstream of it.

An ORF of only 216 bp which may code for a 7.9 kDa (73 aa) polypeptide was identified downstream of ORF24. It was designated ORF8 on the basis that it might have been responsible for a 8 kDa polypeptide detected in *in vitro* transcription - translation reactions which had been mapped to this region (Section 2.3.5). ORF8 was not homologous to any previously described sequence in the nucleic acid or protein databases. Examination of the predicted polypeptide domains using BLAST / BEAUTY computer analysis showed that it shared a helix-turn-helix motif with the *Bacillus subtilis* DNA gyrase B (Moriya et al. 1985)

```
pTF5 ORF8  4  PQVSPIGGLPSRG 16
              PQ+SP GG PS G
Bs        48  PQLSPAGGRPSPG 60
```

Figure 2.14 : Analysis of the N-terminal region of the ORF8 predicted protein showed sequence identity to a number of DNA-binding proteins, represented here by a protein domain of the *B. subtilis* DNA gyrase B enzyme (Bs).

The 1148 bp *rep*A gene of pTF5 was located downstream of ORF8. It has a putative RBS sequence and a predicted polypeptide M(r) of 43.5 kDa. The putative RepA protein was homologous to a RepA protein from an *E. coli* broad host-range Inc A/C plasmid, RA1 (Llanes et al. 1994) as is shown in Figure 2.15. GAP analysis of the two predicted proteins gave an amino acid sequence identity of 24.7% and a similarity of 48.8%. These RepA proteins appear to be poorly conserved between different plasmids as there was a similar identity (20.3%) between the RA1 RepA protein and that from *Synechocystis* sp. plasmid pUH24.

```
pTF5     DSIPRYSISRARMNTMRTADGFLGVASCHSTIVVEHIRQEYIPHRLKINDGQ
         ||||||     |                |          ||   |   |
RA1....  DALPKYIWDQKREHEDLSNAVVTRQCSIRGQQFTVKVKP-AIIEKDDGRTV-

pTF5          WKSYYPSAREELVEYALRKISAEQGAGFF
                   |    |||  |  ||||   |   | |
RA1           --LIYAGQREEILEDALRKL-AVNGKGHIIE
```

Figure 2.15 : Amino acid sequence similarity between regions of the RepA homologs of plasmid RA1 from *E. coli* (Llanes et al. 1994), and pTF5 from *T. ferrooxidans*

```
Long repeated sequences

1     GACAGGGGAAATAGGTGACGGGTCAGCCCTCCG
2     GACAGGGGAAATAGGTGACGGGTCGGCCCTCCG
3     GACAGGGGAAATAGGTGACGGGTCGGCCCTCCG
4     AACAGGGGAAATAGGTGACGGGTCAGCCCTCCG
5     GACAGGGGAAATAGGTGACGGGTC
6     GACAGGGGAAATAGGTGACGGGTCGTCAAAAA
7     GACAGGGGAAATAGGTGACGGGTCGTCAAAGAA

Short repeated sequences

a     TTATCCTTT
b     TTATCCTTT
c     TTAACCTTC
d     TTATCCTTA
```

Figure 2.16 : Plasmid pTF5 *ori*V direct repeats. The seven long repeated sequences contain a highly conserved 24 bp core repeat, followed by an imperfectly conserved 9 bp region. The four short repeated sequences have 7/9 bp conserved within the region.

The origin of replication of pTF5 is immediately downstream of the *rep*A gene. This sequence was highly homologous (98%) to that of pTFI91 (Chakravarty et al. 1995). It consisted of seven direct repeats (iterons) which usually function (in other *oriV*'s) as binding sites for the RepA protein (Kues and Stahl, 1989). Also in common with many

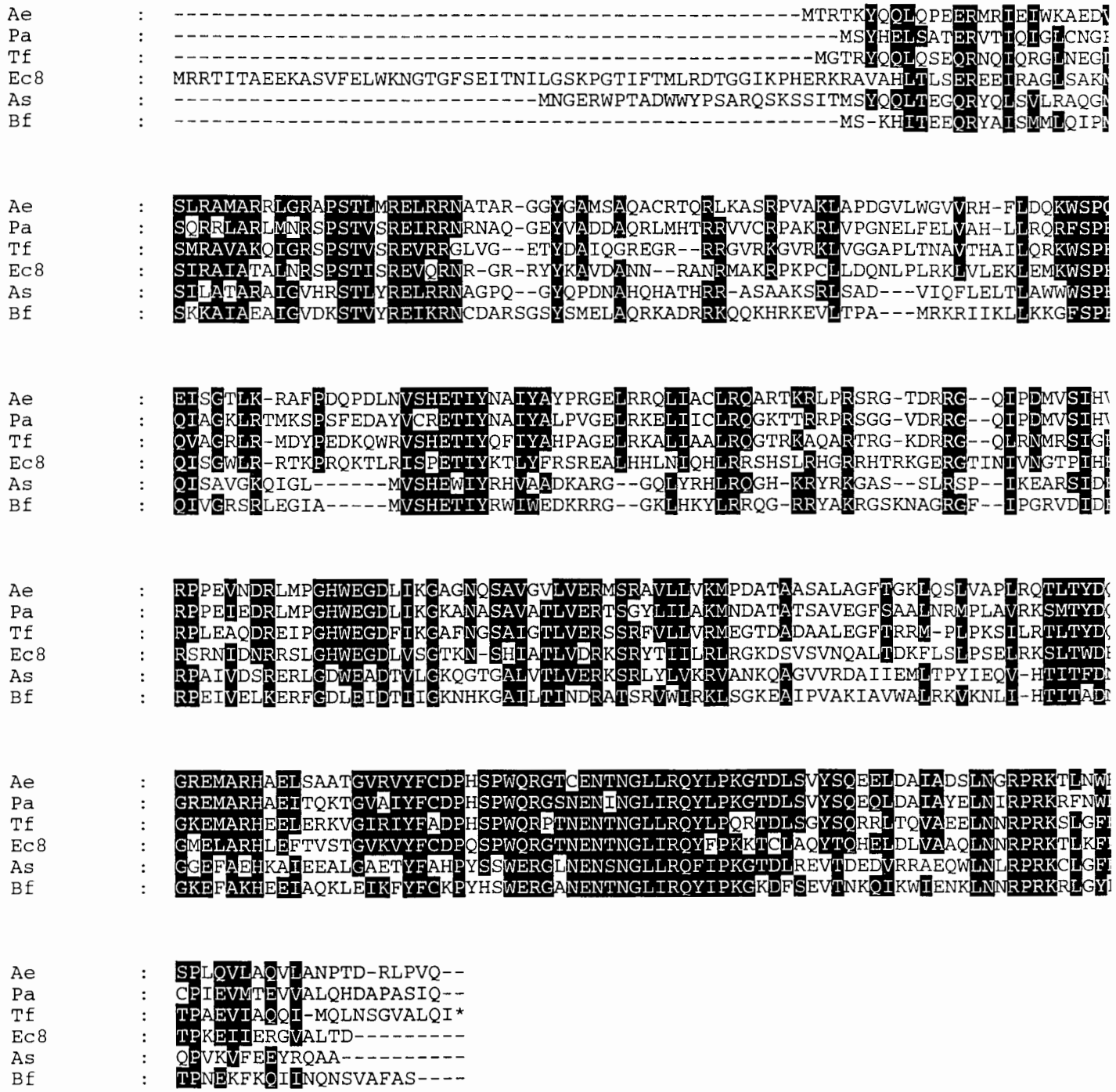
other plasmid origins of replication was the presence of regions with homology to binding sites for the DnaA protein and integration host factor (IHF). This was followed by a series of shorter imperfectly conserved direct repeats (Figure 2.16). Other features common to *oriV*'s was a region of unusually high AT mol % followed by a region of high GC mol % (Figure 1.1, Chakravarty et al. 1995). This is believed to be the region of DNA melting during the initiation of plasmid replication (Thomas and Helinski, 1989)

The 319 bp *Sal*I-*Xba*I fragment which was able to confer metronidazole sensitivity to *E. coli* F19 was located in the high GC mol% region. Analysis of this sequence did not reveal obvious palindromic sequences suitable for DNA binding proteins, or any other unusual features capable of forming secondary structures. Short direct repeated sequences were observed, and consisted of two 4 bp "CTTG" repeats followed by another two 4 bp "CGTT" repeats followed by a final two 5 bp "ACCGT" repeated sequences. The first two sets of repeats in themselves form a "reverse repeat" (Figure 2.17). Downstream of these repeats is another "reverse repeat" sequence. Whether these features have any special biological significance is uncertain.

```
12681    TTTGCGTCGCTTCTAGGGCAAATACGGGCTATTTACGGCGGGGTTCTGCATC

                                                  ⟶  ⟵
AGGCTTCTGGTCGGTTGTGCTGCGAATGATGCTTGGGACCGGCCTTGCTTGCGTTCGTTA
                                                  ⟶ ⟶ ⟶ ⟶

CCGTCACCGTGCGCTGCTGACCGGCTTTCTTCGAGTGCAAAGTGAGCGGTTTCTACTGGG

          XbaI
GGGGCGCGCTCTAGA
```

Figure 2.17 : Direct and reverse repeated sequence within the 319 bp metronidazole active fragment clone (pCD313). The region shown is from position 12681 of pTF5.

An ORF with sequence homology to the transposase gene, *tnpA* from a number of insertion sequence elements (Figure 2.18) was located downstream of the pTF5 *oriV*.


Ae
Pa
Tf
Ec8
As
Bf

This layout was identical to that described for the pTFI91 plasmid (Chakravarty et al.1995). Insertion elements are typically 1.0 - 2.5 kbp in size and have terminal inverted repeat sequences, which recombine into host DNA producing short tandemly repeated target sequences (Galas and Chandler, 1989). Analysis of the region both upstream and downstream of the *tnpA* gene did not reveal any suitable terminal repeats, which would determine the extent of the insertion element in pTF5.

Figure 2.18 (opposite) : Multiple sequence alignments of the *tnpA* gene of plasmid pTF5 against the TnpA homologues (and accession numbers) from IS*1086* (P37248) of *Alcaligenes eutrophus* (Ae), *Pseudomonas alcaligenes* - P37284 (Pa), *T. ferrooxidans* IS*3091* - U73041 (Tf), *E. coli* IS*30* - P37246 (Ec8), *Aeromonas salmonicida* - S46409 (As) and *B. fragilis* IS*4351* - P37247 (Bf).

The remaining 4.5 kbp of the pTF5 genome was sequenced as part of another study (N. Coram, Bsc(Hons) thesis, University of Cape Town, 1996). This region contains an ORF with 37% aa identity to the *stbB* gene product from *Pseudomonas syringae* insertion sequence IS*1240* (Hanekamp et al. unpublished) believed to have a role in plasmid stability. Downstream of this lay a large 2.7kbp ORF which had 57% aa sequence similarity to large bacterial transposases (Chapter 5). Completing the sequence of pTF5 was an ORF with homology to a second invertase gene (*min*-like). It had 40% amino acid identity to the previously described *pin*-like gene, and 52% nucleic acid identity, indicating that it was probably independantly derived from the *pin*-like gene, and not merely the result of a recent duplication event.

2.3.5 *In vitro* protein analysis of pTF5 replicon

The ORFs identified in this region of plasmid pTF5 had, with the exception of ORF8, been identified by either database sequence homology or codon preference patterns (Section 2.3.4). In order to determine whether these ORFs could be expressed in *E. coli, in vitro* transcription-translation reactions (Promega) were used to detect protein products

after SDS-PAG electrophoresis (as described in Appendix C). Figure 2.19 illustrates the physical map of this region along with key subclones generated in order to map putative proteins produced by the *in vitro* reactions onto the DNA sequence.

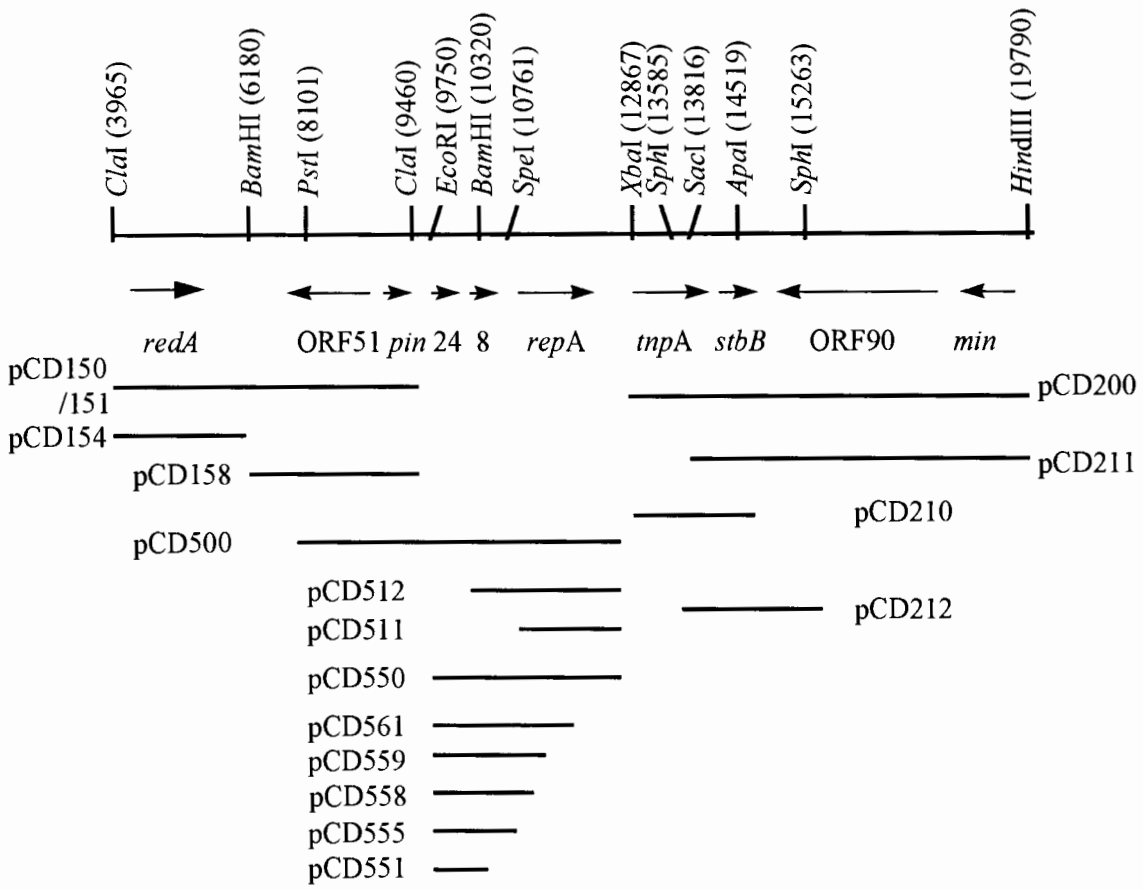


Figure 2.19 : Physical map of the region of pTF5 containing ORFs with homology to replication, recombination and transposition enzymes. Relevant subclones from this region used in the *in vitro* protein analysis are represented as black bars beneath the map. The *in vitro*-derived proteins corresponding to the *redA* gene product are discussed in Chapter 3.

No protein products could be detected for either the ORF51 or pin-like open reading frames (Figure 2.20). Clones pCD150 (lane 2) and pCD151 (lane 3) contain the same *Cla*I restriction fragment which spanned the region encoding ORF51 inserted into pBS-KS in opposite orientations. A *Bam*HI site, internal to the *Cla*I fragment was used to clone the two independent *Cla*I-*Bam*HI fragments, pCD154 (lane 4) and pCD158 (lane5).

Comparison of the putative pTF5 proteins produced by the *in vitro* reactions to the pBS-KS control (lane 1), show that there are no proteins produced in pCD158 (lane5) which contains both ORF51 and the *inv*-like open reading frames. Clone pCD150 and its derivative pCD154 appear to express proteins of 37 and 39 kDa under the control of the *lac*Z promoter of the vector, but this was from a region which will be discussed in Chapter 3.

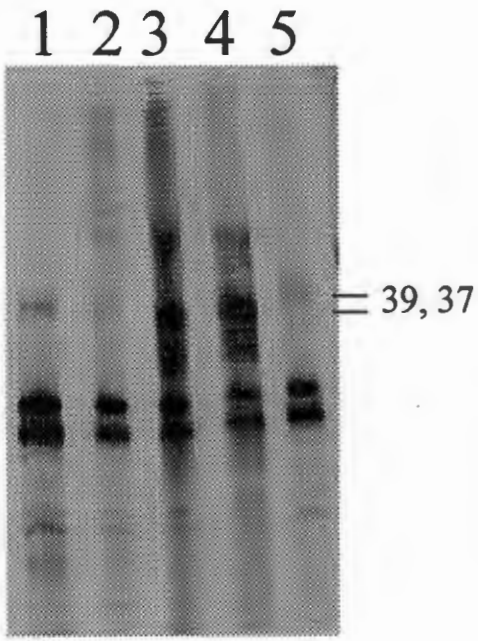


Figure 2.20 : *In vitro* protein analysis of the DNA fragment containing ORF51 and the invertase open reading frame. Lane 1 contains pBS-KS control; lane 2, pC150; lane 3, pCD151; lane 4, pCD154 and lane 5, pCD158. Molecular weight standards (in Kd) are marked with bars. The polypeptides identified by solid bars in the pCD154-derived sample corresponded to ORFs lying outside the region covered in this chapter, and will be discussed in chapter 3.

Clone pCD550, a 3.1 kbp *Eco*RI-*Xba*I fragment downstream of the pin-like ORF was used to investigate *E. coli*-mediated protein expression in the region upstream of the *oriV*. Three highly expressed proteins with estimated sizes of 40 kDa, 26 kDa and 9 kDa were

detected using the *E. coli* system (Figure 2.21). The most probable candidate promoter for the expression of the three polypeptides is the *E. coli* σ^{70}-like promoter recognition sequence identified in Section 2.3.4. Figure 2.21a lane 4 contains *in vitro* translation products from clone pCD550. Exonuclease III digestion from the 3' end of the ORFs (Section 2.2.5) showed that progressively larger nested deletions contain increasingly fewer or truncated proteins (lanes 1,2,5,6,7,8). The size of the C-terminal deleted proteins allowed the mapping of the putative pTF5 products onto the physical plasmid map (Figure 2.19).

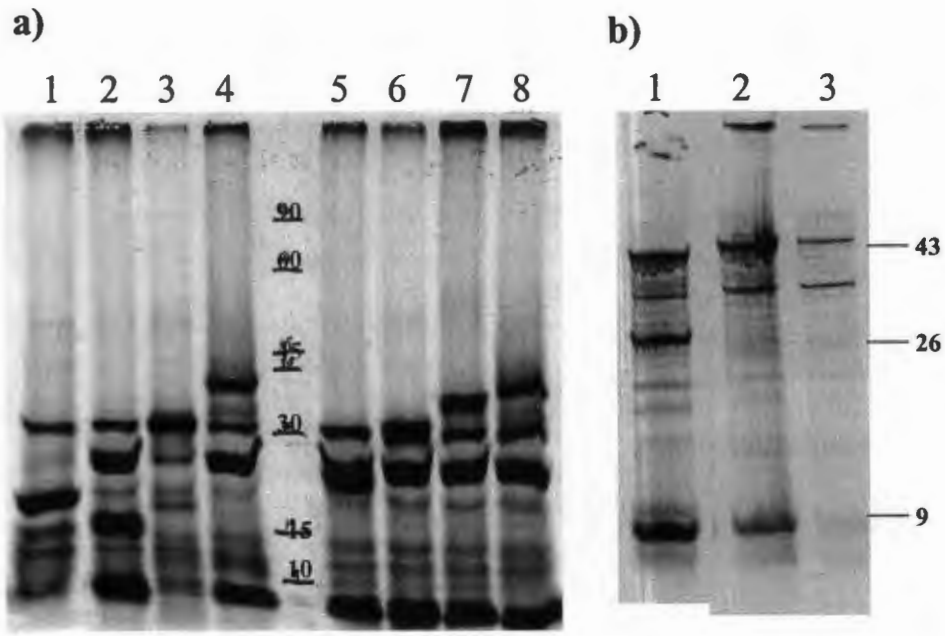


Figure 2.21 : Panel a) contains *in vitro* reaction products of DNA fragments generated from clone pCD550 (lane 3) by exonuclease III digestion. Lane 1 contains pCD551; lane 2, pCD555; lane 3, pBS-KS; lane 4, pCD550; lane 5, pCD558; lane 6 pCD559; lane 7, pCD561. Panel b) - Lane 1, pCD500; lane 2, pCD512; lane 3, pCD511. Protein molecular weights are indicated.

The results were consistent with the sequence positions and coding direction of ORFs 24, 8 and the *rep*A gene. *In vitro* protein analysis of Exonuclease III generated clones indicated that it mapped to a position between *rep*A and ORF26. The orientation of the ORF on the nucleic acid sequence could not be determined due to its small size. Due to the lack of a RBS or promoter-like sequence in the intergenic region between these larger ORFs, clones were constructed to test whether this polypeptide could possibly be a degradation product of the larger 40 kDa *in vitro* proteins.

Figure 2.21b contains *in vitro* transcription - translation reactions from clones pCD500 (lane 1), pCD512 (lane 2), pCD511 (lane 3). The results show (lane 3) that clones constructed to contain the *repA* gene sequence exclusively do not contain any detectable degradation products. Similarly Figure 2.21a demonstrates that the deletion pCD551 which still retains a truncated version of the 26 kDa polypeptide has lost the 9 kDa band (lane 1). The sole evidence of the existence of the 9 kDa polypeptide in *E. coli* thus lies in the *in vitro* protein analysis. The observation that the ORF8 open reading frame, the only suitable ORF to code for the 9 kDa polypeptide, contains a helix-turn-helix motif (Section 2.3.4),. suggests a possible replication associated DNA binding role for this putative gene product.

DNA fragments subcloned from regions downstream of the *oriV* also produced proteins on *E. coli in vitro* transcription - translation gels (Figure 2.22). Clone pCD200 contains a 6.9 kbp *Xba*I-*Hin*dIII fragment cloned into pBS-KS produced three proteins in the *E.coli*-derived system (lane 1). The proteins had estimated sizes of 90 kDa, 17 kDa and 10kDa. Subclones of this region were generated to detect protein products which may have been expressed from the region with homology to the *tnp*A-like ORF (lane 1). No appropriate sized protein (predicted TnpA size = 37 kDa) was produced from pCD1. It was possible however that rather than not being expressed in *E. coli*, the promoter for this gene may have been deleted in the subcloning event due to the *Xba*I site being 48 bp upstream of the *tnp*A-like ORF start site. Furthermore the *E. coli* vector promoter position was in the opposing orientation and could not be used to express the TnpA product.

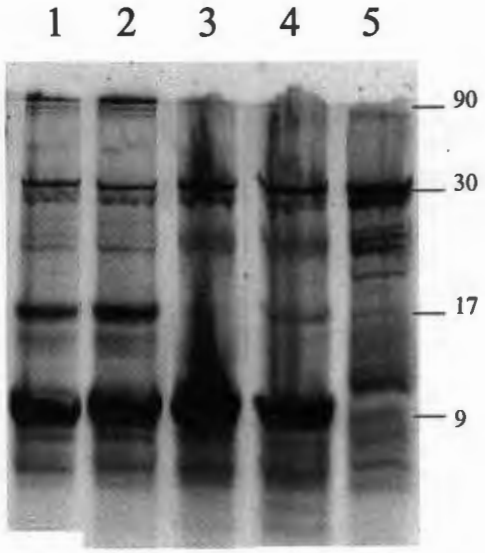
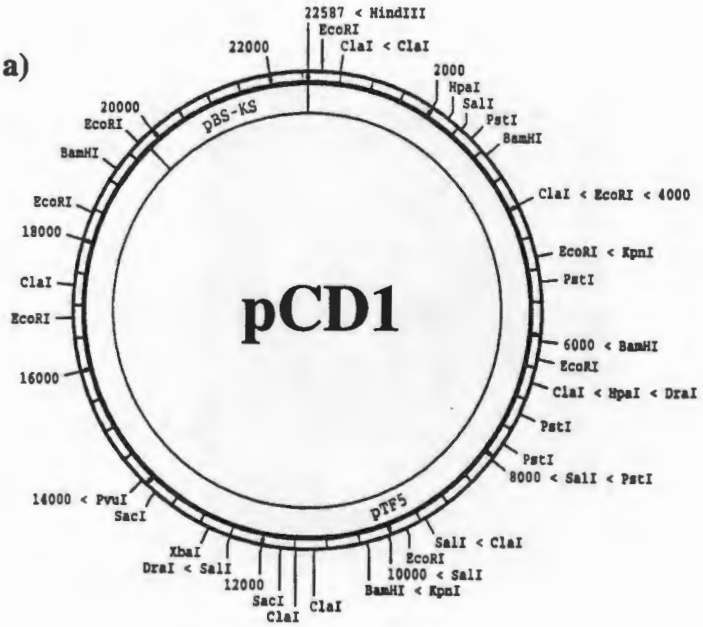


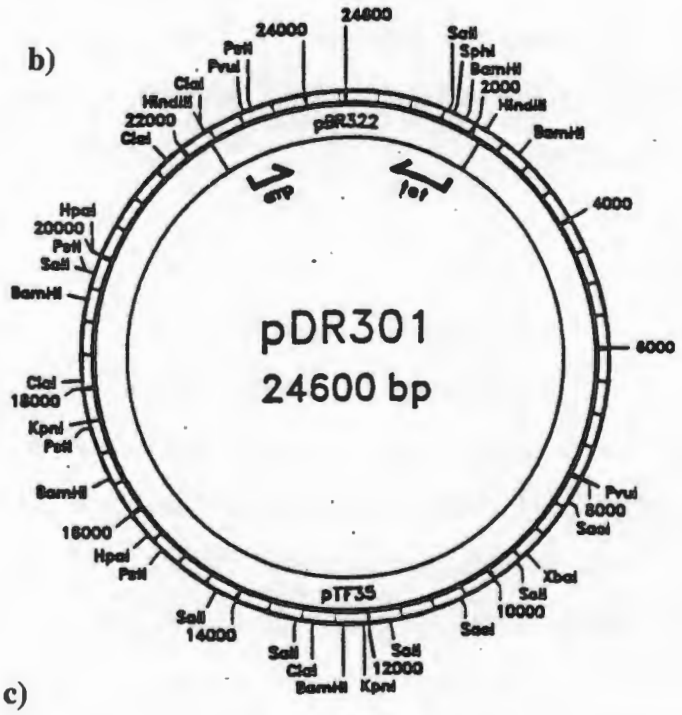
Figure 2.22 : *In vitro* transcription - translation products produced from a 6.9 kbp *Xba*I-*Hin*dIII fragment containing the *tnp*A-like ORF. Lane 1, pCD200; lane 2, pCD211; lane 3, pCD210; lane 4, pCD212; lane 5, pBS-KS (control). Molecular weights are shown as solid bars.

A highly expressed 10 kDa polypeptide was detected in both pCD211 (lane 2) and pCD210 (lane 3).In order for a 10kDa polypeptide to be produced from this region, an ORF of approximately 270 bp would be required. No suitable ORF was observed. The 17 kDa protein could potentially be produced from a gene flanking the *Apa*I restriction site due to its expression from pCD211 (lane 2) and pCD212 (lane 4) clones but not from clone pCD210 (lane 3). A potential candidate gene for this protein would be the 450 bp *stbB* gene from *Pseudomonas syringae* (unpublished, SWISS-Prot Accession no. L48985) which formed part of another investigation (N. Coram BSc. Hons thesis, 1996). The most probable candidate for expression of the 90 kDa protein on pCD200(lane 1) and pCD211 (lane 2) was a 2.7 kbp ORF which lies downstream of the *stb*B gene (N. Coram BSc. Hons thesis, 1996).

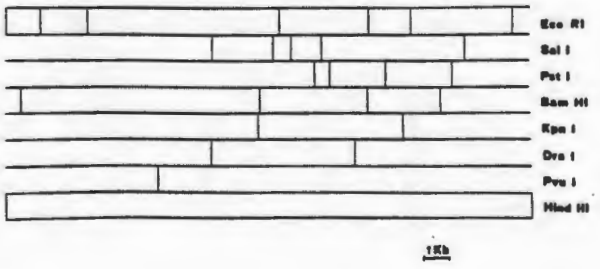
a)


pCD1
pBS-KS
pTF5
22587 < HindIII
EcoRI
ClaI < ClaI
22000
2000
HpaI
SalI
PstI
BamHI
20000
EcoRI
BamHI
EcoRI
18000
ClaI
EcoRI
16000
ClaI < EcoRI < 4000
EcoRI < KpnI
PstI
6000 < BamHI
EcoRI
ClaI < HpaI < DraI
PstI
PstI
8000 < SalI < PstI
14000 < PvuI
SacI
XbaI
DraI < SalI
12000
SacI
ClaI
ClaI
SalI < ClaI
EcoRI
10000 < SalI
BamHI < KpnI

b)


pDR301
24600 bp
pBR322
ori
rep
24600
24000
PstI
PvuI
ClaI
HindIII
22000
ClaI
SalI
SphI
BamHI
2000
HindIII
BamHI
4000
6000
PvuI
8000
SacI
XbaI
SalI
10000
SacI
SalI
12000
KpnI
SalI
ClaI
BamHI
SalI
14000
HpaI
PstI
16000
BamHI
KpnI
PstI
ClaI
18000
HpaI
20000
PstI
SalI
BamHI

c)


Eco RI
Sal I
Pst I
Bam HI
Kpn I
Dra I
Pvu I
Hind III
1Kb

2.3.6 Comparison of pTF5 with plasmid pTF35

Physical maps of pCD1, pDER301 (Rawlings et al. 1983) and pTFO (Valenti et al. 1987) indicated a substantial degree of similarity between the three plasmids. Figure 2.23 shows the comparative restriction maps of the three plasmids. Each plasmid has 5 *Sal*I, 4 *Pst*I, 4 *Bam*HI, 2 *Kpn*I and single *Pvu*I and *Hin*dIII sites in common with one another. Plasmids pCD1 and pTFO both have 7 *Eco*RI and 2 *Dra*I sites, similarly plasmids pCD1 and pDER301 each contained 2 *Hpa*I, 2 *Sac*I and a single *Xba*I site. Plasmid pTFO was not mapped for *Hpa*I, *Sac*I and *Xba*I, and plasmid pDER301 was not mapped for *Eco*RI or *Dra*I.

Figure 2.23 (opposite) : Physical maps of - Panel A) pCD1, Panel B) pDER301 (Rawlings et al. 1983); Panel C) pTFO (Valenti et al. 1993).

It appears on the basis of the above evidence that the plasmids are very similar if not identical.

Note in proof of pTF5-pTF35 near identity

Sequence analysis of the 1.6 kbp *Eco*RI fragment from pTF35 cloned into pBS-SK (M. Mulder, B.Sc (Hons) thesis, 1992) indicated that it was 100% homologous to the corresponding DNA fragment from pTF5 (N. Coram, B.Sc (Hons) thesis, 1996).

2.4 Discussion

The finding that a portion of the origin of vegetative replication (*ori*V) of plasmid pTF5 was associated with the activation of metronidazole, and consequent toxicity to *E. coli* F19 was surprising. A reasonable theory for how this occurred required further

characterisation of the mutant F19. Once it was discovered that two copies of Tn*PhoA* were present in F19 an extensive characterisation was considered to be beyond the scope of the current investigation. Two phenomena would need to be addressed in this investigation. The first is the mechanism of metronidazole resistance by *E. coli* F19, and the second is the manner in which the pTF5 *oriV* confers increased sensitivity to this strain.

Investigating a mechanism of resistance for F19 is complicated by the fact that the function of the *yeh*B gene product is unknown. One may speculate that metronidazole resistance is due to a blocked glycolytic pathway. Glycolytic inhibitors which function at the level of drug uptake and which confer metronidazole resistance to *C. pasteurianum* have been reported (Church and Laishley, 1995). It is difficult to understand how the pTF5 *oriV* could contribute to sensitising F19 to metronidazole at the level of drug uptake into the cell. It would be more likely to act as a binding site for a host-encoded regulatory protein which may be involved in cellular redox reactions. Two previous studies may provide clues as to the sensitising mechanism of F19 to metronidazole. In these studies two fragments of DNA from *C. acetobutylicum* were isolated which resulted in increased sensitivity of F19 to metronidazole. Neither of the two fragments encoded potential metronidazole reducing proteins, one having sequence homology to tRNA genes (V. Sealy, MSc thesis, 1993) while the other encoded the *regA* gene (Davison et al. 1995) discussed in section 2.1. Some members of the LacI family (to which RegA belongs) are repressors of tRNA genes (Leclerc et al. 1990) and it is possible that the cloned tRNA genes had sites which bound and titrated members of this repressor family. One could speculate that DNA fragments responsible for activating metronidazole could possess a binding site for the repressor of an *E. coli* host protein capable of reducing the drug (Dominy et al. submitted for publication). This would have the effect of derepressing the putative gene, resulting in an increased sensitivity to metronidazole.

Interestingly plasmid pTF5 did contain open reading frames with homology to genes encoding electron-transferring proteins (Chapter 3), but unexpectedly these did not contribute to the increase in metronidazole sensitivity of the *E. coli* F19 strain, implying that either they were not expressed in that host, or their redox potentials were lower than -485 mV.

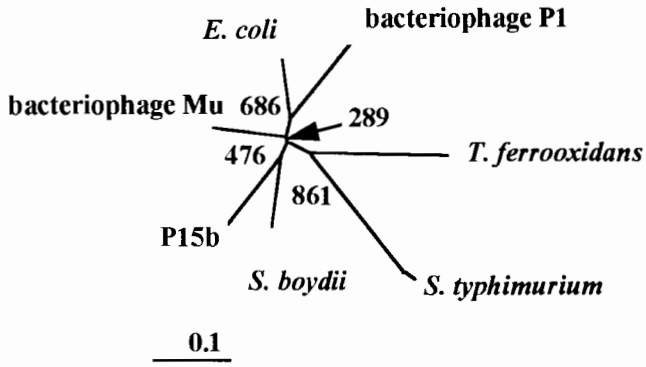


Figure 2.24 : A radial phylogenetic tree of the relationship of the invertase-like open reading frame to its closest homologues as described in Figure 2.12b. Bootstrapping values indicate relative confidence levels within the dendogram.

A dendogram of the invertase-like ORF from pTF5 is shown in Figure 2.24. This ORF shows high amino acid sequence identity to other member of the bacterial invertase/ recombinase family. As is the case with the Cin invertase (Rozsa et al. 1988), a truncated form of the protein is located downstream of the intact gene. Whether this was due to prior recombinational activity by a now inactive invertase is unknown. Bootstrap values show that the pTF5 invertase like ORF clusters close to the Hin invertase of *Salmonella typhimurium* sharing 58% aa similarity (Glasgow et al. 1989). This invertase is located adjacent to the H2 structural flagellum gene and is responsible for inverting H2 in order for the active protein to be either expressed or switched off. This has the advantage of allowing the pathogen to evade cellular immune systems by expressing altered flagellae which are not detected by host antibodies. A similar inversion function cannot be assigned to the invertase-like ORF of pTF5 however, until appropriate inversion sites are located.

Three open reading frames were found which were highly expressed in *E. coli*. The open reading frames encoded predicted proteins of 24, 8, and 43 kDa. This agreed with the SDS-PAGE results from which the proteins were estimated to be 26, 9 and 40 kDa in size. It is also likely that the 43 kDa protein which corresponded to the RepA protein of the Inc A/C plasmid RA1, was the same as that detected in an Italian study. In this study a highly expressed 40 kDa protein which could be isolated from *T. ferrooxidans* strains which harboured pTFO was reported (Valenti et al. 1993).

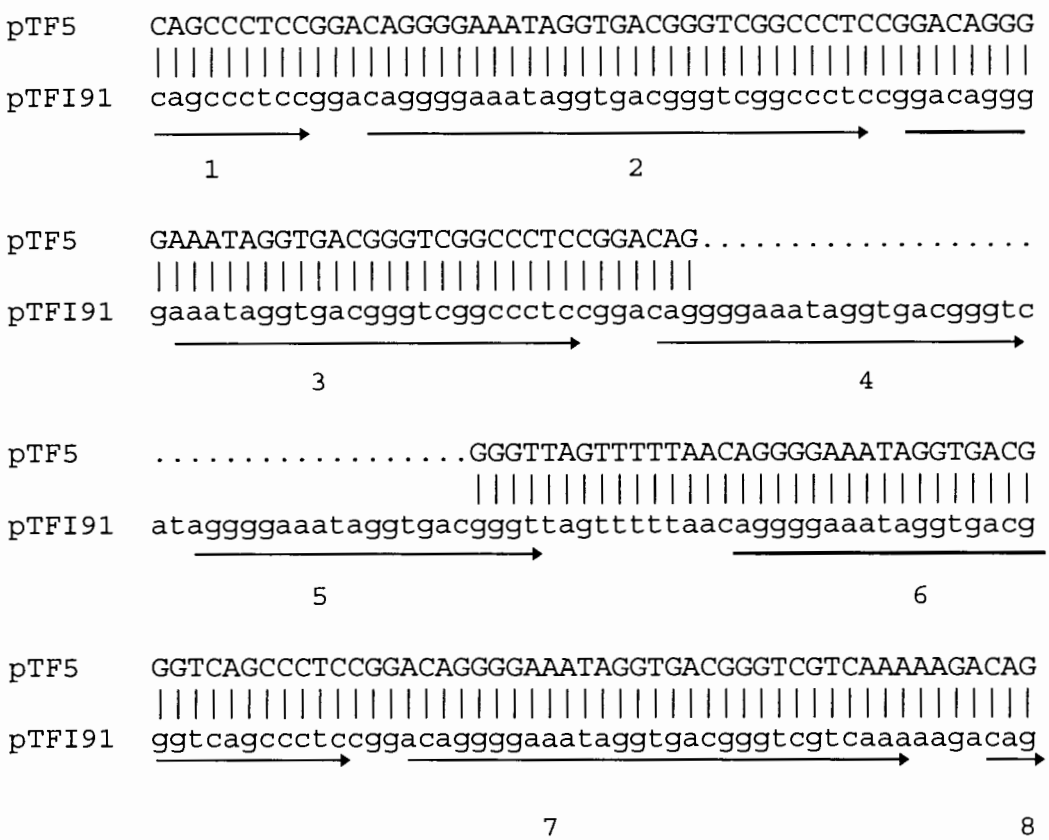


Figure 2.25 : Gap analysis (Needleman and Wunsch, 1970) of the pTF5 and pTFI91 *oriV* regions. The pTFI91 nucleic acid sequence shows the presence of two extra iterons (⟶). The iterons have been numbered according to Chakravarty et al. (1995).

The origin of vegetative replication of pTF5 was very similar to that of pTFI91 (Chakravarty et al. 1995) - 98% identity at the nucleic acid level. The major difference which contributed to the slight divergence was that pTF5 has two fewer iterons than pTFI91 as is illustrated in the GAP analysis of the two sequences (Fig 2.25).

Repeated sequences are frequently involved in the regulation of plasmid copy number within the cell (Thomas and Helinski, 1989). It would be interesting to know whether plasmid pTFI91, with the extra iterons, has a lower copy number than pTF5 within its *T. ferrooxidans* host. Plasmids belonging to the pTFI-like plasmid family cannot be assigned to an incompatibility group due to their inability to replicate in bacteria such as *E. coli.* However the organisation of the pTF5 *oriV* immediately downstream of the *repA* gene is similar to members of the IncP-type plasmids (Thomas and Helinski, 1989, Perri et al. 1991). In incompatibility groups IncN, IncQ, IncW (Couterier et al. 1989), the *oriV* is located upstream of the *repA* gene. Drawing any conclusions from this replicon organisation however isn't possible, as plasmid incompatibility involves the type of iteron, rather than their locations on the plasmid.

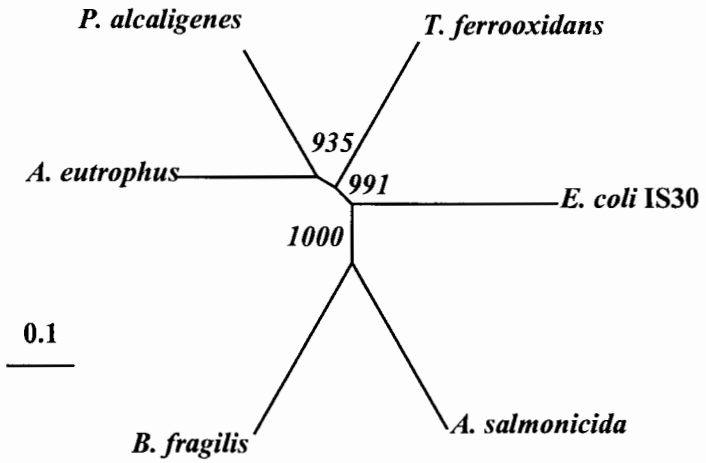


Figure 2.26 : Radial cladogram of the transposase enzymes with homology to pTF5 TnpA-like proteins discussed in Figure 2.18. The scale bar represents 10 % aa divergence.

As is the case with pTFI91, a transposase gene with homology to those described in certain bacterial insertion elements was located downstream of the *oriV* (Fig 2.26). Due to its perfect identity with IS3091 from pTFI91, it has been given the same designation in pTF5. Attempts to locate terminal inverted repeat sequences of IS3091 were not successful, and therefore the extent of the insertion element could not be determined in pTF5. It may be possible that the IS-like *tnpA* gene is not associated with an insertion element in pTF5.

Both restriction and sequence analysis of plasmids pTF5 and pTF35 indicate that they are almost identical. The similarity of the *oriV* of pTF5 to that of pTFI91 from *T. ferrooxidans* TFI91, suggests that both pTF5 and pTF35 could be assigned to the pTFI91-like family of *T. ferrooxidans* plasmids. Plasmid pTF5, like pTFI91 and pTF35 was unable to replicate in an *E. coli polA* mutant (data not shown). Comparison of the pTF5 restriction map to that from the 20.2 kbp pDSM583 plasmid (Chakravarty et al. 1995) has shown that in spite of their sizes being similar, the two restriction maps are very different, and therefore the plasmids are also likely to be different.

Chapter 3

Characterisation of genes from plasmid pTF5 which encode putative iron-sulphur proteins

Contents

3.0 Summary

This chapter describes the sequencing and characterisation of a 5600 bp region of plasmid pTF5. This region contained open reading frames (ORFs) with amino acid identity to iron-sulphur containing proteins. The four ORFs (*fdxA, fnr, psmA, redA*) had homology to ferredoxins, FNR regulatory proteins, prismane proteins and NADH oxidoreductases respectively. Evidence of a transpositional "scar" was detected between the putative ferredoxin- and fnr-like ORFs. This region contained a partial transketolase-like ORF fused to a *T. ferrooxidans* insertion element, IST2. Only the ferredoxin-like ORF, produced a corresponding polypeptide in *E. coli*-derived *in vitro* transcription-translation reactions. The prismane- and NADH oxidoreductase-like ORFs were expressed in the *E. coli* system when placed behind a *tac* promoter, whilst the expression of the *fnr*-like ORF was not detected. In-frame translational *lacZ* fusions were used to investigate the promoter-like element upstream of the *fdxA* gene. Screening of a *T. ferrooxidans* cosmid gene library did not reveal any potential regulators of the *fdxA* gene, suggesting that it may be constitutively expressed. Expression of the *lacZ* phenotype was altered relative to the copy number of the vector in which it had been cloned. Complementation studies in appropriate *E. coli* mutants were negative which was consistent with the finding that neither the truncated transketolase ORF, nor the *fnr*-like ORF were expressed in *E. coli.*

3.1 Introduction

The transfer of electrons between electro-negative donor compounds and more electro-positive acceptor molecules, is catalysed by a large family of redox-active proteins. These oxidation-reduction mechanisms are responsible for many important biological pathways such as respiration and photosynthesis. Figure 3.1 illustrates the diversity of electron transferring proteins which make up this super-family. This introduction will focus on selected types of electron-transferring proteins which contain iron-sulphur (Fe-S) cofactors, homologues of which have been found on plasmid pTF5.

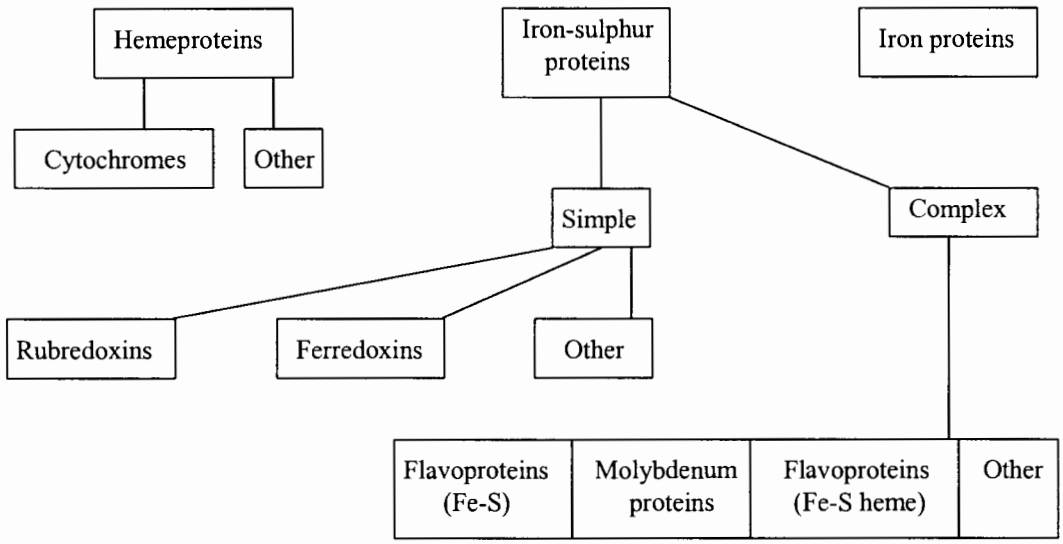


Figure 3.1 : Diagram illustrating the diversity of electron transporting proteins described in biological systems. Plasmid pTF5 contained a ferredoxin-like encoding gene (simple iron-sulphur protein), a *fnr*-like ORF (simple iron-sulphur protein), a prismane protein-like gene (simple iron-sulphur protein) and an NADH oxidoreductase-like gene (iron-sulphur containing flavoprotein).

3.1.1 Ferredoxins

Ferredoxins are low molecular weight iron-sulphur proteins which are responsible for electron transport in a wide range of biological processes (Beinert, 1990, Beinert and Kennedy, 1992). Ferredoxins contain a number of internal cysteine residues which are responsible for binding to the sulphydryl moieties of their iron-sulphur cofactors. These conserved cysteine residues have a characteristic spacing within the amino acid sequence which indicates the type of iron-sulphur cofactor to which the protein can bind. Ferredoxins have been grouped according to these cofactors, and are identified according to a well described member of each group (Figure 3.2; Beinert, 1990).

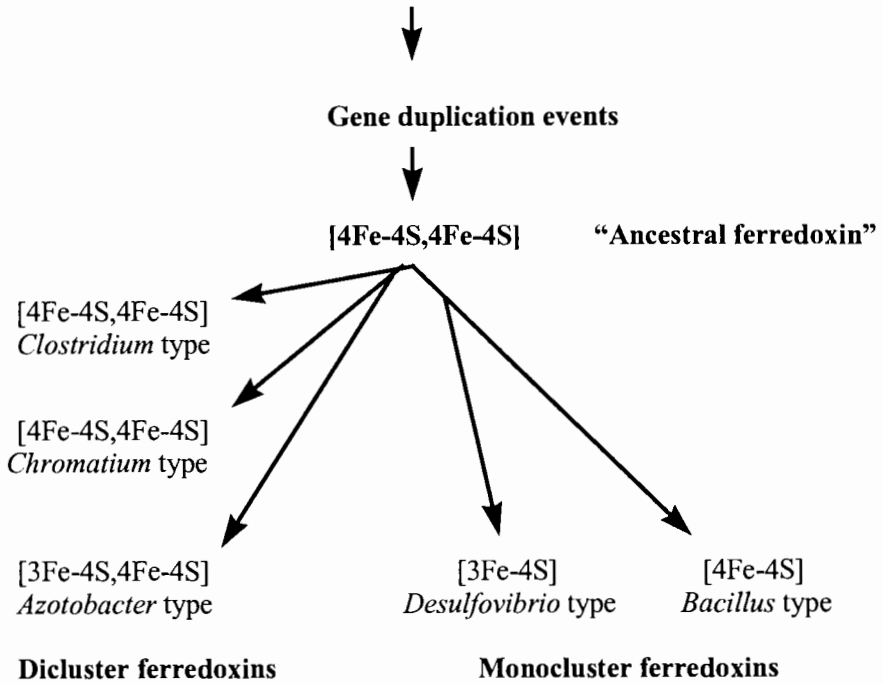


Figure 3.2 : The evolutionary relationship between the various classes of ferredoxin molecule (Beinert, 1990). The [4Fe-4S, 4Fe-4S] dicluster molecule is thought to have been ancestral to the modern ferredoxins. An exception are the [2Fe-2S] ferredoxins, believed to have evolved independently to the above classes.

Ferredoxins containing a [2Fe-2S] cluster are thought to have diverged away from the other ferredoxins before the ancestral molecule proposed in Fig 3.2. Two types have been identified, plant-type ferredoxins being distinguished from bacterial Rieske-type ferredoxins due to a slightly differing cysteine spacing. This results in different redox potential ranges between the two types of protein. The plant-type ferredoxins, which are involved in photosynthesis, exhibit midpoint redox potentials of -230 to -450 mV, whereas the high potential Rieske- and oxygenase-type ferredoxins have redox potentials from +100 to +350 mV (Iwasaki et al. 1994). The Rieske-type ferredoxins have been reported in bacteria, such as *E. coli* (function unknown) and the archaeon *Halobacterium salinarium* where it is involved in the coenzyme A-acetylating 2-oxoacid oxidising system (Seaton et al. 1992; Pfeifer et al. 1993).

The monocluster type ferredoxins are believed to have evolved from the proposed ancestral dicluster molecule by the loss of one iron-sulphur cluster binding site (Beinert, 1990). They have been classed into two groups, the [3Fe-4S] desulfovibrio-type and [4Fe-4S] bacillus-type ferredoxins. Research on the bacillus-type ferredoxin has concentrated on refinements in the crystal structure of the protein (Fukuyama et al. 1989) and no functional analysis has been reported. More progress has been made however with the *Desulfovibrio gigas* ferredoxin where it has been shown to link the evolution of H_2 gas with pyruvate dehydrogenase activity (Kissinger et al. 1991).

[4Fe-4S,4Fe-4S] clostridial ferredoxins play an important role in the phosphoroclastic reaction (Chapter 1) whereby excess reducing equivalents generated by central metabolic processes are converted to hydrogen gas. Other members of the dicluster ferredoxin family are the 2[4Fe-4S] "*chromatium*" type ferredoxins, the functions of which are unknown and the [3Fe-4S,4Fe-4S] "*azotobacter*" type ferredoxins. The ferredoxin from *Azotobacter vinelandii* has a [3Fe-4S,4Fe-4S] dicluster arrangement (Ghosh et al. 1982) and like many other ferredoxins its role within the cell remains uncertain. Mutational analysis has shown that this ferredoxin is not involved in nitrogen fixation in *A. vinelandii* (Morgan et al. 1988). Shen et al. (1994) have suggested that it may function as a regulatory protein by repressing the synthesis of other unknown small acidic proteins. This characteristic, in addition to its proven electron transferring functions would make it unique in this class of proteins. Other [3Fe-4S,4Fe-4S] ferredoxins include those from *Saccharopolyspora erythrea*, possibly involved in electron transport to a hydroxylase (Donadio and Hutchison, 1991); *Thermus thermophilus* (Sato et al. 1981) and *Streptomyces griseus* (Trower et al. 1990) may both function as electron carriers to cytochrome P-450. *Pseudomonas putida* contains a [3Fe-4S,4Fe-4S] ferredoxin, the function of which has not been determined (Hase et al. 1978). Pseudomonads also express additional mono-cluster [2Fe-2S] ferredoxins which are involved in passing electrons to hydroxylase/oxygenases during the degradation of aromatic compounds (Suen and Gibson, 1994; Tan et al. 1994; Section 3.1.4).

Rieske-type ferredoxins with a [2Fe-2S] cluster contain a diagnostic Cys-X_4-Cys-X_2-Cys-X_{29}-Cys cyteine motif (Ta and Vickery, 1992). However, when part of a larger protein structure such as the bacterial oxygenase complex, this cysteine spacing can vary slightly (Mason and Cammack, 1993). Plant-type [2Fe-2S] ferredoxins also have cysteine residues with variable spacing. Within the [3Fe-4S,4Fe-4S] dicluster ferredoxins [4Fe-4S] clusters have Cys-X_{18}-Cys-X_2-Cys-X_2-Cys and [3Fe-4S] clusters contain Cys-X_7-Cys-X_{32}-Cys as invariant cysteines (Trower et al. 1990). [4Fe-4S,4Fe-4S] clostridial-type ferredoxins contain two Cys-X_2-Cys-X_2-Cys-X_3-Cys-Pro signatures (Beinert, 1990). Due to their small size (approximately 100 amino acids) many ferredoxins in the SWISS-Prot database were sequenced at the protein level and little is known of their function within the cell. Consequently many ferredoxins have been categorised according to their cofactors, and not according to physiological function.

It is evident from the brief introduction presented above that the role of a particular ferredoxin cannot be predicted solely from its amino acid sequence. The reduction-oxidation potential of the mature protein within its physiological milieu, will also determine the type of donor and acceptor molecules which the ferredoxin may interact with. In many cases the role of a particular ferredoxin within its host cell can only be determined biologically, either by its position within an operon (Rossi et al. 1993), or by the generation and characterisation of *fdx* mutants.

3.1.2 FNR-like regulators

The FNR protein belongs to a family of transcriptional regulators, and (in *E. coli*) is responsible for inducing genes involved in regulating anaerobic respiration (Gunsalus, 1992). The *E. coli* FNR protein is structurally similar to a number of homologs from other bacteria, and many of these proteins are able to complement *E. coli fnr* mutants (Spiro, 1994). Despite this heterologous complementation, not all *fnr* like proteins are involved in anaerobic respiration. FNR homologues from *Rhizobium meliloti* (Batut et al.

1989), *R. leguminosarum* (Colonna-Romano et al. 1990), *Bradyrhizobium japonicum* (Anthamatten et al. 1992) and *Azorhizobium caulinodans* (Kaminski et al. 1991) are responsible for sensing oxygen levels during nitrogen fixation. FNR-like proteins in *Actinobacillus pleuropneumoniae* and *Bordetella pertussis* are responsible for the (anaerobic) expression of toxin compounds (MacInnes et al. 1990, Bannan et al. 1993) whilst in *Vibrio fischeri*, FNR regulates bioluminesence (Spiro, 1994). An interesting example in the light of the other pTF5 ORFs presented in this study, was the isolation of a FNR-like homolog from *Rhodopseudomonas palustris*. This purple non-sulphur bacterium degrades benzoate under anaerobic conditions, and the FNR homologue is thought to regulate this process (Dispensa et al. 1992). This may well provide clues as to the potential regulatory function of the pTF5 FNR-like ORF.

FNR and its homologues contain an iron-sulphur cluster and are thought to be able to respond to oxygen levels by binding oxygen at this site (Green et al. 1993). Site directed mutagenesis studies have shown that three cysteine residues in the N-terminal portion of the FNR protein, and one internal cystine residue were responsible for binding to the iron-sulphur cluster (Spiro and Guest, 1988, Sharrocks et al. 1990). This presumably produces a conformational change in the protein which either results in the induction or repression of appropriate genes (Spiro, 1994).

FNR-like regulators are very similar to cAMP receptor (CRP) proteins. Two distinguishing characteristics of FNR-like proteins is the ability to bind to specific DNA target sequences known as FNR boxes (See Figure 3.3) and secondly they do not respond to cAMP (Spiro, 1994). A number of residues between positions 81 and 87 in the *E.coli* FNR, have been shown to be important in transcriptional activation (Williams et al. 1991). Site directed mutants in this region were able to bind to DNA at the FNR box, but not activate the appropriate genes. Site-directed mutagenesis was also used to establish the importance of residues Glu-209 and Ser-212 in binding to the FNR target sequence (Spiro et al. 1990).

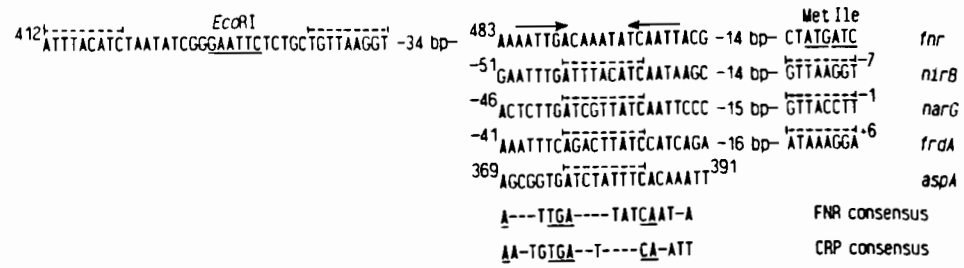


Figure 3.3 : The FNR box consensus sequence (Spiro and Guest, 1987). *E. coli* FNR boxes appear to have a slightly extended consensus region relative to the closely related CRP binding regulatory protein (De Crombrugghe et al. 1984). FNR-regulated genes included here are the *fnr* gene, the *nirB* gene (nitrite reductase), *narG* (nitrate reductase), *frdA* (fumarate reductase) and *aspA* gene (aspartase) of *E. coli*.

3.1.3 Prismane proteins

Prismane proteins are relatively rare iron-sulphur proteins which contain a single [6Fe-6S] cluster (Pierik et al. 1992a). An independent report has suggested the presence of a second [6Fe-6S] cluster in the prismane from *D. desulfuricans* (Moura et al. 1992), but the original researchers have refuted this (Van den Berg et al. 1994).

Apart from being detected in *D. vulgaris* (Hildenborough) (Pierik et al. 1992a) and *D. desulfuricans* (Moura et al. 1992, Stokkermans et al. 1992a), antibodies raised to the *D. vulgaris* prismane, have also shown them to occur in other sulphate reducing bacteria (Stokkermans et al. 1992b). The genomic sequencing project of the archeon *Methanococcus jannaschii* has also revealed an open reading frame with homology to the *Desulfovibrio* prismanes (http://www3.ncbi.nlm.nih.gov).

The iron-sulphur cluster present in the prismane protein exhibits an EPR spectrum which is very similar to the prismoidal core iron-sulphur structure synthesized *in vitro*

(Coucouvanis et al. 1984; Stokkermans et al. 1992c). Both the *D. vulgaris* and *D. desulfuricans* prismane genes have been cloned and sequenced. Expression of the *D. vulgaris* gene behind the inducible *E. coli lacZ* promoter resulted in the expression of an inactive, insoluble protein which did not contain the characteristic iron-sulphur core. Overproduction of both of the prismanes was achieved by expressing them in multicopy in the *Desulfovibrio vulgaris* host (Stokkermans et al. 1992c, Van den Berg et al.1994).

Mossbauer spectroscopy (Pierik et al.1992b), EPR spectroscopy (Van den Berg et al. 1994) and magnetic circular dichroic spectroscopy (Marritt et al. 1995) have been used to establish a spectroscopic model for the prismane proteins. The presence of the [6Fe-6S] iron sulphur cluster allows the prismane to exist in four redox states (+3, +4, +5 and +6). This would give the prismane the ability to accept more than one reducing equivalent in its potential role as an electron transporter (Pierik et al. 1992a). Despite a wealth of spectroscopic evidence on the structure of the prismane, very little is known about its function. It has been shown to be cytoplasmically located within *D. vulgaris*, although due to the fractionation procedure a membrane association cannot be excluded (Pierik et al. 1992b). The purified protein has been tested for the ability to function as an hydrogenase, a dehydrogenase (lactate and formate) as well as a reductase (fumarate, sulphite, nitrite, thiosulphate and APS), all without any significant enzymatic activity. Its widespread presence among sulphate reducing bacteria may ultimately give an indication of its function within these bacteria.

3.1.4 Bacterial oxygenases

Bacterial oxygenases are large enzyme complexes responsible for the degradation of a wide variety of aromatic compounds. Well known examples include the benzene and toluene dioxygenases from the TOL plasmids of pseudomonads (Harayama and Timmes, 1989). The dioxygenase enzyme complex consists of two to three soluble electron transporting proteins which oxidise NADH and use the two electrons to hydroxylate aromatic hydrocarbons (Figure 3.4). This initial hydroxylation reaction allows the

subsequent degradation of these normally recalcitrant aromatic compounds (Mason and Cammack, 1992).

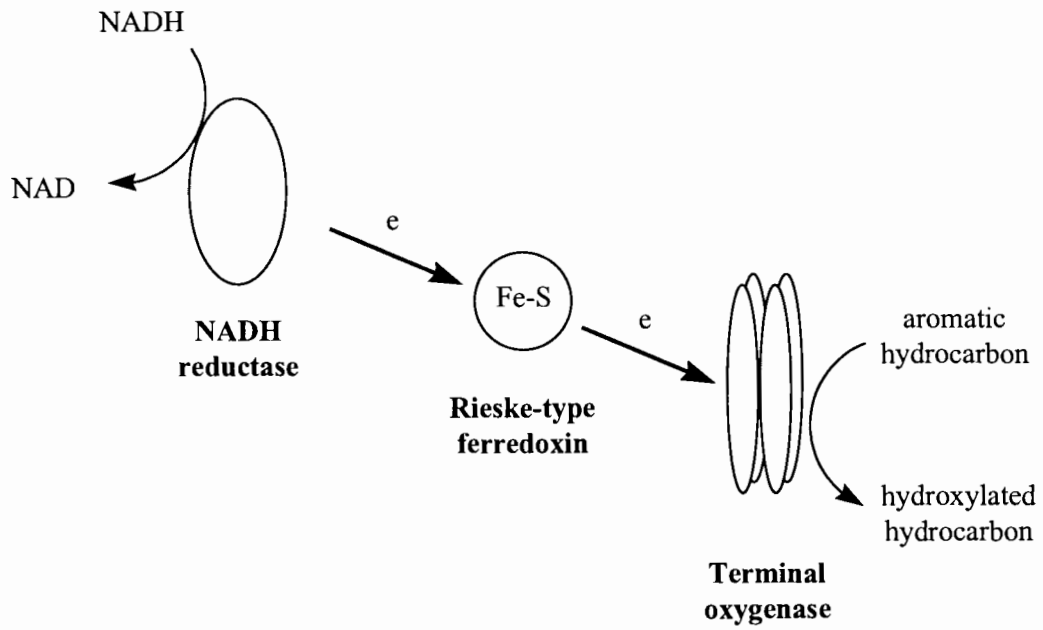


Figure 3.4 : A generalised bacterial oxygenase pathway. A NADH reductase containing a flavin-binding domain and possibly a [2Fe-2S] ferredoxin (Fe-S_{fd}) domain pass electrons either directly to the oxygenase complex, or via a Rieske iron protein (Fe-S_r), to the terminal oxygenase (Mason and Cammack, 1992).

The dioxygenases have been classified according to the number of components in the pathway as well as the type of redox centres employed by the various components (Mason and Cammack, 1992). The class I dioxygenase is a two component system consisting of i) a reductase enzyme which contains both a flavin cofactor as well as a [2Fe-2S] redox center and ii) the oxygenase complex. An example of this is benzoate 1,2 dioxygenase from *Pseudomonas arvilla* (Yamaguchi and Fujisawa, 1978). Class II dioxygenases have three components i) a flavoprotein, ii) a separate ferredoxin protein and iii) the terminal oxygenase. They include the benzene and toluene dioxygenases (Axcell and Geary, 1975, Subramanian et al. 1981). Class III dioxygenases are three component systems which appear to be a combination of the first two classes. They have i) a flavoprotein with a [2Fe-2S] redox centre, ii) a ferredoxin with its own [2Fe-2S]

redox centre and finally iii) the terminal oxygenase. Only one example of this class of dioxygenase has been reported, the napthalene dioxygenase (Haigler and Gibson, 1990). These classes can be further subdivided depending on their flavin cofactors, the position of the [2Fe-2S] redox centres, and the number of subunits which comprise the oxygenase component (Mason and Cammack, 1992).

Dioxygenases function purely as oxygenases by adding two oxygen atoms onto their substrates, while monooxygenases add a single oxygen atom onto the aromatic substrate, and function as an oxidase by adding the second oxygen atom to a free hydroxyl ion to form water (Mason and Cammack, 1992). The flavoprotein present in three component systems functions as an NADH oxidoreductase, by transferring electrons from NADH to the ferredoxin. They have also been detected in other redox systems, functioning as a mercuric reductase and as a glutathione reductase in *Pseudomonas aeruginosa* (Brown et al. 1983). These flavoproteins possess both an FAD and NAD binding sites which are approximately 140 aa apart on the proteins. The iron sulphur flavoproteins of the two component system contain an NAD binding site and either an FAD or FMN binding site as well as a [2Fe-2S] cluster. The protein can be roughly divided into three domains, the N-terminal flavin domain has homology to plant ferredoxin NADP reductases, a middle domain binds NAD and finally a cyanobacterial-type ferredoxin in the C-terminal domain (Fukuyama et al. 1980). The ferredoxin plays an intermediate role in passing electrons from the primary oxidoreductase to the terminal oxygenase. These ferredoxins are of the [2Fe-2S] plant type, and are specific for each oxygenase system as one cannot substitute for another involved in catabolizing a different compound (Subramanian et al. 1985). The final catalytic oxygenase component hydroxylates its appropriate substrate in the presence of both ferrous iron and oxygen (Mason and Cammack, 1992). The oxygenases are generally large oligomeric protein complexes of up to 150-200 kDa. Most of them have two subunits (a and b) of 50 kDa and 20 kDa respectively. The benzene, toluene and napthalene dioxygenases have an a_2b_2 configuration (Gibson and Subramanian, 1984, Sauber et al. 1977, Zamanion and Mason, 1987) while the benzoate dioxygenase has an a_3b_3 structure (Yamaguchi and Fujisawa, 1982).

The observation that a number of these degradative pathways occur on plasmids has been well established in the TOL plasmids of the pseudomonads (Nordlund et al. 1990a, Harayama et al. 1991, and Neidle et al. 1992), as well as the chlorobenzene dioxygenase (Werlen et al. 1996), the phenol hydroxylase (Nordlund et al. 1990b) and the 3-chlorobenzoate 3,4-dioxygenase systems (Nakatsu et al. 1995). Of potential relevance to this study however, is that two of the above mechanisms are intricately associated with mobile genetic elements (Werlen et al. 1996, Nakatsu et al. 1995). The 3-chlorobenzoate 3,4-dioxygenase has been localised within a transposon, Tn*5271*, which is located on an 85 kbp conjugable plasmid in *Alcaligenes* sp strain BR60 (Nakatsu et al. 1995). Werlen et al. (1996) have reported a degradative pathway to chlorobenzene where the chlorobenzene operon is located on a transmissible plasmid in *Pseudomonas* sp strain P51. Interestingly, the genes for the two component oxygenase were not localised with the genes for the rest of the pathway but rather within a transposable element, Tn*5280.* This pathway also has a regulator gene whose role has not been reported. The two genetically distinct parts to the operon also appeared to be related to the toluene benzene pathways rather than to the chlorobenzoate oxygenases. The authors hypothesized that a horizontal gene transfer had been responsible for this occurrence (Werlen et al. 1996).

3.2 Materials and methods

3.2.1 *LacZ* translational fusions

In-frame translational fusions to the promoterless *lacZ* gene of pMC1403 (Casadaban et al. 1983) were constructed to determine whether sequence upstream of the putative ferredoxin start site contained a promoter sequence recognizable by *E. coli* MC1061. A 400 bp pTF5 DNA fragment, pCS4, which included the first 48 bp of the ferredoxin ORF, was fused in-frame to the *lacZ* gene (fxp1). The fusion was excised from pMC1403 with *Xba*I and *Sal*I and ligated into the low copy number vector pACYC184 at the unique *Xho*I and *Nhe*I sites which provided compatible overlapping ends for *Xba*I and *Sal*I

(fxp2). In order to test whether these constructs were actively expressing the *lacZ* phenotype, the constructs were transformed into the *lacZ* deficient *E. coli* mutant MC1061. Transformants were plated on Luria agar (Miller, 1972) containing X-Gal and appropriate antibiotic selectable markers (ampicillin 100 μg/ml for fxp1 in pBS-KS, and chloramphenicol 30 μg/ml for fxp2 in pACYC184).

Regulation of *lacZ* expression : *E. coli* MC1061 which contained the fxp2 fusion was transformed with other subclones of plasmid pTF5 as well as members of *T. ferrooxidans* ATCC33020 gene libraries, to determine whether they expressed genes capable of regulating the *lacZ* phenotype at the ferredoxin promoter. Transformants and transductants were selected on Luria agar supplemented with X-Gal, ampicillin 100 μg/ml and chloramphenicol 30 μg/ml.

3.2.2 PCR template amplification

Polymerase Chain Reaction amplification was carried out using artificially synthesized primers to the predicted start and stop sites of a number of pTF5 ORFs. Primers were synthesized on a Oligo 1000M DNA synthesizer, Beckman Instruments Inc. (W.Brandt, Department of Biochemistry, University of Cape Town). Primers used were

1.) 5'- AACAGCTATGACCATGATTAG-3' (pBS-KS reverse primer)
2.) 5'- CGTAAAAGCCGCATCATGGGG-3'
3.) 5'- CTTGATTGAAATCAAGGATGG-3'
4.) 5'- CGTAGCACAATGGAACGCCGC-3'
5.) 5'- CACGAGTTCCTTCTTGGCTGG-3'

The primer binding sites have been indicated in Fig 3.5. Amplification was carried out in a JDI 2500 thermal cycler using "Red Hot" polymerase (Separations Scientific). The following cycle was used - denaturing temperature 94°C for 2 min, annealing temperature

52°C for 30 sec, elongation temperature 72°C for 1 min, denaturation 94°C for 30 sec. The annealing-elongation-denaturation stages (steps 2 to 4), were repeated thirty times.

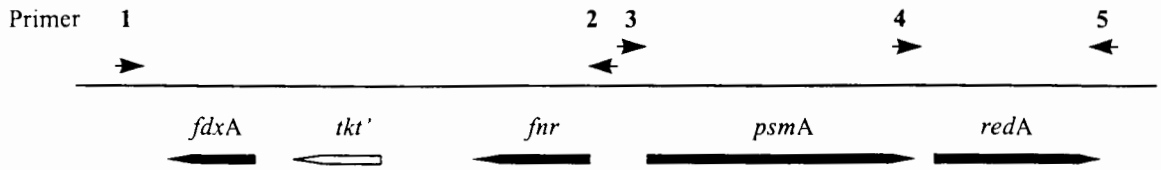


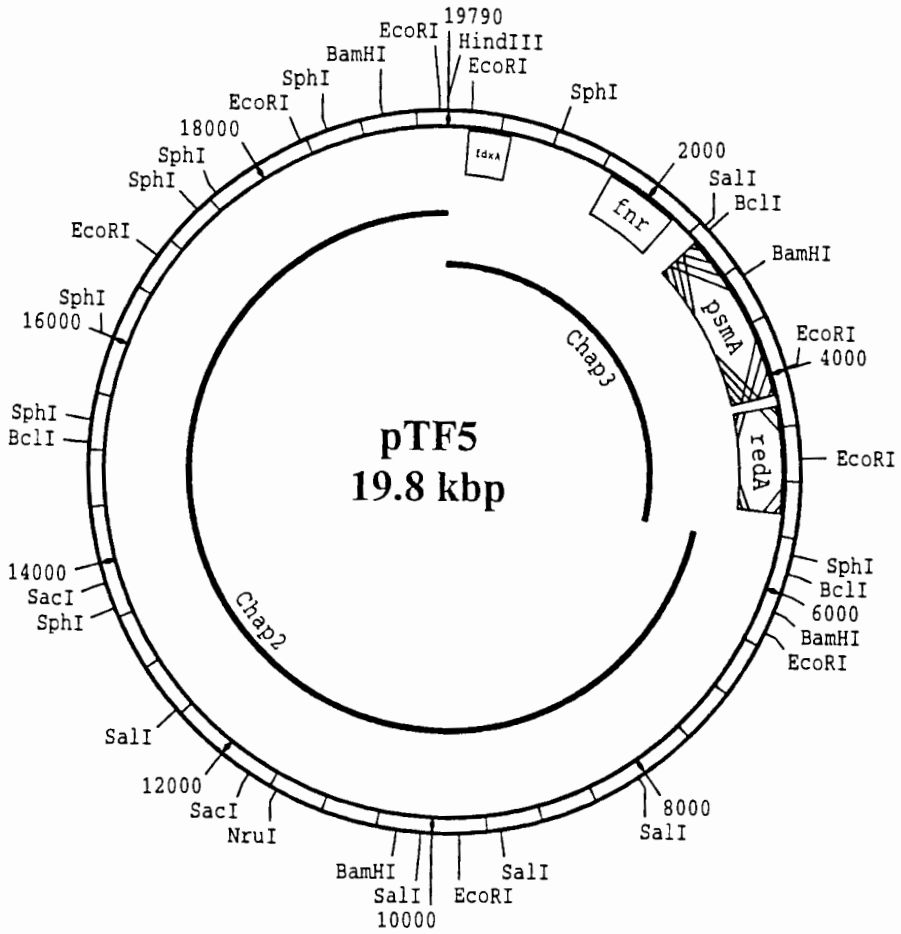
Figure 3.5 : Diagram indicating the positions of the PCR primers relative to the pTF5 iron-sulphur protein encoding ORFs. Primer numbers are the same as used in section 3.2.2.

3.2.3 Heavy metal resistance assays

To test for the expression of a number of heavy metal resistance phenotypes, *E. coli* DH5α was transformed with appropriate subclones expressing pTF5 ORFs under the control of the *E. coli tac* promoter in the pKK223-3 expression vector (Pharmacia Biolabs, Inc.). Expression mixes were plated onto media containing increasing concentrations of uranium or silver ions. The *E. coli* MIC for the particular metal, was taken as the lowest concentration on which colonies were unable to grow.

3.2.4 Complementation assays

Transketolase complementation assay : Two *E. coli* transketolase mutants were obtained with either both *tkt* genes (A and B) present on the *E. coli* chromosome deleted (AI118 - Zhao and Winkler, 1994) or with the dominant *tktA* gene inactivated (GSJ423 - Sprenger et al.1995). Constructs containing the pTF5 incomplete transketolase ORF were transformed into these mutants, and grown on M9 minimal media (Sambrook et al. 1989) supplemented with 0.2% D-ribose (Saarchem) as a carbon source. The *tkt* mutants required an *in trans* active transketolase enzyme to convert the C_5 ribose compound into a


pTF5
19.8 kbp
19790
EcoRI
HindIII
EcoRI
SphI
2000
SalI
BclI
BamHI
EcoRI
4000
EcoRI
SphI
BclI
6000
BamHI
EcoRI
8000
SalI
SalI
EcoRI
BamHI
SalI
10000
NruI
SacI
12000
SalI
14000
SacI
SphI
SphI
BclI
SphI
16000
EcoRI
SphI
SphI
18000
EcoRI
SphI
BamHI
fdxA
fnr
psmA
redA
Chap3
Chap2

metabolizable C_6 sugar via the pentose phosphate pathway. Plates were incubated at 37°C for 3 days and inspected for growth.

FNR complementation assay : *E. coli fnr* mutant RM102 was unable to respire anaerobically using nitrate as an alternate electron acceptor (Birkmann et al. 1987). Plasmid pTF5 constructs to be tested for *fnr* complementation were transformed into RM102 and expressed on M9 minimal media containing 0.5 % glycerol (Saarchem) and 0.5 % $NaNO_3$. Plates were incubated anaerobically at 37°C for 3 days and examined for growth.

3.3 Results

3.3.1 Sequence analysis of the pTF5 ORFs encoding putative iron-sulphur containing proteins

A physical map of pTF5 highlights the region to be discussed in this chapter (Fig 3.6a) The nucleic acid sequence from the unique *Hin*dIII site of pTF5 (position 1 bp) to the *Sph*I site (5641) containing ORFs with homology to iron-sulphur proteins is presented in Fig 3.6b.

Figure 3.6a (opposite) : Physical map of pTF5 indicating the region covering the iron-sulphur protein encoding ORFs discussed in Chapter 3. b (overleaf) : Nucleic acid sequence of the 5.6 kbp *Hin*dIII-*Sph*I fragment containing the ORFs with homology to iron-sulphur proteins. The direction of transcription is indicated by short arrows, potential stem-loop structures are indicated by inverted pairs of arrows, and the FNR consensus binding site is underlined. Potential promoter sequences are in bold. The transposition scar with homology to a truncated IST2 element is italicised and demarcated with bent arrows. Nucleic acid numbering is from the first base of the unique *Hin*dIII restriction site used in the cloning of pTF5 (Section 2.3.2).

*Hin*dIII

1 AAGCTTACGTAGAAACGTGCTGGGCCCAGGAACTCCACCCGGTGCGGCAGATCGGGCGGTATCCATGCAGGCCGTGCCGA

81 CTCAGAATCCAGGCGCGGCGCTGCCGTCTTCCAGGCAATATCGCAAGTTTCCTTCCAGGAGCACGATGCGCCCCCAGACC

161 CCGGCTTTGGTGCGATGCTCATGGAGCAGGCCGGCGGGGGCGTCGTCGTCGGTGAATTCCGGTGTGCGCTGGTATTCCTT

241 CAAGGGCACGGGCGGCTCGGGCAGGGGCAATTCCTCCTCCTCTTCCCCGGTATCCAGATATGGGCGTTTATCCCGCACAT

* P C P R S P C P C N R R R K G P I W I H A N I G C M

321 GCGCCACTGCTCGGCATCGGGCAGGGCAGGCTTTTTTTGAATGATGACGGGCCAGCGGCGTGCCAGGCGCGCGTTGAGTT

R W Q E A D P L A P K K Q I I V P W R R A L R A N L

401 CCGGATATTTTTCCACTGCGTCCGGCAAATCCACGTCGCGGAAGATGGCGTCTACCGGACACTCGGGCACACAGAGGGTG

E P Y K E V A D P L D V D R F I A D V P C E P V C L T

*Cla*I *Cla*I

481 CAATCGATGCATTCGTCTGGATCGATGGCCAAAAAGTTGGGGCCCTCGTGGAAGCAATCCACCGGGCATACTGTGACGCA

C D I C E D P D I A L F N P G E H F C D V P C V T V C

*Eco*RV

561 ATCGGTATACTTACACCGAATACAGGCTTCTGTAACCACATGGGTCATTGGATATCCCCCTTAACGTATGCCCCGGAAAT

D T Y K C R I C A E T V V H T M *fdx* start

641 GCCTTCAACGCCAGATCGGCGGGATTGCGCTCGCCCCATTTCCGAAAATCCGTCAAAACCG**ATCATC**ACGGCGTATAAAA

-10 N_{17}

721 GATG**CCTCAA**TCAGAACTTCATGTCGATAAACGGATAGATTTTATACCTATCCCGCGACTCGCCCGCGTTATAGTAGCCG

-35 * L L R

801 TTGGGCTTCCTGTACCACATTTTCCACAGTGAATCCCAGATGGCGGAAATTTTCTGGATAGGGTGCACTGGCGCCGAAGC

Q A E Q V V N E V T F G L H R F N E P Y P A S A G F

881 GGCTCAGGCCTATGGTGGCGCCCTCTCCCACCCATTCGCACCACCCCAGGGATGCCGCAGCCTCTACGGCGAGCCGGACT

R S L G I T A G E G V W E C W G L S A A A E V A L R V

transketolase homology

*Sph*I

961 TTGATATTCCGTGGGAGCACCGTATCCCGGTAAGCGGCCGGCTGGGCGCGGAATATCTCCCAGCTGGGCATGCTCT*TCGC*

K I N R P L V T D R Y A A P Q A R F I E W S P M S

homology to IST2

1041 *CACTGCGCCTTGAGGCGTCCCAATGCCGCAGGGGAAAGTCCCTTGGCCTCTTCGCCCAGGAGGATAGAGAGCGCCCCCTG*

*Eco*RV

1121 *CATGTGGCCAGAAGATATCCCATGCAGATACAGCCAGGGGACTGTGGCCGCCACCGTCCGCGATTTACGCACATACGGTG*

1201 *GCATCAGTGTCGAATTGAATTTAATCCCCGAACCCGCGCGATCTCGAACCTTGGGGACCTTCACTGACACAGGACCCAGG*

1281 *GCCGTCATGATCTCTCGCTCCGGCAGGTGGCCGTTACGAACGACCGCCTGACGACCATTGACCATCCGTACAGCAGAAAA*

1361 *CTCTTCCAGCACTTGGTCTGCAAACCCtTTGTAGCAACAGTACGCCACCTTGCTCAGGTCACTTGTAACCCGCCTATACA*

1441 *CCAGTTTCGATCATAGCTCT*CCGCATTGGCGTTCCTTCAGGACAATGCGGTGCGCGATGTGGCCCCAGTTTGGCTGACAG

IST2 homology

1521 AGCCAGGCCCCCTCCGTGTAGCTGATGAGATCGTTATTGATTGGATGACGCTTGCGGCGAACGGTGAAACCTATTTCCTG

1601 TCATCTACCGGGTAGACGATTTGCTGTTCGCCCATTGCCGTAACTCAGAAGGGCGGTGGATGTATATCTTGAGTTTTTCA

* R T S S K S N A W Q R L E S P R H I Y I K L K E

1681 ACAGCTATCAAATGCTCCTCTCGCAATCGCTTCAACACACGAGAAAATGTTTCCGGGGTTAAACCTAGTCGGGCCGCGAT

V A I L H E E R L R K L V R S F T E P T L G L R A A I

*Eco*RV

1761 CATGGCTTTTTTGGCAGGGAGAAGCAAACTCACGGGAGAATTTCCCGATGGGCAGAGTTCCAGCAGATATCCTGCAACCC

M A K K A P L L L S V P S N G S P C L E L L Y G A V

1841 GTTGGTCAGCGGATTCCACAGTAAGATGTCGCAACTCCATTATTAAATGACGTAAACGCAGACTCAGTTCGCCCAACATT

R Q D A S E V T L H R L E M I L H R L R L S L E G L M

1921 TGCAGCATCAGCTCCTGATGCTGCTGTAGGGTGACGAGAAAATTATCGTAAGGGGAATTTCCAGAACCCGGCCTCGCATG

Q L M L E Q H Q Q L T V L F N D Y P S N G S G P R A H

*Eco*RV

2001 GCCGCTATTGCCTGCACGGGATAACGATATCCCAAAAAAGCTACAGCTTCGGCAAAAAAGGTCTCCGCCTTGCAAAATAT

G S N G A R S L S I G F F S C S R C F L D G G Q L I

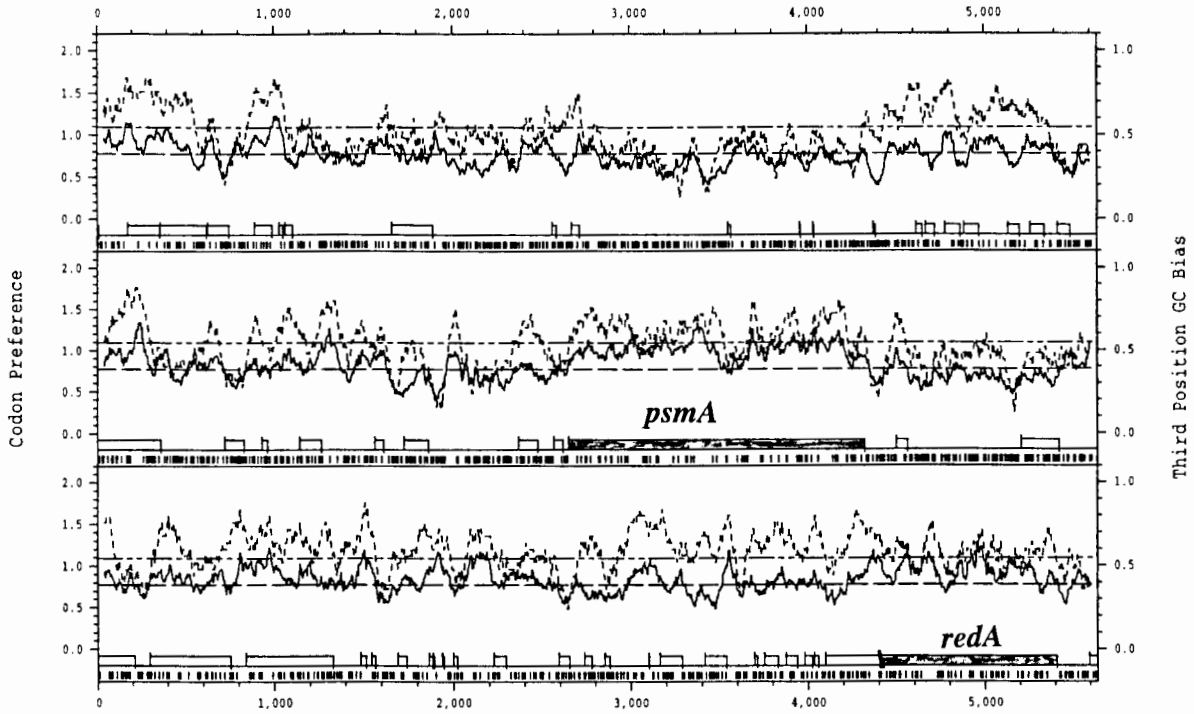
2081 CAACTATTTTCTCGTGGCCGTCAGCAGCCGACGTCACCAGCTTCACTTGTCCATCCAGAAGAAAATAGAAAGCGGTAGCC

D V I K E H G D A A S T V L K V Q G D L L F Y F A T A

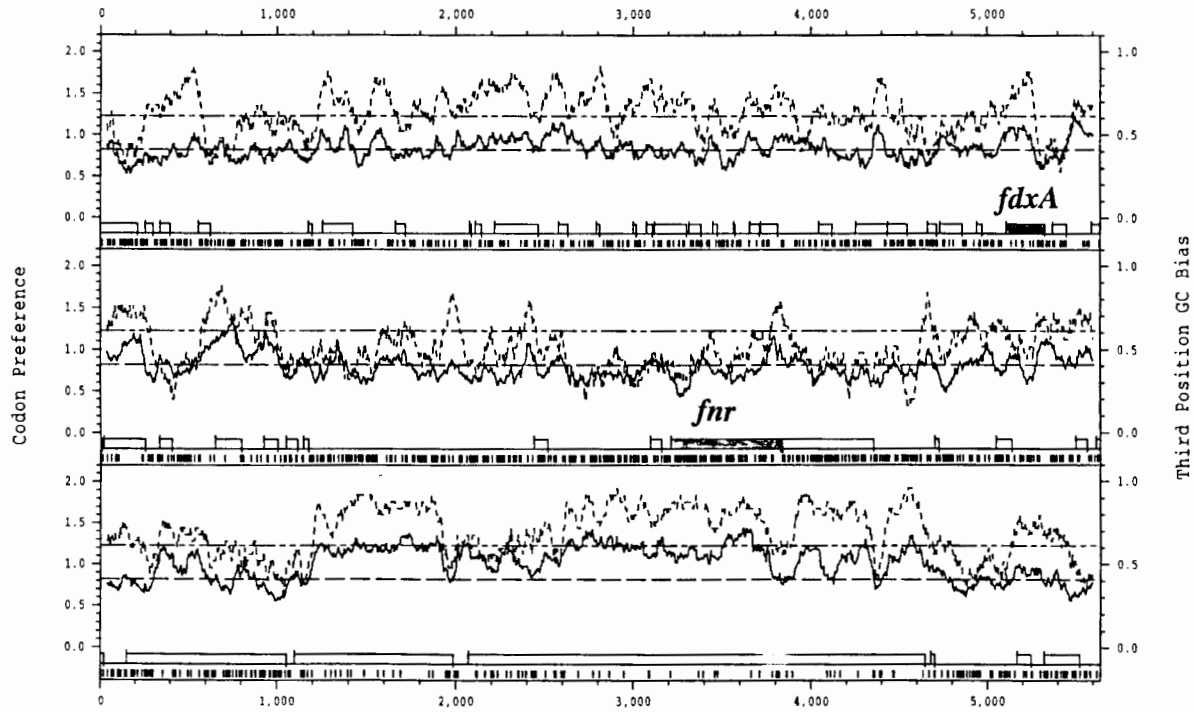
```
2161  GCATCGCCTTCAAGGAAAAGGATTTGCCCGACACTCATCTCCAGAACACGTGCGCCCTTAAATAAATGCGAAAGGTCCTC
       A  D  G  E  L  F  L  I  Q  G  V  S  M  E  L  V  R  A  G  K  F  L  H  S  L  D  E

2241  ATCCGACCAGACAGAGAATATTTCTGTTCGGCGCAATCGCTCCATTTCTGTTAACTCTACTTGGCCGATTGTCATCAATC
       D  S  W  V  S  F  I  E  T  R  R  L  R  E  M  E  T  L  E  V  Q  G  I  T  M
                                                                    fnr start
2321  TGGCTTTTGGGCGAACAAAACTGTGTCTAAAACCATCTTTACCCCATGATGCGGCTTTTACGGACCATAACTCAGCCTGT
                                                                              -12
2401  TATGTGCCTGCGGGTGGCGTCTCGCAGGCGGATTCAGCCGCCGGATTCCAGACAATTGATCGGGATAGGCGGAGCGTAGT
       N5  -24
2481  GTGACGTTAAATACGTTACGCCGAATCTCCGGGCAGGGTCAATCGCCTGTCTTTTGCTTAGGGTTTCCCGTCGACGATCA
                              FNR box                          -35      N17        -10
2561  TGAGGAATGCGCAGCCACTTGATTGAAATCAAGGATGGTCCTGATCACAGCAGATATTGTTGACAAACAAGCACCACAAC
2641  CTCAGGAGACACCATGATGTTTTGCTATCAATGCGAACAAACCACCCGCAGCCCCGCCGGAATAGGCTGTACCAGTGAGC
       start psmA   M  M  F  C  Y  Q  C  E  Q  T  T  R  S  P  A  G  I  G  C  T  S  E  P
                                              PstI
2721  CGGGAACGTGCGGCAAGGATGAGGCCACTGCGGTTCTGCAGGACATTCTCACGCACCTTATGAAGGGCATTGCCAGTATG
       G  T  C  G  K  D  E  A  T  A  V  L  Q  D  I  L  T  H  L  M  K  G  I  A  S  M

2801  CGCGCCGGGCGGGCGATGGGTGTTGCCGACAGACGTACGGACGACTTCATTTTTTATGGGCTTTTCACCACGCTCACCAA
       R  A  G  R  A  M  G  V  A  D  R  R  T  D  D  F  I  F  Y  G  L  F  T  T  L  T  N

2881  TGTGAACTTCACTGCCACACGTTTTGTCCACCTGATCCAGGAGGCCAGCAAAAGGCGTGAACGGATCAAATTATTGTACG
       V  N  F  T  A  T  R  F  V  H  L  I  Q  E  A  S  K  R  R  E  R  I  K  L  L  Y  E

2961  AGGAAGCGGCGCGAGAGCAGGGCAAGACTCCCGAAATCCTGTCCGGTCCGGCCCTCTTCCAACCTGCCGATAGTCTGGAA
       E  A  A  R  E  Q  G  K  T  P  E  I  L  S  G  P  A  L  F  Q  P  A  D  S  L  E

3041  CAACTGCTGCGTCAAGCCCCTAGCGTCGCCATAAACGCCGACGTAGAACATCTCGGATCCGATGTGATCGGTGCTCGTGC
       Q  L  L  R  Q  A  P  S  V  A  I  N  A  D  V  E  H  L  G  S  D  V  I  G  A  R  A

3121  CCTCATCCTCTACGGCATGAAAGGCGTAGCGCTTATTGCTCAACATGCCCGCGTCCTGGGTTATCAGAGCGACGAAGTCA
       L  I  L  Y  G  M  K  G  V  A  L  I  A  Q  H  A  R  V  L  G  Y  Q  S  D  E  V  M

3201  TGCCGCAGGCAGAGGAAATCCTGGACTACTTGGCAAGTAATCCAACGGACCTCGATGAAATGTTGGAAGAATCATTGGAA
       P  Q  A  E  E  I  L  D  Y  L  A  S  N  P  T  D  L  D  E  M  L  E  E  S  L  E

3281  GTCGGCCGTTTGAACCTGAAGGTGATGGAGTTGCTTGACGTTGCCAATACGGACAGTTTTGGAGCGCAGGAAATTACCTC
       V  G  R  L  N  L  K  V  M  E  L  L  D  V  A  N  T  D  S  F  G  A  Q  E  I  T  S

3361  TGTGCGCATCTCCCCGATCCAGGGCAAGGCAATCCTGGTCAGCGGTCACGATCTCCATGATCTCAAACAAATCCTGGAGC
       V  R  I  S  P  I  Q  G  K  A  I  L  V  S  G  H  D  L  H  D  L  K  Q  I  L  E  Q

3441  AGACCAAGGATCAGGGGATCAATGTTTACACCCATGGGGAGATGTTGCCAGCCAACGCCTATCCCTTGCTCAAGGCATAC
       T  K  D  Q  G  I  N  V  Y  T  H  G  E  M  L  P  A  N  A  Y  P  L  L  K  A  Y

3521  CCTCATCTGGCCGGGAATCTAGGGGGAGCATGGCAAGATCAGCAGCGTGAATTTGCAGATTTCCCTGGGCCCATTGTCAT
       P  H  L  A  G  N  L  G  G  A  W  Q  D  Q  Q  R  E  F  A  D  F  P  G  P  I  V  M

3601  GACTTCCAACTGTATCATTGAGCCGGGCAGAAGTTATAAAAATCGAATTTTTACCCTTGGCCCAGTGGGGTGGCCCGGCG
       T  S  N  C  I  I  E  P  G  R  S  Y  K  N  R  I  F  T  L  G  P  V  G  W  P  G  V
3681  TCCGCCATATCGACAATGGGGATTTCACCCCGGTGATCCAGGCCGCCAAGGCATTACCGGGATTTACTGCCGATGCAAAA
       R  H  I  D  N  G  D  F  T  P  V  I  Q  A  A  K  A  L  P  G  F  T  A  D  A  K

3761  GAGCAGCGTATCACCATCGGTTTTGGGCATCACACTCTCCTGGGTGTGGCGGATAAAATCGTCGATGCGGTGAAACACGG
       E  Q  R  I  T  I  G  F  G  H  H  T  L  L  G  V  A  D  K  I  V  D  A  V  K  H  G

3841  AGATATTCGCCACTTCTTCCTCGTCGGTGGGTGCGATGGTGTTTCTCCGGCGCGCAACTACTTCACGGAGGTGGCGGATA
       D  I  R  H  F  F  L  V  G  G  C  D  G  V  S  P  A  R  N  Y  F  T  E  V  A  D  N
                                                    ClaI
3921  ATGCCCCCGCCGATTCGGTAGTGATGACCCTGGGATGCGGTAAGTATCGATTCAACAAGCATGAATTCGGAGATATTGGA
       A  P  A  D  S  V  V  M  T  L  G  C  G  K  Y  R  F  N  K  H  E  F  G  D  I  G
```

```
4001  GGCATCCCCCGCCTGCTGGATATAGGCCAATGCAATGATGCCCACTCCGCCATCCGGGTAGCGGGTGCCCTGGCCGAGGC
        G  I  P  R  L  L  D  I  G  Q  C  N  D  A  H  S  A  I  R  V  A  G  A  L  A  E  A

4081  ATTCAATTGTGGGGTCAATGACCTGCCGTTGTCGATCATGCTCTCCTGGTTTGAACAAAAGGCTACCGCCATTCACCTTT
        F  N  C  G  V  N  D  L  P  L  S  I  M  L  S  W  F  E  Q  K  A  T  A  I  H  L  S

4161  CCTTGCTGGCTCTGGGCATCAAGGGGATCAAGCTGGGGCCCACCCTGCCTGCCTATCTCACACCAACTTTGGTGCAGAAG
          L  L  A  L  G  I  K  G  I  K  L  G  P  T  L  P  A  Y  L  T  P  T  L  V  Q  K

4241  CTCCAATCACGTTTCGATCTTGATCTCGATCTTATTGGCGAGGCGCAAGCGGATTTGCAGGCGGCTCTGGCGCATACGGC
        L  Q  S  R  F  D  L  D  L  D  L  I  G  E  A  Q  A  D  L  Q  A  A  L  A  H  T  A
                     ──────►                                                    EcoRV
4321  GTAGCACAATGGAACGCCGCTTCGTTCTGCAACGGACGAAGCGGCATAAGGATGAGCAGATCATGACCAACTACGATATC
        *                                ◄──────          start redA  M  T  N  Y  D  I

4401  ACTATCCATACCCGAGATAAGCAACAGGTGTCCTTCGTCTGTTCTGAGGCGGAAGATTTACTCTCCGCTGCGGACCGGGG
        T  I  H  T  R  D  K  Q  Q  V  S  F  V  C  S  E  A  E  D  L  L  S  A  A  D  R  G

4481  AAGCATTCTGCTTCCTTCCCAATGCCGGAAGGGAACTTGCGGAGCCTGCGTGGCCACCGTCACCGCAGGCACTTATCACC
        S  I  L  L  P  S  Q  C  R  K  G  T  C  G  A  C  V  A  T  V  T  A  G  T  Y  H  L

4561  TGGGTGAAGTCAGTATGGAAGCCCTGCCAGAGAAAGCGCAGGCACGAGGCGATGTACTTCTGTGCCGTACCTATCCACGA
          G  E  V  S  M  E  A  L  P  E  K  A  Q  A  R  G  D  V  L  L  C  R  T  Y  P  R

4641  GCGGATCTGATTCTGGAGGCGCCTTATGACTACAACTACATCCGCTTTGAGCGCATTCCGGAGCGCGAGGCAGAAGTGAT
        A  D  L  I  L  E  A  P  Y  D  Y  N  Y  I  R  F  E  R  I  P  E  R  E  A  E  V  M

4721  GGACGTCACTATGGTGGCTACGGGTACGCGACGACTGTTGTTACGCCTGCAACCCGATGAGCAAGGGGGAGCTGCGGAAT
        D  V  T  M  V  A  T  G  T  R  R  L  L  L  R  L  Q  P  D  E  Q  G  G  A  A  E  F

4801  TCGAAGCTGGGCAATTTATGGAAATCCAGGTACCTGGCAGCGATGCGCGTCGCCCTAATTCTCTAGCCAACAACACCAAC
          E  A  G  Q  F  M  E  I  Q  V  P  G  S  D  A  R  R  P  N  S  L  A  N  N  T  N

4881  TGGAATGGCGACCTGGAATTTTTCATCACGCTACGACCAGGTGGTGCCTTCTCTACCTATCTGGAATCCGCCTTGGTGGG
        W  N  G  D  L  E  F  F  I  T  L  R  P  G  G  A  F  S  T  Y  L  E  S  A  L  V  G

4961  CGACCGCCTGAACATACGTGGACCCCTGGGAACCTTTACCCTGAGAGAAAACGGCCTGCGCACGTTTGGTTCATTGGTGG
        D  R  L  N  I  R  G  P  L  G  T  F  T  L  R  E  N  G  L  R  T  F  G  S  L  V  A

5041  CGGGACAAGGTCTCGTGCCGCTACTGTCGATGCTGCGACGCATGGCAGATTGGGGAGAAATGCTTCCGGCACGTCTGTAC
          G  Q  G  L  V  P  L  L  S  M  L  R  R  M  A  D  W  G  E  M  L  P  A  R  L  Y
                                                                          PstI
5121  TTCGGCGCAAGATATGAAGACGAACTCTTCTGTCAGGAGGAAATCCGTCAGATCCAGGATAAACTTCCCCAACTGCAGGT
        F  G  A  R  Y  E  D  E  L  F  C  Q  E  E  I  R  Q  I  Q  D  K  L  P  Q  L  Q  V

5201  GAAGATATGCCTTTCACGTCCTGGCAATCACTGGATGGATTATCGAGGCAGCGTGGTGGATGCCTTGCGCGACGATCTGG
        K  I  C  L  S  R  P  G  N  H  W  M  D  Y  R  G  S  V  V  D  A  L  R  D  D  L  G

5281  GAAGCCTTGCCACTTTACCGGACTTGTATGTTTGCGGATCAACGCGATTGGTCCAGGGGGTGACGGAATTAGCTCTGAGC
          S  L  A  T  L  P  D  L  Y  V  C  G  S  T  R  L  V  Q  G  V  T  E  L  A  L  S
5361  CAAGGGTTGCCGGACTCCTGTTTGCAGTTCGAACGGTTTTTGTCCGCATAGTGCAATGTCTATAGACTTGCTAGAACCAT
        Q  G  L  P  D  S  C  L  Q  F  E  R  F  L  S  A  *
5441  TGATTTCGTCCAGCCAAGAAGGAACTCGTGGTAACCATAGAAACCAATAAAGACTATCGTTTTTCCCTTCTCCATCTTGG
5521  GTTCCGCCTCTTTTTTTTGTTAGGGGCGGTATTCTCTGCTTTTGCCATCTCTGCCTGGTTCTGGTTCTATTATCATGGAA
                                                  SphI
5601  CGCAGTTTCTTGATAATGGCAATTTCCCGATTATCCTGTGGCATGC
```

Forward strand



Reverse strand



The ORFs identified in Figure 3.6b have been graphically displayed in Figure 3.7 showing the UW-GCG (Devereaux et al. 1984) CODONPREFERENCE results and GC bias using codon preference tables used by chromosomally-associated *T. ferrooxidans* genes (Rawlings et al. 1991).

b.

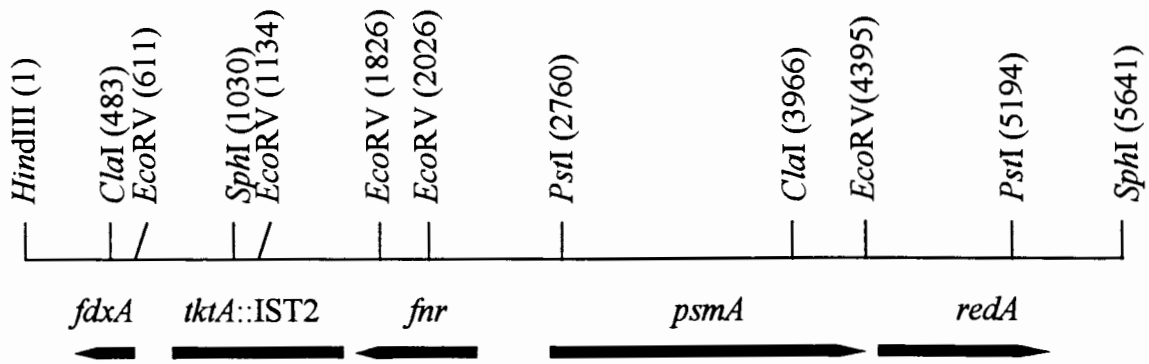


Figure 3.7a (opposite) ORFs of the 5.6 kbp fragment, as predicted by known codon preference patterns and GC bias from previously sequenced *T. ferrooxidans* ATCC33020 chromosomal genes (Rawlings et al. 1991). A physical map (Fig 3.7b) shows relevant restriction sites along with a summary of the ORFs.

The 366 bp *fdxA* gene, encoding a 122 aa protein (14 kDa), was identified on the basis of amino acid sequence identity to ferredoxins in the SWISS-Prot database. Multiple sequence alignments of the deduced FdxA sequence to other ferredoxin proteins (Fig 3.8) showed the 9 highly conserved internal cysteine residues responsible for binding to the two iron-sulphur moieties. The spacing of conserved cysteines within the predicted protein is characteristic of a [3Fe-4S, 4Fe-4S] binding sequence (Trower et al. 1990).

The *fdxA* gene had a putative RBS sequence, AGGGGG, 7 bp upstream of the translation start site. A putative *E. coli* like σ^{70} promoter recognition sequence occurred 92 bp upstream of the ORF. A *Stu*I- *Acc*I fragment from this region which included the first 16 aa of the ferredoxin protein (Fig 3.9), was fused in-frame to a promoterless *lacZ*

[4Fe-4S]
[3Fe-4S, 4Fe-4S]
[4Fe-4S, 4Fe-4S]
[4Fe-4S]
[3Fe-4S]
[2Fe-2S]
Sa VGVDFDL C IADGS.. C ITA CP VNVFQWYET PGHPASEK.. .....KADPV NQQA C IF C MA C VNV CP VAAI
Ta VAVDWDC C IADGA.. C MDV CP VNLYEWNLN PGKSGTGNDH KIQKGSEEWN KYRTDKCDPV RESD C IF C MA C ESV CP VRAI
Mt VTIDYDK C KGPECAE C VNA CP MEVFE.... .... .IQGDKVVVA KEDD C TF C MV C VDV CP TDAI
Ms TYVIAEP C VD VKDKA C IEE CP VD C I... ..... YEGAR..MLY IHPDE C VD C.....G...A C EPV CP VEAI
Se TYVIAEP C VD VLDKA C IEE CP VD C I... ..... YEGGR..MLY IHPDE C VD C.....G...A C EPV CP VEAI
Sg TYVIAQP C VD VKDKA C IEE CP VD C I... ..... YEGQR..SLY IHPDE C VD C.....G...A C EPV CP VEAI
Pp TFVVTDN C IK CKYTD C VEV CP VD C F... ..... YEGPN..FLV IHPDE C ID C.....A...L C EPE CP AQAI
Ps TFVVTDN C IK CKYTD C VEV CP VD C F... ..... YEGPN..FLV IHPDE C ID C.....A...L C EPE CP AQAI
Tt PHVICQP C IG VKDQS C VEV CP VE C I... ..... YDGGD..QFY IHPEE C ID C.....G...A C VPA CP VNAI
Tf THVVTEA C IR CKYTD C VTV CP VD C F... ..... HEGPN..FLA IDPDE C ID C.....T...L C VPE CP VDAI
Ca AYVINEA C IS C.GA C EPE CP VNAI..... ..... SSGDD..RYV IDADT C ID C.....G...A C AGV CP VDAP
Cs AYVINDS C IS C.GA C EPE CP VNAI..... ..... TAGDD..KYV IDAAT C ID C.....G...A C AGV CP VDAP
Cb AFVINDS C VS C.GA C AGE CP VSAI..... ..... TQGDT..QFV IDADT C ID C.....G...N C ANV CP VGAP
Cp AYKIADS C VS C.GA C ASE CP VNAI..... ..... SQGDS..IFV IDADT C ID C.....G...N C ANV CP VGAP
Ct AYFITDA C IS C.GA C ESE CP VSPI..... ..... SPGDS..VYV IDADA C IE C.....G...A C ANV CP VDAP
Csp AYKITDG C IN C.GA C EPE CP VEAI..... ..... SESDA..VRV IDADK C ID C.....G...A C ANT CP VDAI
Cv ALMITDQ C IN C.NV C QPE CP NGAI..... ..... SQGDE..TYV IEPSL C TE C VGHYETS..Q C VEV CP VDCI
Bs TIVDKET C IA C.GA C GAA AP DIYD..... ..... YDEDGIAYVT LDDNQGIVEV PDILIDDMMD AFEG CP TESI
Bt TIVDKET C IA C.GA C GAA AP DIYD..... ..... YDEDGIAYVT LDDNQGIVEV PDILIDDMMD AFEG CP TDSI
Pf VSVDQDT C IG D.AI C ASL CP DVFE..... ..... MNDEGKA... ..QPKVEVIE DEELYN C AKEAMEA CP VSAI
Tl VSVDKDA C IG C.GV C ASI CP DVFE..... ..... MDDDGKA... ..KALV..... AETDLE C AKEAAES CP TGAI
Dg IEVNDD C MA C.EA C VEI CP DVFE..... ..... MNEEGD..... ....KAVVIN PDSDLD C VEEAIDS CP AEAI
Hh GTME VA EGEY ILEAAEAQGY DWPFS C.RAGA C AN C ASIVK EGEIDMDMQQ ILSDEEVEEK DV.....RLT C IGSPAADEV
Hs GSLE VN EGEY ILEAAEAQGY DWPFS C.RAGA C AN C AAIVL EGDIDMDMQQ ILSDEEVEDK NV.....RLT C IGSPDADEV
Lg KEFE CP DDVY ILDQAEELGI ELPYS C.RAGS C SS C AGKLV EGDLDQSDQS FLDDEQIEEG WV.....LT C AAYPRSDVV
Sn QEFE CP DDVY ILEHAEELGI DIPYS C.RAGS C SS C AGKLV AGSVDQSDQS FLDDEQIEEG WV.....LT C VAYPKSDVT
Msa QEFE CP DDVY ILDHAEEEGI VLPYS C.RAGS C SS C AGKVA AGEVNQSDGS FLDDDQIEEG WV.....LT C VAYAKSDVT
Ce QEFQ CP DDVY ILDQAEEVGI DLPYS C.RAGS C SS C AGKVK VGDVDQSDGS FLDDEQIGEG WV.....LT C VAYPVSDGT
Wh VELE VP DDVY ILDQAEEEGI DLPYS C.RAGS C SS C AGKLV SGEIDQSDQS FLDDDQMEAG WV.....LT C HAYPKSDIV
Ec V.LE AN SGET ILDAALRNGI EIEHA C EKSCA C TT C HCIVR EGFDSLPESS EQEDDMLDKA WGLEPESRLS C QARVTDEDL

Figure 3.8 (opposite): Partial multiple sequence alignments of representatives of each class of ferredoxin protein, showing the conserved cysteine motifs responsible for binding the various Fe-S clusters. The ferredoxins used in the analysis were isolated from the following organisms with database accession numbers; *Sulfolobus acidocaldarius* - P00219 (Sa), *Thermoplasma acidophilum* - P00218 (Ta), *Methanococcus thermolithotrophicus* - P21305 (Mt), *Mycobacterium smegmatis* - P00215 (Ms), *Saccharopolyspora erythraeus* - P24496 (Se), *Streptomyces griseus* - P13279 (Sg), *Pseudomas putida* - P00213 (Pp), *P. stutzeri* - P08811 (Ps), *Thermus thermophilus* - P03942 (Tt), *T. ferrooxidans* - U73041 (Tf), *Clostridium acidi-urici* - P00198 (Ca), *C. sticklandii* - P80168 (Cs), *C. butyricum* - P00196 (Cb), *C. pasteurianum* - P00195 (Cp), *C. thermocellum* - P07508 (Ct), *Clostridium sp* strain M-E - P00197 (Csp), *Chromatium vinosum* - P00208 (Cv), *Bacillus stearothermophilus* - P00212 (Bs), *B. thermoproteolyticus* - P10245 (Bt), *Pyrococcus furiosus* - P29603 (Pf), *Thermococcus litoralis* - P29604 (Tl), *Desulfovibrio gigas* - P00209 (Dg), *Halobacterium halobium* - P00216 (Hh), *Halobacterium sp.* - P00217 (Hs), *Leucaena glauca* - P00225 (Lg), *Sambucus niagra* - P00226 (Sn), *Medicago sativa* - P00220 (Msa), *Colocasia esculenta* - P00222 (Ce), *Triticum aestivum* - P00228 (Wh) and *E. coli* - P25528 (Ec).

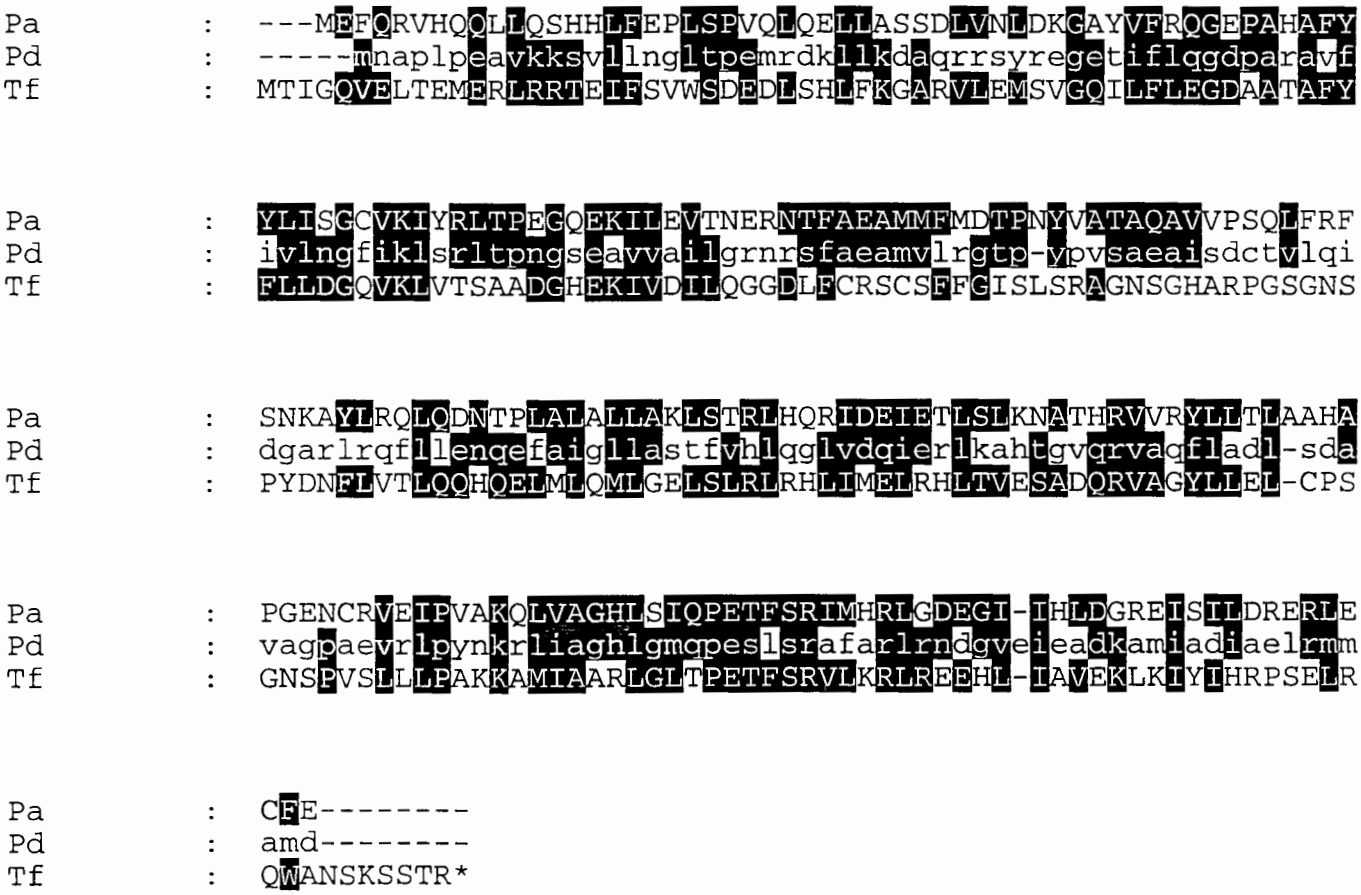
gene. The *lacZ* phenotype was able to be expressed in both pMC1403 and pACYC184 cloning vectors with colour intensities which corresponded approximately with the respective copy numbers of the vectors. The fusion, fxp2, was cloned into pACYC in order for it to be compatible with vectors pHC79 and pECOR251 used in the construction of the cosmid and plasmid gene libraries respectively. Both of these gene libraries were transduced/ transformed into *E. coli* DH5α which harboured fxp2. However no clones (chromosomal or pTF5-associated) could be identified which regulated the expression of the *lacZ* phenotype, suggesting that this promoter, if recognised in *T. ferrooxidans*, may be constitutively expressed.

TTGAGG - N_{17} - GATGAT - N_{92} - ATG ACC CAT GTG GTT ACA GAA
-35 -10 Met Thr His Val Val Thr Glu

GCC TGT ATT CGG TGT AAG TCC **GTC GTT TTA**
Ala Cys Ile Arg Cys Lys **Ser Val Val Leu**

Figure 3.9 : Nucleic acid sequence of the putative *E. coli* σ^{70} promoter sequence believed to be responsible for the expression of the *lacZ* phenotype. The sequence shows 4/6 bp homology with the -35 consensus sequence followed by appropriate spacing (17 bp - Harley and Reynolds, 1987) to the presumptive -10 consensus region (4/6 bp). The fusion site of the putative *fdxA* ORF with the promoterless *lacZ* gene of pMC1403 (Casadaban et al. 1983) is shown downstream of the putative ATG start site in bold. The *fdxA* genes *Acc*I (GTCGAC) and *Eco*RI (GAATTC - vector) restriction sites used were flushed off by T_4 polymerase (Appendix C) and were thus destroyed in the cloning event.

Immediately upstream of the *fdxA* gene a truncated ORF of 82 aa with 48% predicted amino acid sequence similarity to the transketolase genes from *E. coli* was found (Sprenger et al. 1993). This truncated ORF was fused to a truncated IST2-like insertion element (81% nucleic acid identity) similar to that found in *T. ferrooxidans* ATCC19859 (Yates et al. 1988). The truncated IST2-like element of pTF5 had homology to the first 510 bp of the IST*2* sequence, with a 48 bp deletion corresponding to the first 16 aa of the presumptive IST*2* transposase enzyme (Fig 3.10). This "scar" is evidence of recombinational / transpositional activity in this region of pTF5. Stem-loop structures have been detected at the fusion sites of this "scar" region, but the significance of these features is as yet unknown.


Pa : ---MEFQRVHQQLLQSHHLFEPLSPVQLQELLASSDLVNLDKGAYVFRQGEPAHAFY
Pd : -----mnaplpeavkksvllngltpemrdkllkdaqrrsyregetiflqgdparavf
Tf : MTIGQVELTEMERLRRTEIFSVWSDEDLSHLFKGARVLEMSVGQILFLEGDAATAFY
Pa : YLISGCVKIYRLTPEGQEKILEVTNERNTFAEAMMFMDTPNYVATAQAVVPSQLFRF
Pd : ivlngfiklsrltpngseavvailgrnrsfaeamvlrgtp-ypvsaeaisdctvlqi
Tf : FLLDGQVKLVTSAADGHEKIVDILQGGDLFCRSCSFFGISLSRAGNSGHARPGSGNS
Pa : SNKAYLRQLQDNTPLALALLAKLSTRLHQRIDEIETLSLKNATHRVVRYLLTLAAHA
Pd : dgarlrqfllenqefaigllastfvhlqglvdqierlkahtgvqrvaqfladl-sda
Tf : PYDNFLVTLQQHQELMLQMLGELSLRLRHLIMELRHLTVESADQRVAGYLLEL-CPS
Pa : PGENCRVEIPVAKQLVAGHLSIQPETFSRIMHRLGDEGI-IHLDGREISILDRERLE
Pd : vagpaevrlpynkrliaghlgmqpeslsrafarlrndgveieadkamiadiaelrmm
Tf : GNSPVSLLLPAKKAMIAARLGLTPETFSRVLKRLREEHL-IAVEKLKIYIHRPSELR
Pa : CFE--------
Pd : amd--------
Tf : QWANSKSSTR*

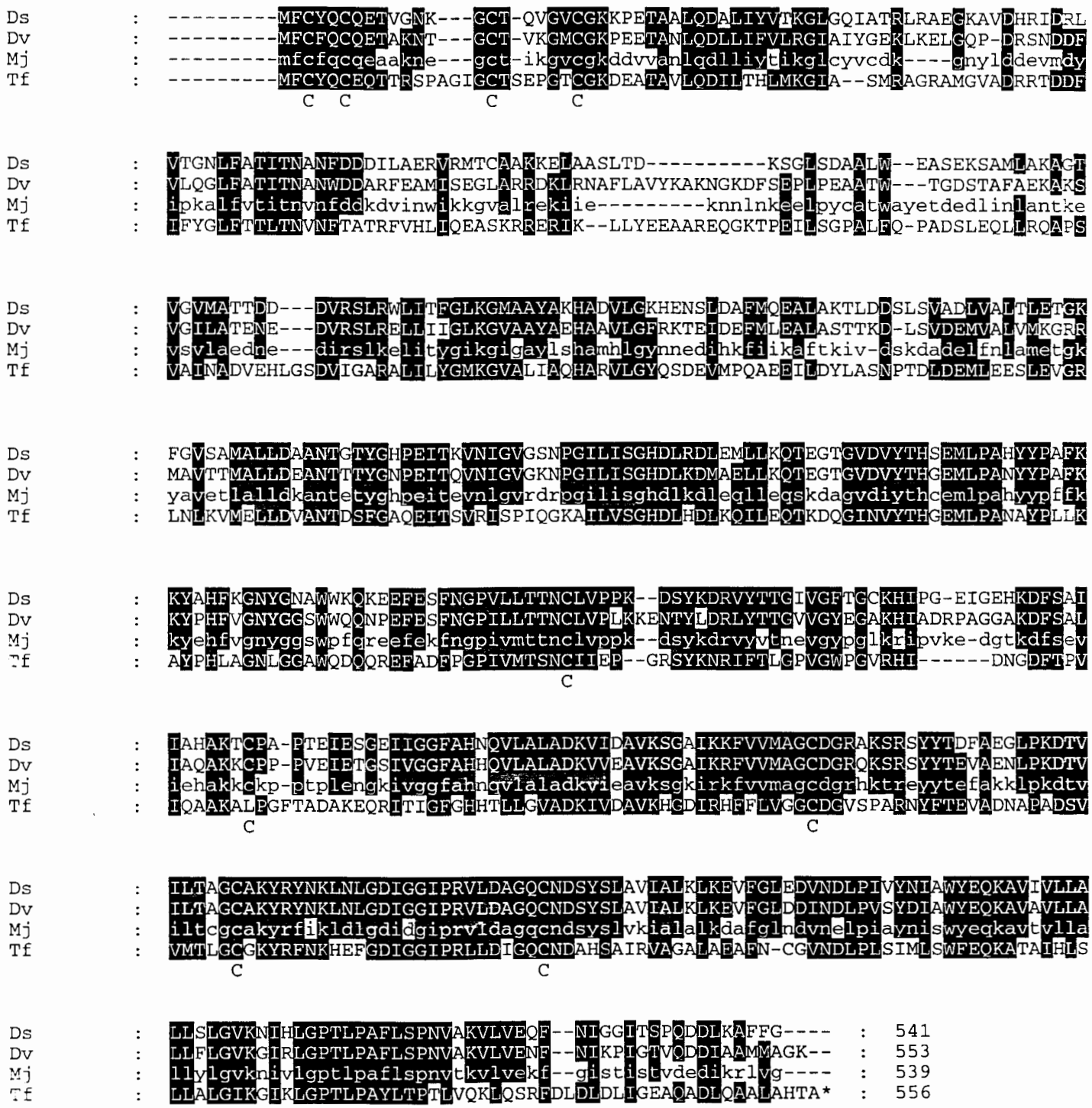
```
pTF5  CCGTTACGAACGACCGCCTGACGACCATTGACCATCCGTACAGCAGAAAA
      ||  |||||  |||||||||||||||  ||||  ||||||||  ||  |   | |||
IST2  CCATTACGCACGACCGCCTGACGCCCATCAACCATCCGCACCGTCGCAAA

pTF5  CTCTTCCAGCACTTGGTC
      ||||||||||
IST2  TTCTTCCAGCAATACCGCCACCTCTGCCTCTATGGCCTGCTCAATGCAGA

IST2  TGCCGCGCGTCGAAGTATCCCTTCAATGCCCAGACCCAACTCTCCCATCC

IST2  CACCTGCTATTACGGTATTCTTTTCCACGGCGTACTCCTTATGTTGCTCT

pTF5        GCAAACCCTTTGTAGCAACAGTACGCCACCT
            | ||||||||||||||||||||||||||||
IST2  TTGAGTCGGAAACCCTTTGTAGCAACAGTACGCCACCTCATTCAAGCCAG

IST2  TTGTAACGCGCCCATACACCACTTTCGAGCATAGCTCGGGAATGATCAAG

IST2  TCGGCATGAATTACCG
```

Figure 3.10 : Partial alignment of the pTF5 truncated IST2 element with that from *T. ferrooxidans* ATCC19859 (Yates et al. 1988).

Downstream of this "scar" region is a 609 bp ORF (26.5 kDa) with amino acid sequence similarity to the FNR family of regulatory proteins. The 203 aa product has highest homology to the DNR/NNR proteins of *P. aeruginosa* (28.9% aa identity) and *Paracoccus denitrificans* (27.5% aa identity) (Arai et al. 1995, Van der Oost, 1995) as shown in Fig 3.11. The DNR protein regulates the nitric oxide (*norCB*) and nitrite reductase (*nirS*) genes in *P. aeruginosa* in a similar manner to the NNR homologue in *P. denitrificans*.

Figure 3.11 (opposite) : Multiple sequence alignment of the pTF5 predicted FNR protein with that from *P. aeruginosa* (Pa)- accession number D50019 and *P. denitrificans* (Pd) - accession number U17435.

A σ^{54}-like promoter recognition sequence was observed upstream of the *fnr* ORF with the sequence CAGGCAC - N_5 - AGGCT which agrees over 6/7 bp and 2/5 bp with the σ^{54} -24/-12 promoter consensus sequence. Furthermore a perfectly conserved FNR

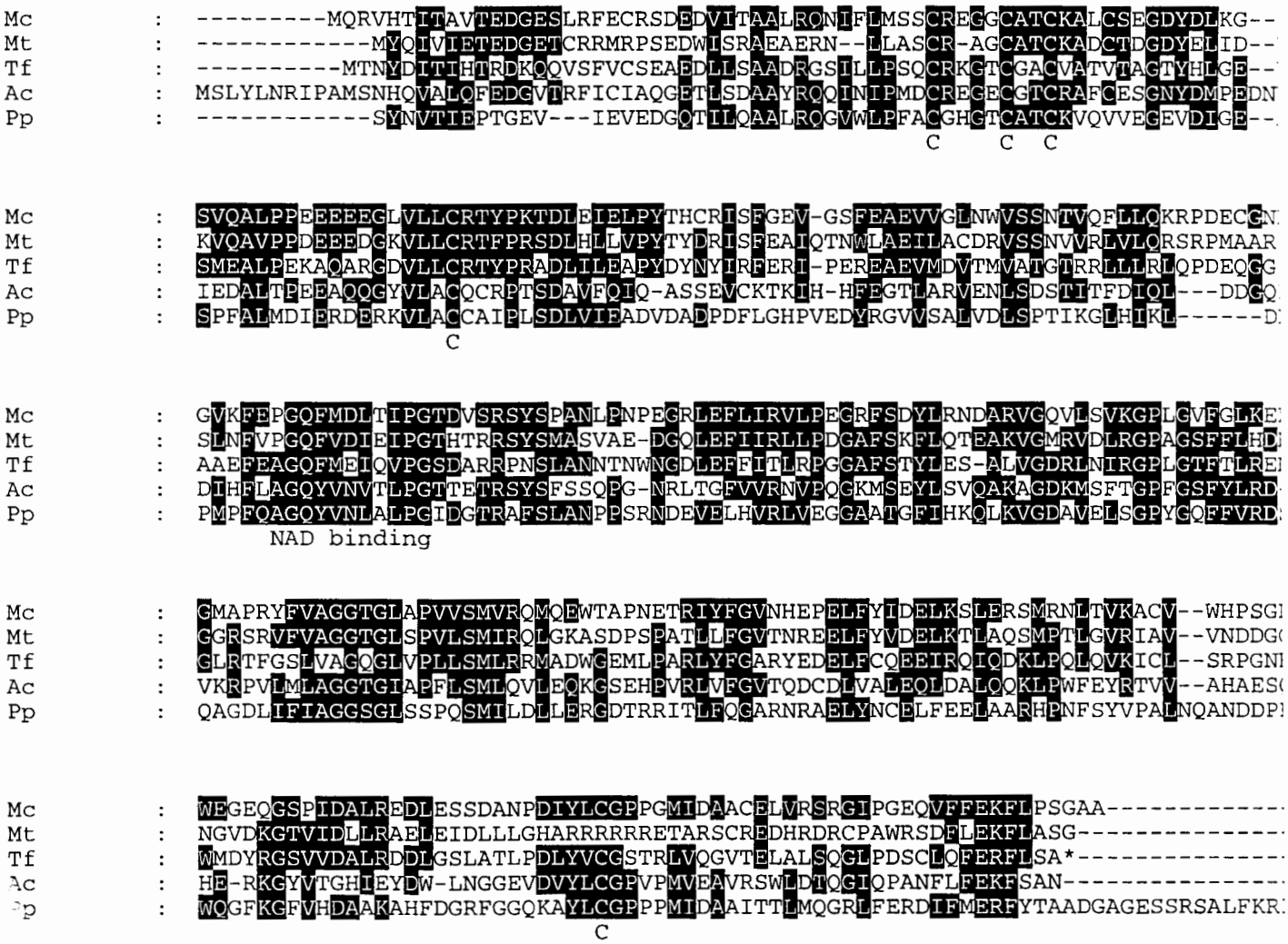


consensus binding site "TTGAT-N_4-ATCAA" was detected upstream of the prismane-like ORF, (*psmA)* adjacent to the *fnr*-like ORF. This implies that the prismane gene is regulated by FNR in *T. ferrooxidans*, although whether it responds to oxygen limited conditions could not be determined due to the lack of internal cysteine residues in the pTF5 homologue, which are required for this function.

The 1668 bp prismane-like ORF corresponds to a 556 aa protein with a predicted M_r of 60.3 kDa. Three other prismane proteins have been described, two from *Desulfovibrio (D. vulgaris* 46% aa identity, *D. desulfuricans* 43% aa identity) and one from the archeon *Methanococcus jannaschii* (41% aa identity). As is the case with the ferredoxins, conserved cysteine residues serve to bind the characteristic [6Fe-6S] prismoidal moiety to the protein. The three previously reported prismanes have nine of these conserved cysteines, the first four of which have a spacing of Cys-X_2-Cys-X_7-Cys-X_5-Cys, whilst the deduced pTF5 prismane has eight cyteine residues, the first four with a Cys-X_2-Cys-X_{11}-Cys-X_6-Cys spacing (Fig 3.12). Whether this new configuration would alter the structure of the iron-sulphur moiety could not be established without purifying the prismane from *T. ferrooxidans*. A weakly conserved *E. coli*-like σ70 promoter recognition sequence was observed 90 bp upstream of the *psmA* start codon. The predicted promoter sequence TTGCTT - N_{17} - GATCAT shares 3/6 and 4/6 bp homology with the *E. coli* σ^{70} consensus sequence.

Figure 3.12 : Multiple sequence alignments of the prismane proteins showing the conserved cysteine residues (*) between homologues from *D. vulgaris* (Dv) accession number P31101, *D. desulphuricans* (Ds) - accession number Q01770, *Methanococcus jannaschii* (Mj) - accession number U67522 and *T. ferrooxidans* (Tf).

Immediately downstream of the prismane-like ORF is an 1029 bp NADH oxidoreductase-like ORF encoding a 343 aa protein with a predicted Mr of 38.1 kDa. The ORF has a 40.5% amino acid sequence identity with subunit c of the *Methylococcus*


Mc
Mt
Tf
Ac
Pp
NAD binding

capsulatus methane monooxygenase (Stainthorpe et al. 1991). Homology was also noted (Fig 3.13) to the *Methylosinus trichosporium* methane monooxygenase subunit c (36.2% aa identity) (Cardy et al. 1991) as well as to NADH reductases from the phenol hydroxylase pathway of *Pseudomonas* sp strain CF600 (27.7% aa identity) (Nordlund et al. 1990), and the benzoate degradation pathway from *Acinetobacter calcoaceticus* (26.4% aa identity) (Neidle et al. 1991). Subunit c in the methanogenic complex reduces NADH and donates two electrons to the oxygenase subunits a,b which are responsible for oxidising methane to methanol in *M. capsulatus*.

Figure 3.13 : Multiple sequence alignments of the pTF5 *redA* gene against its closest homologues in the GenEMBL databases. Homologues with accession numbers include *Methylococcus capsulatus* - P22868 (Mc), *Methylosinus trichosporium* - C48360 (Mt), *Acinetobacter calcoaceticus* - P07771 (Ac) and *Pseudomonas putida* - P19734 (Pp). Conserved "Rieske"-type cyteine residues are indicated by C, the predicted NADH binding site (Mason and Cammack, 1992) is also shown.

In common with other bacterial oxygenase NADH oxidoreductases the pTF5 ORF contains a Cys-X_3-Cys-X_2-Cys-X_{31}-Cys motif which agrees closely with the "Rieske" ferredoxin motif (Fig 3.13). Furthermore regions within the NADH oxidoreductase-like ORF can be identified as FAD/NAD(P) binding regions based on conserved aa similarity GQ-N_6-PG with other homologues in the SWISS-Prot database (Neidle et al. 1991). An interesting feature was a 10 bp inverted repeat sequence which may form a stem-loop structure (ΔG -14.6 kcal.mol), in the 58 bp intergenic region between the prismane- like ORF and the NADH oxidoreductase- like ORF. Whether the stem-loop structure would serve as a translational terminator between the two ORFs is uncertain.

3.3.2 *In vitro* protein analysis

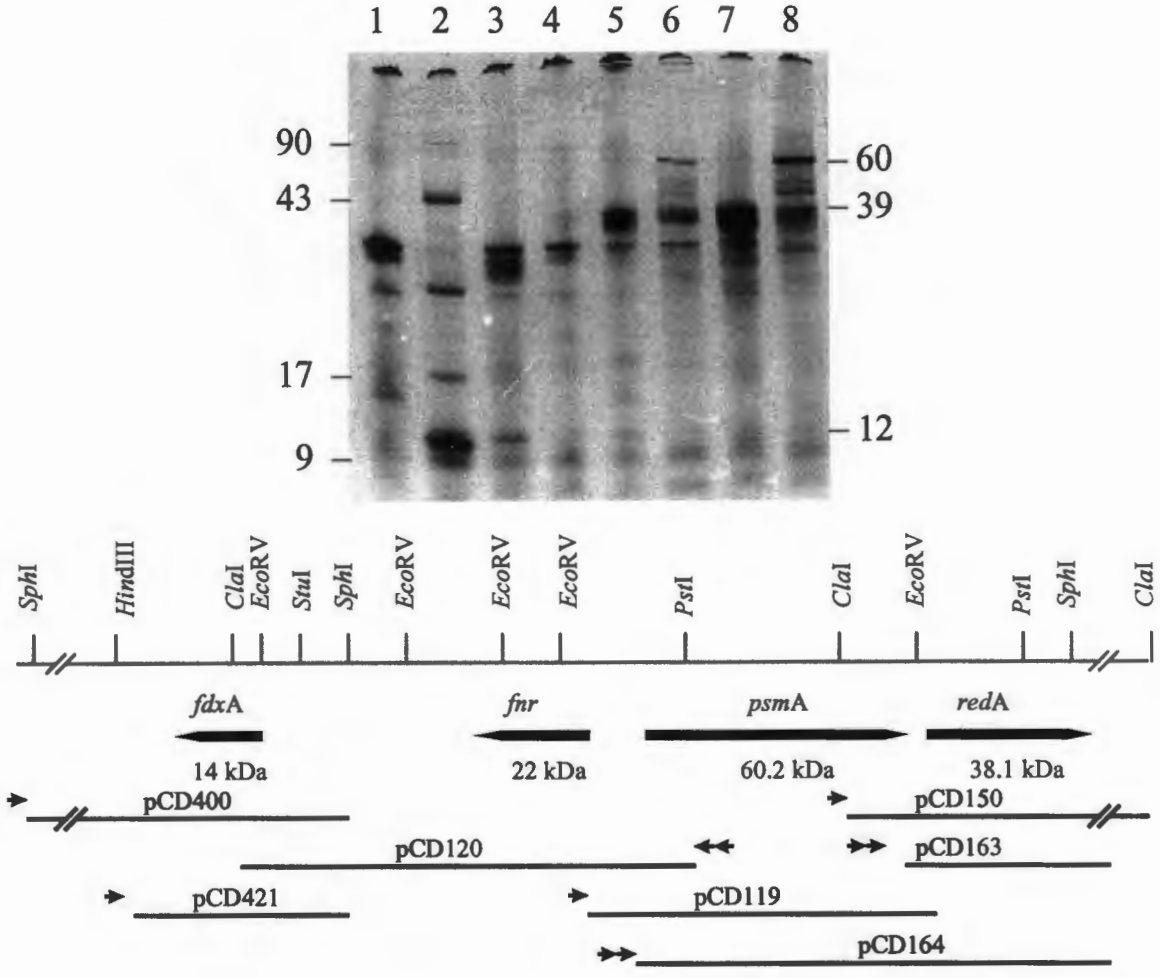


Figure 3.14 : SDS-PAGE analysis of polypeptides translated *in vitro* from pCD1 (pTF5 cloned into pBS-KS) and pTF5 subclones. The location of the constructs tested is shown below the PAGE gel and the direction of the vector promoter indicated by one arrowhead, *lacZ* promoter and two arrowheads, *tac* promoter. Lane 1 pBS-KS vector control, lane 2 pCD1, lane 3 pCD421, lane 4 and 5 pCD150 in both vector orientations, lane 6 pCD119, lane 7 pCD163, lane 8 pCD164.

In order to determine whether the iron-sulphur protein encoding genes were expressed in *E. coli, in vitro* transcription-translation reactions were carried out on selected pTF5 fragments cloned into *E. coli* vectors (Figure 3.14). Due to the lack of pTF5 protein

specific antibodies, a heterologous system had to be used for this purpose. This had the obvious disadvantage, that the *E. coli* system may not contain all of the factors required for the expression of *T. ferrooooxidans* genes. Polypeptides produced from pTF5 are shown in Fig 3.12. When compared to the pBS-KS control (lane 1), plasmid pCD1 (pTF5 cloned into pBS-KS, lane 2) produced 90, 40, 26, 17 and at least two 14 - 9 kDa protein products discussed in Chapter 2. The only polypeptide produced by the iron-sulphur-protein encoding region was produced from pCD421 (lane 3) which contained the ferredoxin gene. The protein was expressed from a pTF5 promoter-like sequence due to the expression of the ORF in the opposite orientation to the pBS-KS *lacZ* promoter. It should be noted however that the *lacZ* promoter of the vector may well have been active as a larger 28 kDa polypeptide was detected (lane 3) which did not correspond to a recognisable ORF within pCD421. The ferredoxin gene in all probability, was expressed from the same promoter sequence identified in the *lacZ* fusion experiment described in Section 3.3.1. A 60 kDa protein was expressed (lane 6) from a PCR product of the prismane-like ORF which had been cloned behind a *lacZ* promoter or the *tac* promoter (pCD164, lane 8), but not from pCD1 (lane 2). This implies that the endogenous prismane promoter was not recognised by the *E. coli* system.

A similar result was obtained for the NADH oxidoreductase-like ORF. A protein doublet of 39 and 37 kDa was detected in clones which oriented the ORF in the same direction as the *lacZ* promoter (lane 5) or *tac* promoter (lane 7) but not when cloned in the opposite orientation (lane 4). Interestingly, when the region containing both the prismane- and NADH oxidoreductase-like ORFs was cloned (following PCR amplification) behind a tac promoter (pCD164) and the expression of the proteins tested, the 60 kDa and 39, 37 kDa proteins were produced, arguing against the existence of a terminator element between the two ORFs (lane 8). Although the proteins produced by the *in vitro* kit can not be assigned to specific ORFs without Western antibody analysis, the proteins all corresponded closely with the predicted sizes of the ORFs.

A clone of the PCR product of the *fnr*-like ORF failed to produce an appropriate polypeptide on *E. coli*-derived *in vitro* transcription-translation kits (pCD120). This does not imply that it is not expressed within *T. ferrooxidans*, as not all of the components involved in its expression may be present in the *E. coli* system.

3.3.3 Investigation of an expressed phenotype from the iron-sulphur protein encoding ORFs of plasmid pTF5.

As has been discussed in Chapter 1, plasmids carry non-essential genes which may give their hosts a selective advantage in their ecological niches. A number of plasmids have been described for other bacteria which confer heavy metal resistance on their host cells (Novick and Roth, 1968; Chopra, 1975). To test whether plasmid pTF5 might encode a metal resistance mechanism (the reason for this speculative experiment is explained in the discussion), Luria agar containing an increasing percentage of either uranyl salts or silver cations was used to determine whether *E. coli* DH5α expressing the pTF5 ORFs, had an altered MIC to these heavy metals. Constructs expressing either the prismane-like gene, the NADH oxidoreductase-like gene, or both were unable to raise strain DH5α's tolerance to these heavy metals above that of the vector-containing controls (0.25% $UNO_3.6H_2O$, 0.005% $AgNO_3$). Complementation assays were carried out to determine whether the truncated pTF5 transketolase-like ORF and the complete *fnr*-like ORF could complement appropriate *E. coli* mutants. As expected the truncated transketolase ORF did not express in either *E. coli* AI118 (Zhao and Winkler, 1994) or GSJ423 (Sprenger et al. 1995). The pTF5 *fnr*-like ORF was also unable to complement an *fnr*-requiring *E. coli* mutant.

3.4 Discussion

It is possible that the cluster of iron-sulphur protein-encoding ORFs present on plasmid pTF5 may form an electron transport chain, or part thereof. In order to establish this, a

thorough biochemical investigation of the proteins expressed from this plasmid would be required. For example the redox potential of the transported electrons, and therefore the likely order in which these proteins occur within the putative chain would need to be established. A second important aspect and even greater challenge would be to determine the role that the pTF5 proteins might play (if any) in conferring a selective advantage to the *T. ferrooxidans* cells which harbour the plasmid.

General regulator genes (Werlen et al. 1996) and a FNR-like regulators (Dispensa et al. 1992) have been described in some oxygenase pathways indicating that at least some of these systems are subject to regulation. If expressed in *T. ferrooxidans*, the FNR-like product is most likely to bind at the perfectly conserved FNR box upstream of the *psmA* gene in pTF5. On the basis of the heterologous expression studies it would seem that the *fdxA* gene is constitutively expressed, at least in *E. coli*. The *fdxA* gene product may not necessarily be physiologically linked to the *psmA-redA* gene products due to the presence of the intervening transposition scar. The physical linkage of the *fdxA* gene may be coincidental and it may well have an alternate role in the cell. However the possibility that the *fdxA* gene was deliberately recruited by transpositional activity cannot be excluded. One can envisage a scenario whereby the *fdxA* gene was spliced into its current orientation, and then the entire *fdxA-redA* region inserted into an ancestral plasmid as a single cassette.

The observation that many of the bacterial oxygenase pathways are carried on plasmids (Harayama and Rekik, 1990) is noteworthy especially when some of the plasmid associated genes have been found on mobile DNA elements (Nakatsu et al. 1995, Werlen et al. 1996). This indicates that oxygenase complex can be inherited in both a vertical (replication) and horizontal (conjugation, transposition) fashion. In the light of the transposase and recombinase-like ORFs on plasmid pTF5 (Chapter 2), it would appear as if the pTF5 cluster could have been transferred in such a manner.

A highly speculative model based on a comparison of the pTF5 ORFs with bacterial oxygenase systems is outlined in Fig 3.15. The prismane-like ORF, (*psmA*) and the NADH oxidoreductase ORF (*redA*) are likely to be encoded by a monocistronic mRNA species, and that this potential operon is regulated in all probability by FNR (due to the presence of the FNR box). This suggests a simple pathway whereby NAD(P)H is oxidised by RedA and two electrons are passed to the prismane protein, PsmA.

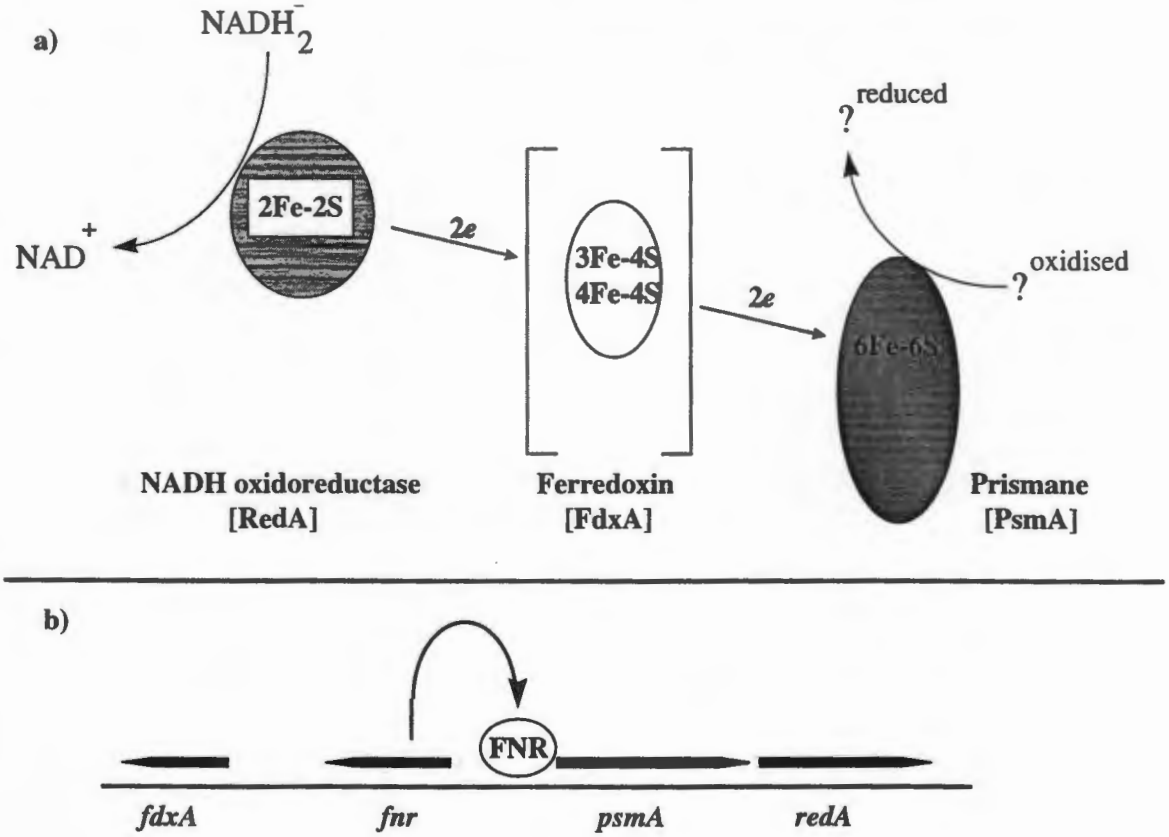


Figure 3.15 : A proposed electron transport mechanism for the proteins whose ORFs are found on plasmid pTF5. a) A schematic diagram of the pathway at the protein level, showing the possible involvement of the FdxA protein. b) The proposed regulation of the pathway based on evidence observed at the nucleic acid level.

An understanding of the physiological function of the prismane protein is probably the key to determining the active phenotype of plasmid pTF5 in *T. ferrooxidans*

ATCC33020. Studies of research on the prismane proteins of the *Desulfovibrio* will be monitored with interest.

As stated previously, it cannot be determined at present whether the ferredoxin protein, FdxA, acts as an intermediate between the RedA and PsmA proteins, or whether it performs another role in the cell. Although ferredoxins in bacterial hydroxylase/ oxygenase systems tend to have Rieske-type [2Fe-2S] clusters, it should be noted that in the case of *S. erythraea* a [3Fe-4S,4Fe-4S] ferredoxin has been implicated in donating electrons to a cytochrome P-450 hydroxylase. This in turn catalyses the conversion of 6-deoxyerythronolide B to the mature erythronolide B during erythromycin biosynthesis (Shafiee and Hutchison, 1987, Donadio and Hutchison, 1991). It is interesting to note therefore that the high potential "Rieske"-type ferredoxins are not always appropriate in these electron transport chains, and lower potential ferredoxins such as the [3Fe-4S,4Fe-4S] homologue referred to above, have been described.

Testing the constructs in a heterologous host such as *E. coli* for phenotypes such as resistance to heavy metal elements, is of limited value due to the possibility that the mature folded polypeptides may not be functionally active in that host. This was the case for the *D. vulgaris* prismane gene which did not bind to its [6Fe-6S] moiety in *E. coli* and thus would probably have been inactive (Stokkermans et al. 1992b). Nevertheless the pTF5 ORFs, *psmA* and *redA*, were expressed in *E. coli* when placed in front of the *tac* promoter.

The possibility of pTF5 conferring resistance to uranium and silver was investigated in *E. coli* cells over-expressing the *fdxA, psmA* and *redA* ORFs. A number of lines of evidence led to the inclusion of uranium in the study. Firstly, the *T. ferrooxidans* strain from which pTF5 was isolated, ATCC33020 had been reported to be resistant to uranium (Holmes et al. 1983). Secondly sulphate reducing bacteria, such as the prismane-bearing *Desulfovibrio* have been used to detoxify uranium-contaminated water sources in the USA (Lovley, 1993). Finally, there was the rather weak circumstantial evidence referred

to in Chapter 1 that Canadian *T. ferrooxidans* isolates which had become sensitised to uranium had also been observed to have lost a 20 kbp plasmid.

The reason for the inclusion of silver in the study was due to the suspicion of the Italian researchers working on pTFO that the plasmid conferred tolerance to silver on its host organism (Valenti et al. 1990, Visca et al. 1990). These researchers also investigated mercury resistance in these strains, but concluded that the phenotypes were not associated with presence of plasmid DNA. This confirmed the findings of Inoue et al. (1989) who cloned and characterised the mercury resistance operon and found it to be chromosomally associated, as discussed in Chapter 1.

Definitive proof of pTF5 containing markers for uranium (or any other metal) resistance would require the reintroduction of endogenous plasmid back into a uranium sensitive bacterium and its successful regeneration on a heavy metal containing solid medium. As discussed in Chapter 1, the tools required for this experiment are still unreliable in *T. ferrooxidans.*

Chapter 4

Distribution analysis and characterisation of pTF5 and related plasmids in *T. ferrooxidans* isolates.

Table of Contents

4.0 Summary

Several aspects of plasmid pTF5 were characterised by pulsed field gel electrophoresis (PFGE) and Southern hybridisation. The plasmid contained up to 6 kbp of sequence which was also present on the chromosome of at least two *T. ferrooxidans* strains, ATCC33020 (Japan) the strain from which pTF5 was cloned and FC1 (South Africa). This region covered the portion of pTF5 which contained the iron-sulphur protein-encoding genes (Chapter 3). Genomic walking along this region, enabled the junction sites between exclusively plasmid encoded sequence and that which was also present on the chromosome to be located within a region of 0.2 kbp and 0.6 kbp from the ends of the *fdxA* and *redA* genes respectively. PFGE showed the existence of pTF5-like plasmids in a number of strains of *T. ferrooxidans* isolated from within South Africa. The plasmid was not detected in all isolates of *T. ferrooxidans*, nor in the *T. thiooxidans* and *L. ferrooxidans* strains tested. The copy number of pTF5 in *T. ferrooxidans* ATCC33020 was determined to be between 2 and 4 copies per genome.

4.1 Introduction

4.1.1 Pulsed Field Gel Electrophoresis

Pulsed field gel electrophoresis has provided researchers with an efficient means to resolve very large DNA fragments of between 10 kbp and 10000 kbp on agarose gels (Smith and Condemine, 1990, Fonstein and Haselkorn, 1995). A number of variations of PFGE have been developed which all rely on slightly differing electric fields being applied to DNA samples to separate out these large genomic fragments (Southern and Elder, 1995).

This is a well-used technology and the main aspects of PFGE will be reviewed briefly. The first, and simplest, form of PFGE was field inversion gel electrophoresis (FIGE). This relied on the application of two electric fiels of unequal intensity from opposite ends

of the agarose gel (Fig 4.1a). The stronger electric field resulted in a net forward migration of the DNA sample. A disadvantage of this system was that DNA molecules of different sizes often had the same mobility, and thus an accurate sizing of specific bands was not always possible (Southern and Elder, 1995). Consequently crossed field gel (CFG) systems were developed which overcame these difficulties.

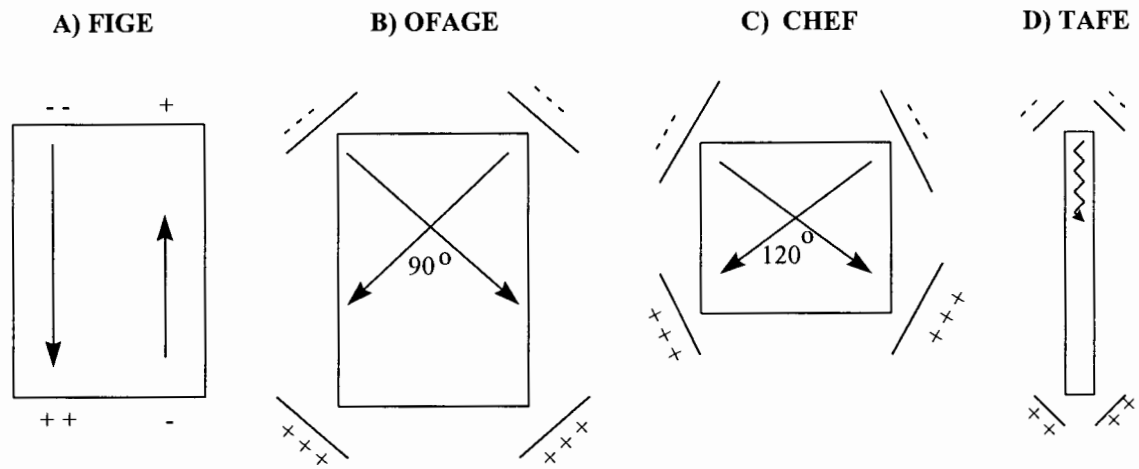


Figure 4.1 : Diagram illustrating the application of the electric field in A) field inversion gel electrophoresis (FIGE) B) orthogonal field gel electrophoresis (OFAGE) C) contour-clamped homogenous electric field electrophoresis (CHEF) D) transverse alternating gel electrophoresis (TAFE).

The main differences between the various CFG systems is i) the manner in which the electric fields are applied to the gel, ii) the angle of the DNA direction changes, and iii) the shape of the electric field (Smith and Condemine, 1990). Orthoganol field gel electrophoresis (OFAGE) was originally developed in order to separate out bacteriophage genomes in the size range 50-750 kbp (Carle and Olson, 1984). It had two electric fields set at 45° angles to the agarose gel (Fig 4.1b). In a similar manner contour-clamped homogenous electric field (CHEF) electrophoresis applies two electric fields at fixed angles of 120° (Fig 4.1c) across the width of the agarose gel to separate out very large fragments of over 100 kbp - up to 3 Mbp (Ougen and Cohen, 1995). It has been used in

the mapping of YAC's in a number of genomic sequencing projects (Fonstein and Haselkorn, 1995). Transverse alternating field electrophoresis (TAFE) differs from CHEF (Fig 4.1d) in that the electric fields are applied across the depth of the gel in a vertical electrophoresis tank (Gardiner and Patterson, 1988). This has the advantage of eliminating distortion between the lanes on the gel, whilst DNA fragments of up to 1 Mbp can be adequately resolved. By varying the pulse times ie. the time between the switching of electric fields, researchers can control the resolution parameters for specific sizes of DNA.

PFGE complements other techniques such as gene encyclopaedia's and total DNA sequencing in constructing high resolution physical maps of entire genomes (Fonstein and Haselkorn, 1995). It functions primarily as a rapid means of ordering genes on a genome, and provides valuable information to aid the more labour intensive encyclopaedia/ sequencing methods mentioned above. Many prokaryote genomes have been mapped in this way, a synopsis of which is beyond the scope of this introduction. However some interesting phenomena have been uncovered by these investigations and deserve further discussion. Firstly, genome size can vary considerably within a genus, with the spirochaetes containing between 1 and 5 Mbp of genome, while the cyanobacteria vary between 2.7 and 6.5 Mbp. Large (2 Mbp) chromosomal deletions have been reported in some strains of *Streptomyces ambofaciens* leading to questions as to why these bacteria can "afford" to carry these large, apparently non-essential blocks of sequence (Leblond et al. 1991). Secondly the notion of prokaryote chromosomal organisation consisting of a single circular genome has been disproven in a number of bacteria, most of which belong to the α sub-division of *Proteobacteria. Rhodobacter capsulatus* has been shown to contain a 3.7 Mbp chromosome as well as a 130 kbp plasmid, while the closely related bacterium *R. sphaeroides* has two chromosomes of 3.05 and 0.9 Mbp. Linear chromosomes have been reported in *Borrelia burgdorferi* (Ferdouws and Barbour, 1989), streptomycetes (Leblond et al. 1991) and *Agrobacterium tumefaciens* (Allerdet-Servant et al. 1993).

PFGE has also provided researchers with an efficient means of identifying and isolating large plasmids which are harboured by both prokaryotes and eukaryotes. Tightly supercoiled covalently closed circular plasmid DNA (cccDNA) migrates independantly of pulse time in pulsed field gels (Hightower et al. 1987, Beverley, 1987, Mathew et al. 1988). This offered a simple diagnostic test for the identification of ccc plasmid DNA. Once nicked, or enzymatically relaxed, the plasmid DNA behaves similarly to digested chromosomal DNA fragments. Mathew et al. (1988) attempted to identify plasmid DNA with differing topologies by PFGE. They proposed that tightly supercoiled ccc plasmid, relaxed (nicked) plasmid and linearised plasmid DNA would have differing end-to-end lengths within the agarose gel matrix. It was thought that the relaxed plasmid DNA would have approximately half the end-to-end length of the linear species, although in practise this did not appear to hold for the smaller plasmids (2.7 - 9.0 kbp) used in the study. They suggested that other undetermined factors such as either tertiary structure or plasmid "impalement" on the agarose matrix, may play a role along with pulse time during PFGE. Similar studies with larger plasmids 4-16 kbp (Hightower et al. 1987) and 30-85 kbp plasmids (Beverley 1988) showed that the larger covalently closed molecules co-migrated independently of pulse time suggesting that plasmid size plays a minor role in the resolution of these molecules.

Another use of PFGE has been the investigation of gene amplification / gene copy number. OFAGE has been used for this purpose in drug resistant *Leishmania major* protozoa (Garvey and Santi 1986, Beverley 1988, Hightower et al. 1988). Circular extrachromosomal elements were identified which carried resistance to the drug methotrexate (MTX) (Garvey and Santi, 1986). Protozoa sensitive to MTX contained undetectable amounts of extrachromosomal DNA as observed on OFAGE, but once exposed to the drug this DNA was amplified to an estimated 20 copies per genome (Hightower et al. 1988)

In this chapter PFGE is used to provide evidence that pTF5 and pTF5-like plasmids have a wide geographic distribution in the *T. ferrooxidans* species, and may integrate either wholly or partially into the genomes of a number of these isolates. Interestingly although this technology has been used to detect plasmids in other studies (Hightower et al. 1987, Beverley, 1988) this is thought to be the first report of PFGE being employed to investigate this integrative feature of some bacterial plasmids.

4.2 Materials and Methods

4.2.1 Bacterial strains and plasmids

South African isolates of *T. ferrooxidans* and other bioleaching bacteria used in the distribution study were supplied by Gencor Ltd. - South Africa. The *T. ferrooxidans* isolates were FC1, FC6, Black Mountain Pb, Bateman Reno. Bioleaching bacteria from international culture collections included *T. ferrooxidans* type strain ATCC23270 (USA), ATCC33020 (Japan), *T. thiooxidans* type strain ATCC19377 (Libya), and *L. ferrooxidans* type strain DSM 2705 (Armenia).

4.2.2 Pulsed Field Gel Electrophoresis

PFGE was conducted using a Beckman Geneline Transverse Alternating Field system. Total genomic DNA was prepared by resuspending cell cultures in SET Buffer to OD_{600} = 1. Cells were lysed with 50 µg/ml Proteinase K (Merck) for 30 minutes at 37°C. Samples were embedded in a low temperature gelling agarose (Hispanagar) 1.1% final concentration. Samples were sliced into 20x6x2mm plugs and TE containing 1% (w/v) SDS and 50 µg/ml Proteinase K added. Samples were incubated for 48 h at 45° C and washed (3x15 min) in ES solution (Na-lauroyl sarcosine 10 g/l, 168 g/l EDTA, pH8) and transferred to TE. Plugs were incubated in Pefabloc proteinase inhibitor (Boehringer Mannheim) at 4° C overnight and then washed in TE. Samples for digestion with *Xba*I

were incubated in 1x restriction buffer H (Boehringer Mannheim) containing 1 μg/ml bovine serum albumin (BSA) for 1 h, followed by digestion in 1x H buffer containing 30 units of *Xba*I for 4 h. Digested samples were washed in TE and the plugs were stored at 4° C until used. Plugs were equilibrated in 1xPFGE running buffer prior to use. Agarose gels (1 % w/v) were made with 1x PFGE running buffer. Gels were electrophoresed at 150 mA for 14 h with an 8 s pulse time.

4.2.3 Southern hybridisation

Probe preparation : DNA fragments to be used as probes were purified from larger constructs by excision from low temperature gelling agarose. Gel slices were melted at 65° C for 15 min in 1x agarase buffer and digested with 1 unit agarase (Boehringer Mannheim) for 1 h at 45° C. The DNA fragment was resuspended in 15 μl sterile H_2O and denatured in a beaker of boiling water for 10 min. The sample was flash-cooled in an ice-ethanol bath, and 2 μl hexanucleotide primer mix, 2 μl dNTP labelling mix and 1 μl Klenow polymerase (all Boehringer Mannheim) added to the solution. The labelling reaction proceeded for 24 h at room temperature whereupon it was terminated with 2 μl 0.2 M EDTA (Saarchem) and precipitated with 2.5 μl 4M LiCl (Merck) and 75 μl absolute ethanol (Merck) and resuspended in 100 μl TE. The probe was placed in a boiling waterbath for 10 min and flash-cooled in ice-ethanol immediately prior to hybridisation.

Gel preparation : Pulsed field gels to be used for Southern hybridisation were placed in a 5% HCl (Merck) for 30 min to facilitate depurination. The gel was rinsed with deionised water and neutralised in a 0.4N NaOH (Saarchem). The gel was capillary blotted onto Hybond N^+ membrane (Amersham) according to Sambrook et al. (1989) in 0.4N NaOH transfer buffer for 24 h, whereupon excess buffer was removed and the transfer completed by dry blotting the gel for a further 24 h. The membrane was pre-hybridised for 4 h at 65° C in hybridisation fluid (5xSSC, 5% low-fat milk powder (Elite), 0.1% N-

lauroylsarcosine, 0.02% SDS) whereupon the denatured probe was added to the solution and hybridised to the membrane for 16 h at 65° C.

Detection of hybridisation signal : After hybridisation the membrane was washed in a high salt solution (2xSSC, 0.1% SDS) at room temperature for 2x15 min, and in a high stringency low salt wash (0.1xSSC, 0.1% SDS) for 2x15 min at 65° C. The membrane was immersed in wash buffer (11.6 g/l maleic acid, 8.7 g/l NaCl, 1ml Tween 20, pH7.5) at room temperature for 5 min and placed in blocking buffer (11.6 g/l maleic acid, 8.7 g/l NaCl, 10 g/l milk powder, pH7.5) for 30 min. Anti-DIG antibody (2 µl) was added to 20 ml blocking buffer and incubated for 30 min at room temperature. The membrane was washed in wash buffer for 30 min and then incubated in DEA buffer (Diethanolamine 10 ml/l, pH10) containing 50 µl CSPD for 15 min. The membrane was sealed in a plastic bag and incubated for 30 min at 37° C. Autoradiography with Xonix XR film (Kodak) was typically 1-2 h in duration.

4.2.4 Determination of pTF5 copy number

Dot blot analysis of *T. ferrooxidans* ATCC33020 total genomic DNA and cloned pTF5 DNA (pCD200) was used to determine the number of plasmids per genome of *T. ferrooxidans*. The amounts of genomic and plasmid DNA to be hybridised were standardised according to the relative sizes of the *T. ferrooxidans* chromosome and plasmid pTF5 assuming a copy number of one per chromosome. Two fold dilutions of each sample were bound to Hybond N^+ membrane (Amersham) and hybridised with pCD200 DIG-dUTP labelled DNA. Copy number was determined by comparing the hybridisation signal intensities of the genomic and plasmid dilution series.

4.3 Results

To confirm that plasmid pCD1 contained nucleic acid sequence of plasmid origin, PFGE and Southern hybridisation were used to probe total genome of *T. ferrooxidans* ATCC33020 as well as other South African strains of bioleaching bacteria. Probes homologous to each of the putative iron-sulphur protein-encoding genes as well as the origin of vegetative replication *oriV* were hybridised to both digested and undigested samples.

4.3.1 Identification of a region of pTF5 which is also located on the chromosome of *T. ferrooxidans* ATCC33020

Undigested total DNA from *T. ferrooxidans* ATCC33020 cells as well as fragments of *Xba*I digested total DNA were separated using PFGE. The gels were blotted and probed with DIG-labelled fragments of pTF5 (Figure 4.2). Probes C, D, E, F and G gave identical hybridisation signals to undigested plasmid (lane 3) as well as to the 20 kbp linearised plasmid and a 60 kbp chromosomal fragment (lane 4). Probes B and H hybridised to the covalently closed undigested sample (lanes 1, 5) and to the linearised plasmid band only (lanes 2, 6). No hybridisation to the chromosomal band at 60 kbp was observed with probes B and H. This indicated that the region of pTF5 which covers the iron-sulphur protein encoding genes has become incorporated into the chromosome of strain ATCC33020. The signal obtained from the DNA in the well of the undigested samples (lanes 1, 3, 5) could be due to entangled plasmid DNA within the chromosomal DNA and was absent from *Xba*I digested DNA (lanes 2, 4, 6). The experiment shown in Figure 4.2 also enabled the approximate position of the junction between plasmid specific DNA, and the DNA present on both plasmid and chromosome to be determined. The junction site downstream of the *fdxA* gene could be localised to a region within probe C. The most likely site is on a 215 bp *Eco*RI-*Hin*dIII fragment immediately downstream of the *fdxA* gene. Similar analysis of the region downstream of the *redA* gene on probe G identified a 531 bp *Sph*I-*Bam*HI fragment likely to contain the junction site. No transposon termini,

inverted repeats, direct repeats or other features frequently identified with transposon activity were obvious within either junction region.

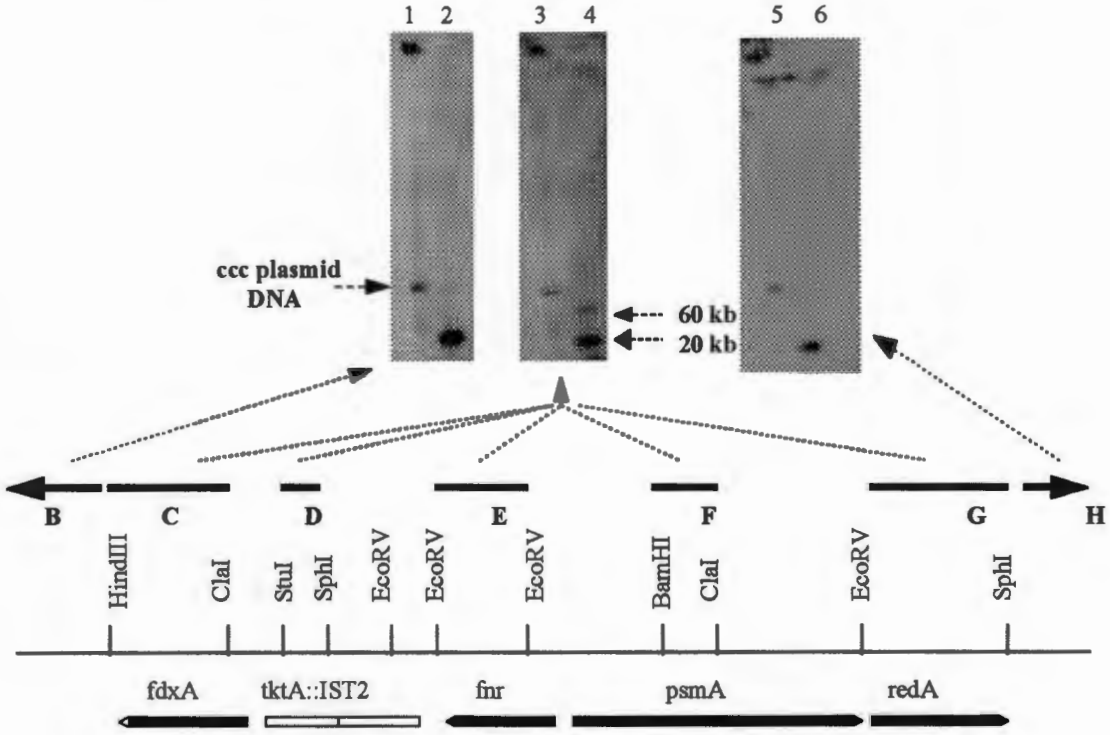


Figure 4.2 : Southern hybridisation of *T. ferrooxidans* ATCC33020 total genomic DNA run under PFGE conditions. Lanes 1, 3, 5 contain undigested genomic DNA, lanes 2, 4, 6, genomic DNA digested with *Xba*I. The location of probes B to H is indicated in bold above the restriction map. The centre panel represents hybridisation probes C to G since each probe gave an similar result. λ concatamers were used as molecular weight .

4.3.2 Determination of plasmid pTF5 copy number

The copy number of pTF5 was estimated by hybridisation of a plasmid probe pCD200, to known amounts of purified plasmid and total DNA prepared from *T. ferrooxidans* in a dotblot experiment. Based on the relative intensities of the hybridisation signals, pTF5 was estimated to be present at between 2 and 4 copies per chromosome (Figure 4.3). This estimate was supported by the observation that the intensity of the plasmid signal to

hybridisation probes C to G (Fig 4.2) was approximately 3 to 4 times greater than chromosomal signal. It should be noted however that the signal intensities may have been a product of divergence between the two target sequences.

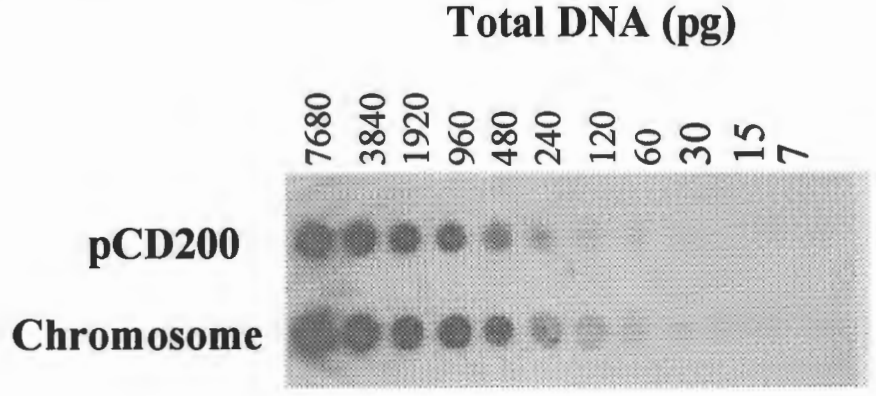


Figure 4.3 : Estimation of pTF5 copy number by dot blot hybridisation. Total *T. ferrooxidans* total genomic DNA (row 2) was standardised relative to pCD200 (row 1) assuming a copy number of 1. Serial 2-fold dilutions of both samples are shown with the total DNA shown in pg.

4.3.3 Distribution analysis of pTF5-like plasmids

The presence of pTF5-like plasmids among 5 *T. ferrooxidans* strains, including 4 strains which were isolated in South Africa was examined. Undigested and *Xba*I digested total DNA was subjected to PFGE and a Southern hybridisation experiment was carried out using probe G (Fig 4.2). Positive hybridisation signals to plasmids present in *T. ferrooxidans* strains ATCC23270 (USA), ATCC33020 (Japan) and FC1 (South Africa) were obtained (Figure 4.4A, lanes 3, 4, 6). The hybridisation signal to DNA remaining in the well of each of these strains was due to non-mobile undigested chromosomal DNA and possibly some entrapped plasmid DNA. In the case of *T. ferrooxidans* Black Mountain Pb and Bateman Reno (Fig 4.4A, lanes 8, 9), a hybridisation signal for DNA remaining in the wells was obtained but there was no evidence of plasmid DNA. The *Xba*I digested DNA samples from all strains confirmed these results. Two bands were obtained for *T. ferrooxidans* strains ATCC23270,

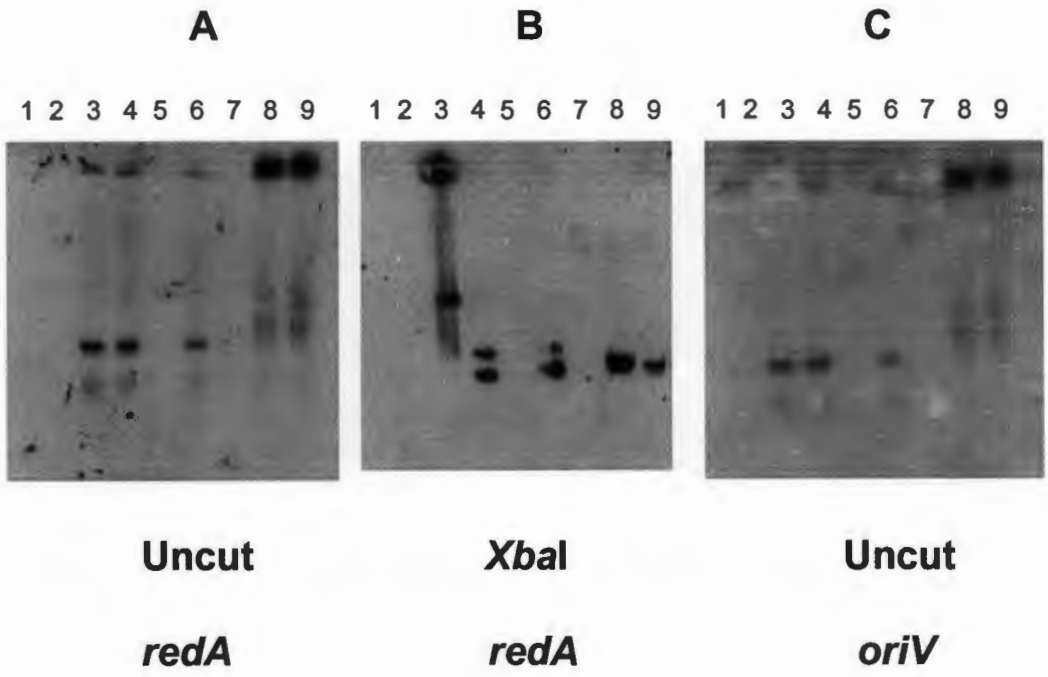


Figure 4.4 : Pulsed field gels after Southern hybridisation of a number of bacteria present in bioleaching environments. Panels A and C contain undigested total genomic DNA, panel B) contains *Xba*I digested samples. Samples in all cases contain lane 1 *L. ferrooxidans* DSM2705, lane 2 *T. thiooxidans* ATCC19377, lane 3 *T. ferrooxidans* ATCC23270, lane 4 *T. ferrooxidans* ATCC33020, lane 5 molecular weight standards, lane 6 *T. ferrooxidans* FC1, lane 7 *T. ferrooxidans* FC6, lane 8 *T. ferrooxidans* "Black Mountain Pb", lane 9 *T. ferrooxidans* "Bateman Reno". Panels A and B were probed with probe G, panel C was probed with pCD300 (the 2.2 kbp *oriV*-containing *Sac*I fragment).

ATCC33020 and FC1 (Fig 4.4B, lanes 3, 4, 6), one signal due to the copy of the *redA* gene on the plasmid and the other due to the chromosomal copy. In the case of strain ATCC23270 (lane 3), the chromosomal DNA migrated only a short distance from the well as it was on an *Xba*I fragment of greater than 500 kbp and fragments of this size do not penetrate the gel under the running condition used to resolve smaller fragments. From the *Xba*I digested samples it may be seen that the the plasmid in strain ATCC23270 is clearly of a larger size than the plasmids in strains ATCC33020 and FC1 (Fig 4.4B, lanes

3, 4, 6). *T. ferrooxidans* strains Black Mountain Pb and Bateman Reno gave only a single hybridisation signal at approximately 30 kbp (Fig 4.4B, lanes 8, 9), slightly larger than the digested plasmid in strains ATCC33020 and FC1 and smaller than the signal from the chromosomal copy in those strains (Fig 4.4B, lanes 4, 6). No signal was obtained from either *T. ferrooxidans* FC6, *L. ferrooxidans* type strain DSM2705 or the *T. thiooxidans* type strain ATCC19377 (Fig 4.4A and B, lanes 1, 2).

The same filter shown in Figure 4.4A was stripped and probed with labelled DNA from the plasmid origin of replication (pCD300). As predicted a hybridisation signal to a single plasmid band in strains ATCC23270, ATCC33020 and FC1 was obtained (Fig 4.4C, lanes 3, 4, 6). Unexpectedly hybridisation signals to strains Black Mountain Pb and Bateman Reno were obtained to chromosomal DNA in the wells (Fig 4.4C, lanes 8, 9). This implies that a plasmid with homology to pTF5 had become integrated into the chromosome of these two strains.

4.4 Discussion

Plasmid pTF5 is a low copy number plasmid which belongs to the pTFI91-like family of *T. ferrooxidans.* Two features of this plasmid family that are especially noteworthy are its wide geographic distribution among *T. ferrooxidans* strains, and the discovery that it has a segment of DNA which is also present on the chromosome of strain ATCC33020. In this study the geographical distribution of the plasmid was extended from *T. ferrooxidans* strains found in the northern hemisphere to strains found in South Africa. Even though it is widely distributed, pTF5 is not completely ubiquitous as *T. ferrooxidans* strain FC6 and several northern hemisphere strains (Valenti et al. 1989, Dominy et al. submitted for publication) do not possess the plasmid, indicating that it is not essential to the growth of the organism. Furthermore the *T. ferrooxidans* type strain ATCC23270 possesses a much larger plasmid than the 19.8 kbp pTF5 which also has homology to the *redA* gene probe. Strains "Black Mountain Pb" and "Bateman Reno" did not contain detectable ccc plasmid DNA on pulsed field gels when probed with either the *redA* or *oriV* containing sequence.

Since the DNA from which the *redA* and *oriV* probes were made are located on opposite sides of pTF5 it is likely that these two isolates harbour an entire pTF5-like plasmid integrated into their chromosome (Fig 4.4b, lanes 8, 9). These two strains were different from the other *T. ferrooxidans* strains examined in that they possessed the iron-sulphur protein-encoding ORFs solely on their chromosomes. Plasmid pTF5 was not detected in the type strains of *T. thiooxidans* or *L. ferrooxidans*. However a greater number of isolates of these bacteria will need to be screened before an appropriate comment can be made on the prevalence of pTF5-like plasmids in these species.

The transketolase enzyme is an essential component of the carbon dioxide fixing pathways of both photosynthetic and autotrophic organisms (Tabitha, 1988). This implies that a second intact copy of the transketolase ORF is present elswhere on the chromosome of *T. ferrooxidans* ATCC33020. However probes with homology to the partial transketolase ORF of pTF5 hybridised to the plasmid band and the 60 kbp chromosomal band. This suggests that either an active chromosomal transketolase gene has diverged away from the pTF5-associated copy and was not detected under the stringent hybridisation conditions used, or the active transketolase gene was present on the same chromosomal fragment as the partial transketolase ORF.

A "long shot" attempt was made to amplify the region flanking the partial transketolase ORF (Fig 4.5). This was done to investigate whether the chromsomal organization of the transketolase - *fnr* region differed from that of pTF5. Opposing primers from the two genes were synthesized, and using these in a thermocycling reaction should result in a 600bp amplified product from the pTF5 fragment (Fig 4.5a) which contains the truncated IST*2* sequence. If the *fnr*-like sequence was not present near the chromosomal copy of the transketolase gene, two amplification products (Fig 4.5 a, b) might have indicated that a larger, functionally-active chromosomal transketolase was present in this region. Only a single 600 bp product corresponding to the plasmid-amplified DNA was observed (data not shown). A number of unknown parameters in this experiment does not allow a clear inference to be made as there are several possibilities for the lack of a second

chromosomally amplified product. Firstly the tkt and fnr genes and hence primer binding sites might have been too far apart on the *T. ferrooxidans* chromosome (Fig 4.5c) and thus no amplification product would be obtained from the thermo-cycling reaction. Secondly the *tkt* and *fnr* genes may be arranged differently on the *T. ferrooxidans* chromosme such that primers used in the experiment, may not be opposing one another over the intact transketolase gene (Fig 4.5 d). To obtain an unequivocal answer, the 60 kbp fragment would need to be cloned from the chromosome of strain ATCC33020, and further characterised.

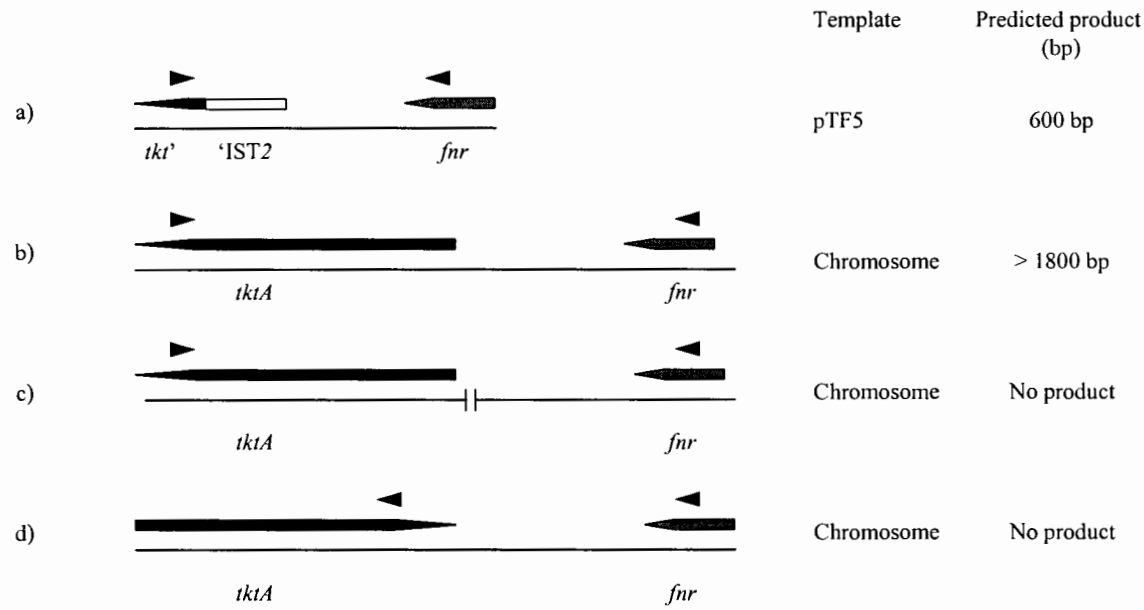


Figure 4.5 Possible organization of the 60 kbp *tkt'-fnr*-containing chromosomal band of *T. ferrooxidans* ATCC33020. a) The pTF5 gene sequence amplified the predicted 600 bp fragment. b) Assuming that the intact transketolase had been disrupted by an IST2-like element, a larger fragment would have been amplified. No product would have been produced however, if c) the primers were situated more than 4 kbp apart, or d) chromosomal rearrangements had inverted one of the primers.

Chapter 5

General Conclusions

This study has focussed on the 19790 bp plasmid pTF5 (Figure 5.2a), from *Thiobacillus ferrooxidans* ATCC33020. Evidence has been presented, showing that pTF5 is a new member of the pTFI91-like plasmid family. In agreement with previous studies of plasmids of this family, pTF5 could not replicate in heterotrophic bacteria which are commonly used for genetic studies (Rawlings et al. 1983, Holmes et al. 1984, Valenti et al. 1989, Chakravarty et al. 1995). Furthermore the plasmid was not detected in the strains of *T. thiooxidans* and *L. ferrooxidans* tested (this study). The main aspects discussed in this concluding chapter are a) the extension of the plasmid family described by Chakravarty et al. (1995) by the identification of a pTF5-like sub-group within the pTFI91 plasmid family and b) the possible origin of pTF5 and the evolutionary implications of its GC mol% and codon usage patterns.

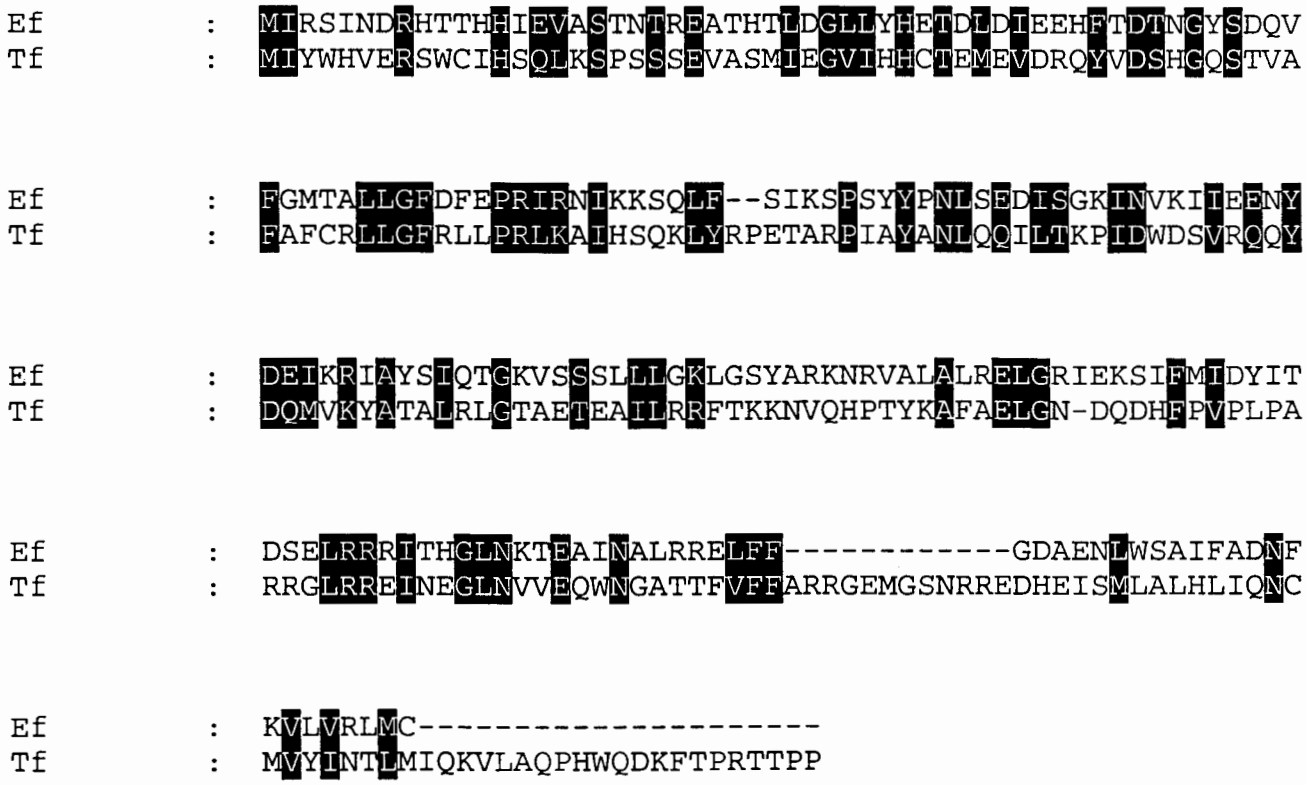
Expression of the exclusively plasmid-associated genes : It is interesting to note that the portion of pTF5 not present on the chromosome of strain ATCC33020 contained a number of open reading frames, whose products are involved in the rearrangement and possible transposition of DNA. These ORFs included at least two insertion elements (IS*3091* and IST*2*- like), two invertase/recombinase-like ORFs (*inv* and *min*), one transposase-like ORF (*tnpA*) and a large recombinase-like ORF (ORF90), and may explain the occurrence of pTF5-like sequence on both plasmid and chromosomal DNA. Whether any of these ORFs are expressed in *T. ferrooxidans* is unknown. Antibodies against the purified products of the ORFs and Western blot experiments on *T. ferrooxidans* protein extracts would be required in order to answer this question. Certainly the PFGE distribution analysis has indicated that the genes (both the replication region and *redA* gene) exist in both plasmid and chromosomal form in a number of the South African strains tested. Thus the iron-sulphur protein-encoding ORFs have been

widely spread amongst *T. ferrooxidans* isolates, but the whether this was due to the pTF5-encoded ORFs or other trans-acting factors is unknown. It is possible that these transfer-related ORFs are present on most members of the pTFI91 family as based on restriction map comparisons, they would be located within the "lesser" well conserved region defined by Chakravarty et al. (1995).

Due to the absence of the *redA* gene in some *T. ferrooxidans* strains ie FC6 (Fig 4.4, lane 6) the iron-sulphur protein-encoding genes may have been recruited from a bacterial species other than *T. ferrooxidans*. Consequently the genes may not have been acquired by plasmid pTF5 from the chromosome of *T. ferrooxidans* strain ATCC33020, but instead may have been deposited there. The various pTFI91-like plasmids may well contain different "insert" DNA fragments and may play very different physiological roles in their host strains. It is interesting to note that in the case of pTF5 only 6 kbp of "insert" DNA has been acquired with approximately 14 kbp of pTF5 being associated only with the plasmid (Chapter 4). However smaller plasmids than this, specifically pTFI91 (9.8 kbp) have been assigned to this family of plasmids, indicating that not all the family members can contain all of the exclusively plasmid-associated DNA of pTF5.

It has been demonstrated that pTF5, or at least part thereof, is capable of integrating into the chromosome of its host strain. Evidence from PFGE indicating that pTF5-like *oriV* sequences have also become inserted into the genomes of their respective hosts was obtained (Fig 4.4, lanes 8, 9). This study has identified the previously described pTF35 and pTFO plasmids as members of the pTFI91 family. These plasmids together with pTF5 have a wide geographic distribution and presumably the electron transfer protein encoding genes present on this pTF5 sub-family give the host organisms a selective advantage in their habitats.

A disadvantage with the approach involving plasmid pTF5 is that it lacks a selectable marker. Consequently attempts to determine a phenotype for pTF5 were limited to studies in *E. coli*. A predicted domain of ORF90 had homology to an erythromycin


Ef : MIRSINDRHTTHHIEVASTNTREATHTLDGLLYHETDLDIEEHFTDTNGYSDQV
Tf : MIYWHVERSWCIHSQLKSPSSSEVASMIEGVIHHCTEMEVDRQYVDSHGQSTVA
Ef : FGMTALLGFDFEPRIRNIKKSQLF--SIKSPSYYPNLSEDISGKINVKIIEENY
Tf : FAFCRLLGFRLLPRLKAIHSQKLYRPETARPIAYANLQQILTKPIDWDSVRQQY
Ef : DEIKRIAYSIQTGKVSSSLLLGKLGSYARKNRVALALRELGRIEKSIFMIDYIT
Tf : DQMVKYATALRLGTAETEAILRRFTKKNVQHPTYKAFAELGN-DQDHFPVPLPA
Ef : DSELRRRITHGLNKTEAINALRRELFF----------GDAENLWSAIFADNF
Tf : RRGLRREINEGLNVVEQWNGATTFVFFARRGEMGSNRREDHEISMLALHLIQNC
Ef : KVLVRLMC--------------------
Tf : MVYINTLMIQKVLAQPHWQDKFTPRTTPP

resistance marker from *Enterococcus faecalis* (Fig 5.1), but although a 90 kDa polypeptide corresponding to this ORF was expressed in *E. coli,* no erythromycin resistance was observed. Speculative attempts to define a function for the putative electron transfer mechanism by testing for growth on the potentially toxic heavy metals, uranium and silver, did not cause an increased tolerance to these metals in *E. coli.* This does not imply however that they are not responsible for this, or a similar function in *T. ferrooxidans.*

Figure 5.1 : (opposite) Alignment of an ORF90 protein domain with 46 % aa sequence similarity with that of the erythromycin resistance protein from *Enterococcus faecalis* (Shaw and Clewell, 1985).

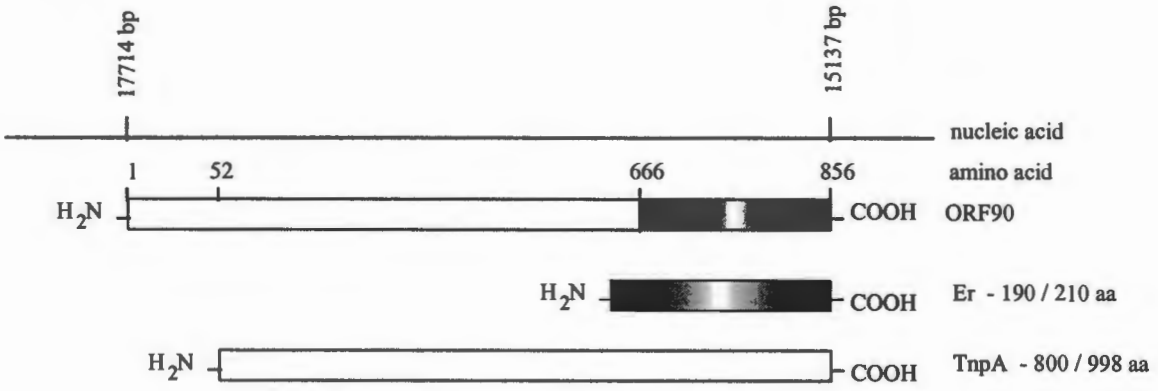
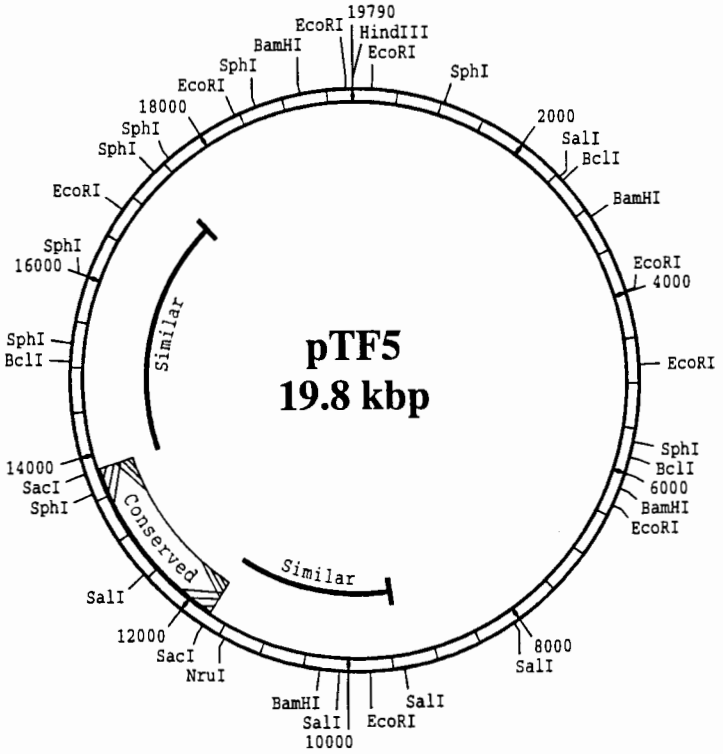


Figure 5.2 : A graphical representation of the ORF90 protein showing the location of protein domains with varying homologies to transposases and antibiotic resistance proteins. The 190 aa domain of ORF90 which has homology to the 210 aa *E. faecalis* erythromycin resistance protein (Fig 5.1) is shown below ORF90. Approximately 800 aa of ORF90 shared 57% aa sequence similarity to the 998 aa IS*1546* transposase from *Enterococcus faecium* (Arthur et al. 1993). This region extended over almost the entire ORF including the region which had homology to the erythromycin resistance protein.

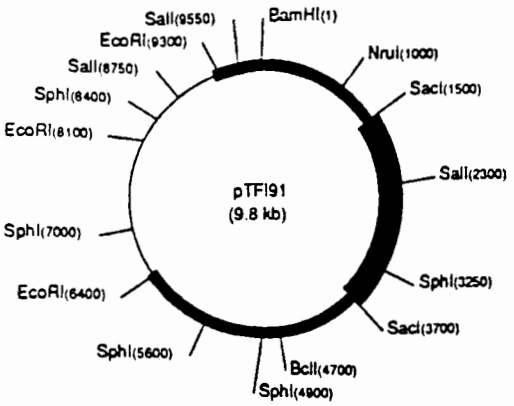
The pTFI91 plasmid family : As pTF5 has been completely sequenced it is possible to compare its physical map to those published in other studies - Figure 5.3 a,b (Rawlings et al. 1983, Valenti et al. 1990, Chakravarty et al. 1995). This has also helped to define the regions of pTF5 identified by Chakravarty et al. (1995) as being less homologous to one another among the members of the family. Inspection of the pTFI91 restriction maps suggest that the pTFI91, pTF35 and pDSM583 are unlikely to have acquired the same iron-sulphur protein encoding genes as pTF5. Based on the hybridisation of the near identical pTF5, pTF35, pTFO as well as the pTF5-like plasmids detected in this study to the *redA* probe, a sub-family of plasmids can be defined within the pTFI91 family. Included in this sub-group of *redA*- hybridising plasmids is the 200 - 250 kbp plasmid in the *T. ferrooxidans* type strain, ATCC23270. It is clear that this large plasmid must contain a lot more DNA than the other members of the sub-family. Whether this extra sequence is from chromosomally-duplicated sources flanking the *redA* gene or from sequence encoding plasmid-related functions (such as a conjugation system) has not been determined.

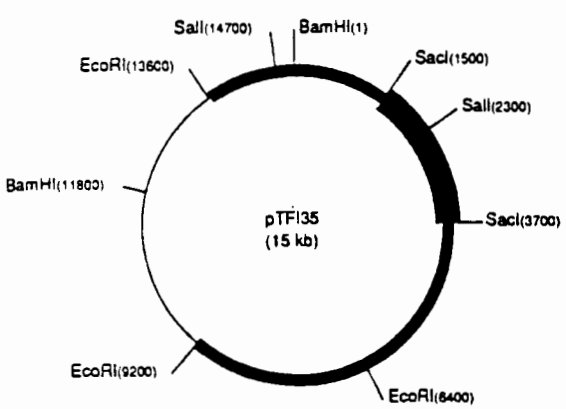
An attempt to decipher the evolutionary history of this plasmid family within *T. ferrooxidans* is complicated by the limited members of the family described to date. It appears as though the homologous *oriV* sequences encode an intrinsic replication "engine" which is shared by all members of the family. Inspection of the physical maps of the pTFI91-like family show the characteristic 2.2 kbp *Sac*I fragment originally used to define the family (Chakravarty et al. 1995) as well as a lesser conserved flanking region. The spacing of the restriction enzyme sites within all members of the family appears to be conserved from the *Bam*HI site (position 1) on pTFI91 to the second *Sac*I site (position 3700) (Fig 5.3a,b). By analogy to the pTF5 plasmid this region would be predicted to cover the *repA* gene - *oriV* and most of the *tnpA*-like ORF present on pTF5. It cannot be determined whether the sequence covering ORF24, ORF8 is present on the other plasmids as their physical maps are not detailed enough. Consequently one cannot speculate on exactly which ORFs comprise the minimal replicon of this plasmid family.

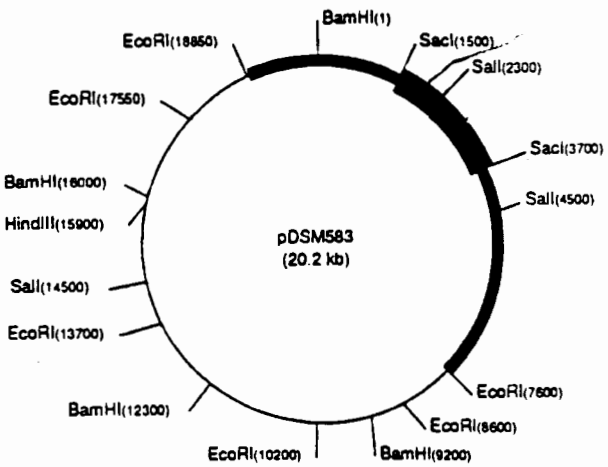
a)


pTF5
19.8 kbp
19790
EcoRI
HindIII
EcoRI
BamHI
SphI
EcoRI
18000
SphI
SphI
EcoRI
SphI
16000
SphI
BclI
14000
SacI
SphI
SalI
12000
SacI
NruI
BamHI
SalI
EcoRI
10000
SalI
8000
SalI
SphI
BclI
6000
BamHI
EcoRI
EcoRI
4000
BamHI
SalI
BclI
2000
SphI
Similar
Conserved
Similar

b)


pTFI91
(9.8 kb)
SalI(9550)
BamHI(1)
EcoRI(9300)
NruI(1000)
SalI(8750)
SacI(1500)
SphI(8400)
EcoRI(8100)
SalI(2300)
SphI(7000)
SphI(3250)
EcoRI(6400)
SacI(3700)
SphI(5600)
BclI(4700)
SphI(4900)


pTFI35
(15 kb)
SalI(14700)
BamHI(1)
EcoRI(13600)
SacI(1500)
SalI(2300)
BamHI(11800)
SacI(3700)
EcoRI(9200)
EcoRI(6400)

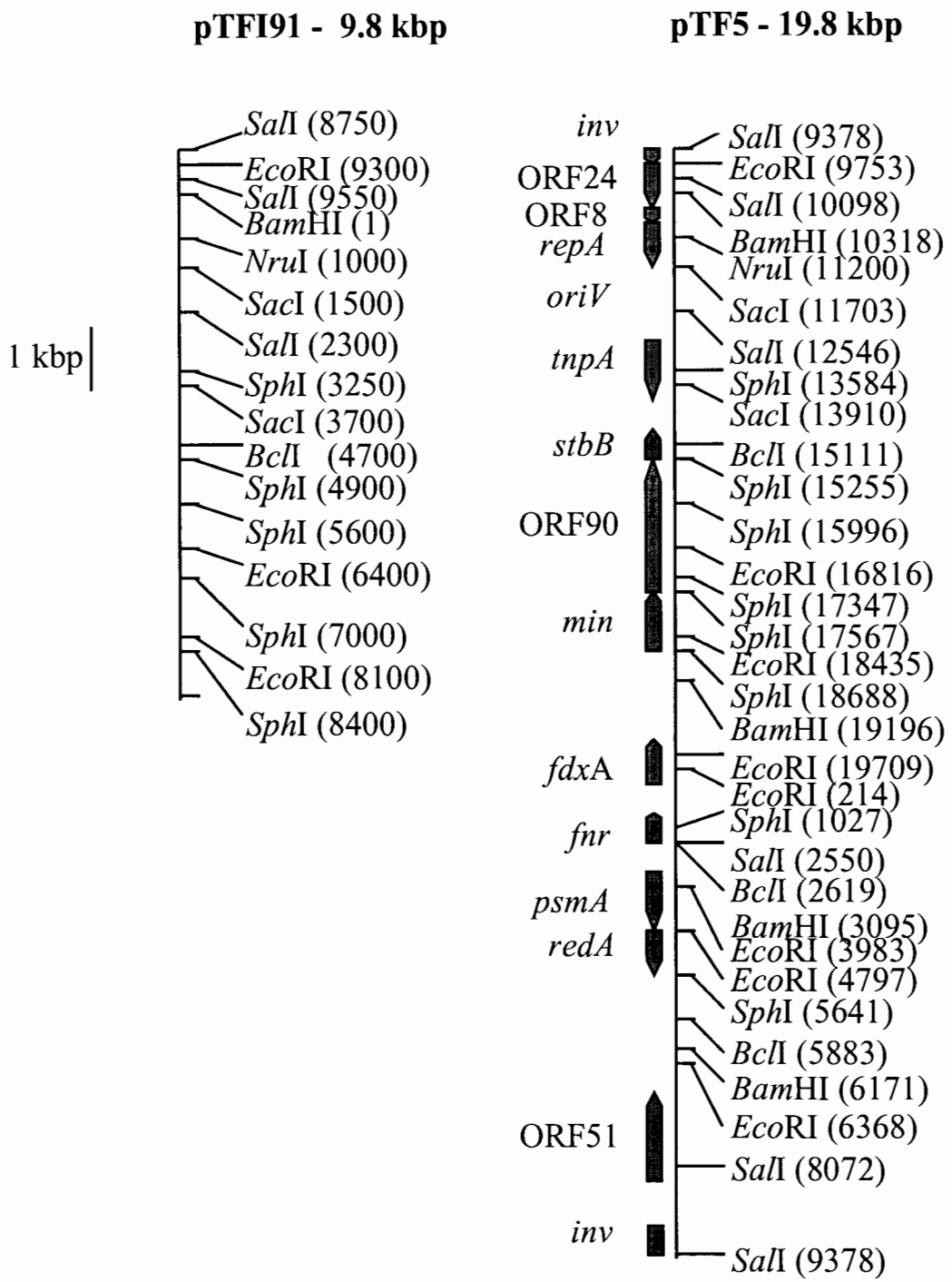

pDSM583
(20.2 kb)
BamHI(1)
EcoRI(18850)
SacI(1500)
SalI(2300)
EcoRI(17550)
SacI(3700)
BamHI(16000)
SalI(4500)
HindIII(15900)
SalI(14500)
EcoRI(13700)
EcoRI(7600)
BamHI(12300)
EcoRI(8600)
EcoRI(10200)
BamHI(9200)

As will be discussed later pTF5 appears more homologous to pTFI91, than to the other described members, pDSM583 and pTFI35, within the pTFI91 plasmid family.

Figure 5.3a (opposite) : Physical restriction map of plasmid pTF5. Sites shown are for the common restriction endonucleases used in the study. The highly conserved 2.2 kbp *Sac*I fragment as defined by Chakravarty et al. (1995) has been shown as a block, whilst regions with restriction enzyme site similarity to pTFI91 have been indicated with a solid black line. 5.3b: Physical maps of members of the pTFI91-like plasmid family of *T. ferrooxidans* (Chakravarty et al. 1995). The highly conserved 2.2 kbp *Sac*I fragment as defined by Chakravarty et al. (1995) has been shown as a thick bold line. Lesser conserved regions are represented by a thinner bold line. For the purposes of comparison, plasmids pTF5, pTF35 (Rawlings 1985) and pTFO (Valenti et al. 1990) have been considered to be identical, as have pTFI91 (Chakravarty et al. 1995) and pTfA-4 (Sanchez et al. 1986).

Plasmid pTFI91 appears to be particularly closely related to pTF5 over most of its length. A comparison of the linearised restriction enzyme maps of pTFI91 and pTF5 (Figure 5.4) clearly illustrates this close relationship. With the exception of an SphI site (position 17567) on pTF5 there did not appear to be any major discrepancies between the two physical maps (Fig 5.3 a,b and Fig 5.4.). It is thus possible that pTFI91 sequence is contained entirely within pTF5, a region which corresponds to the invertase ORF (*inv*) through the plasmids replicon to the second invertase ORF (*min*).

The evolutionary origin of pTF5 : This chapter concludes with an attempt to establish whether pTF5 has had a long association with *T. ferrooxidans* or whether it was likely to have been recruited relatively recently in its evolutionary history. Gene transfer in natural bacterial populations is facilitated mostly by promiscuous plasmids.



5.4 : A linear restriction enzyme map of pTF5 with open reading frames and appropriate enzyme sites is compared to linearised maps of pTFI91 (Chakravarty et al. 1995).

Genes which confer a selective advantage on the host cell are acquired by plasmids in a process (usually) facilitated by insertion elements (Thomas and Helinski, 1989). Transfer to another bacterium can proceed via either conjugative mechanisms or by natural transformation. Once in the new host the genes are often integrated into the chromosome whereupon they may be expressed by the host organism. This is particularly important in the case of narrow host-range plasmids which may be acquired by bacterial cells. If the plasmid is unable to replicate within the new host, the integration of its genes into the chromosome provides an ideal mechanism by which they may be expressed in the new bacterium. These integrative (or episomal) plasmids have played a significant role in directing bacterial evolution. By allowing new combinations of proteins to be expressed together they allow new biochemical pathways to be developed. The best known example of this class of an episome is the F-plasmid of *E. coli* (Meynell, 1972). It is a conjugative plasmid which once transferred to a recipient cell, may integrate into its chromosome by homologous recombination between insertion sequence elements (IS2 and IS3) present on both plasmid and chromosome (Reimmann et al. 1989). Gene "pick-up" mechanisms by the F plasmid and other integrative plasmids have been well described and may proceed by either an excision mechanism or by a duplication event (Figure 5.5). The excision mechanism results in a truncated or deleted chromosomal segment, whilst the duplication method leaves both copies of the gene intact (Meynell, 1972). The identification of insertion elements contained on plasmid sequences explained the ability of plasmids to acquire DNA from two separate sources (Shapiro et al. 1977, Galas and Chandler, 1989).

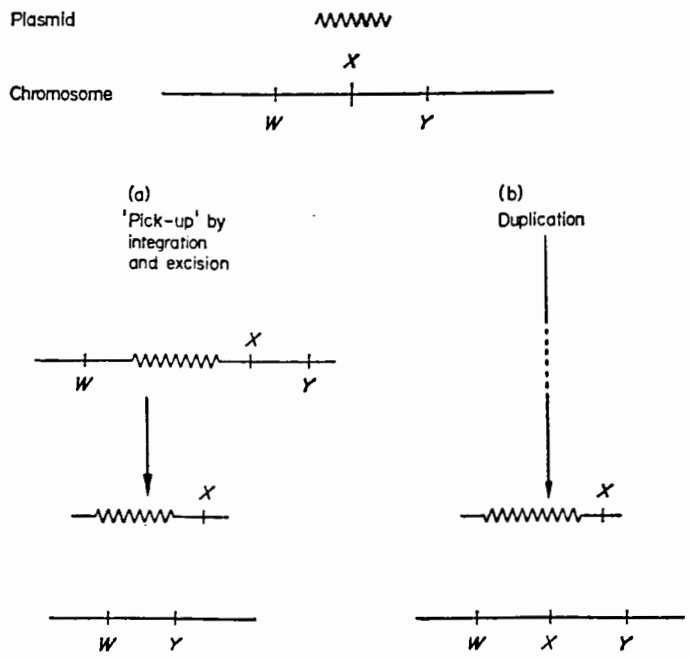


Figure 5.5 : A proposed mechanism by which plasmids can acquire chromosomal DNA fragments. a) An excision mechanism which removes a gene (X) from its host chromosome, and b) a duplication event which results in both a chromosomal copy of X as well as a plasmid-encoded copy (Meynell, 1972).

The transpositional scar between the *fdxA* gene and the *fnr*-like ORF may indicate that the fdxA gene and the rest of the ORFs have been spliced together in the formation of pTF5. Probes designed to both of these regions hybridised to a 60 kbp chromosomal fragment on pulsed field gel electrophoresis, but there was no indication as to whether the two genes have the same organization on the chromosome as they do on pTF5. The chromosome of *T. ferrooxidans* is thought to require an intact transketolase gene, which since only one chromosomal band hybridised to the tkt probe, would imply that either the chromosomal *tkt* gene is intact and functional, or that there is a second active tkt gene on the 60 kbp fragment which is responsible for the fixation of carbon dioxide. The most parsimonious explanation is that the chromosomal organisation does not include the transpositional scar ie it is an artefact formed by an integration event during the acquisition of chromosomal

DNA (not necessarily from *T. ferrooxidans*) by plasmid pTF5. This is highly speculative and would require several carefully designed Southern hybridisation experiments in order to establish this. Integrative plasmids can be used as crude mechanisms to regulate gene copy number. Gene amplification by the chromosomal excision of appropriate genes on an extrachromosomal element has been used as a means of expressing drug resistance markers (Garvey and Santi, 1986) as discussed in Chapter 4.

Codon usage patterns and GC mol% of extrachromosomal elements often differ from that of the host organism especially if the element has not had a long evolutionary association with the host (Rawlings et al. 1991). The codon preference graphs of pTF5 (Fig 2.10a, Fig 3.7) were not as close to the *T. ferrooxidans* chromosomal codon usage pattern as might have been expected if pTF5 had had a long evolutionary association with the bacterium. This raises the possibility that the plasmid may have originally been associated with another host. *T. ferrooxidans* species have a reported overall GC mol % of 56 - 59 % (Holt et al. 1994), based on the kinetic dissociation of DNA. Plasmid pTF5 has an overall GC mol% of 55% which is close to the reported range. The region covering the iron-sulphur protein-encoding ORFs discussed in Chapter 3 has a 53 % GC mol% whilst the nucleic acid sequence found exclusively on pTF5 in strain ATCC33020 (Chapter 2) has a 57 % GC mol %. Organisms have been known to alter the GC mol % content of the non-coding regions of plasmids in order to correct an overall discrepancy between that of the plasmid and the host.

Consequently codon usage tables of each of the ORFs on pTF5 were compared to previously published *T. ferrooxidans* chromosomal genes. The codon usage tables were analysed by the CORRESPOND program in the UWGCG analysis package (Devereaux et al. 1984) and are shown in Table 5.1. The frequency (D) with which the pTF5 genome used a particular codon triplet was compared relative to previously reported chromosomally derived data. The closer the D^2 value was to 1, the more similar the patterns of codon usage, D^2 values greater than 4 are considered to be noticeably

different. It should be noted however that very short coding regions would have fewer amino acids than longer ones and may not be as statistically accurate. Table 5.1 shows the D^2 values of the pTF5 ORFs relative to the codon usage tables of the rubisco large subunit gene (*rbcL*, Kusano et al. 1992), the glutamine synthetase gene (*glnA*, Barros et al. 1984) and the α and β subunit genes of the F_1F_0 ATPase (Brown et al. 1994) from *T. ferrooxidans*. A composite codon usage pattern consisting of (Rawlings et al. 1991) (the same used to generate Fig 2.10a, and Fig 3.7), was also included.

Table 5.1 : Correspond table comparing pTF5 codon frequencies

pTF5 ORFs	*rbcL*	*glnA*	*atp*-α	*atp*-β	Tfchrom
fdxA	**3.45**	**4.06**	**5.84**	**4.62**	**3.31**
fnr	**6.11**	**6.76**	**4.18**	**5.52**	**4.42**
psmA	**3.31**	**3.71**	**1.45**	**3.12**	**1.97**
redA	**3.41**	**4.24**	**1.94**	**3.11**	**2.24**
ORF51	**1.69**	**1.16**	**3.21**	**3.14**	**1.25**
pin	**2.38**	**1.88**	**3.99**	**4.13**	**1.93**
ORF24	**4.98**	**5.19**	**5.04**	**2.54**	**3.1**
ORF8	**4.47**	**6.43**	**7.39**	**5.96**	**4.98**
repA	**4.87**	**5.78**	**3.46**	**4.08**	**3.61**
tnpA	**1.27**	**2.19**	**3.88**	**3.74**	**1.78**
FeS total	**2.59**	**3.11**	**1.83**	**2.68**	**1.53**
Plasmid-only total	**1.47**	**1.79**	**2.57**	**2.12**	**0.9**

The iron-sulphur protein encoding genes : It was of interest to determine if the codon usage patterns gave any indication as to whether the iron-sulphur protein-encoding genes may have had a different origin to the rest of plasmid pTF5. Comparing the values of the iron-sulphur protein-encoding genes with those from the chromosomally based standards, the smaller genes *fdxA* and *fnr* have markedly higher values, ranging from 3.31 to 6.76, than the larger genes (*psmA, redA*) which ranged from 1.94 to 4.24.

In order to gain a better statistical representation of this region, all four open reading frames were combined to produce a composite table (FeS). This removed any statistical bias of the smaller ORFs and brought the D^2 values down to the range 1.53 to 3.11. This indicates that the iron-sulphur protein encoding ORFs displayed slightly differing codon usage patterns to the chromosomal ORFs, but the D^2 values were not high enough to clearly indicate that the origin of these genes was from a bacterium other than *T. ferrooxidans*.

The plasmid-associated genes : Comparing the plasmid-associated ORFs with the same chromosomal genes as a standard, the smaller ORFs (ORF24, ORF8) displayed similarly biased D^2 values (2.54 to 7.39). As expected the larger ORFs tended to produce lower D^2 values in the range 1.27 to 5.78. Interestingly the 1148bp *repA* gene had noticeably higher D^2 values (ranging from 3.46 to 5.78) compared with other large ORFs, the significance of this is uncertain. The composite table of all the plasmid-associated genes and ORFs displayed lower D^2 values, suggesting that the plasmid has a codon usage pattern even more similar to the *T. ferrooxidans* chromosome than the iron-sulphur protein-encoding ORFs.

This indicates that the pTF5 sequence taken as a whole displays relatively normal codon usage patterns as well as similar GC mol% values to *T. ferrooxidans* chromosomal DNA. Therefore, even though the codon usage patterns of individual plasmid genes differs from what would be expected for chromosomally encoded genes, the more statistically accurate average codon usage pattern and GC ratio for the entire plasmid, may well be interpreted to reflect a long-standing association between pTF5 and *Thiobacillus ferrooxidans*.

Appendix A

Complete nucleotide sequence of pTF5, with deduced amino acid translation of the ORFs. Predicted ORF start and stop sites are highlighted in bold and restriction sites have been underlined.

```
          HindIII                  ApaI
1         AAGCTTACGTAGAAACGTGCTGGGCCCAGGAACTCCACCCGGTGCGGCAGATCGGGCGGT
61        ATCCATGCAGGCCGTGCCGACTCAGAATCCAGGCGCGGCGCTGCCGTCTTCCAGGCAATA
121       TCGCAAGTTTCCTTCCAGGAGCACGATGCGCCCCCAGACCCCGGCTTTGGTGCGATGCTC
                                                 EcoRI
181       ATGGAGCAGGCCGGCGGGGGCGTCGTCGTCGGTGAATTCCGGTGTGCGCTGGTATTCCTT
           fdxA stop
241       CAAGGGCACGGGCGGCTCGGGCAGGGGCAATTCCTCCTCCTCTTCCCCGGTATCCAGATA
            *  P  C  P  R  S  P  C  P  C  N  R  R  R  K  G  P  I  W  I
301       TGGGCGTTTATCCCGCACATGCGCCACTGCTCGGCATCGGGCAGGGCAGGCTTTTTTTGA
            H  A  N  I  G  C  M  R  W  Q  E  A  D  P  L  A  P  K  K  Q
361       ATGATGACGGGCCAGCGGCGTGCCAGGCGCGCGTTGAGTTCCGGATATTTTTCCACTGCG
            I  I  V  P  W  R  R  A  L  R  A  N  L  E  P  Y  K  E  V  A
421       TCCGGCAAATCCACGTCGCGGAAGATGGCGTCTACCGGACACTCGGGCACACAGAGGGTG
            D  P  L  D  V  D  R  F  I  A  D  V  P  C  E  P  V  C  L  T
              ClaI                ClaI                ApaI
481       CAATCGATGCATTCGTCTGGATCGATGGCCAAAAAGTTGGGGCCCTCGTGGAAGCAATCC
            C  D  I  C  E  D  P  D  I  A  L  F  N  P  G  E  H  F  C  D
541       ACCGGGCATACTGTGACGCAATCGGTATACTTACACCGAATACAGGCTTCTGTAACCACA
            V  P  C  V  T  V  C  D  T  Y  K  C  R  I  C  A  E  T  V  V
                     EcoRV
601       TGGGTCATTGGATATCCCCCTTAACGTATGCCCCGGAAATGCCTTCAACGCCAGATCGGC
            H  T  M    fdxA start
661       GGGATTGCGCTCGCCCCATTTCCGAAAATCCGTCAAAACCGATCATCACGGCGTATAAAA
721       GATGCCTCAATCAGAACTTCATGTCGATAAACGGATAGATTTTATACCTATCCCGCGACT
781       CGCCCGCGTTATAGTAGCCGTTGGGCTTCCTGTACCACATTTTCCACAGTGAATCCCAGA
841       TGGCGGAAATTTTCTGGATAGGGTGCACTGGCGCCGAAGCGGCTCAGGCCTATGGTGGCG
901       CCCTCTCCCACCCATTCGCACCACCCCAGGGATGCCGCAGCCTCTACGGCGAGCCGGACT
961       TTGATATTCCGTGGGAGCACCGTATCCCGGTAAGCGGCCGGCTGGGCGCGGAATATCTCC
                   SphI
1021      CAGCTGGGCATGCTCTTCGCCACTGCGCCTTGAGGCGTCCCAATGCCGCAGGGGAAAGTC
                                                          EcoRV
1081      CCTTGGCCTCTTCGCCCAGGAGGATAGAGAGCGCCCCCTGCATGTGGCCAGAAGATATCC
1141      CATGCAGATACAGCCAGGGGACTGTGGCCGCCACCGTCCGCGATTTACGCACATACGGTG
1201      GCATCAGTGTCGAATTGAATTTAATCCCCGAACCCGCGCGATCTCGAACCTTGGGGACCT
1261      TCACTGACACAGGACCCAGGGCCGTCATGATCTCTCGCTCCGGCAGGTGGCCGTTACGAA
1321      CGACCGCCTGACGACCATTGACCATCCGTACAGCAGAAAACTCTTCCAGCACTTGGTCTG
1381      CAAACCCTTTGTAGCAACAGTACGCCACCTTGCTCAGGTCACTTGTAACCCGCCTATACA
1441      CCAGTTTCGATCATAGCTCTCCGCATTGGCGTTCCTTCAGGACAATGCGGTGCGCGATGT
1501      GGCCCCAGTTTGGCTGACAGAGCCAGGCCCCCTCCGTGTAGCTGATGAGATCGTTATTGA
```

```
1561    TTGGATGACGCTTGCGGCGAACGGTGAAACCTATTTCCTGTCATCTACCGGGTAGACGAT
                                                 fnr stop   *  R  T  S  S
1621    TTGCTGTTCGCCCATTGCCGTAACTCAGAAGGGCGGTGGATGTATATCTTGAGTTTTTCA
         K  S  N  A  W  Q  R  L  E  S  P  R  H  I  Y  I  K  L  K  E
1681    ACAGCTATCAAATGCTCCTCTCGCAATCGCTTCAACACACGAGAAAATGTTTCCGGGGTT
         V  A  I  L  H  E  E  R  L  R  K  L  V  R  S  F  T  E  P  T
1741    AAACCTAGTCGGGCCGCGATCATGGCTTTTTTGGCAGGGAGAAGCAAACTCACGGGAGAA
         L  G  L  R  A  A  I  M  A  K  K  A  P  L  L  L  S  V  P  S
                                   EcoRV
1801    TTTCCCGATGGGCAGAGTTCCAGCAGATATCCTGCAACCCGTTGGTCAGCGGATTCCACA
         N  G  S  P  C  L  E  L  L  Y  G  A  V  R  Q  D  A  S  E  V
1861    GTAAGATGTCGCAACTCCATTATTAAATGACGTAAACGCAGACTCAGTTCGCCCAACATT
         T  L  H  R  L  E  M  I  L  H  R  L  R  L  S  L  E  G  L  M
1921    TGCAGCATCAGCTCCTGATGCTGCTGTAGGGTGACGAGAAAATTATCGTAAGGGGAATTT
         Q  L  M  L  E  Q  H  Q  Q  L  T  V  L  F  N  D  Y  P  S  N
                                                        EcoRV
1981    CCAGAACCCGGCCTCGCATGGCCGCTATTGCCTGCACGGGATAACGATATCCCAAAAAAG
         G  S  G  P  R  A  H  G  S  N  G  A  R  S  L  S  I  G  F  F
2041    CTACAGCTTCGGCAAAAAAGGTCTCCGCCTTGCAAAATATCAACTATTTTCTCGTGGCCG
         S  C  S  R  C  F  L  D  G  G  Q  L  I  D  V  I  K  E  H  G
2101    TCAGCAGCCGACGTCACCAGCTTCACTTGTCCATCCAGAAGAAAATAGAAAGCGGTAGCC
         D  A  A  S  T  V  L  K  V  Q  G  D  L  L  F  Y  F  A  T  A
2161    GCATCGCCTTCAAGGAAAAGGATTTGCCCGACACTCATCTCCAGAACACGTGCGCCCTTA
         A  D  G  E  L  F  L  I  Q  G  V  S  M  E  L  V  R  A  G  K
2221    AATAAATGCGAAAGGTCCTCATCCGACCAGACAGAGAATATTTCTGTTCGGCGCAATCGC
         F  L  H  S  L  D  E  D  S  W  V  S  F  I  E  T  R  R  L  R
2281    TCCATTTCTGTTAACTCTACTTGGCCGATTGTCATCAATCTGGCTTTTGGGCGAACAAAA
         E  M  E  T  L  E  V  Q  G  I  T  M     fnr start
2341    CTGTGTCTAAAACCATCTTTACCCCATGATGCGGCTTTTACGGACCATAACTCAGCCTGT
2401    TATGTGCCTGCGGGTGGCGTCTCGCAGGCGGATTCAGCCGCCGGATTCCAGACAATTGAT
2461    CGGGATAGGCGGAGCGTAGTGTGACGTTAAATACGTTACGCCGAATCTCCGGGCAGGGTC
2521    AATCGCCTGTCTTTTGCTTAGGGTTTCCCGTCGACGATCATGAGGAATGCGCAGCCACTT
2581    GATTGAAATCAAGGATGGTCCTGATCACAGCAGATATTGTTGACAAACAAGCACCACAAC
2641    CTCAGGAGACACCATGATGTTTTGCTATCAATGCGAACAAACCACCCGCAGCCCCGCCGG
        psmA start    M  M  F  C  Y  Q  C  E  Q  T  T  R  S  P  A  G
                                                                PstI
2701    AATAGGCTGTACCAGTGAGCCGGGAACGTGCGGCAAGGATGAGGCCACTGCGGTTCTGCA
         I  G  C  T  S  E  P  G  T  C  G  K  D  E  A  T  A  V  L  Q
2761    GGACATTCTCACGCACCTTATGAAGGGCATTGCCAGTATGCGCGCCGGGCGGGCGATGGG
         D  I  L  T  H  L  M  K  G  I  A  S  M  R  A  G  R  A  M  G
2821    TGTTGCCGACAGACGTACGGACGACTTCATTTTTTATGGGCTTTTCACCACGCTCACCAA
         V  A  D  R  R  T  D  D  F  I  F  Y  G  L  F  T  T  L  T  N
2881    TGTGAACTTCACTGCCACACGTTTTGTCCACCTGATCCAGGAGGCCAGCAAAAGGCGTGA
         V  N  F  T  A  T  R  F  V  H  L  I  Q  E  A  S  K  R  R  E
2941    ACGGATCAAATTATTGTACGAGGAAGCGGCGCGAGAGCAGGGCAAGACTCCCGAAATCCT
         R  I  K  L  L  Y  E  E  A  A  R  E  Q  G  K  T  P  E  I  L
3001    GTCCGGTCCGGCCCTCTTCCAACCTGCCGATAGTCTGGAACAACTGCTGCGTCAAGCCCC
         S  G  P  A  L  F  Q  P  A  D  S  L  E  Q  L  L  R  Q  A  P
                                                  BamHI
3061    TAGCGTCGCCATAAACGCCGACGTAGAACATCTCGGATCCGATGTGATCGGTGCTCGTGC
         S  V  A  I  N  A  D  V  E  H  L  G  S  D  V  I  G  A  R  A
3121    CCTCATCCTCTACGGCATGAAAGGCGTAGCGCTTATTGCTCAACATGCCCGCGTCCTGGG
         L  I  L  Y  G  M  K  G  V  A  L  I  A  Q  H  A  R  V  L  G
```

```
3181    TTATCAGAGCGACGAAGTCATGCCGCAGGCAGAGGAAATCCTGGACTACTTGGCAAGTAA
         Y  Q  S  D  E  V  M  P  Q  A  E  E  I  L  D  Y  L  A  S  N
3241    TCCAACGGACCTCGATGAAATGTTGGAAGAATCATTGGAAGTCGGCCGTTTGAACCTGAA
         P  T  D  L  D  E  M  L  E  E  S  L  E  V  G  R  L  N  L  K
3301    GGTGATGGAGTTGCTTGACGTTGCCAATACGGACAGTTTTGGAGCGCAGGAAATTACCTC
         V  M  E  L  L  D  V  A  N  T  D  S  F  G  A  Q  E  I  T  S
3361    TGTGCGCATCTCCCCGATCCAGGGCAAGGCAATCCTGGTCAGCGGTCACGATCTCCATGA
         V  R  I  S  P  I  Q  G  K  A  I  L  V  S  G  H  D  L  H  D
3421    TCTCAAACAAATCCTGGAGCAGACCAAGGATCAGGGGATCAATGTTTACACCCATGGGGA
         L  K  Q  I  L  E  Q  T  K  D  Q  G  I  N  V  Y  T  H  G  E
3481    GATGTTGCCAGCCAACGCCTATCCCTTGCTCAAGGCATACCCTCATCTGGCCGGGAATCT
         M  L  P  A  N  A  Y  P  L  L  K  A  Y  P  H  L  A  G  N  L
                                                        ApaI
3541    AGGGGGAGCATGGCAAGATCAGCAGCGTGAATTTGCAGATTTCCCTGGGCCCATTGTCAT
         G  G  A  W  Q  D  Q  Q  R  E  F  A  D  F  P  G  P  I  V  M
3601    GACTTCCAACTGTATCATTGAGCCGGGCAGAAGTTATAAAAATCGAATTTTTACCCTTGG
         T  S  N  C  I  I  E  P  G  R  S  Y  K  N  R  I  F  T  L  G
3661    CCCAGTGGGGTGGCCCGGCGTCCGCCATATCGACAATGGGGATTTCACCCCGGTGATCCA
         P  V  G  W  P  G  V  R  H  I  D  N  G  D  F  T  P  V  I  Q
3721    GGCCGCCAAGGCATTACCGGGATTTACTGCCGATGCAAAAGAGCAGCGTATCACCATCGG
         A  A  K  A  L  P  G  F  T  A  D  A  K  E  Q  R  I  T  I  G
3781    TTTTGGGCATCACACTCTCCTGGGTGTGGCGGATAAAATCGTCGATGCGGTGAAACACGG
         F  G  H  H  T  L  L  G  V  A  D  K  I  V  D  A  V  K  H  G
3841    AGATATTCGCCACTTCTTCCTCGTCGGTGGGTGCGATGGTGTTTCTCCGGCGCGCAACTA
         D  I  R  H  F  F  L  V  G  G  C  D  G  V  S  P  A  R  N  Y
3901    CTTCACGGAGGTGGCGGATAATGCCCCCGCCGATTCGGTAGTGATGACCCTGGGATGCGG
         F  T  E  V  A  D  N  A  P  A  D  S  V  V  M  T  L  G  C  G
              ClaI                 EcoRI
3961    TAAGTATCGATTCAACAAGCATGAATTCGGAGATATTGGAGGCATCCCCCGCCTGCTGGA
         K  Y  R  F  N  K  H  E  F  G  D  I  G  G  I  P  R  L  L  D
4021    TATAGGCCAATGCAATGATGCCCACTCCGCCATCCGGGTAGCGGGTGCCCTGGCCGAGGC
         I  G  Q  C  N  D  A  H  S  A  I  R  V  A  G  A  L  A  E  A
4081    ATTCAATTGTGGGGTCAATGACCTGCCGTTGTCGATCATGCTCTCCTGGTTTGAACAAAA
         F  N  C  G  V  N  D  L  P  L  S  I  M  L  S  W  F  E  Q  K
                                                              ApaI
4141    GGCTACCGCCATTCACCTTTCCTTGCTGGCTCTGGGCATCAAGGGGATCAAGCTGGGGCC
         A  T  A  I  H  L  S  L  L  A  L  G  I  K  G  I  K  L  G  P
4201    CACCCTGCCTGCCTATCTCACACCAACTTTGGTGCAGAAGCTCCAATCACGTTTCGATCT
         T  L  P  A  Y  L  T  P  T  L  V  Q  K  L  Q  S  R  F  D  L
4261    TGATCTCGATCTTATTGGCGAGGCGCAAGCGGATTTGCAGGCGGCTCTGGCGCATACGGC
         D  L  D  L  I  G  E  A  Q  A  D  L  Q  A  A  L  A  H  T  A
4321    GTAGCACAATGGAACGCCGCTTCGTTCTGCAACGGACGAAGCGGCATAAGGATGAGCAGA
          * psmA stop
         redA start     EcoRV
4381    TCATGACCAACTACGATATCACTATCCATACCCGAGATAAGCAACAGGTGTCCTTCGTCT
           M  T  N  Y  D  I  T  I  H  T  R  D  K  Q  Q  V  S  F  V  C
4441    GTTCTGAGGCGGAAGATTTACTCTCCGCTGCGGACCGGGGAAGCATTCTGCTTCCTTCCC
           S  E  A  E  D  L  L  S  A  A  D  R  G  S  I  L  L  P  S  Q
4501    AATGCCGGAAGGGAACTTGCGGAGCCTGCGTGGCCACCGTCACCGCAGGCACTTATCACC
           C  R  K  G  T  C  G  A  C  V  A  T  V  T  A  G  T  Y  H  L
4561    TGGGTGAAGTCAGTATGGAAGCCCTGCCAGAGAAAGCGCAGGCACGAGGCGATGTACTTC
           G  E  V  S  M  E  A  L  P  E  K  A  Q  A  R  G  D  V  L  L
```

```
4621    TGTGCCGTACCTATCCACGAGCGGATCTGATTCTGGAGGCGCCTTATGACTACAACTACA
          C  R  T  Y  P  R  A  D  L  I  L  E  A  P  Y  D  Y  N  Y  I
4681    TCCGCTTTGAGCGCATTCCGGAGCGCGAGGCAGAAGTGATGGACGTCACTATGGTGGCTA
          R  F  E  R  I  P  E  R  E  A  E  V  M  D  V  T  M  V  A  T
                                                                EcoRI
4741    CGGGTACGCGACGACTGTTGTTACGCCTGCAACCCGATGAGCAAGGGGGAGCTGCGGAAT
          G  T  R  R  L  L  L  R  L  Q  P  D  E  Q  G  G  A  A  E  F
                                      KpnI
4801    TCGAAGCTGGGCAATTTATGGAAATCCAGGTACCTGGCAGCGATGCGCGTCGCCCTAATT
          E  A  G  Q  F  M  E  I  Q  V  P  G  S  D  A  R  R  P  N  S
4861    CTCTAGCCAACAACACCAACTGGAATGGCGACCTGGAATTTTTCATCACGCTACGACCAG
          L  A  N  N  T  N  W  N  G  D  L  E  F  F  I  T  L  R  P  G
4921    GTGGTGCCTTCTCTACCTATCTGGAATCCGCCTTGGTGGGCGACCGCCTGAACATACGTG
          G  A  F  S  T  Y  L  E  S  A  L  V  G  D  R  L  N  I  R  G
4981    GACCCCTGGGAACCTTTACCCTGAGAGAAAACGGCCTGCGCACGTTTGGTTCATTGGTGG
          P  L  G  T  F  T  L  R  E  N  G  L  R  T  F  G  S  L  V  A
5041    CGGGACAAGGTCTCGTGCCGCTACTGTCGATGCTGCGACGCATGGCAGATTGGGGAGAAA
          G  Q  G  L  V  P  L  L  S  M  L  R  R  M  A  D  W  G  E  M
5101    TGCTTCCGGCACGTCTGTACTTCGGCGCAAGATATGAAGACGAACTCTTCTGTCAGGAGG
          L  P  A  R  L  Y  F  G  A  R  Y  E  D  E  L  F  C  Q  E  E
                                            PstI
5161    AAATCCGTCAGATCCAGGATAAACTTCCCCAACTGCAGGTGAAGATATGCCTTTCACGTC
          I  R  Q  I  Q  D  K  L  P  Q  L  Q  V  K  I  C  L  S  R  P
5221    CTGGCAATCACTGGATGGATTATCGAGGCAGCGTGGTGGATGCCTTGCGCGACGATCTGG
          G  N  H  W  M  D  Y  R  G  S  V  V  D  A  L  R  D  D  L  G
5281    GAAGCCTTGCCACTTTACCGGACTTGTATGTTTGCGGATCAACGCGATTGGTCCAGGGGG
          S  L  A  T  L  P  D  L  Y  V  C  G  S  T  R  L  V  Q  G  V
5341    TGACGGAATTAGCTCTGAGCCAAGGGTTGCCGGACTCCTGTTTGCAGTTCGAACGGTTTT
          T  E  L  A  L  S  Q  G  L  P  D  S  C  L  Q  F  E  R  F  L
5401    TGTCCGCATAGTGCAATGTCTATAGACTTGCTAGAACCATTGATTTCGTCCAGCCAAGAA
          S  A  *  redA stop
5461    GGAACTCGTGGTAACCATAGAAACCAATAAAGACTATCGTTTTTCCCTTCTCCATCTTGG
5521    GTTCCGCCTCTTTTTTTTGTTAGGGGCGGTATTCTCTGCTTTTGCCATCTCTGCCTGGTT
5581    CTGGTTCTATTATCATGGAACGCAGTTTCTTGATAATGGCAATTTCCCGATTATCCTGTG
        SphI
5641    GCATGCCCATGAAATGATATTTGGATATTCCATGGCCGTGGCTGTTGGTTTTTTGTTGAC
5701    CGCAGTCAAAAACTGGACGGGTCACCAGACGCTTCGCGGATATGGATTATTATTACTAGG
5761    TTTGTTCTGGCTTATGGCACGCATAATGCCGTTCACCGATTTTCCTGGCGCACTGATGAT
5821    CATGGCGCTGTTCGACCTGTTGTTTGATGCCACTGTTTGCATCGCCTTACTGTATCCTTT
5881    GGTCCGCAGCAAGCAATGGAAACAGCTTGGCATATGGTCCATATTATTACTAATGACGAC
5941    AGGTAATACATTGTTCTATATAGGACTCATCTGGAAGCATCCTGAGCTGACGCAGATTGG
6001    GTTGTTTTTCGGCCTCTATCTGATTATCGCGCTGATTATGCTGATGGGACAGCGGGTAAT
6061    GCCGTTCTTTATTGAAAGAGGCGTTGGTTATCCGGTGGTTTTGATGGTTCGGCCGTGGGT
                                                              BamHI
6121    AAACGTCAGTAGCGTGATTCTTATGATGTTTTTTATAGTGGCTGTTGTATGGGATCCGAT
6181    TTCACTGATTACGGCTTTGGTGGCCATCGCGTTAGTTTTTTTGCAGGTTCTACAAATGAT
6241    CGGATGGTATACTCCGGGCATATGGCGTAACCTTTGTTATGGAGTCTTTACCTGGCTTAT
6301    GGATGGATATTCATTGGATTTATTTTGACTGCGCTTGTTCATTGGATACCCATCAATCCA
              EcoRI
6361    ATGTTAGGAATTCATGCTTTTGCATATGGTGGGGTAGGCATGATGACCATTGGCATGATG
6421    GCGAGAGTCTCTCTAGGGCACACAGGAAGAAACATCTTTACACCCCCACAGGCCTTGAAC
6481    TGGATGTTCTTTTCGTTGTTTCTTGGTGCCATTGTGCGCGTGATTATTCCAATAATATTG
```

```
6541    CCCATGCAGTACCGAATGTGGATTGAGGTGTCGCAAGGCCTGTGGATTATTGCCTTCGTC
                                                          ClaI
6601    GGTTTTATCATCGTTTATGCGCCCATGCTTATAAAGCCACGGATCGATGGCCGATATGGA
6661    TAATGAAATAAATAGAACAACGTGGATTGTTATTTATGGAATCACAAGTGATGGGGTAAA
6721    ATTCCGCCCTTCGTCGTGGGCGGAAATGCTTATTGAGGGGGTTGGGTTAACATCCTTTGG
6781    CAGAGACCACAAAATTCATTATACAAGACATATTTGGACAATATTAACGAATGGCAATGC
                                  PmeI
6841    CGGCATTATAATGAATGGCTGGTTTAAACATGAGCAGCCTATAGCATTTGCTGAGATTAT
6901    GCGCTTTGCCAATAATAATAGATTAACAGTCAATACGTATAACAATACCATCGCTCTTCT
6961    ATAACCACTGAAGTTTAGACATGCAGCCTCCTCGCCAGCTTCTATGTGTGGCGATATGGC
7021    GTAGTAAATGTGGAGCAGGATTGAGTCAGGCGACCTTTTTGCGGGGTGGGTCATCGGTAT
                        orf51 stop    A  V  K  K  R  P  P  D  D  T
7081    CGGCCCAAAAGGCGGGGGCGGGGCACTGGAATTTCCCGGACTTCGGGTCGTAGCCCTTCC
        D  A  W  F  A  P  A  P  C  Q  F  K  G  S  K  P  D  Y  G  K
7141    AGAATGGAACCCGCAGAAATTGCAGATGAAGCCATTGGGCCATGAGTTCTGCGGTGACAG
        W  F  P  V  R  L  F  Q  L  H  L  W  Q  A  M  L  E  A  T  V
                                  PstI
7201    GAGTAGCGATTACCCAGGGCGCTACTGCAGTCATGGGGAAATCTTCGGATAGGGTTTCGG
        P  T  A  I  V  W  P  A  V  A  T  M  P  F  D  E  S  L  T  E
7261    CGAGGAGTTGCACCATGGACCAAGCGATGCAGCGGATTTGCATCCAGCGTTCCAAGACGG
        A  L  L  Q  V  M  S  W  A  I  C  R  I  Q  M  W  R  E  L  V
7321    TGCGTTTTTGCTGCCAGAGGTTATGGATGCCCCACCAGCGTTTGAGATTGTGGAAGATCG
        T  R  K  Q  Q  W  L  N  H  I  G  W  W  R  K  L  N  H  F  I
7381    GTTCGATCTGCCAGCGGTTGCTGTAGAGGACGACGATTTCGCTGGCGGTGAGATCCGTCT
        P  E  I  Q  W  R  N  S  Y  L  V  V  I  E  S  A  T  L  D  T
7441    CGGAAGCGAGATAGAGGCGGGTGGGAGACCACCGATGATGCTTTTCATCGTAGATCGCAC
        E  S  A  L  Y  L  R  T  P  S  W  R  H  H  K  E  D  Y  I  A
7521    TCCATACGGCACGGACCGGATGGCCGTTCAGGAATCGTGCATTGGCGATGATGGATTGAA
        S  W  V  A  R  V  P  H  G  N  L  F  R  A  N  A  I  I  S  Q
7581    GACGGACCTTCTGCCGTTTCCCATAGAGATCCAGAGAGACGATCTCGGTGGGCAGCGCAG
        L  R  V  K  Q  R  K  G  Y  L  D  L  S  V  I  E  T  P  L  A
7641    CCATGTCTTCCACGGTGTACTTGTTTCCATATATCCGAGGACGCCCGCGTTGGGATGTCC
        A  M  D  E  V  T  Y  K  N  G  Y  I  R  P  R  G  R  Q  S  T
7701    TCGGCGGTGGGTCGCGGAGCAGTACAGTATCTTTGCGCACCTGGCCGATGAACTGCTGCC
        R  P  P  P  D  R  L  L  V  T  D  K  R  V  Q  G  I  F  Q  Q
           PstI
7761    GGCTGCAGCGTTGCTGACGCAACAGGGGCAAGAGCAGGCGCCGCCGCATGAACCAAGCAT
        R  S  C  R  Q  Q  R  L  L  P  L  L  L  R  R  R  M  F  W  A
7821    CGGTGAGCAACCGGACGGGTACGGGGATGTGGGGTATAAAGGCGCGCAGCAATGCTCCGG
        D  T  L  L  R  V  P  V  P  I  H  P  I  F  A  R  L  L  A  G
7881    CGATTTTCAGCTTGTGGGTGTTGCCGGTACTGGGCACCAGGCGCAAGCGGATGGGAAAGG
        A  I  K  L  K  H  T  N  G  T  S  P  V  L  R  L  R  I  P  F
7921    AAATGGCTTTGCCGGAACGGGTACGCAGGGTCAACGCCAGACCCACCCAATTCTGGGCGA
        S  I  A  K  G  S  R  T  R  L  T  L  A  L  G  V  W  N  Q  A
7981    GGACATAACTGGGGCGGTTGTGCTTATGACTGTGGTCGAAGCGGATGGCCGCGCCTGGGG
        L  V  Y  S  P  R  N  H  K  H  S  H  D  F  R  I  A  A  G  P
8041    CCTTGGTAGAGTGCCGCAGGACCATCGTGTCGTCGACGACTAACTCGACGATCAGGGGCT
        A  K  T  S  H  R  L  V  M  T  D  D  V  V  L  E  V  I  L  P
                 PstI
8101    TGGGCAAGGACTGCAGGAGCCAGCGTACCTGGGTTATGGCCAACGCCTGTGTACGCAAGG
        K  P  L  S  Q  L  L  W  R  V  Q  T  I  A  L  A  Q  T  R  L
8161    ACGCACGCTCCAGGAGTTTGTAATAGGTGGTCCAGTGGCAGCGGAGGGAGATGCTGCTGA
        S  A  R  E  L  L  K  Y  Y  T  T  W  H  C  R  L  S  I  S  S
```

```
8221    GGGCGCGGGTGACCCAGCCTTCGGGCGAGATCAGGCAACCCGCAAAGCAGCTCCACAAAA
        L  A  R  T  V  W  G  E  P  S  I  L  C  G  A  F  C  S  W  L
8281    CTGCGCCGCGAGCGGGCGGCACGACCGTCAGAAAGTAGGAAATCCACTCCGAGAGGCGGA
        V  A  G  R  A  P  P  V  V  T  L  F  Y  S  I  W  E  S  L  R
8341    GGGCAACGAGACGATGAAAGGGCATGACAGTGATTCCGAGAAGCCTAAAACGCTTCTCAT
        L  A  V  L  R  H  F  P  M  orf51 start
8401    TGGCCGACCCTTTTTTACAAGGGTCGGCGACACGGACTATTCACTGTGTCAAGCACTTTC
8461    GTAGCAAAAATAGAAAAAACCCGGCGGAGCTGGTAATGTCTAAACTTCAGTATAACCATA
8521    TGAAGTTGAGCAGGTATTTGGTGTCCAGGCCCAGCATAGTAAAAATGCCCAACTTGCTGC
8581    GCGGGCCAAGATGGAGGGAGTATCCGCTCAGTAATGGGTAATCCCTACCGCGATACCGGT
8641    TTGCAGTGGGCGGTACTGAAACTTGCCCACCCTGGGGCTACCTCCCGTTCTGTCCGGGTA
8701    GATTCCGTCAAACGATCATTAGTCGGGCCATGTCGGACAATGTCCGGGAACGTGCTTGTT
8761    TAACGGGTTTATCGGACATTAAGTTAGACGGACGGAAGGACGGACAAAAGGGGCGATCAT
                                                  invertase start  M
8821    GGCACTGATCGGTTATGCGCGGGTATCGACGGCGGAACAGGACACCGCTTTACAGACGGA
         A  L  I  G  Y  A  R  V  S  T  A  E  Q  D  T  A  L  Q  T  D
8881    TGCGTACGGCAAGGCGGGATGCGAGCGCGTTTTCGAGGACCGGCTTCCGGGGTTAAGTCA
         A  Y  G  K  A  G  C  E  R  V  F  E  D  R  L  P  G  L  S  Q
8941    GACCCGGCACGCTTTGGCCGCCGCGCTGGCCTACCTGCGCGACGGCGACGTGCTAGCTGT
         T  R  H  A  L  A  A  A  L  A  Y  L  R  D  G  D  V  L  A  V
9001    CTGGCGGTTAGATCGCTTTGGCCGCTCCCTGCCGCACCTGATCGAAACGATAAGCGCTCT
         W  R  L  D  R  F  G  R  S  L  P  H  L  I  E  T  I  S  A  L
9061    GGAAGCTCGTGGTGTCCGGTTCCGCTCGCTGACCGACAGCATCGACACCACCACGCCGGG
         E  A  R  G  V  R  F  R  S  L  T  D  S  I  D  T  T  T  P  G
9121    CGGACGGCTCATCTTCCATGTGTTCGGCGCACTGGGCCAGTTCGAGCGCGACTTGATCCG
         G  R  L  I  F  H  V  F  G  A  L  G  Q  F  E  R  D  L  I  R
9181    CGAGCGCACCAAGGCCGGGGTTGAGCGCTGCCGCGCTCGCGGGCGCAAGGGCGGGCGCAA
         E  R  T  K  A  G  V  E  R  C  R  A  R  G  R  K  G  G  R  K
9241    GCCGGTTATTACTGCCGACAAGTTGCAAAGGGCGCGGGAGCATATCGCCAATGGGCTGAA
         P  V  I  T  A  D  K  L  Q  R  A  R  E  H  I  A  N  G  L  N
9301    CGTCCGGGAGGCTGCGGCACGGCTCAAGGTTGGGAAAACCGCCCTTCTTCCTTGCAGGCC
         V  R  E  A  A  A  R  L  K  V  G  K  T  A  L  L  P  C  R  P
9361    GGCTGGTTTCCGCAGTGCCGTCGACTCCTGATTTCCGTGCCATCGCCTTCTGAAAATGAC
         A  G  F  R  S  A  V  D  S  *   invertase stop
                                                  ClaI
9421    CGCAGGCTCGGGAGATTCAGTTCAAACGCTTGACCGGCGCCATCGATACCGGGACACCCT
9481    CGGGACGGTTCTTCTTTCATGTCATGGTCAGGTTGGCCGAGATGGAGCGGGATCTGACCA
9541    TTGAGCGCAGCTGTGCCGGTTTGGAAGTCGCCCGCAAGTTCGGGTGGATGCCGGGGAAGA
9601    AGCGACTAATGACGGAAAGCAAGGTGGCATTGGCCCGGAAGCCCCCGGATAACGATACTC
9661    CACGTCGAGAGGTGGCTGAGCACATCGTGGCGTCACTGCTGACGCTATATCGTTGGATTC
                                                  EcoRI
9721    CGGGGGCTTCACACTCTTAACGGGCTTCATATTCCGAATTCTCAAGTGCGCAATCAATAT
9781    TGCCAAACTTGACAACTATGCTTGCATACAAATATGATTGCATAATGTTTACTTGTATTG
9841    GAGCTTATTTGCATGTTCCGTATTGTCTGCAATTCCCAGAAGGGGGGTAGCGGAAAATCT
       orf26 start  M  F  R  I  V  C  N  S  Q  K  G  G  S  G  K  S
9901    ACGGTTTGCCGCGTCCTTTCCGTTCACGCAGCCCGCCTTGGGTTTTCTGTATATTTGGTG
        T  V  C  R  V  L  S  V  H  A  A  R  L  G  F  S  V  Y  L  V
9961    GACACCGACACGCAAGGCACTCTGACCCAATGGCACGAAGCCAGAGAGACTGAGGAGCCT
        D  T  D  T  Q  G  T  L  T  Q  W  H  E  A  R  E  T  E  E  P
10021   CGACGCGTGGTAGTTGACCAGAGACGTCTTGGGAAAAGGCATGAACGTCACTTGCGGCAC
        R  R  V  V  V  D  Q  R  R  L  G  K  R  H  E  R  H  L  R  H
10081   AGGGCGCGGATTTTGTTTTTGTCGACACACCACCGAATGCCCAGTGAACATTTAGACGAT
        R  A  R  I  L  F  L  S  T  H  H  R  M  P  S  E  H  L  D  D
```

```
10141   GTCTTTGAACTGGCAGATTTGGTGCTTGTACCCATAAAACCGACACCGGACGACCTGAAA
          V  F  E  L  A  D  L  V  L  V  P  I  K  P  T  P  D  D  L  K
10201   GCAGCACTGGTGACCGTTTATCGCCTAAAGAGCTTGGGAGTGCCATTTCTATTGGTAATT
          A  A  L  V  T  V  Y  R  L  K  S  L  G  V  P  F  L  L  V  I
10261   ACGCAGGCGATCCAAAACACAAACATCACTGCTCAGGCAATTGCGGCTCTCTCCCATCAC
          T  Q  A  I  Q  N  T  N  I  T  A  Q  A  I  A  A  L  S  H  H
                BamHI
10321   GGATCCGTCGCTGAAACCATCCTTGTGAATCGCGTCGCCTACCCGTCGGCATTCACAGAC
          G  S  V  A  E  T  I  L  V  N  R  V  A  Y  P  S  A  F  T  D
10381   GGACGCACACCACAGGAGATTGAGCCAAGAGCACCAGCGGCGCGGGAAATTGCGTCTTTA
          G  R  T  P  Q  E  I  E  P  R  A  P  A  A  R  E  I  A  S  L
                                      KpnI
10441   TGGGATAACATACTATCATACTTGCATACTGGTACCGTTGAGACAAGGAGTGAAGCGCGT
          W  D  N  I  L  S  Y  L  H  T  G  T  V  E  T  R  S  E  A  R
10501   GGCTAAACCATCTGCTCTAACAGTTGACGTGATGAACGCACCGCAGGTCTCGCCTATAGG
          G  *  orf26 stop      orf9 start M  N  A  P  Q  V  S  P  I  G
10561   TGGCTTACCATCTCGCGGCTTTGCCGAGCCCCACCAGAACAAACAGCAAGCAGCACCTAA
           G  L  P  S  R  G  F  A  E  P  H  Q  N  K  Q  Q  A  A  P  K
10621   AAATGATGCGCTCGTGCAAATCCGTTGTTCTAAGGCGGCAGCTAAAGAGGTGAAGCGCGC
           N  D  A  L  V  Q  I  R  C  S  K  A  A  A  K  E  V  K  R  A
10681   TGCCGTTGAAGCGGAAATGACGATTAGCGAGTTCATGCTTGTATGCTTTCATGCTTATAT
           A  V  E  A  E  M  T  I  S  E  F  M  L  V  C  F  H  A  Y  M
                                   SpeI
10741   GAAACGATGATTTTTTGTTGGCATACTAGTATGCTAGTATGTGCGAGCAAGGATAATAAA
           K  R  *     orf9 stop                              repA start   M
10801   TGGCAGACAACAATGTGACCAAGAAAAATCCCCCTACACGCAGATCAGTAGCGCCGGTGT
             A  D  N  N  V  T  K  K  N  P  P  T  R  R  S  V  A  P  V  S
10861   CTGATACCGCTTTTGCCGGTTGGCAGCTAAGCCTTTTCCAGGGCTTCCTAGCCAACACGG
             D  T  A  F  A  G  W  Q  L  S  L  F  Q  G  F  L  A  N  T  D
10921   ACGACCAGGTCGAATCACTCTCAAACGCCGTTGACTTGTGGGACAGCATACCGCGTTACT
             D  Q  V  E  S  L  S  N  A  V  D  L  W  D  S  I  P  R  Y  S
10981   CTATTTCACGAGCCAGAATGAACACCATGCGAACAGCTGATGGGTTTCTAGGCGTTGCGA
             I  S  R  A  R  M  N  T  M  R  T  A  D  G  F  L  G  V  A  S
11041   GTTGTCATTCCACTATCGTGGTAGAGCATATACGGCAAGAATATATCCCGCACAGGTTAA
             C  H  S  T  I  V  V  E  H  I  R  Q  E  Y  I  P  H  R  L  K
11101   AAATTAATGATGGCCAATGGAAAAGCTATTATCCGAGCGCGCGTGAAGAGCTTGTCGAAT
             I  N  D  G  Q  W  K  S  Y  Y  P  S  A  R  E  E  L  V  E  Y
                                                          ClaI
11161   ATGCCTTGCGGAAAATCTCTGCTGAACAGGGCGCTGGCTTCTTTGATCGATCAACTTATC
             A  L  R  K  I  S  A  E  Q  G  A  G  F  F  D  R  S  T  Y  R
11221   GCAGCGGAGCCCGATTTTCTTTATATCAGCTCCGTAAAGAGTTGGAGCAACAAGGACACC
             S  G  A  R  F  S  L  Y  Q  L  R  K  E  L  E  Q  Q  G  H  Q
                                         EcoRV
11281   AGCTTGCGTATGACCAGATCATCGAGGCGCTCGATATCCTCTCTCTGAGTAGTATCGAAA
             L  A  Y  D  Q  I  I  E  A  L  D  I  L  S  L  S  S  I  E  I
11341   TTGAATGCGCAACAGACAGTGGAGATGGGGCCTTCGCTCGCTCCACATATTTTGCTGCCT
             E  C  A  T  D  S  G  D  G  A  F  A  R  S  T  Y  F  A  A  L
11401   TAAGCGGTGTTAAACGCAAGGATTACGAAACCCATCGCGACACTCGATGGATAGCACAAT
             S  G  V  K  R  K  D  Y  E  T  H  R  D  T  R  W  I  A  Q  F
                                   ClaI
11461   TCCATCCGTTAGTGACGCAGAGTATCGATCATGTGACCTATCGACAGTTTAACTATCAGA
             H  P  L  V  T  Q  S  I  D  H  V  T  Y  R  Q  F  N  Y  Q  R
```

```
11521   GAATGATGACATGCAGCACTCAGCTTGCCCGATGGTTGATCGGACAATTGGTCTTGAAAT
          M  M  T  C  S  T  Q  L  A  R  W  L  I  G  Q  L  V  L  K  Y
11581   ATACCCAAGCTGCCATGCTCAACAGTTTCGAAATACGTTATAGCACCATCAAGCGTGATA
          T  Q  A  A  M  L  N  S  F  E  I  R  Y  S  T  I  K  R  D  S
11641   GCGCCCTATTGGCAGGGTACAAACTGGATCGTCAGGCGATAGCAGCCTTGGACCAAGCGT
          A  L  L  A  G  Y  K  L  D  R  Q  A  I  A  A  L  D  Q  A  W
              SacI
11701   GGGATGAGCTCAAAAGCCTGGGCGCGCTATCTACCGTAAAAAAGATCGAGCAGCGGGGAG
          D  E  L  K  S  L  G  A  L  S  T  V  K  K  I  E  Q  R  G  A
11761   CGCGCAGCAAACTAGAGGATGTTATTTATACGTTACATCCGACGCAGGAATTTGTTGCCG
          R  S  K  L  E  D  V  I  Y  T  L  H  P  T  Q  E  F  V  A  E
11821   AGCAGAAGGCAGCGAACCGTCGTCAGAACAATGCGAAAGATGGTATCAGCAGCGCGGTAG
          Q  K  A  A  N  R  R  Q  N  N  A  K  D  G  I  S  S  A  V  E
11881   AGATGCAAAATCGAGTAGAGCAGCTTAACAAAAAACATCCTGTTCGTGCGACGAGACAGG
          M  Q  N  R  V  E  Q  L  N  K  K  H  P  V  R  A  T  R  Q  G
11941   GGAAATAGGTGACGGGTCAGCCCTCCGGACAGGGGAAATAGGTGACGGGTCGGCCCTCCG
          K  *     repA stop
12001   GACAGGGGAAATAGGTGACGGGTCGGCCCTCCGGACAGGGGTTAGTTTTTAACAGGGGAA
12061   ATAGGTGACGGGTCAGCCCTCCGGACAGGGGAAATAGGTGACGGGTCGTCAAAAAGACAG
12121   GGGAAATAGGTGACGGGTCGTCAAAAAGACAGGGGAAATAGGTGACGGGTCGTCAAAGAA
12181   TCGCTCTCAATGACCCCCAATCTGTGGATAACTGCTAAGTCTTTGCCTTAAATTCATACC
12241   TCGGCAGGGGAAATGGGTGACGGATGAAAATCATAACAAAATAATTTTGTTGGAAAATTT
                                                  PacI
12301   CGCCATTTTTCAGCCCTTATCCTTTTTATCCTTTAATTAACCTTCTTTTATCCTTAGGCT
12361   TCGCAAACTCCGCCGCCGTCTTCGTGTCGCCCCCACCGACTCGAACCCACGCCCTTCTGC
12421   TACGCGTCCGCCTGCGTGTCCCACCCGGCACCTCCCAACTCTTTAAAGAGCGCCCACTTG
12481   CCTCCGGCCAGGCGCGAACATCCCGGCGGGACCCAGCAGGTGCCCCTGTTCCCGCTTCGG
12541   CTTGCCGGGTCGACCCCCTTTGGCTGCGTCCCCCAACTCCTATACGGCAGCGGCTCCCCC
12601   TAAAATTTTCCAAAAAGCGGGTGATTTTAGCTCAGGGCGCATTTTTGGTGACCGGGTGAT
12661   GGAAATATACCGGAATGGAATTTGCGTCGCTTCTAGGGCAAATACGGGCTATTTACGGCG
12721   GGGTTCTGCATCAGGCTTCTGGTCGGTTGTGCTGCGAATGATGCTTGGGACCGGCCTTGC
12781   TTGCGTTCGTTACCGTCACCGTGCGCTGCTGACCGGCTTTCTTCGAGTGCAAAGTGAGCG
                                  XbaI
12841   GTTTCTACTGGGGGGGCGCGCTCTAGAATGAAGTGCAACATCTCCGGTCGAAGGATGGCG
                                  EcoRV
12901   GAGGTGAGTGAATGGGGACGAGATATCAGCAGTTGCAGTCGGAGCAGAGGAACCAGATTC
        tnpA start M  G  T  R  Y  Q  Q  L  Q  S  E  Q  R  N  Q  I  Q
12961   AGCGAGGGCTGAACGAGGGGTTGAGTATGCGGGCCGTGGCCAAGCAGATAGGGAGGAGTC
          R  G  L  N  E  G  L  S  M  R  A  V  A  K  Q  I  G  R  S  P
13021   CCAGCACGGTCAGCCGGGAGGTACGCCGGGGTTTGGTGGGAGAAACCTACGATGCGATAC
          S  T  V  S  R  E  V  R  R  G  L  V  G  E  T  Y  D  A  I  Q
13081   AAGGCCGGGAGGGACGGCGCAGAGGCGTTCGTAAGGGGGTTAGAAAGCTGGTGGGAGGCG
          G  R  E  G  R  R  R  G  V  R  K  G  V  R  K  L  V  G  G  A
13141   CGCCCTTAACCAACGCGGTGACACACGCTATCCTGCAAAGGAAATGGTCACCAGAGCAGG
          P  L  T  N  A  V  T  H  A  I  L  Q  R  K  W  S  P  E  Q  V
13201   TGGCCGGGAGGTTGCGGATGGACTATCCCGAGGACAAGCAGTGGCGTGTCTCCCATGAGA
          A  G  R  L  R  M  D  Y  P  E  D  K  Q  W  R  V  S  H  E  T
13261   CCATTTATCAGTTCATCTATGCCCACCCGGCCGGTGAGCTGCGTAAGGCGCTGATAGCGG
          I  Y  Q  F  I  Y  A  H  P  A  G  E  L  R  K  A  L  I  A  A
13321   CGCTACGCCAGGGGACACGCAAAGCGCAAGCGCGCACACGCGGAAAGGACCGGCGCGGAC
          L  R  Q  G  T  R  K  A  Q  A  R  T  R  G  K  D  R  R  G  Q
13381   AACTGCGGAACATGCGTTCCATCGGGGAGCGTCCCTTGGAGGCCCAAGACCGCGAGATAC
          L  R  N  M  R  S  I  G  E  R  P  L  E  A  Q  D  R  E  I  P
```

```
13441   CCGGCCACTGGGAAGGAGACTTCATCAAAGGGGCTTTCAACGGCAGCGCCATTGGTACTC
          G  H  W  E  G  D  F  I  K  G  A  F  N  G  S  A  I  G  T  L
13501   TGGTGGAGCGCAGCAGCCGTTTCGTGCTTCTGGTCAGGATGGAAGGCACCGATGCCGACG
          V  E  R  S  S  R  F  V  L  L  V  R  M  E  G  T  D  A  D  A
                                  SphI
13561   CGGCCCTGGAGGGGTTCACCAGGCGCATGCCCTTGCCCAAGTCCATCCTGCGGACCCTCA
          A  L  E  G  F  T  R  R  M  P  L  P  K  S  I  L  R  T  L  T
13621   CCTATGACCAGGGCAAGGAGATGGCACGGCACGAGGAGCTGGAGCGCAAGGTGGGCATCC
          Y  D  Q  G  K  E  M  A  R  H  E  E  L  E  R  K  V  G  I  R
13681   GTATCTACTTTGCCGACCCGCATAGTCCTTGGCAGCGCCCAACCAACGAGAACACCAATG
          I  Y  F  A  D  P  H  S  P  W  Q  R  P  T  N  E  N  T  N  G
13741   GTCTCCTGCGCCAGTATCTTCCCCAAAGGACGGATTTATCAGGATATTCACAACGCCGCT
          L  L  R  Q  Y  L  P  Q  R  T  D  L  S  G  Y  S  Q  R  R  L
                          SacI
13801   TGACGCAGGTGGCGGAAGAGCTCAACAATCGCCCAAGAAAATCTTTGGGATTCCGAACGC
          T  Q  V  A  E  E  L  N  N  R  P  R  K  S  L  G  F  R  T  P
13861   CAGCAGAAGTAATAGCACAGCAAATCATGCAGTTAAACAGTGGTGTTGCGCTTCAAATTT
          A  E  V  I  A  Q  Q  I  M  Q  L  N  S  G  V  A  L  Q  I  *
                                                          tnpA stop
13921   GAAACCGCCGGGTGTAAGCCGGAATCCCAGATTTTTTCGGCAGGAGGTTCTGCGCCAGAA
13981   TCGCGGGGTTTGTCCGCTTCCGGACGGTTTGACGACTCTGAAACGCTTGCCCCTGCTCGA
14041   TCTGTGTAGTAACGTATCTACATTGATCCAGCATACGGGGTGCCATCATGACCATTACTA
14101   CCTTATCCAGCCGCGAGTTCAACCAGGGGGCGAGCCAGGCAAAACGGGCCGCAAACAATG
14161   GACCAGTGTTCATCACCGATCGAGGCCGACCGGCCCATGTGCTGATGAGCTTTGAGGATT
14221   ATCAGCGGCTCACCAAGCAGCGACGCAACATTGCCGATGCGCTGGCTATGCCGGGTATTG
14281   CCGACATCGAGTTCGAGCCGCCGCGTGTGACGATTGAAACGCGTCCGGCGGATTTCTCAT
                                                    stbB start   M
14341   GAGGTTTGTTCTCGATACCAACGTGGTGTCTGAACTGCGCAAGGTGCGGCTTGGTAAGGC
          R  F  V  L  D  T  N  V  V  S  E  L  R  K  V  R  L  G  K  A
14401   CGATGTGAACGTGACGGCATGGACGGAAAGCGTGGATGCCGCCGATCTCTTTGTGTCAGC
          D  V  N  V  T  A  W  T  E  S  V  D  A  A  D  L  F  V  S  A
14461   CATCACCATCCTGGAACTGGAGCTTGGCGTTCTGTCGATGGAGCGAAAGGACGCCACCCA
          I  T  I  L  E  L  E  L  G  V  L  S  M  E  R  K  D  A  T  Q
            ApaI
14521   GGGGCCCTGTTGCGCTCGTGGCTAGAGCAGCACGTCTTGCCCGAGTTCTCCGGGCGCACG
          G  P  C  C  A  R  G  *  S  S  T  S  C  P  S  S  P  G  A  R
14581   CTGCCTGTGGATACCGCCGTGGCACAACGCTGCGCCAGGCTGCATGTGCCCGACAAGCGC
          C  L  W  I  P  P  W  H  N  A  A  P  G  C  M  C  P  T  S  A
14641   GGCGACGAATTTATGCACTCATCGCGGCAACCGCCCTGGTGCATGGCATGACGGTGGTCA
          R  R  I  Y  A  L  I  A  A  T  A  L  V  H  G  M  T  V  V  T
14701   CTCGTAACGTCGCTGATTTTAAGCCTACGGGAGTGCCCCTCATCAATCCGTGGGCGGTGT
          R  N  V  A  D  F  K  P  T  G  V  P  L  I  N  P  W  A  V  S
14761   CGCAGTAACGGGAACCCGTTGTTTCCCGACGCCACTGACCACTTGGCAACTGCCGCTCAA
          Q  *   stbB stop
14821   AATTCAGCAAACCGTGTATGGAACACCATGTCGGTCAGGCGTAGTGCCCGGAAATCCTGC
14881   CGCCGATGTTTCCCGACGGAATTTTCTAGGGTTCCGGCTTACACCCTCTACTACGGAAAC
14941   TGTTTCTGATGTGCCCGACAGTATGCCCGCCCACCAAATCGGTCCTTGCGCGTAAAACAA
15001   AAAGAGGCTTCCTCCTGGCTAATGGGGGTAGAACAAGCAGTCGTGTCACCTGCCGCCATC
15061   TTGATCACGGCAGATCAAGTCGGGTATTCATGTCCAGGTCGAACCGCCCATATGGGTTGA
15121   CGTGTTCCCATATCAGCGGCGTCAGGGCGGCGTAGTCCGCGGCGTGAACTTGTCTTGCCA
                orf90 stop   *  P  P  T  T  R  P  T  F  K  D  Q  W
```

```
15181   ATGCGGCTGGGCCAGCACCTTCTGGATCATCAGCGTGTTGATATAGACCATGCAGTTCTG
          H  P  Q  A  L  V  K  Q  I  M  L  T  N  I  Y  V  M  C  N  Q
                          SphI
15241   GATCAGGTGCAGCGCGAGCATGCTGATCTCGTGGTCCTCACGACGGTTGCTCCCCATCTC
          I  L  H  L  A  L  M  S  I  E  H  D  E  R  R  N  S  G  M  E
15301   GCCACGGCGGGCGAAGAATACGAAAGTCGTCGCGCCGTTCCACTGCTCGACCACATTCAG
          G  R  R  A  F  F  V  F  T  T  A  G  N  W  Q  E  V  V  N  L
15361   CCCCTCGTTGATTTCCCGGCGCAGGCCTCGTCGTGCAGGTAGCGGCACAGGAAAATGGTC
          G  E  N  I  E  R  R  L  G  R  R  A  P  L  P  V  P  F  H  D
15421   TTGATCGTTTCCCAACTCAGCAAATGCCTTGTAGGTCGGATGCTGCACGTTCTTCTTGGT
          Q  D  N  G  L  E  A  F  A  K  Y  T  P  H  Q  V  N  K  K  T
15481   GAAGCGGCGCAGAATGGCTTCGGTTTCCGCCGTCCCCAGGCGTAGCGCGGTGGCATACTT
          F  R  R  L  I  A  E  T  E  A  T  G  L  R  L  A  T  A  Y  K
15541   GACCATCTGGTCGTACTGTTGCCGCACCGAGTCCCAGTCGATGGGCTTGGTCAGAATCTG
          V  M  Q  D  Y  Q  Q  R  V  S  D  W  D  I  P  K  T  L  I  Q
15601   TTGCAGGTTGGCGTAGGCAATCGGCCTTGCGGTTTCCGGCCGGTACAGCTTCTGTGAGTG
          Q  L  N  A  Y  A  I  P  R  A  T  E  P  R  Y  L  K  Q  S  H
15661   GATCGCCTTCAACCGGGGTAGCAACCGGAAGCCGAGTAGTCGGCAGAAGGCGAACGCGAC
          I  A  K  L  R  P  L  L  R  F  G  L  L  R  C  F  A  F  A  V
15721   CGTGCTCTGGCCGTGCGAATCGACATACTGCCGATCCACCTCCATCTCGGTGCAATGGTG
          T  S  Q  G  H  S  D  V  Y  Q  R  D  V  E  M  E  T  C  H  H
15781   GATCACGCCCTCGATCATCGACGCCACTTCCGACGACGACGGCGACTTGAGCTGCGAATG
          I  V  G  E  I  M  S  A  V  E  S  S  S  P  S  K  L  Q  S  H
15841   GATGCACCAGCTGCGCTCGACGTGCCAGTAAATCATGATGCCGCGCCGCCGTAGCGGACG
          I  C  W  S  R  E  V  H  W  Y  I  M  I  G  R  R  R  L  P  R
15901   TGCCATTGCGTGGTGAGGTTCTGATCCGCACGCGCCGAAGTGTTTGGAGTCCGAGCGCGC
          A  M  A  H  H  P  E  S  G  C  A  G  F  H  K  S  D  S  R  A
                                                 SphI
15961   AGGCGGTGGTGCCGCTACCCCAGATCGCCGGGTTGCGGCATGCAGTGTGCCGTCTGCGAC
          P  P  P  A  A  V  G  S  R  R  T  A  A  H  L  T  G  D  A  V
16021   GATGGCAATCGCGCAACGCATTGCGTCCACGCTGATGTAACGGCGGCGCACATAGGCTAG
          I  A  I  A  C  R  M  A  D  V  S  I  Y  R  R  R  V  Y  A  L
16081   GTCGCGCGCCGTGGTGCCGGAATCCAGCCCGGCCATGCGCTGCAATCCGGCATTGGTGCC
          D  R  A  T  T  G  S  D  L  G  A  M  R  Q  L  G  A  N  T  G
16141   CAGGCCGTGCAGACACAGCAGCAGGCGCGGCTGCAACACCGAGCGATCCATCGTCTCGTA
          L  G  H  L  C  L  L  L  R  P  Q  L  V  S  R  D  M  T  E  Y
16201   GGAGGTCGGGCTTTTCAGGGCATCGGTGAAGCTCAGCCGCAGATCGGTTTCCTTGACCAT
          S  T  P  S  K  L  A  D  T  F  S  L  R  L  D  T  E  K  V  M
16261   ATCGAGCAGGCTGGTCATCGGCCAGATGGTGTTGAGTTCGGCCTTGAGCGCAGTCAGATT
          D  L  L  S  T  M  P  W  I  T  N  L  E  A  K  L  A  T  L  N
16321   GGGCGGATCTGGTTGCGCATCGAGCGGCGTCAGCGTGATCCAGCCACCGCCCTTGCTGCT
          P  P  D  P  Q  A  D  L  P  T  L  T  I  W  G  G  G  K  S  S
16381   CAGCCGGACGGATAGATTCTTCTTCAAGCCGACGTCGAAAGTGGACCACGCCTCGCGCAT
          L  R  V  S  L  N  K  K  L  G  V  D  F  T  S  W  A  E  R  M
16441   TTCGGCCTGCAAGTCGGCGATGAAGCGCTCCGCATCGAGCGGCAGACTCAGCGCCCGGTA
          E  A  Q  L  D  A  I  F  R  E  A  D  L  P  L  S  L  A  R  Y
16501   GTAGTCCTCGCGGTTTTGCTCGAAGTCGGTCGGCAGATCGTCGTCGGGGTTGCGGTAGCG
          Y  D  E  R  N  Q  E  F  D  T  P  L  D  D  D  P  N  R  Y  R
16561   GTTTGCGCCGACCACCCATATCTCCTTGCAGCGCAGCCGCTCGGGCAGTGCTTCCAGCAC
          N  A  G  V  V  W  I  E  K  C  R  L  R  E  P  L  A  E  L  V
16621   CGCGATCTCATACGTGACGCGGTTGATGCGATCCCGACCGGCGGCATCCTTCTCCATCAC
          A  I  E  Y  T  V  R  N  I  R  D  R  G  A  A  D  K  E  M  V
```

```
16681   GGCTTTGCGCCATAGGCCGCGTACCACGCCATCGAGCGGCACCTCCACGTCGGCCGGGAA
          A  K  R  W  L  G  R  V  V  G  D  L  P  V  E  V  D  A  P  F
16741   GGTGTGCACCTTGGTGTCGGCAAAGCGCTTCACCAAGTCGAGTGCATCCATCACCGGACG
          T  H  V  K  T  D  A  F  R  K  V  L  D  L  A  D  M  V  P  R
                                 EcoRI
16801   ATGGTGGTCGTTGTTGGAGCGGAATTCCAGCGCGGCCAGTAAGGTCGGCACCATGCGGCG
          H  H  D  N  N  S  R  F  E  L  A  A  L  L  T  P  V  M  R  R
16861   GTAGTGCCCCTGGTACGAATTGCGGATCACCGTGCGCAGCGTGATGCGGTAGGTCGCCAG
          Y  H  G  Q  Y  S  N  R  I  V  T  R  L  T  I  R  Y  T  A  L
16921   TGGCTTCCACTCCTTGACCAGATCGCGCAGCGTCTGCTCGCCGACCACCGGAAACACTAC
          P  K  W  E  K  V  L  D  R  L  T  Q  E  G  V  V  P  F  V  V
16981   ATCGCGCACCACACCATCCGGCTGGGCCAGCGTAGCGTCGGCCAGGTTGAACAGCAGGTT
          D  R  V  V  G  D  P  Q  A  L  T  A  D  A  L  N  F  L  L  N
17041   CTGCTTGCCGGACACGCGCTTGAGGTCTTCCAGCAGTTTGCGCTCGACCTTGCGCTCAGC
          Q  K  G  S  V  R  K  L  D  E  L  L  K  R  E  V  K  R  E  A
17101   ACGCGCGCCGATTTGGTGAATGGTTTCGATCAACAAGTCCACCAGGTCGTCGGTCAGGCT
          R  A  G  I  Q  H  I  T  E  I  L  L  D  V  L  D  D  T  L  S
17161   GCGGGCGCGCAGGTAAACGAAAGCGGCCAGCCAAGTGAGGCGCGCTGCATCAGGATGCCG
          R  A  R  L  Y  V  F  A  A  L  W  T  L  R  A  A  D  P  H  R
17221   ACGAAGATCGCGGGGAACCTCGACCGACACACGCTGGCGACAGCGCCCCAAGTCACGCGG
          R  L  D  R  P  V  E  V  S  V  R  Q  R  C  R  G  L  D  R  P
                                          ClaI
17281   CGAAGTCCGGTCGAACAGATCGACAGGCAGATCGATCCCCTGGATCAGTTCCCATCGCCA
          S  T  R  D  F  L  D  V  P  L  D  I  G  Q  I  L  E  W  R  W
                SphI
17341   ACTCGTCCTGCATGCTGGCAAGGCTGGGGCGGCCGGGACTGCCGCGTAGTTTTCAGCAGA
          S  T  R  C  A  P  L  A  P  A  A  P  V  A  A  Y  N  E  A  S
17401   ACGGCCGGCGCGCTGCCTGCGGCTGTCGCCTTGAGCATCTTCAACGGCGCTTTCCCCGTC
          R  G  A  R  Q  R  R  S  D  G  Q  A  D  E  V  A  S  E  G  D
17461   GCCCGGTTTTTCGGGGCGCACAAGGGCATCCAAGCGTTCGCGGGGTCGCCGGCGGCAGCC
          G  P  K  E  P  R  V  L  A  D  L  R  E  R  P  R  R  R  C  G
                                                          SphI
17521   TTTATGGACTTCGTTATGAAAGCGGTCTTCGTGAGCACGCAACGGCTGCGCATGCGCTCC
          K  H  V  E  N  H  F  R  D  E  H  A  R  L  P  Q  A  H  A  G
17581   CATCCGGTCTGGTGTCGGAGGTTCAATGGCGAGTTCGCGGCAACGTCCTTCCAGCCGTTC
          M  R  D  P  T  P  P  E  I  A  L  E  R  C  R  G  E  L  R  E
17641   GATCATCGGCTCAATGTCACCGCCGACTTCTCCGGCAACATGATCGCGCAACCACTCCGT
          I  M  P  E  I  D  G  G  V  E  G  A  V  H  D  R  L  W  E  T
17701   CAGCATCTCGGCATCGGCCATCGTCGCTTCGTGAAAACCAAAACGCACACGGATTTCGCC
          L  M  E  A  D  A  M   orf90 start
17761   GCGCAACCGCTCGGCAGTGCGGCCCGTAAGGGAAGTTCGCCGTCAATGGGCGCGGGTGCG
                                          min stop    *  H  A  R  T  R
17821   TCTAGTTGTTTGACGAGCGCGGGCTATGCCCTGTGCCTCGACTTCGGTTTCGTCGCGCGG
          R  T  T  Q  R  A  R  A  I  G  Q  A  E  V  E  T  E  D  R  P
17881   GAAACGGGCGCGCTCGCGGAAGAAGGTCAGCAGGAGAGCAAAGCCCAGCCGGGTCGCCCT
          F  R  A  R  E  R  F  F  T  L  L  L  A  F  G  L  R  T  A  R
17941   GTTCTTGGTCATCACCAGCTCGCGTTCGGTCGGTGCCAACAGCCACTGATCGTCCGCGTT
          N  K  T  M  V  L  E  R  E  T  P  A  L  L  W  Q  D  D  A  N
18001   GCCGTCCGCCATGCCGCTACACTCCTGGTTGCGATGGCTCGGGCAGCCGGCCCGCGCCAG
          G  D  A  M  G  S  C  E  Q  N  R  H  S  P  C  G  A  R  A  L
18061   GGTTTCGATCAAGGTGGTGCGCTTGACGCCAAAGGTTCGGCACACCGCCGCCTTGGACAT
          T  E  I  L  T  T  R  K  V  G  F  T  R  C  V  A  A  K  S  M
```

```
18121    GCCGCCATCGAGCGCGGCAGTGATGGCGTCCAGCTTCTCGCCGGTGATCGCTTGCGGGTC
           G  G  D  L  A  A  T  I  A  D  L  K  E  G  T  I  A  Q  P  D
18181    GGCCGCCGATCCGGCGCGTTTGCGGGGCTGCGCCAGGCCCGCAACGACACGCTCCTGGAT
           A  A  S  G  A  R  K  R  P  Q  A  L  G  A  V  V  R  E  Q  I
18241    CAAGGCGCGCTCGTACTGTGCGAGCGCGCCGAACACCTGGAACAGGAACTCGCCCGAGGG
           L  A  R  E  Y  Q  A  L  A  G  F  V  Q  F  L  F  E  G  S  P
18301    TGTCGTGGTGTCCAGGCTCTCGGTCAGCGAGCGAAACGCCACCTGCTTGTCCTTGAGCGA
           T  T  T  D  L  S  E  T  L  S  R  F  A  V  Q  K  D  K  L  S
18361    GGTCACGATGGCAAGCAGGTGCGACAGCGAAAGACCGAGCCGGTCGAGCTTCCACACGAC
           T  V  I  A  L  L  H  S  L  S  L  G  L  R  D  L  K  W  V  V
                                  EcoRI
18421    CAGCACGTCGCCAGCGCAGACGAATTCGGAGGTTGCCCAAACGGCGCGGTCATCCTTCGC
           L  V  D  G  A  C  V  F  E  S  T  A  W  V  A  R  D  D  K  A
18481    GCCGGAAGCGTGATCCTCGAACAGGTGACGCGGATCAACGCCGGCGGGCAGCGCATCGCG
           G  S  A  H  D  E  F  L  H  R  P  D  V  G  A  P  L  A  D  R
18541    CTGCAAGTCCGTGCTCTGCCGGTCGGAATCCGACGACACGCGCATGTAGCCTACCAACAT
           Q  L  D  T  S  Q  R  D  S  D  S  S  V  R  M  Y  G  V  L  M
18601    AAACGGATAACCATCAGAACAAGGTTTCCGTATGATAACGCATAGCGACAGAGTTTTCCG
                                        SphI                  min start
18661    CACAATTTTGAGGCTACCCAGAGGTTGGTGCATGCGACGGTTGCAGATGTGTCGCTAAAC
18721    AAGTGTTTTGCGACATCATCCTTTCAGGTGTTGCACTTTCCGCCTTACCTGAACAGGTAT
18781    GAATTTAGGAGCGCGCCTCCCTGAAAATACGGATGCTGCGCAAGTGGCTGAGCAAATGAG
18841    CGCGTCCGGCATAGCAGGAACGCATCGACTAGGTACGAGACAAGCCACGTTGATCGGCTG
18901    ATGGAATCTTCGGCGGTTCCGGCTTGCTCCAAGGTGGCGGCAGGTGACACGACTGCTTGT
18961    TCTACCCCAGGTGGGTTCCTGAACTTTATGGGCGCCTGGCCAGGCGCGAATTGGTGGGCA
19021    GGATTTCTTCCCAGAGACGGCGCTGATCTTTGTCCAGAATGAAGCGCATGGTCGTGATGC
19081    GTGGATTCTCCAGGGCCGGCCGGAGCAAGGCGTCAAACAAGGCCGGGGGCTTGTACATGG
19141    AAAGGCAGACATTGAACAGATCATTTCCGCTGGCGTGGACCAAGAACTCCGTGCTGACTT
           BamHI
19201    GCGGATCCGGTTCGGCCCGATGAGTACGGCATCGGGAGGCAGGAGGCCCACTTCGATTTC
19261    ACGCAGGCTTACCTGGACGGCCGCAAGTCGGCCCTCCAAACGCTCCATGGCCCCTTCCCG
19321    GCGCATATCCCGTAGAAAGAGCAGGGCGATGAGCGCAACAGTCAGGGTAAGCAATACCTC
19381    ACCCTGAATCAGGTGCAGGAGTTCCATGACGATGGCCGTGACCGCCGCGATAACGCCTGC
19441    GATGGCATCCCATTCCAAATTGAGCAGGCGTTTCGTATGCACCATTACCCCTCCATGATG
19501    TCGGCTGCCCCCTCCGTAAGGCCTAGATGCATCAGTCCCGCGATCCTTCCTCAGCGAGAC
19561    GCAGGATGCCGCCCTTATGATAGGCCAACATGGCGGCTTTGTCGCTTTGGATGGCGTATT
19621    GCGGGTCTTCCGGGCTCGCATGGTGGGTATAGCCGTGGACCGGAAAATCCTGCTCGTGGA
                                              EcoRI
19681    TGGCAACGATCCGCCCCGTTACCCTGCCTGCCTCCGAATTCCAGGTAACGTGATCGCCCA
19741    CGGCAAAGCGTGGTGCCGGCGGCTTTTTCATGAGGTGTCCTCAGCGCCAA
```

Appendix B

Media, buffers and solutions

B.1 Media

All media, buffers and solutions were sterilized by autoclaving at 121°C for 20 min unless otherwise stated. Heat labile substances were sterilized by filtration through 0.22 μm membrane filters (Millipore).

B.1.1 Luria Broth (1000 ml)

Bactotryptone	10 g
Yeast Extract (Difco)	10 g
NaCl (Saarchem)	5 g
Distilled water	to 1000 ml

Solid media contained 1.5% (w/v) agar

B.1.2 M9 minimal glycerol media (1000 ml)

Salts solution :

K_2HPO_4	10.5 g
KH_2PO_4	4.5 g
Na_3-citrate	2.5 g
$(NH_4SO_4)_2SO_4$	1 g
Distilled water to 200 ml	
$MgSO_4.7H_2O$	0.2 g
Distilled water to 10 ml	
Glycerol (v/v)	0.5 %
Distilled water to 10 ml	
Agar (Oxoid No. 1)	15 g
Distilled water to 780 ml	

Each solution was autoclaved separately, and combined together with 1 ml of a sterilised solution of thiamine (0.005 g/ml stock).

B.1.3 9K media (Silverman and Lundgren, 1959)

Solution A :

$(NH_4)_2SO_4$	3 g
KCl	0.1 g
K_2HPO_4	0.5 g
$MgSO_4.7H_2O$	0.5 g
$Ca(NO_3)_2$	0.01 g
Distilled water to 500 ml	

Solution B :

$FeSO_4.7H_2O$	45 g
Distilled water to 500 ml	

Solution A pH was adjusted to 1.9 and autoclaved, allowed to cool and mixed aseptically to filter sterilized solution B (pH 1.1). After mixing, the pH was adjusted to 1.7 with H_2SO_4.

B.1.4 Tetrathionate media (Powles et al. 1995)

Mineral salt solution

$(NH_4)_2SO_4$	3.0 g/l
KCl	0.1 g/l
K_2HPO_4	0.5 g/l
$Ca(NO_3)_2$	0.01 g/l

The pH was adjusted to 2.5 with H_2SO_4 and autoclaved.

Trace elements solution

$FeCl_3.6H_2O$	11 mg/l
$CuSO_4.5H_2O$	0.5 mg/l
HBO_3	2 mg/l
$Na_2MoO_4.2H_2O$	0.8 mg/l
$CoCl_2.6H_2O$	0.6 mg/l
$ZnSO_4.7H_2O$	0.9 mg/l

Trace element solution (1ml) was added to 100 ml mineral salts solution. To this was added 50 mM $K_2S_4O_6$ which had been filter sterilised. The final pH was adjusted to 2.5.

B.2 Media additives

B.2.1 Antibiotics

Antibiotic stock solutions were as follows

Ampicillin (Amp) in water	100 mg/ml
Chloramphenicol (Cm) in ethanol	20 mg/ml
Kanamycin (Km) in water	25 mg/ml
Erythromycin (Er) in ethanol	10 mg/ml

All antibiotics were filter sterilized and stored at -20°C.

B.2.2 IPTG (isopropyl-β-D-thio-galactopyranoside)

IPTG (100 mM)	23.4 mg
Distilled water	1 ml

The solution was stored in aliquots at -70°C

B.2.3 X-Gal (5-bromo-4-chloro-3-indolyl-β-galactoside)

X-Gal (2% w/v)	0.2 g
Dimethylformamide	10 ml

The solution was stored in aliquots at -70°C

B.3 Buffers and solutions

B.3.1 DNA loading buffer (6x)

Bromophenol blue	0.25 g
Sucrose	40 g
Distilled water to 100 ml	

B.3.2 EDTA (Ethylene diamine tetracetic acid, 0.5 M pH8)

$EDTA._2H_2O$	168.1 g/ml
Distilled water to 1000 ml	
pH was adjusted to 8 with NaOH (10 N)	

B.3.3 Ethidium bromide

A solution of 10 mg/ml (2,7-diamino-10-ethyl-9-phenyl-phenanthridinium bromide) was made in distilled water and stored in the dark.

B.3.4 Exonuclease III shortening solutions (Henikoff, 1987)

B.3.4.1 ExoIII buffer

Tris/HCl (1M, pH 8)	660 µl
$MgCl_2$ (0.1M)	66.4 µl
Distilled water to 9.27 ml	

B.3.4.2 Klenow mix

Tris/HCl (0.1M, pH8)	3 µl
$MgCl_2$ (1M)	6 µl
Distilled water	20 µl

B.3.4.3 S1 buffer (10x)

KOAc (3M)	1.1 ml
NaCl (5M)	5 ml
Glycerol	5 ml
$ZnSO_4$	30 mg

B.3.4.4 S1 stop

Tris	0.3 M
EDTA (pH8)	0.05 M

B.3.5 Hybridisation solution (100 ml)

20x SSC	25 ml
Milk powder (Elite)	5 g
10% N-lauroylsarcosine	1 ml
10 % SDS	0.2 ml
Distilled water to 100 ml	

B.3.6 Na-phosphate running buffer (1M, pH7)

Solution A: $NaH_2PO_4.H_2O$ (2M) 276 g/l
Solution B: $Na_2HPO_4.12H_2O$ (2M) 71.6 g/l
Solution A (195 ml), solution B (305 ml), and distilled water (500 ml) were combined, and the pH adjusted to 7.

B.3.7 20x SSC

NaCl (3M)	175.3 g
Sodium citrate (0.3 M)	88.2 g

Distilled water to 1000 ml
Adjusted pH to 7 with NaOH (10 N). Autoclaved.

B.3.8 TE (100x)

Tris	121 g
EDTA (0.5 M, pH8)	200 ml

Distilled water to 1000 ml
Adjusted pH to 8, autoclaved, and diluted with sterile water before use.

B.3.9 10x TBE

Tris	108 g
Boric acid	55 g
EDTA (0.5M, pH8)	40 ml

Distilled water to 1000 ml

B.3.10 20x PFGE buffer

Tris	24.2 g
Glacial acetic acid	5 ml
EDTA	2.9 g

Deionised water to 1000 ml

B.3.11 ES solution

Na-lauroyl sarcosine (Saarchem)	1 g
EDTA (0.5M, pH8)	16.8 g

Distilled water to 100 ml

B.3.12 SET Buffer (1x)

Sucrose (25% w/v)	50 g
EDTA (0.5 M)	0.8 ml
Tris-Cl (1 M, pH8)	10 ml

Appendix C

General techniques

General techniques such as agarose gel electrophoresis, and the transformation and transduction of *E. coli* were carried out as stated in Sambrook et al. (1989).

C.1 DNA isolation and restriction endonuclease digestion

Small scale (5 ml overnight cultures) and large scale (200 ml overnight cultures) methods of isolating *E. coli* plasmid DNA were done according to the method of Ish-Horowicz and Burke (1981). Restriction digests of both chromosomal and plasmid DNA were conducted using restriction enzymes (Boehringer Mannheim) in the appropriate 1x restriction buffer at the recommended temperature. Where necessary DNA was precipitated with 10% (v/v) Na-citrate (1M, pH 5.2), 1 volume of isopropanol (Merk) and centrifuged for 10 min at 14000 g.

C.2 Cloning and ligation protocols

Digested DNA fragments to be subcloned were electrophoresed in low temperature gelling agarose (Hispanagar) containing 1xTBE buffer. Fragments were stained briefly in ethidium bromide and excised under long UV (365 nm) light. Gel slices were melted at 65°C for 5 minutes and combined according to the type of ligation required. All ligations were done at room temperature with 1 unit of ligase (Boehringer Mannheim) in the supplied 1x ligation buffer. Cloning protocols involving different restriction endonucleases which produced at least one overhanging fragment were carried out at a vector to insert ratio of 1:1, with 1 pmol/ml final total DNA concentration and in a 100 µl reaction volume. Where identical, compatible or blunt ended restriction endonucleases were used, ligations were adjusted to 20 pmol/ml final total DNA concentration and a 10 µl reaction volume.

C.3 DNA sequencing

Sequencing of double stranded plasmid DNA templates was carried out using the Thermosequanase fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP (Amersham. Corp., Ohio) with the modifications of Dr H. Voss (EMBL, Heidelberg, Germany). Templates were labelled on a Hybaid thermocycler using the dideoxynucleotide chain termination method of Sanger et al. (1977). Templates were run on a Pharmacia Alf Express automated sequencer for 12 h at 60 W and 25 mA at 55°C. All preparative steps were carried out according to the manufacturers standard operating procedures.

C.4 *In vitro* transcription-translation reactions

The *in vitro* transcription-translation of plasmid DNA using an *E. coli* derived kit (Promega Corp., Madison, Wisconsin) consisted of the following components.

3 μg DNA in 4.7 μl distilled water
5 μl premix
1.7 μl amino acids without methionine
1 μl S35 methionine
5 μl S30 fraction

The reaction was allowed to proceed at 37°C for 1 h and was terminated by the addition of an equal volume (17.4 μl) of stop dye. The reaction was placed at -20°C until required for use. Immediately prior to loading, samples were boiled for 2 min and placed on ice. SDS-PAGE was carried out according to the method of Laemmli (1970) in a 15% polyacrylamide resolving gel, with a 7.5% stacking gel.

C.5 T4 DNA polymerase "filling in" reactions

Where necessary 3' or 5' protruding DNA fragment ends were "filled in" using T4 DNA polymerase (Boehringer Mannheim). Typically 200 - 800 ng of DNA was resuspended in 12 µl distilled water, 2 µl BSA, 2 µl dNTP's, 2 µl A restriction buffer (Boehringer Mannheim) and 2 µl (2 units) T4 DNA polymerase. The reaction was allowed to proceed at 37°C for 15 min and was terminated by heating the sample at 65°C for 5 min. The sample was then precipitated (Appendix C.1) and used as required.

C.6 Computer analysis

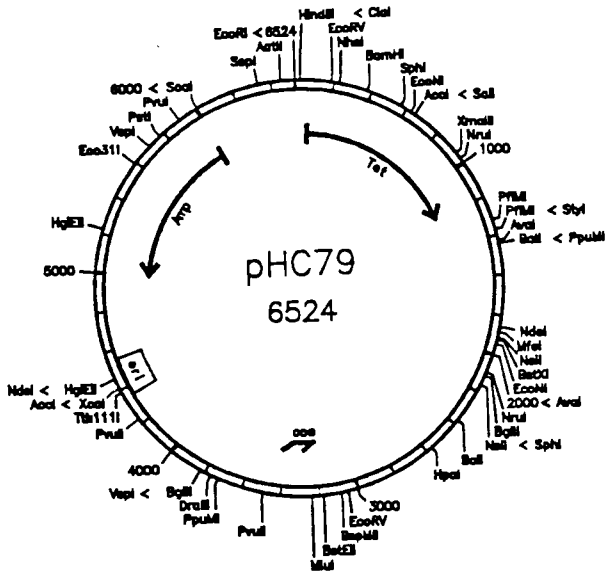
Computer analysis of nucleic acid sequence as well as deduced protein sequence was carried out using the GCG-Wisconsin ver 9 suite of programs (Devereaux et al. 1984). Codon preferences and third position bias were graphically represented using the CODONPREFERENCE program and the *T. ferrooxidans* chromosomal codon preference table. Multiple sequence alignments generated using the neighbor-joining method by the Clustal W program for the Vax / Vms. These were imported into Genedoc v 1.1.004 (Nicholas and Nicholas, 1996) for boxing and shading of conserved regions. Conserved regions were defined as those having more than 66% of the amino acids at a particular alignment site, conserved according to the Blosum 62 comparison matrix. Dendograms were compiled using Treeview v1.3 (Page, 1996). The calculation of GC mol% and the prediction of possible secondary structures in the nucleic acid sequence was done using DnaMAN v 2.2 (Lynnon Biosoft, 1995). BLAST searches (Altschul et al. 1990) were conducted from within GCG with the exception of BLAST with BEAUTY post-processing searches (Worley et al. 1995) which were done on the World-Wide-Web (http://www.ncbi.nlm.nih.gov) using Netscape 3.0.

Appendix D Amino acid codes

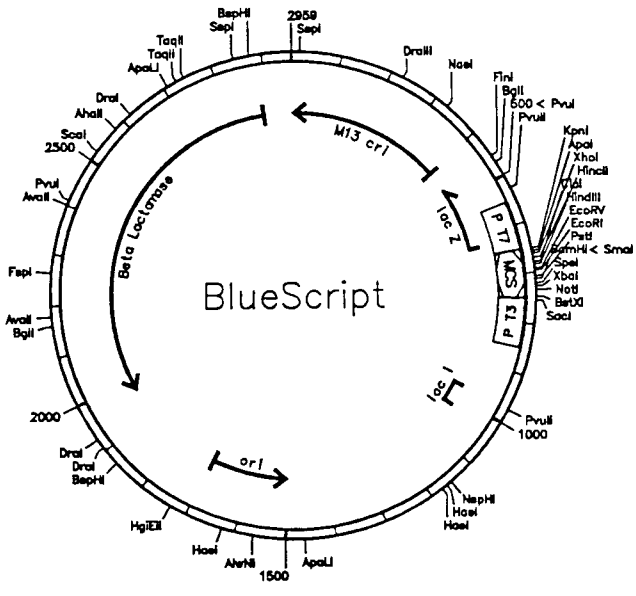
Amino acid	Codes	
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Appendix E

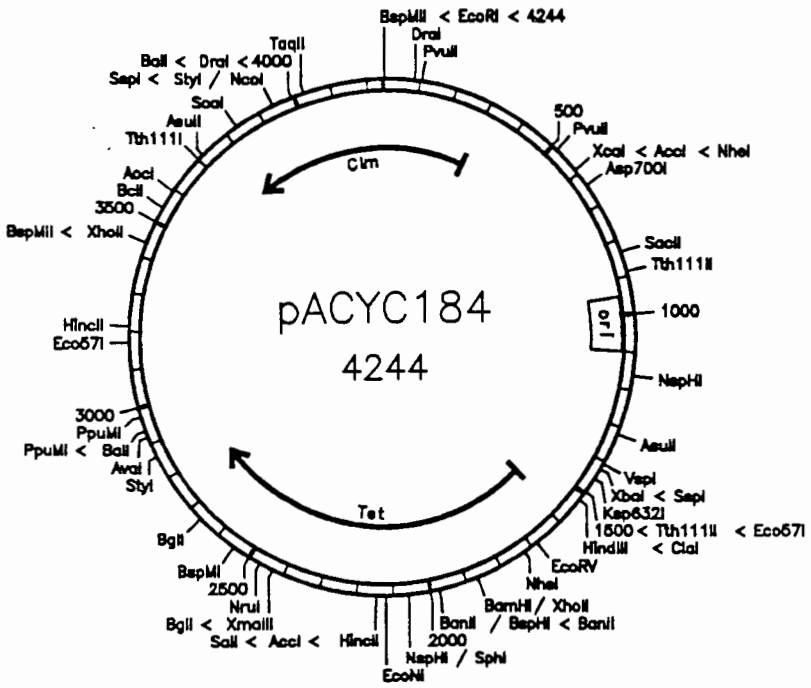
Physical maps of vectors and transposons



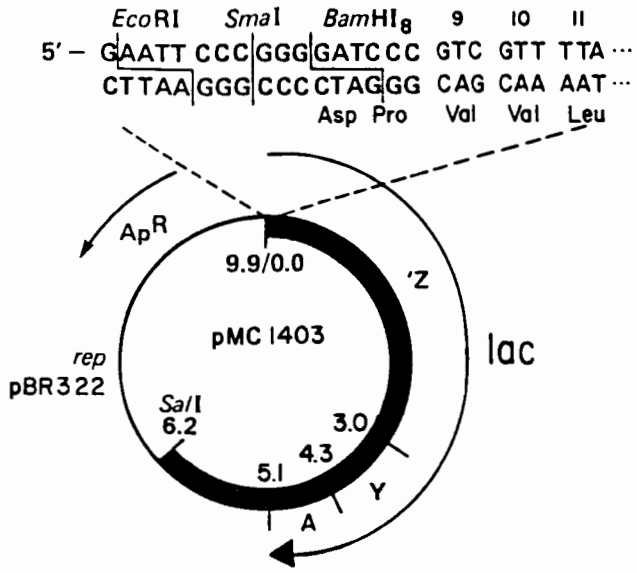
pHC79 cosmid cloning vector (Hohn and Collins, 1980)



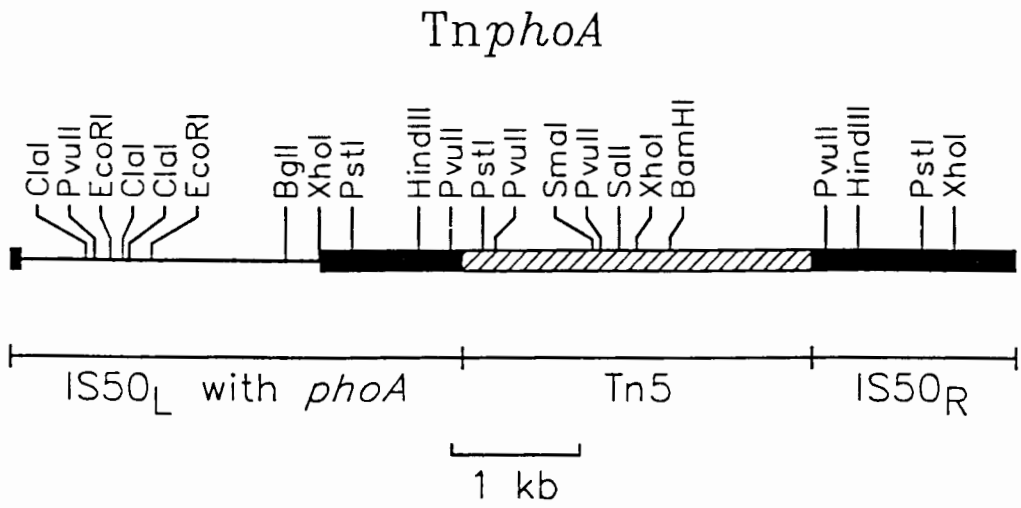
Map of Bluescript KS+ (pBS-KS+). Bluescript SK+ differs in the orientation of the MCS polylinker between the KpnI and SacI sites. Stratagene, San Diego, CA, USA.



pACYC184 (Chang and Cohen, 1978)



pMC 1403 (Casadaban et al. 1983)



Tn*PhoA* (Manoil and Beckwith, 1985)

REFERENCES

Allardet-Servant A., Michaux-Charachon S., Jumas-Bilak E., Karayan L. and Ramuz M. (1993) Presence of one linear and one circular chromosome in *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. 175: 7869-7874.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215: 403-410

Anonymous (1996) TIGR's genomic tale is growing. ASM News 62(3): 126-127.

Anthamatten D., Scherb B. and Hennecke H. (1992) Characterization of a *fixLJ*-regulated *Bradyrhizobium japonicum* gene sharing similarity with the *Escherichia coli fnr* and *Rhizobium meliloti fixK* genes. J. Bacteriol. 174: 2111-2120.

Arai H, Igarashi Y, Kodama T (1995) Expression of the *nir* and *nor* genes for denitrifcation of *Pseudomonas aeruginosa* requires a novel CRP/FNR-related transcriptional regulator, DNR, in addition to ANR. FEBS Lett 371: 73-76

Arthur M., Molinas C., Depardieu F. and Courvalin P. (1993) Characterisation of Tn*1546*, a Tn*3*-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. J. Bacteriol. 175: 117-127.

Asturias J.A., Diaz E. and Timmis K.N. (1995) The evolutionary relationship of biphenyl dioxygenase from Gram-positive *Rhodococcus globerulus* P6 to multicomponent dioxygenases from Gram-negative bacteria. Gene 156: 11-18.

Axcell B.C. and Geary P.J. (1975) Purification and some properties of a soluble benzene-oxidising system from a strain of *Pseudomonas*. Biochem J. 146: 173-183.

Bannan J.D., Moran M.J., MacInnes J.I., Soltes G.A. and Friedman R.L. (1993) Cloning and characterization of *btr*, a *Bordetella pertussis* gene encoding a FNR-like transcriptional regulator. J. Bacteriol. 175: 7228-7235.

Barros M.E., Rawlings D.E. and Woods D.R. (1985) Cloning and expression of the *Thiobacillus ferrooxidans* glutamine synthetase gene in *Escherichia coli*. J. Bacteriol. 164: 1386-1389.

Bartosik D., Baj J. and Wlodarczyk M. (1995) Construction and preliminary characterisation of mini-derivatives of a large (107-kb) cryptic plasmid of the sulfur bacterium *Thiobacillus versutus*. FEMS Microbiol. Lett. 129: 169-174.

Batut J., Daveran-Mingot M.L., David M., Jacobs J., Garnerone A.M. and Kahn D. (1989) *fixK*, a gene homolous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. EMBO J. 8: 1279-1286.

Bednarska M., Jagusztyn-Krynicka E.K., Popowski J. and Wlodarczyk M. (1983) Extrachromosomal DNA in *Thiobacillus A2*. FEMS Microbiol. Lett. 16: 183-185.

Beinert H. (1990) Recent developments in the field of iron-sulphur proteins. FASEB J. 2483-2491.

Beinert H. And Kennedy MC (1992) Engineering of protein bound iron-sulphur clusters. Eur. J. Biochem. 186: 5-15.

Berger D.K., Woods D.R. and Rawlings D.E. (1990) Complementation of *Escherichia coli* σ54 (NtrA)-dependent formate hydrogenlyase activity by a cloned *Thiobacillus ferrooxidans ntrA* gene. J. Bacteriol. 172: 4399 - 4406.

Beverley S.M. (1988) Characterisation of the 'unusual' mobility of large circular DNA's in pulsed field-gradient electrophoresis. Nucleic Acids Res. 16: 925-939.

Birkmann A, Sawers RG, Bock A (1987) Involvement of the *ntr*A gene product in the anaerobic metabolism of *Escherichia coli*. Mol. Gen. Genet. 210: 535-542

Blake R.C. and Shute E.A. (1994) Respiratory enzymes of *Thiobacillus ferrooxidans*. Kinetic properties of an acid stable Iron:Rusticyanin oxidoreductase. Biochem. 33: 9220-9228.

Blake R.C., Shute E.A., Greenwood M.M., Spencer G.H. and Ingledew W.J. (1993) Enzymes of aerobic respiration on iron. FEMS Microbiol. Rev. 11: 9-18.

Blake R.C., Shute E.A. and Howard G.T. (1994) Solubilization of minerals by bacteria: Electrophoretic mobility of *Thiobacillus ferrooxidans* in the presence of iron, pyrite, and sulphur. Appl. Environ. Microbiol. 60: 3349-3357.

Blake R.C., White K.J. and Shute E.A. (1988) Electron transfer from Fe(II) to rusticyanin is catalysed by an acid-stable cytochrome. in Biohydrometallurgy: Proceedings of the International symposium. Norris P.R. and Kelly D.P. (eds) pp 103-110.

Blaylock B.A. and Nason A. (1963) Electron transport systems of the chemoautotroph *Ferroobacillus ferrooxidans*. I. Cytochrome c containing iron oxidase. J. Biol. Chem. 238: 3453-3462.

Booth J.E. and Williams J.W. (1984) The isolation of a mercuric ion-reducing flavoprotein from *Thiobacillus ferrooxidans*. J. Gen. Microbiol. 130: 725-730.

Britz M.L. and Wilkinson R.G. (1979) Isolation and properties of metronidazole-resistant mutants of *Bacteroides fragilis*. Antimic. Agents Chemother. 16: 19-27.

Brown L.D., Dennehy M.E. and Rawlings D.E. (1994) The F1 genes of the F1F0 ATP synthase from the acidophilic bacterium *Thiobacillus ferrooxidans* complement *Escherichia coli* F1 *unc* mutants. FEMS Microbiol. Lett. 122: 19-25.

Brown N.L., Ford S.J., Pridmore R.D. and Fritzinger D.C. (1983) Nucleotide sequence of a gene from the *Pseudomonas* transposon Tn*501* encoding mercuric reductase. Biochemistry 22: 4089-4095.

Cadiz R., Gaete L., Jedlicki E., Yates J., Holmes D.S. and Orellana O. (1994) Transposition of IST*2* in *Thiobacillus ferrooxidans*. Molecular Microbiology 12(1): 165-170.

Cardy D.L.N., Laidler V., Salmond G.P.C. and Murrell J.C. (1991) The methane monooxygenase gene cluster of *Methylosinus trichosporium*: cloning and sequencing of the *mmoC* gene. Arch. Microbiol. 156: 477-483.

Carle G.F. and Olson M.V. (1984) Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. Nucleic Acids Res. 12: 5647-5664.

Casadaban M.J., Martinez-Arias A., Shapira S.K. and Chou J. (1983) B-Galactosidase gene fusions for analyzing gene expression in *Escherichia coli* and Yeast. Meth. Enzymol. 100:293-308.

Casimiro D.R., Toy-Palmer A., Blake R.C. and Dyson H.J. (1995) Gene synthesis, high level expression, and mutagenesis of *Thiobacillus ferrooxidans* rusticyanin: His 85 is a ligand to the blue copper centre. Biochem. 34: 6640-6648.

Cavalcanti S.M.B., Domingues R.M.C.P., Ferreira M.C.de S. and Filho F.C.S. (1991) Effect of metronidazole on surface properties of *Bacteroides fragilis*. J. Antimicrob. Chemother. 28: 819-826.

Cavazza C. and Bruschi M. (1995) Iron oxidation by *Thiobacillus ferrooxidans*: Characterization of two electron transfer proteins. in: Biohydrometallurgical processing. Vargas T., Jerez C.A., Wiertz J.V. and Toledo H. (eds), University of Chile, Santiago pp 97-108.

Cavazza C., Guigliarelli B., Bertrand P. and Bruschi M. (1995) Biochemical and EPR characterization of a high potential iron-sulphur protein in *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. 130: 193-200.

Chakravarty L., Zupancic T.J., Baker B., Kittle J.D., Fry I.J. and Tuovinen O.H. (1995) Characterisation of the pTFI91-family replicon of *Thiobacillus ferrooxidans* plasmids. Can. J. Microbiol. 41: 354-365.

Chang A.C.Y. and Cohen S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. 134: 1141-1166.

Chopra I. (1975) Mechanism of plasmid mediated resistance to cadmium in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 7: 8-14

Chrystal E.J.T., Koch R.L., McLafferty M.A. and Goldman P. (1980) Relationship between metronidazole metabolism and bacteriocidal activity. Antimic. Agents. Chemother. 18: 566 - 573.

Church D.L. and Laishley E.J. (1995) Reduction of metronidazole by hydrogenase from Clostridia. Anaerobe 1: 81-92.

Church D.L., Rabin H.R. and Laishley E.J. (1990) Reduction of 2-,4- and 5-nitroimidazole drugs by hydrogenase 1 in *Clostridium pasteurianum*. J. Antimic. Chemother. 25: 15-23.

Clennel A-M., Johnston B. and Rawlings D.E. (1995) Structure and function of Tn*5467*, a Tn*21*-like transposon located on the *Thiobacillus ferrooxidans* broad-host-range plasmid pTF-FC2. Appl. Environ. Microbiol. 61: 4223-4229.

Colonna-Romano S., Arnold W., Schluter A., Boistard P., Puhler A. and Priefer U.B. (1990) An FNR-like protein encoded in *Rhizobium leguminosarum* biovar viciae shows structural and functional homology to *Rhizobium meliloti*. Mol. Gen. Genet. 223: 138-147.

Coucouvanis D., Kanatzidis M.G., Dunham W.R. and Hagen W.R. (1984) Oxidative transformation of the $[Fe_4S_4S_4X_4]^{2-}$ cubanes to the $[Fe_6S_6X_6]^{2-}$ 'prismane' clusters (X= Cl, Br). The crystal and molecular structure of $[(C_6H_5)_4P]_2Fe_6S_6Cl_6$ J. Am. Chem. Soc. 106: 7998-7999.

Couterier M., Bex F., Bergquist P.L. and Maas W.K. (1988) Identification and classification of bacterial plasmids. Microbiol. Rev. 52: 375-395.

Cox J.C. and Boxer D.H. (1978) The purification and some properties of rusticyanin, a blue copper protein, involved in iron (II) oxidation from *T. ferrooxidans*. Biochem. J. 174: 497-502.

Dachs G.U., Abratt V.R. and Woods D.R. (1995) Mode of action of metronidazole and a *Bacteroides fragilis metA* resistance gene in *Escherichia coli*. J. Antimicrob. Chemother. 35: 483-496.

Davidson M.S. Roy P. and Summers A.O. (1985) Transpositional mutagenesis of *Thiobacillus novellus* and *Thiobacillus versutus*. Appl. Environ. Microbiol. 49: 1436-1441

Davidson M.S. and Summers A.O. (1983) Wide-host-range plasmids function in the genus *Thiobacillus*. Appl. Environ. Microbiol. 46: 565-572.

Davison S.P., Santangelo J.D., Reid S.J. and Woods D.R. (1995) A *Clostridium acetobutylicum* regulator gene (*regA*) affecting amylase production in *Bacillus subtilis*. Microbiol. 141: 989 - 996.

De Crombrugghe B., Busby S. and Buc H. (1984) Cyclic AMP receptor: role in transcription activation. Science 224: 831-838.

De Siloniz M.I., Lorenzo P., Murua M. and Perera J. (1993) Characterisation of a new metal-mobilising *Thiobacillus* isolate. Arch. Microbiol. 159: 237-243.

Devereux J, Haeber P, Smithies O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395.

Dispenza M., Thomas C.T., Kim M-K., Perrotta J.A., Gibson J. and Harwood C.S. (1992) Anaerobic growth of *Rhodopseudomonas palustris* on 4-hydroxybenzoate is dependent on AadR, a member of the cyclic AMP receptor protein family of transcriptional regulators. J. Bacteriol. 174: 5803-5813.

Dispirito A.A and Tuovinen O.H. (1982a) Uranous ion oxidation and carbon dioxide fixation by *Thiobacillus ferrooxidans*. Arch. Microbiol. 133: 28-32.

Dispirito A.A and Tuovinen O.H. (1982b) Kinetics of uranous ion and ferrous iron oxidation by *Thiobacillus ferrooxidans*. Arch. Microbiol. 133: 33-37.

Dolan K.M. and Tuovinen O.H. (1986) Characterisation and mapping of plasmid DNA of *Thiobacillus ferrooxidans* by restriction endonuclease digestion. in: Fundamental and applied Biohydrometallurgy, Lawrence R.W., Branion R.M.R. and Ebner H.G. (eds) pp 457-469.

Dominy C.N., Deane S.M. and Rawlings D.E. (1997) A geographically widespread plasmid from *Thiobacillus ferrooxidans* has genes for ferredoxin-, FNR-, prismane- and NADH oxidoreductase-like proteins which are also present on the chromosome. Submitted for publication.

Donadio S. and Hutchison C.R. (1991) Cloning and characterisation of the *Saccharopolyspora erythraea fdxA* gene encoding ferredoxin. Gene 100: 231-235.

Dorrington R.A. and Rawlings D.E. (1989) Identification and sequence of the basic replication region of a broad-host-range plasmid isolated from *Thiobacillus ferrooxidans*. J. Bacteriol. 171: 2735-2739.

Dorrington R.A. and Rawlings D.E. (1990) Characterisation of the minimum replicon of the broad-host-range plasmid pTF-FC2 and similarity between pTF-FC2 and the IncQ plasmids. J. Bacteriol. 172: 5697-5705.

Dorrington R.A., Bardien S. and Rawlings D.E. (1991) The broad-host-range plasmid pTF-FC2 requires a primase-like protein for autonomous replication in *Escherichia coli*. Gene 108: 7-14.

Drobner E., Huber H. and Stetter K.O. (1990) *Thiobacillus ferrooxidans*, a facultative hydrogen oxidiser. Appl. Environ. Microbiol. 56: 2922-2923.

Drolet M. and Lau P.C.K. (1992) Mobilisation protein-DNA binding and divergent transcription at the transfer origin of the *Thiobacillus ferrooxidans* pTF1 plasmid. Mol. Microbiol. 6 : 1061 - 1071.

Drolet M., Zanga P. and Lau P.C.K. (1990) The mobilisation and origin of transfer regions of a *Thiobacillus ferrooxidans* plasmid: relatedness to plasmids RSF1010 and pSC101. Mol. Microbiol. 4: 1381 - 1391.

Edwards D.I. (1980) Mechanisms of selective toxicity of metronidazole and other nitroimidazole drugs. Br. J. Vener. Dis. 56: 285-290.

Edwards D.I. (1993) Nitroimidizole drugs - action and resistance mechanisms. II Mechanisms of resistance. J. Antimic. Chemother. 31: 201-210.

Edwards DI and Mathison GE (1970) The mode of action of metronidazole against *Trichomonas vaginalis*. J Gen Microbiol 63:297-302

Elbehti A. and Lemesle-Meunier D. (1996) Identification of membrane bound c-type cytochromes in an acidophilic ferrous ion oxidising bacterium *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. 136: 51 - 56.

Espejo R.T., Pizarro J., Jedlicki E., Orellana O. And Romero J. (1995) Bacterial populations in the bioleaching of copper as revealed by analysis of DNA obtained from leached ores and leaching solutions. in: Biohydrometallurgical processing. Vargas T., Jerez C.A., Wiertz J.V. and Toledo H. (eds), University of Chile, Santiago pp 1-8.

Ferdows M.S. and Barbour A.G. (1989) Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. Proc. Natl. Acad. Sci. USA 86: 5969-5973.

Fonstein M. and Haselkorn R. (1995) Physical mapping of bacterial genomes. J. Bacteriol. 177: 3361-3369.

Fonstein M., Koshy E.G., Nikolskaya T., Mourachov P. and Haselkorn R. (1995) Refinement of the high resolution physical map of *Rhodobacter capsulatus* and genome surveys using blots of the cosmid encyclopaedia. EMBO J. 14: 1827-1841.

Fry I.V., Lazaroff N. and Packer L. (1986) Sulphate dependent iron oxidation by *Thiobacillus ferrooxidans* : Characterization of a new EPR detectable transport component on the reducing side of rusticyanin. Arch. Biochem. Biophys. 246:2, 650-654.

Fujiwara T., Fukumori Y. and Yamanaka T. (1992) A novel terminal oxidase, cytochrome baa3 purified from aerobically grown *Pseudomonas aeruginosa*: it shows a clear difference between resting state and pulsed state. J. Biochem. 112: 290-298.

Fukumori Y., Yano T., Sato A. and Yamanaka T. (1988) Fe(II)-oxidising enzyme purified from *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. 50: 169-172.

Fukuyama K., Hase T., Matsumoto S., Tsukihara T. and Katsube Y. (1980) Structure of *S. platensis* [2Fe-2S] ferredoxin and evolution of the chloroplast-type ferredoxins. Nature 286: 522-524.

Fukuyama K., Matsubara H., Tsukihara T. and Katsube Y. (1989) Structure of [4Fe-4S] ferredoxin from *Bacillus thermoproteolyticus* refined at 2.3A resolution. J. Mol. Biol. 210: 383-398.

Galas D.J. and Chandler M. (1989) Bacterial insertions sequences in: Mobile DNA, Berg D.E. and Howe M.M. (eds) pp 109 - 162.

Gallie D.R. and Kado C.I. (1987) *Agrobacterium tumefaciens* pTAR *parA* promoter region involved in autoregulation, incompatibility and plasmid partitioning. J. Mol. Biol. 193: 465 - 478.

Gardiner K. and Patterson D. (1988) Transverse alternating electrophoresis. Nature 331: 371-372.

Garvey E.P. and Santi D.V. (1986) Stable amplified DNA in drug-resistant *Leishmania* exists as extrachromosomal circles. Science 233: 535-540.

Ghosh D., O'Donnell S., Furey W., Robbins A.H. and Stout C.D. (1982) Iron-sulphur clusters and protein structure of *Azotobacter* ferredoxin at 2.0 Å resolution. J. Mol. Biol. 158: 73-109.

Gibson D.T. and Subramanian V. (1984) Microbial degradation of aromatic hydrocarbons. In: Microbial degradation of organic compounds. Gibson D.T. (ed) pp181-252. New York: Marcel Dekker.

Glasgow A.C., Hughes K.T. and Simon M.I. (1989) Bacterial DNA inversion systems in: Mobile DNA, Berg D.E. and Howe M.M. (eds) pp 637 - 660.

Green J., Sharrocks A.D., Green B., Geisow M. and Guest J.R. (1993) Properties of FNR proteins substituted at each of the five cysteine residues. Mol. Microbiol. 8: 61 - 68.

Grossmann J.G., Ingledew W.J., Harvey I., Strange R.W. and Hasnain S.S. (1995) X-ray absorption studies and homology modeling define the structural features that specify the nature of the copper site in rusticyanin. Biochem. 34: 8406-8414.

Gunsalus R.P. (1992) Control of electron flow in *Escherichia coli*: Coordinated transcription of respiratory pathway genes. J. Bacteriol. 174: 7069 - 7074.

Haggoud A., Trinh S., Moumni M. and Reysett G. (1995) Genetic analysis of the minimal replicon of plasmid pIP417 and comparison with the other encoding 5-nitroimidazole resistance plasmids from *Bacteroides spp.* Plasmid 34: 132 - 143.

Haigler B.E. and Gibson D.T. (1990) Purification and properties of ferredoxin NAP, a component of napthalene dioxygenase from *Pseudomonas sp.* strain NCIB9816. J. Bacteriol. 172: 465-468.

Hall J.F., Hasnain S.S. and Ingledew W.J. (1996) The structural gene for rusticyanin from *Thiobacillus ferrooxidans*: cloning and sequencing of the rusticyanin gene. FEMS Microbiol. Lett. 137: 85 - 89.

Harayama S. and Rekik M. (1990) The meta cleavage operon of TOL degradative plasmid pWWO comprises 13 genes. Mol. Gen. Genet. 221: 113-120.

Harayama S., Rekik M., Bairoch A., Neidle E. and Ornston L.N. (1991) Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal benABC and *Pseudomonas putida* TOL pWWO plasmid xylXYZ, genes encoding benzoate dioxygenases. J. Bacteriol. 173: 7540-7548.

Harayama S. and Timmis K.N. (1989) Catabolism of aromatic hydrocarbons by *Pseudomonas*. In: Genetics of bacterial diversity. Hopwood D.A. and Chater K.F. (eds) pp151-174. London. Academic.

Harley and Reynolds (1987) Analysis of *E. coli* promoter sequences. Nucleic Acids Res. 15: 2343-2361.

Harrison A.P. (1982) Genomic and physiological diversity amongst strains of *Thiobacillus ferrooxidans*, and genomic comparison with *Thiobacillus thiooxidans*. Arch. Microbiol. 131: 68-76.

Harrison A.P. (1984) The acidophilic thiobacilli and other acidophilic bacteria that share their habitat. Ann. Rev. Microbiol. 38: 265 - 292.

Hase T. , Wakabayashi S, Matsubara H (1978) *Pseudomonas ovalis* ferredoxin: similarity to *Azotobacter* and *Chromatium* ferredoxins. FEBS Lett. 91: 315 - 319.

Hattori T., Takahashi K., Nakanishi T., Ohta H., Fukui K., Taniguchi S. and Takigawa M. (1996) Novel FNR homologues identified in four representative oral facultative anaerobes: *Capnocytophaga ochracea, Capnocytophagna sputigena, Haemophilus aphrophilus,* and *Actinobacillus actinomycetemcomitans.* FEMS Microbiol. Lett. 137: 213 - 220.

Hayakawa T., Tanaka T., Sakaguchi K., Otake N. and Yonehara H. (1979) J. Gen. Appl. Microbiol. 25, 225-260.

Henikoff S. (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Meth. Enzymol. 155: 156-165.

Hightower R.C., Metge D.W. and Santi D.V. (1987) Plasmid migration using orthogonal-field-alternation gel electrophoresis. Nucleic Acids Res. 15: 8387-8398.

Hightower R.C., Ruiz-Perez L.M., Wong M.L. and Santi D.V. (1988) Extrachromosomal elements in the lower eukaryote *Leishmania.* J. Biol. Chem. 263: 16970-16976.

Hirayama T., Muranaka T., Ohkawa H. and Oka A. (1988) Organization and characterisation of the *virCD* genes from *Agrobacterium rhizogenes.* Mol. Gen. Genet. 213: 229 - 237.

Hohn B. and Collins J. (1980) A small cosmid for efficient cloning of large DNA fragments. Gene 11: 291 - 298.

Holmes D.S. and Ul Haq R. (1989) Adaptation of *Thiobacillus ferrooxidans* for industrial applications. in Biohydrometallurgy - 1989. Salley J., McCready R.G.L. and Wichlacz P.L. (eds) pp 115-128.

Holmes D.S., Lobos J.H., Bopp L.H. and Welch G.C. (1983) Setting up a genetic system *de novo* for studying the acidophilic thiobacillus *T. ferrooxidans.* in Progress in Biohydrometallurgy Cagliari May 1983 Rossi G. and Torma A.E. (eds) pp 541-554.

Holmes D.S., Lobos J.H., Bopp L.H. and Welch G.C. (1984) Cloning of a *Thiobacillus ferrooxidans* plasmid in *Escherichia coli.* J. Bacteriol. 157: 324-326.

Holmes D.S., Yates J.R. and Schrader J. (1987) Mobile repeated DNA sequences in *Thiobacillus ferrooxidans* and their significance for biomining. in Biohydrometallurgy: Proceedings of the International symposium. Norris P.R. and Kelly D.P. (eds) pp 153-160.

Holt J.G., Krieg N.R., Sneath P.H.A., Staley J.T. and Williams S.T. (1994) Bergey's manual of determinative bacteriology 9th ed. pg 436. Published by Williams and Wilkens, MD, USA.

Hurtubise Y., Barriault D. and Sylvestre M. (1996) Characterisation of active recombinant His-tagged oxygenase component of *Comamonas testosteroni* B-356 biphenyl dioxygenase. J. Biol. Chem. 271: 8152 - 8156.

Iida A., Teshiba S. and Mizobuchi K. (1993) Identification and charactrisation of the *tktB* gene encoding a second transketolase in *Escherichia coli* K-12. J. Bacteriol. 175: 5375-5383.

Iida S., Sandmeier H., Hubner, Hiestand-Nauer R., Schneitz K. and Arber W. (1990) The Min DNA inversion enzyme of plasmid p15B of *Escherichia coli* 15T-: a new member of the Din family of site specific recombinases. Mol. Microbiol. 4: 991 - 997.

Ingledew W.J. and Cobley J.G. (1980) A potentiometric and kinetic study on the respiratory chain of ferrous iron grown *Thiobacillus ferrooxidans*. Biochim. Biophys. Acta 590: 141-158.

Inoue C., Sugawara K. and Kusano T. (1990) *Thiobacillus ferrooxidans mer* operon: sequence analysis of the promoter and adjacent genes. Gene 96: 115-120.

Inoue C., Sugawara K. and Kusano T. (1991) The *merR* regulatory gene in *Thiobacillus ferrooxidans* is spaced apart from the *mer* structural genes. Mol. Microbiol. 5: 2707-2718.

Inoue C., Sugawara K., Shiratori T., Kusano T. and Kitagawa Y. (1989) Nucleotide sequence of the *Thiobacillus ferrooxidans* chromosomal gene encoding mercuric reductase. Gene 84: 47-54.

Ippen-Ihler K. (1989) Bacterial conjugation in: Gene transfer in the environment. Levy S.B. and Miller R.V. (eds) pg 33-72.

Ivarson K.C. (1980) Enhancement of uranous-ion oxidation by *Thiobacillus ferrooxidans* Curr. Microbiol. 3: 253 - 254.

Iwasaki T., Wakagi T., Isogai Y., Tanaka K., Iizuka T. and Oshima T. (1994) Functional and evolutionary implications of a [3Fe-4S] cluster of the dicluster type ferredoxin from the thermoacidophilic archeon, *Sulpholobus sp.* strain 7. J. Biol. Chem. 269: 29444-29450.

Jackson D., Salem A. and Coombs G.H. (1984) The *in vitro* activity of metronidazole against strains of *Escherichia coli* with impaired DNA repair systems. J. Antimicrob. Chemother. 13: 227-236.

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

Johnson P.J. (1993) Metronidazole and drug resistance. Parasitology Today 9: 183-186.

Kai M., Yano T., Fukumori Y. and Yamanaka T. (1989) Cytochrome oxidase of an acidophilic iron-oxidising bacterium, *Thiobacillus ferrooxidans*, functions at pH3.5. Biochem. Biophys. Res. Comm. 160: 839 - 843.

Kai M., Yano T., Fukumori Y. and Yamanaka T. (1992) *Thiobacillus ferrooxidans* cytochrome c oxidase: purification, and molecular and enzymatic features. J. Biochem (Tokyo) 112: 816-821.

Kaminski P.A., Mandon K., Arigoni F., Desnoues N. and Elmerich C. (1991) Regulation of nitrogen fixation in *Azorhizobium caulinodans*: identification of a *fixK*-like gene, a positive regulator of *nifA*. Mol. Microbiol. 5: 1983-1989.

Kilkenny C.A., Berger D.K. and Rawlings D.E. (1994) Isolation of the *Thiobacillus ferrooxidans ntrBC* genes using a *T. ferrooxidans nifH-lacZ* fusion. Microbiol. 140: 2543 - 2553.

Kinashi H., Shimaji M. and Sakai A. (1987) Giant linear plasmids in *Streptomyces* which code for antibiotic synthesis genes. Nature 328: 454-456.

Kissinger C.R., Sieker L.C., Adman E.T. and Jensen L.H. (1991) Refined crystal structure of ferredoxin II from *Desulphovibrio gigas* at 1.7 A. J. Mol. Biol. 219: 693-715.

Knight E., D'Estachio A.J. and Hardy R.W.F. (1966) Flavodoxin: flavoprotein with with ferredoxin activity from *Clostridium pasteurianum*. Biochem. Biophys. Acta 113: 626-628.

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

Kues U. and Stahl U. (1989) Replication of plasmids in gram-negative bacteria. Microbiol. Rev. 53: 491-516.

Kulda J., Tachezy J. and Cerkasovova A. (1993) *In vitro* induced anaerobic resistance to metronidazole in *Trichomonas vaginalis*. J. Euk. Microbiol. 40: 262-269.

Kulpa C.F., Roskey M.T. and Travis M.T. (1983) Transfer of plasmid RP1 into chemolithotrophic *Thiobacillus neapolitanus*. J. Bacteriol. 156: 434-436.

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

Kusano T., Takeshima T., Sugawara K., Inoue C., Shiratori T., Yano T., Fukumori Y. and Yamanaka T. (1992b) Molecular cloning of the gene encoding *Thiobacillus ferrooxidans* Fe(II) oxidase. J. Biol. Chem. 267: 11242-11247.

Kusano T., Takeshima T., Inoue C. and Sugawara K. (1991) Evidence for two sets of structural genes coding for ribulose bisphosphate carboxylase in *Thiobacillus ferrooxidans*. J. Bacteriol. 173: 7313-7323.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685

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

Lau P.C.K., Drolet M., Zanga P. and Holmes D.S. (1993) Plasmid technology for biomining applications. Biohydrometallurgical Technologies 635 - 643, edited by A.E. Torma, M.L. Apel, and C.L. Brierley.

Lau P.C.K., Zanga P. and Drolet M. (1989) Characterization of the mobilization region of a *Thiobacillus ferrooxidans* plasmid in Biohydrometallurgy - 1989. Salley J., McCready R.G.L. and Wichlacz P.L. (eds) pp: 149 - 157.

Leblond P., Demuyter P., Simonet J.M. and Decaris B. (1991) Genetic instability and associated genome plasticity in *Streptomyces ambofaciens*: pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. J. Bacteriol. 173: 4229-4233.

Leclerc G., Noel G., Drapeau GR (1990) Molecular cloning, nucleotide sequence, and expression of *shl*, a new gene in the 2-minute region of the genetic map of *Escherichia coli* J Bacteriol 172: 4696-4700.

Levene S.D. and Zimm B.H. (1987) Separations of open-circular DNA using pulsed-field electrophoresis. Proc. Natl. Acad. Sci. USA 84: 4054-4057.

Lindmark D.G. and Muller M. (1973) Hydrogenosme, a cytoplasmic organelle of the anaerobic flagellate *Trichomonas foetus*, and its role in pyruvate metabolism. J. Biol. Chem. 248: 7724-7728.

Llanes C., Gabant P., Couturier M. and Michel-Briand Y. (1994) Cloning and characterisation of the Inc A/C plasmid RA1 replicon. J. Bacteriol. 176: 3403 - 3407.

Lockerby D.L., Rabin H.R., Bryan L.E. and Laishley E.J. (1984) Ferredoxin-linked reduction of metronidazole in *Clostridium pasteurianum*. Antimic. Agents Chemother. 26: 665-669.

Lorenzo P., Velasco A., Aparicio T., de Siloniz M.I. and Perera J. (1995) Plasmid pT3.21 of *Thiobacillus T3.2* in: Biohydrometallurgical processing. Vargas T., Jerez C.A., Wiertz J.V. and Toledo H. (eds), University of Chile, Santiago pp 129-134.

Lovley D.R. (1993) Dissimilatory metal reduction. Annu. Rev. Microbiol. 47: 263 - 290.

Macedo A.L., Moura I., Sureras K.K., Papaefthymiou V., Liu M-Y., LeGall J., Munck E. and Moura J.J.G. (1994) Thiol/disulfide formation associated with the redox activity of the [Fe3S4] cluster of *Desulfovibrio gigas* ferredoxin II. J. Biol. Chem. 269: 8052 - 8058.

MacInnes J.I., Kim J.E., Lian C.J. and Soltes G.A. (1990) *Actinobacillus pleuropneumoniae hlyX* gene homology with the *fnr* gene of *Escherichia coli*. J. Bacteriol. 172: 4587-4592.

Malliaros D.P. and Goldman P. (1991) Interaction of metronidazole with *Escherichia coli* deoxyribonucleic acid. Biochem. Pharmacol. 42: 1739-1744.

Manoil C. and Beckwith J. (1985) Tn*phoA*: A transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA 82: 8129 - 8133.

Mansch R. and Sand W. (1992) Acid stable cytochromes in ferrous ion oxidising cell-free preparations from *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. 92: 83-88.

Mao M.W.H., Dugan P.R., Martin P.A.W. and Tuovinen O.H. (1980) Plasmid DNA in chemoorganotrophic *Thiobacillus ferrooxidans* and *T. acidophilus*. FEMS Microbiol. Lett. 8: 121-125.

Mason J.R. and Cammack R. (1992) The electron transport proteins of hydroxylating bacterial dioxygenases. Annu. Rev. Microbiol. 46: 277-305.

Marritt S.J., Farrar J.A., Breton J.L.J., Hagen W.R. and Thomson A.J. (1995) Characterization of the prismane protein from *Desulphovibrio vulgaris* (Hildenborough) by low-temperature magnetic circular dichroic spectroscopy. Eur. J. Biochem. 232: 501-505.

Martin, P.A.W., Dugan, P.R. and Tuovinen, O.H. (1981) Plasmid DNA in acidophilic, chemolithotrophic thiobacilli. Can. J. Microbiol. 27: 850-853.

Mathew M.K., Hui C-F., Smith C.L. and Cantor C.R. (1988) High resolution separation and accurate size determination in pulsed-field gel electrophoresis of DNA. 4. Influence of DNA topology. Biochem. 27: 9222-9226.

McFadzean J.A. (1986) Flagyl. The story of a pharmaceutical discovery. The Parthenon Publishing Group Ltd., London.

Meyer T.E., Fitch J., Bartsch R.G., Tollin D. and Cusanovitch M.A. (1990) Unusual high redox potential ferredoxins and soluble cytochromes from the moderately halophilic purple phototrophic bacterium *Rhodospirillum salinarum*. Biochim. Biophys. Acta 1017: 118-124.

Meynell G.G. (1972) Bacterial plasmids pg 6-10. Published by MacMillan Press, New York, USA.

Miller JH (1972) Experiments in molecular genetics. p352-355, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Mjoli N. and Kulpa C.F. (1988) Identification of a unique outer membrane protein required for iron oxidation in *Thiobacillus ferrooxidans*. in Biohydrometallurgy: Proceedings of the International symposium. Norris P.R. and Kelly D.P. (eds) pp 89-102.

Morgan T.V., Lundell D.J. and Burgess B.K. (1988) *Azotobacter vinelandii* ferredoxin I: cloning, sequencing, and mutant analysis. J. Biol. Chem. 263: 1370-1375.

Moriya S., Ogasawara N. And Yoshikawa H. (1985) Structure and function of the region of the replication origin of the *Bacillus subtilis* chromosome. III Nucleotide sequence of some 10000 base pairs in the origin region. Nucleic Acids Res. 13: 2251-2265.

Moura I., Tavares P., Moura J.J.G., Ravi N., Huynh B.H., Liu M-Y. and LeGall J. (1992) Direct spectroscopic evidence for the presence of a 6Fe cluster in an iron-sulphur protein isolated from *Desulphovibrio desulphuricans* (ATCC 27774). J. Biol. Chem. 267: 4489-4496.

Muller M. (1983) Mode of action of metronidazole on anaerobic bacteria and protozoa. Surgery 1983, 165-171.

Muller M. and Lindmark D.G. (1976) Uptake of metronidazole and its effects on viability in trichomonads and *Entamoeba invadens* under anaerobic and aerobic conditions. Antimicrob. Agents Chemother. 9: 696-700.

Nakatsu C.H., Straus N.A. and Wyndham R.C. (1995) The nucleotide sequence of the Tn*5271* 3-chlorobenzoate 3,4-dioxygenase genes (*cbaAB*) unites the class IA oxygenases in a single lineage. Microbiology 141: 485-495.

Narikawa S. (1986) Distribution of metronidazole susceptibility factors in obligate anaerobes. J. Antimicrob. Chemother. 18: 565-574.

Narikawa S., Suzuki T., Yamamoto M. and Nakamura M. (1991) Lactate dehydrogenase activity as a cause of metronidazole resistance in *Bacteroides fragilis* NCTC 11295. J. Antimicrob. Chemother. 28: 47-53.

Needleman S.B. and Wunsch C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.

Neidle E., Hartnett C., Ornston L.N., Bairoch A., Rekik M. and Harayama S. (1991) Nucleotide seuences of the *Acinetobacter calcoaceticus benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. J. Bacteriol. 173: 5385-5395.

Neidle E., Hartnett C., Ornston L.N., Bairoch A., Rekik M. and Harayama S. (1992) Cis-diol dehydrogenases encoded by the TOL pWWO plasmid xylL gene and the *Acinetobacter calcoaceticus* chromosomal *benD* gene are members of the short-chain alcohol dehydrogenase superfamily. Eur. J. Biochem. 204: 113-120.

Nielsen A.M. and Beck J.V. (1972) Chalcocite oxidation and coupled carbon dioxide fixation by *Thiobacillus ferrooxidans*. Science 175: 1124-1126.

Nordlund I., Powlowski J. and Shingler V. (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas sp.* strain CF600. J. Bacteriol. 172: 6826-6833.

Nordlund I. and Shingler V. (1990) Nucleotide sequences of the meta-cleavage pathway enzymes 2-hydroxymuconic semialdehyde dehydrogenase and 2-hydroxymuconic semialdehyde hydrolase from *Pseudomonas* CF600. Biochim. Biophys Acta 1049: 227-230.

Novick R.P. and Roth C. (1968) Plasmid linked resistance to inorganic salts in *Staphylococcus aureus*. J. Bacteriol. 95: 1335-1342.

Nunzi F., Woudstra M., Campese D., Bonicel J., Morin D. and Bruschi M. (1993) Amino acid sequence of rusticyanin from *Thiobacillus ferrooxidans* and its comparison with other blue copper proteins. Biochim. Biophys Acta 1162: 28-34.

Nunzi F., Guerlesquin F., Shepard W., Guigliarelli B. and Bruschi M. (1994) Active site geometry in the high oxido-reduction potential rusticyanin from *Thiobacillus ferrooxidans*. Biochem. Biophys. Res. Comm. 203: 1655-1662.

O'Brien R.W. and Morris J.G. (1971) The ferredoxin-dependent reduction of chloramphenicol by *Clostridium acetobutylicum*. J. Gen. Microbiol. 67: 265-271.

Olson G.J., Porter F.D., Rubinstein J. and Silver S. (1982) Mercuric reductase enzyme from a mercury-volatilizing strain of *Thiobacillus ferrooxidans*. J. Bacteriol. 151: 1230-1236.

Ougen P. and Cohen D. (1995) Yeast artificial chromosome cloning using PFGE. In: Pulsed field gel electrophoresis : a practical approach. Monaco A.P. (ed), Oxford University Press, UK. pg 95-118.

Page R.D.M. (1996) Treeview: An application to display phylogenetic tree's on personal computers. Comp. Applic. Biosciences 12: 357-358.

Palmer G. and Reedijk J. (1992) Nomenclature of electron-transfer proteins. Recommendation 1989. J. Biol. Chem. 267: 665-677.

Peng J., Yan W-M. and Bao X-Z. (1994) Plasmid and transposon transfer to *Thiobacillus ferrooxidans*. J. Bacteriol. 176: 2892-2897.

Perri S. and Helinski D.R. (1993) DNA sequence requirements for interaction of the RK2 replication initiation protein with plasmid origin repeats. J. Biol. Chem. 268: 3662-3669.

Peterson J.B. (1988) The interaction of metronidazole with electron carriers of the *Azotobacter vinelandii* respiratory particle fraction. FEMS Microbiol. Lett. 56: 135 - 138.

Pfeifer F., Griffig J. and Oesterhelt D. (1993) The *fdx* gene encoding the [2Fe-2S] ferredoxin of *Halobacterium salinarum* (*H. halobium*). Mol. Gen. Genet. 239: 66-71.

Pierik A.J., Wolbert R.B.G., Mutsaers P.H.A., Hagen W.R. and Veeger C. (1992a) Purification and biochemical characterisation of a putative [6Fe-6S] prismane-cluster-containing protein from *Desulphovibrio vulgaris* (Hildenborough) Eur. J. Biochem. 206: 697-704.

Pierik A.J., Hagen W.R., Dunham W.R. and Sands R.H. (1992b) Multi-frequency EPR and high resolution Mossbauer spectroscopy in a putative [6Fe-6S] prismane-cluster-containing protein from *Desulphovibrio vulgaris* (Hildenborough) : Characterisation of a supercluster and superspin model protein. Eur. J. Biochem 206: 705-719.

Polidoro M., Buonfiglio V., Lazzarini M., Valenti P. and Orsi N. (1993) Functional characterization of a 20 kb plasmid present in different *Thiobacillus ferrooxidans* strains. in Biohydrometallurgical Technologies, Edited by Torma A.E., Apel M.L. and Brierly C.L. pg 595-602.

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

Pramila T., Rao G.R., Natarajan K.A. and Rao C.D. (1996) Differential influence of ions on the copy number of plasmids in *Thiobacillus ferrooxidans*. Curr. Microbiol. 32: 57 - 63.

Pretorius I-M., Rawlings D.E. and Woods D.R. (1986) Identification and cloning of *Thiobacillus ferrooxidans* structural *nif* genes in *Escherichia coli*. Gene 45: 59 - 65.

Pronk J.T., De Bruyn J.C., Bos P. and Kuenen J.G. (1992) Anaerobic growth of *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 58: 2227-2230.

Pronk J.T., Liem K., Bos P. and Kuenen J.G. (1991) Energy transduction by anaerobic ferric iron respiration in *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 57: 2063-2068.

Pronk J.T., Meulenberg R., Hazeu W., Bos P. and Kuenen J.G. (1990) Oxidation of reduced inorganic sulphur compounds by acidophilic thiobacilli. FEMS Microbiol. Rev. 75: 293-306.

Quon D.V.K., D'Olivera C.E. and Johnson P.J. (1992) Reduced transcription of the ferredoxin gene in metronidazole resistant *Trichomonas vaginalis*. Proc. Natl. Acad. Sci. USA 89: 4402-4406.

Rawlings D.E. (1995) Restriction enzyme analysis of 16S rDNA genes for the rapid identification of Thiobacillus ferrooxidans, *Thiobacillus thiooxidans* and *Leptospirillum ferrooxidans* strains in leaching environments. in: Biohydrometallurgical processing. Vargas T., Jerez C.A., Wiertz J.V. and Toledo H. (eds), University of Chile, Santiago pg 9-18.

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

Rawlings D.E., Gawith C., Petersen A. and Woods D.R. (1983). Characterisation of plasmids and potential genetic markers in *Thiobacillus ferrooxidans*. Progress in Biohydrometallurgy Cagliari May 1983: Rossi G. and Torma A.E. (eds) pp 555 - 570.

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

Rawlings D.E., Pretorius I. and Woods D.R. (1984a) Expression of a *Thiobacillus ferrooxidans* origin of replication in *Escherichia coli*. J. Bacteriol. 158: 737-738.

Rawlings D.E., Pretorius I. and Woods D.R. (1984b) Construction of arsenic resistant *Thiobacillus ferrooxidans* recombinant plasmids and the expression of autotrophic plasmid genes in a heterotrophic cell-free system. J. Biotech. 1: 129-133.

Rawlings D.E., Pretorius I. and Woods D.R. (1986a) Expression of a *Thiobacillus ferrooxidans* plasmid functions and the development of genetic systems for the Thiobacilli. Biotechnology and Bioengineering Symp. No. 16 pg 281-287 published by John Wiley and Sons Inc.

Rawlings D.E., Sewcharan R. and Woods D.R. (1986b) Characterization of a broad-host-range mobilizable *Thiobacillus ferrooxidans* plasmid and the construction of *Thiobacillus* cloning vectors. in: Fundamental and applied Biohydrometallurgy, Lawrence R.W., Branion R.M.R. and Ebner H.G. (eds) pp 419-427

Rawlings DE, Silver S (1995) Mining with microbes. Bio/Technology 13:773-778

Rawlings D.E. and Woods D.R. (1985) Mobilization of *Thiobacillus ferrooxidans* plasmids among *Escherichia coli* strains. Appl. Environ. Microbiol. 49: 1323-1325.

Rawlings D.E., Woods D.R. and Mjoli N.P. (1991) The cloning and structure of genes from the autotrophic biomining bacterium *Thiobacillus ferrooxidans*. in Advances in Gene Technology Vol.2, Greenaway P.J. (ed) pp 215-237.

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

Reysset G., Haggoud A., Su W.J. and Sebald M. (1992) Genetic and molecular analysis of pIP417 and pIP419: *Bacteroides* plasmids encoding 5-nitroimidazole. Plasmid 27: 181-190.

Reysset G., Haggoud A. and Sebald M. (1993) Genetics of resistance of *Bacteroides* species to 5-nitroimidazole. Clin. Infect. Dis. 16 Suppl. 4: S401-S403.

Rezsohazy R., Hallet B. and Delcour J. (1992) IS*231D, E and F*, three new insertion sequences in *Bacillus thuringiensis*: extension of the IS231 family. Mol. Microbiol. 6: 1959-1967.

Rohrer J. and Rawlings D.E. (1992) Sequence analysis and characterisation of the mobilisation region of a broad-host-range plasmid, pTF-FC2, isolated from *Thiobacillus ferrooxidans*. J. Bacteriol. 174: 6230-6237.

Ronk M., Shively J.E., Shute E.A. and Blake R.C. (1991) Amino acid sequence of the blue copper protein Rusticyanin from *Thiobacillus ferrooxidans*. Biochem. 30: 9435-9442.

Rossi M., Pollock W.B.R., Reij M., Keon R.G., Fu R. and Voordouw G. (1993) The *hmc* operon of *Desulfovibrio vulgaris subsp vulgaris* Hildenborough encodes a potential transmembrane redox protein complex. J. Bacteriol. 175: 4699-4711.

Rowley D.A., Knight R.C., Skowlimowski I.M. and Edwards D.I. (1979) The effect of nitroheterocyclic drugs on DNA: An in vitro model of cytotoxicity. Biochem. Pharmacol. 28: 3009-3013.

Rozsa F.W., Viollier P., Fussenegger M., Hiestand-Nauer R. and Arber W. (1995) Cin-mediated recombination at secondary crossover sites on the *Escherichia coli* chromosome. J. Bacteriol. 177: 1159 - 1168.

Sambrook J, Maniatis T, Fritsch F (1989) Molecular cloning: a laboratory manual. 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sanchez H., Hevia E., Caceres B. and Venegas A. (1986) Studies on Native strains of *Thiobacillus ferrooxidans*. IV. Isolation, physical map, and partial cloning of a cryptic plasmid. Biotechnol. Appl. Biochem. 8, 300-308.

Sand W., Gerke T., Hallmann R. and Schippers A. (1995) Sulphur chemistry, biofilm, and the (in)direct attack mechanism - a critical evaluation of bacterial leaching. Appl. Microbiol. Biotechnol. 43: 961-966.

Santangelo J.D., Jones D.T. and Woods D.R. (1991) Metronidazole activation and isolation and cloning of *Clostridium acetobutylicum* electron transport genes J. Bacteriol. 173: 1088-1095.

Sato A., Fukumori Y., Yano T., Kai M. and Yamanaka T. (1989) *Thiobacillus ferrooxidans* cytochrome c-552: purification and some of its molecular features. Biochim. Biophys Acta 976: 129-134.

Sato S., Nakazawa K., Hon-nami K. and Oshima T. (1981) Purification, some properties and amino acid sequence of *Thermus thermophilus* HB8 ferredoxin. Biochim. et Biophys. Acta 668: 277-289.

Sauber K., Frohner C., Rosenberg G., Eberspacher J. and Lingens F. (1977) Purification and properties of pyrazon dioxygenase from pyrazon-degrading bacteria. Eur. J. Biochem. 74: 89-97.

Sawers R.G. (1991) Identification and molecular characterization of a transcriptional regulator from *Pseudomonas aeruginosa* PAO1 exhibiting structural and functional similarity to the FNR protein of *Escherichia coli*. Mol. Microbiol. 5: 1469 - 1481.

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

Schmehl M., Jahn A., zu Vilsendorf A.M., Hennecke S., Masepohl B., Schuppler M., Marxer M., Oelze J. and Klipp W. (1993) Identification of a new class of nitrogen fixation genes in *Rhodobacter capsulatus*: a putative membrane complex involved in electron transport to nitrogenase. Mol. Gen. genet. 241: 602-615.

Schrader J.A. and Holmes D.S. (1988) Phenotypic switching of *Thiobacillus ferrooxidans*. J. Bacteriol. 170: 3915-3923.

Schwartz D.C. and Cantor C.R. (1984) Separation of yeast chromosome-sized DNA's by pulsed field gradient gel electrophoresis. Cell 37: 67-75.

Sealy VR (1993) A cluster of two serine transfer RNA genes from *Clostridium acetobutylicum* P262. MSc thesis, University of Cape Town.

Seaton B.L. and Vickery L.E. (1994) A gene encoding a DnaK/hsp70 homolog in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 91: 2066-2070.

Shafiee A. and Hutchison C.R. (1988) Purification and reconstitution of the electron transport components for 6-deoxyerythronolide B hydroxylase, a cytochrome P-450 enzyme of macrolide antibiotic (erythromycin) biosynthesis. J. Bacteriol. 170: 1548-1553.

Shapiro J.A., Adhya S.L. and Bukhari A.I. (1977) Introduction: New pathways in the evolution of chromosome structure. in: DNA, insertion elements, plasmids and episomes. Bukhari A.I., Shapiro J.A. and Adhya S.L. (eds) pg 3-12. Published by Cold Spring Harbor Laboratory.

Shen B., Jollie D.R., Stout C.D., Diller T.C., Armstrong F.A., Gorst C.M., La Mar G.N., Stephens P.J. and Burgess B.K. (1994) *Azotobacter vinelandii* ferredoxin I: Alteration of individual surface charges and the [4Fe-4S]2+/+ cluster reduction potential. J. Biol. Chem. 269: 8564-8575.

Shiratori T., Inoue C., Numata M. and Kusano T. (1991) Characterization and cloning of plasmids from iron-oxidising bacterium *Thiobacillus ferrooxidans*. Curr. Microbiol. 23: 321 - 326.

Shiratori T., Inoue C., Sugawara K., Kusano T. and Kitagawa Y. (1989) Cloning and expression of *Thiobacillus ferrooxidans* mercury ion resistance genes in *Escherichia coli*. J. Bacteriol. 171: 3458 - 3464.

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

Silverman M.P. and Lundgren D.G. (1959) Studies on the chemoautotrophic iron bacterium *Ferrobacillus ferrooxidans*, I. An improved medium and a harvesting procedure for securing high cell yields. J. Bacteriol. 77: 642-647.

Sindar P., Britz M.L. and Wilkinson R.G. (1982) Isolation and properties of metronidazole resistant mutants of *Clostridium perfringens*. J. Med. Microbiol. 15: 503-509.

Smith C.L. and Condemine G. (1990) New approaches for physical mapping of small genomes. J. Bacteriol. 172: 1167-1172.

Southern E.M. and Elder J.K. (1995) Theories of gel electrophoresis of high molecular weight DNA. In: Pulsed field gel electrophoresis : a practical approach. Monaco A.P. (ed), Oxford University Press, UK. pg 1-28.

Spiro S. (1994) The FNR family of transcriptional regulators. Antonie van leeuwenhoek 66: 23-36.

Spiro S. and Guest J.R. (1987) Regulation and overexpression of the *fnr* gene of *Escherichia coli*. J. Gen. Microbiol. 133: 3279 - 3288.

Spiro S. and Guest J.R. (1988) Inactivation of the FNR protein of *Escherichia coli* by targeted mutagenesis in the N-terminal region. Mol. Microbiol. 2: 701 - 707.

Spiro S., Gaston K.L., Bell A.I., Roberts R.E., Busby S.J.W. and Guest J.R. (1990) Interconversion of the DNA-binding specificities of two related transcription regulators, CRP and FNR. Mol. Microbiol. 4: 1831 - 1838.

Sprenger G.A., Schorken U., Sprenger G. and Sahm H. (1995) Transketolase A of *Escherichia coli* K12. Eur. J. Biochem. 230: 525-532.

Stainthorpe A.C., Lees V., Salmond G.P.C., Dalton H. and Murrell J.C. (1990) The methane monooxygenase gene cluster of *Methylococcus capsulatus* (Bath). Gene 91: 27-34.

Stokkermans J.W.P.G., van den Berg W.A.M., van Dongen W.M.A.M. and Veeger C. (1992) The primary structure of a protein containing a putative [6Fe-6S] prismane cluster from *Desulphovibrio desulphuricans* (ATCC 27774). Biochim. Biophys. Acta 1132: 83-87.

Stokkermans J.W.P.G., Pierik A.J., Wolbert R.B.G., Hagen W.R., van Dongen W.M.A.M. and Veeger C. (1992) The primary structure of a protein containing a putative [6Fe-6S] prismane cluster from *Desulphovibrio vulgaris* (Hildenborough) Eur. J. Biochem. 208: 435-442.

Stokkermans J.W.P.G., Houba P.H.J., Pierik A.J., Hagen W.R., van Dongen W.M.A.M. and Veeger C. (1992) Overproduction of prismane protein in *Desulphovibrio vulgaris* (Hildenborough): evidence for a second S = 1/2-spin system in the one electron reduced state. Eur. J. Biochem. 210: 983-988.

Subramanian V., Liu T.N., Yeh W.K., Narro M. and Gibson D.T. (1981) Purification and properties of NADH-ferredoxin TOL reductase. A component of toluene dioxygenase from *Pseudomonas putida*. J. Biol. Chem. 256: 2723-2730.

Subramanian V., Liu T.N., Yeh W.K., Serdar C.M., Wackett L.P and Gibson D.T. (1985) Purification and properties of ferredoxin TOL. A component of toluene dioxygenase from *Pseudomonas putida* F1. J. Biol. Chem. 260: 2355-2363

Suen W-C. and Gibson D.T. (1994) Recombinant *Escherichia coli* strains synthesize active forms of napthalene dioxygenase and its individual α and β subunits. Gene 143: 67-71.

Sugio T., Domatsu C., Munakata O., Tano T. and Imai K. (1985) Role of ferric ion-reducing system in sulphur oxidation of *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 49: 1401-1406.

Sugio T., Hirayama K., Inagaki K., Tanaka H. and Tano T. (1992) Molybdenum oxidation by *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 58: 1768-1771.

Sugio T., Hirose T., Li-Zhen Y. and Tano T. (1992) Purification and some properties of sulphite:ferric ion oxidoreductase from *Thiobacillus ferrooxidans*. J. Bacteriol. 174: 4189-4192.

Sugio T., Wada K., Mori M., Inagaki K. and Tano T. (1988a) Synthesis of an Iron-oxidising system during growth of *Thiobacillus ferrooxidans* on sulphur basal salts medium. Appl. Environ. Microbiol. 54: 150-152.

Sugio T., Katagiri T., Moriyama M., Li Zhen Y., Inagaki K. and Tano T. (1988b) Existence of a new type of sulphite oxidase which utilizes ferric ions as an electron acceptor in *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 54: 153-157.

Sugio T., Mizunashi W., Inagaki K. and Tano T. (1987) Purification and some properties of sulphur:ferric ion oxidoreductase from *Thiobacillus ferrooxidans*. J. Bacteriol. 169: 4916-4922.

Sugio T., Tanaka K., Matsugi S., Kanao T. and Nagasawa T. (1995) Purification and some properties of NADH-dependent sulphite reductase from *Thiobacillus ferrooxidans*. in: Biohydrometallurgical processing. Vargas T., Jerez C.A., Wiertz J.V. and Toledo H. (eds), University of Chile, Santiago pp 109-118.

Sugio T., Tsujita Y., Inagaki K. and Tano T. (1990) Reduction of cupric ions with elemental sulphur by *Thiobacillus ferrooxidans*. Appl. Environ. Microbiol. 56: 693-696.

Sugio T., White K.J., Shute E., Choate D. and Blake R.C. (1992) Existence of a hydrogen sulphide:ferric ion oxidoreductase in Iron oxidising bacteria. Appl. Environ. Microbiol. 58: 431-433.

Summers A.O. and Silver S. (1978) Microbial transformations of metals. Annu. Rev. Microbiol. 32: 637-672.

Suwanto A. and Kaplan S. (1989) Physical and genetic mapping of *Rhodobacter sphaeroides* 2.4.1 genome: Presence of two unique circular chromosomes. J. Bacteriol. 171: 5850-5859.

Suzuki H., Tanaka T., Tano T. and Sugio T. (1993) Existence of sulphide binding protein in iron-oxidising bacteria. in: Bioleaching processes, Torma A.E., Wey J.E. and Lakshmanan V.I. (1993) pp 423-431.

Suzuki I., Takeuchi T.L., Yuthasastrakosol T.D. and Oh J.K. (1990) Ferrous iron and sulphur oxidation and ferric iron reduction activities of *Thiobacillus ferrooxidans* are affected by growth on ferrous iron, sulphur, or a sulphide ore. Appl. Environ. Microbiol. 56: 1620 - 1626.

Ta D.T. and Vickery L.E. (1992) Cloning, sequencing, and overexpression of a [2Fe-2S] ferredoxin gene from *Escherichia coli*. J. Biol. Chem. 267: 11120-11125.

Ta D.T., Seaton B.L. and Vickery L.E. (1992) Localization of the ferredoxin (*fdx*) gene on the physical map of the *Escherichia coli* chromosome. J. Bacteriol. 174: 5760-5761.

Tabita F.R. (1988) Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. Microbiol. Rev. 52: 155 - 189.

Tan H-M., Joannou C.L., Cooper C.E., Butler C.S., Cammack R. and Mason J. (1994) The effect of ferredoxinBED overexpression on benzene dioxygenase activity in *Pseudomonas putida* ML2. J. Bacteriol. 176: 2507-2512.

Tamegai H., Kai M., Fukumori Y. and Yamanaka T. (1994) Two membrane-bound c-type cytochromes of *Thiobacillus ferrooxidans*: Purification and properties. FEMS Microbiol. Lett. 119: 147-154.

Thomas C.M. and Helinski D.R. (1989) Vegetative replication and stable inheritance of IncP plasmids in: Promiscuous plasmids of Gram-negative bacteria. Thomas C.M. (ed) pg 1- 26.

Tocher J.H. and Edwards D.I. (1992) The interaction of reduced metronidazole with DNA bases and nucleosides. I. J. Radiation Oncology Biol. Phys. 22: 661-663

Townson S.M, Hanson G.R., Upcroft J.A. and Upcroft P. (1994) A purified ferredoxin from *Giardia duodenalis*. Eur. J. Biochem. 220: 439-446.

Trageser M. and Unden G. (1989) Role of cysteine residues and of metal ions in the regulatory functioning of FNR, the transcriptional regulator of anaerobic respiration in *Escherichia coli*. Mol. Microbiol. 3: 593-599.

Trinh S., Haggoud A., Reysset G. and Sebald M. (1995) Plasmids pIP419 and pIP421 from *Bacteroides*: 5-nitroimidazole resistance genes and their upstream insertion sequence elements. Microbiol. 141: 927-935.

Trower M.K., Marshall J.E., Doleman M.S., Emptage M.H. and Sariaslani F.S. (1990) Primary structure of a 7Fe ferredoxin from *Streptomyces griseus*. Biochim. et Biophys. Acta 1037: 290-296.

Tuovinen O.H. and Kelly D.P. (1974a) Studies on the growth of *Thiobacillus ferrooxidans* II. Toxicity of uranium to growing cultures and tolerance conferred by mutation, other metal cations and EDTA. Arch. Microbiol. 95: 153-164.

Tuovinen O.H. and Kelly D.P. (1974b) Studies on the growth of *Thiobacillus ferrooxidans* IV. Influence of monovalent metal cations on ferrous iron oxidation and uranium toxicity in growing cultures. Arch. Microbiol. 98: 167-174.

Valcova-Valchanova M.B. and Chan S.H.P. (1994) Purification and characterization of two new c-type cytochromes involved in Fe2+ oxidation from *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. 121: 61-70.

Valenti P., Polidoro M., Buonfiglio V., Visca P and Orsi N. (1989). Comparative analysis of *Thiobacillus ferrooxidans* strains. in Biohydrometallurgy - 1989. Salley J., McCready R.G.L. and Wichlacz P.L. (eds) pp 187-201.

Valenti P., Polidoro M., Buonfiglio V., Visca P and Orsi N. (1990). Plasmid DNA profiles in *Thiobacillus ferrooxidans*. J. Gen. Appl. Microbiol. 36: 351 - 355.

Van den Berg W.A.M., Stevens A.A.M., Verhagen M.F.J.M., van Dongen W.M.A.M. and Hagen W.R. (1994) Overproduction of the prismane protein from *Desulphovibrio desulphuricans* ATCC 27774 in *Desulphovibrio vulgaris* (Hildenborough) and EPR spectroscopy of the [6Fe-6S] cluster in different redox states. Biochim. Biophys. Acta 1206: 240-246.

Van der Oost J (1995) Nitrite and nitric oxide reduction in *Paracoccus denitrificans* is under the control of NNR, a regulatory protein that belongs to the FNR family of transcriptional activators. FEBS Lett 360: 151-154

Vernon L.P., Mangum J.H., Beck J.V. and Shafia F.M. (1960) Studies on a ferrous-ion-oxidising bacterium. II Cytochrome composition. Arch. Biochem. Biophys. 88: 227-231.

Visca P., Valenti P. and Orsi N. (1986) Characterisation of plasmids in *Thiobacillus ferrooxidans* strains. in: Fundamental and applied Biohydrometallurgy, Lawrence R.W., Branion R.M.R. and Ebner H.G. (eds) pp 429-441.

Visca P., Valenti P., Bianchi E., Cosentino S. and Orsi N. (1988) Comparative plasmid analysis in *Thiobacillus ferrooxidans* strains in Biohydrometallurgy: Proceedings of the International symposium. Norris P.R. and Kelly D.P. (eds) pp 565-568.

Visca P., Bianchi E., Dalmastri C., Valenti P. And Orsi N. (1990) Characterisation of genetic markers in *Thiobacillus ferrooxidans*. In: Biohydrometallurgy, Karavaiko G.I., Rossi G. and Avakyan Z.A. Moscow. pg 326-334.

Wall J.D., Love J. and Quinn S.P. (1984) Spontaneous *nif-* mutants of *Rhodopseudomonas capsulata*. J. Bacteriol. 159: 652 - 657.

Wehnert G.U., Abratt V.R. and Woods D.R. (1990) Cloning of *Bacteroides fragilis* plasmid genes affecting metronidazole resistance and ultraviolet survival in *Escherichia coli*. Plasmid 23: 155-158.

Wehnert G.U., Abratt V.R. and Woods D.R. (1992) Molecular analysis of a gene from *Bacteroides fragilis* involved in metronidazole resistance in *Escherichia coli*. Plasmid 27: 242-245.

Werlen C., Kohler H-P.E. and Van der Meer J.R. (1996) The broad substrate chlorobenzene dioxygenase and cis-chloro benzene dihydridiol dehydrogenase of *Pseudomonas sp* strain P51 are linked evolutionarily to the enzymes for benzene and toluene degradation. J. Biol. Chem. 271: 4009-4016.

Williams R., Bell A., Sims G. and Busby S. (1991) The role of two surface exposed loops in transcription activation by the *Escherichia coli* CRP and FNR proteins. Nucleic Acids Res. 19: 6705-6712.

Willison J.C. and Vignais P.M. (1982) The use of metronidazole to isolate Nif- mutants of *Rhodopseudomonas capsulata*, and the identification of a mutant with altered regulatory properties of Nitrogenase. J. Gen. Microbiol. 128: 3001-3010.

Wlodaczyk M. and Nowicka B. (1988) Preliminary evidence for the linear nature of *Thiobacillus versutus* pTAV2 plasmid. FEMS Microbiol. Lett. 55: 125-128.

Worley KC, Wiese BA, Smith RF (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. Genome Research 5: 173-184.

Yamaguchi M. and Fujisawa H. (1978) Characterisation of NADH-cytochrome c reductase, a component of benzoate 1,2 dioxygenase system from *Pseudomonas arvilla* c-1. J. Biol. Chem. 253: 8848-8853.

Yamaguchi M. and Fujisawa H. (1982) Subunit structure of oxygenase component in benzoate-1,2-dioxygenase system from *Pseudomonas arvilla* C-1. J. Biol. Chem. 257: 12497-12502.

Yamanaka T. and Fukumori Y. (1995) Molecular aspects of the electron transfer system which participates in the oxidation of ferrous iron by *Thiobacillus ferrooxidans*. FEMS Microbiol. Rev. 17: 401-413.

Yamanaka Y., Yano T., Kai M., Tamegai H. and Fukumori Y. (1993) The electron transport system coupled to the oxidation of Fe2+ in *Thiobacillus ferrooxidans*. in Biohydromet

Yankofsky S.A., Gurevich R., Grimland N. and Stark A. (1983). Genetic transformation of obligately chemolithotrophic thiobacilli. J. Bacteriol. 153: 652-657.

Yano T., Fukumori Y. and Yamanaka T. (1991) The amino acid sequence of rusticyanin isolated from *Thiobacillus ferrooxidans*. FEBS Letters 288: 159-162.

Yates J.R. and Holmes D.S. (1987) Two families of repeated DNA sequences in *Thiobacillus ferrooxidans*. J. Bacteriol. 169: 1861-1870.

Yates J.R., Cunningham R.P. and Holmes D.S. (1988) IST*2*: An insertion sequence from *Thiobacillus ferrooxidans*. Proc. Natl. Acad. USA 85: 7284-7287.

Yeung T-C., Beaulieu B.B., McLafferty M.A. and Goldman P. (1984) Interaction of metronidazole with DNA repair mutants of *Escherichia coli.* Antimicrob. Agents Chemother. 25: 65-70.

Zamanion M. and Mason J.R. (1987) Benzene dioxygenase in *Pseudomonas putida*. Subunit composition and immuno-cross-reactivity with other aromatic dioxygenases. Biochem. J. 244: 611-616.

Zhao H. and Holmes D.S. (1993) Insertion element IST*1* and associated phenotypic switching in *Thiobacillus ferrooxidans*. Biohydrometallurgical Technologies, Torma A.E., Wey J.E. and Lakshmanan V.I. (eds) pp 667-671.

Zhao G. and Winkler M.E. (1994) An *Escherichia coli* K-12 *tktA tktB* mutant deficient in transketolase activity requires pyridoxine (Vitamin B6) as well as the aromatic amino acids and vitamins for growth. J. Bacteriol. 176: 6134-6138.