Genetic studies on the region downstream of the *unc* operon of

*Thiobacillus ferrooxidans.*

by

Joseph Cruick-Shank Oppon.

Submitted in partial fulfilment of the requirements for the degree of Master of Science in the department of Microbiology, University of Cape Town.

April, 1996.
The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Chapter 1: General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2: Cosmid p818.1 and its subclones</td>
<td>37</td>
</tr>
<tr>
<td>Chapter 3: <em>T. ferrooxidans</em> glucosamine synthetase gene</td>
<td>53</td>
</tr>
<tr>
<td>Chapter 4: Tn7-like transposon of <em>T. ferrooxidans</em></td>
<td>81</td>
</tr>
<tr>
<td>Chapter 5: General Conclusions</td>
<td>124</td>
</tr>
<tr>
<td>Appendix 1: Media</td>
<td>130</td>
</tr>
<tr>
<td>Appendix 2: Buffers and Solutions</td>
<td>131</td>
</tr>
<tr>
<td>Appendix 3: Bacterial strains</td>
<td>138</td>
</tr>
<tr>
<td>Appendix 4: Standard Techniques</td>
<td>139</td>
</tr>
<tr>
<td>References</td>
<td>143</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS.

I am most grateful to my supervisor, Prof. Doug Rawlings for his excellent supervision, guidance and encouragement throughout the research. Without him, this work would not have been accomplished. I am also indebted to my colleagues in the Thiobacillus Research Unit: Ant Smith, Ros Powles, Cliff Dominy, Shelly Deane and the rest of my colleagues in the department who in one way or the other helped me throughout the period of study. I am grateful to all the members of the Department of Microbiology for their assistance and support during the course of this research. Special thanks to Di James, Anne Marie; who is always there when you need her and Di DeVillers for her amazing concern to let it be there when you need it.

My friends, Kojo, Joyce, Pee, Chateaux, Vic, Bob, Felix, Mawanda, Moses, the list of which I cannot exhaust, you have not been forgotten. Thank you for your encouragement which became the pillar of my strength. Last but not the least, I acknowledge the financial support from Foundation for Research and Development of CSIR and the General Mining Corp. of South Africa.
ABBREVIATIONS

A adenine
Ala L-alanine
amp ampicillin
Arg L-Arginine
Asn L-Asparagine
Asp L-aspartic acid
ATCC American Type Culture Collection
ATP adenosine triphosphate
bp base pair(s)
C cytosine
CGSC Coli Genetic Stock Centre
Cys L-cysteine
DNA deoxyribonucleic acid
dNTP(s) deoxyribonucleotide triphosphates
DSM Deutsche Sammlungvon Mikroorganismen (The German Collection).
EDTA ethylenediamine tetraacetic acid
g grams
G-6-P Glucosamine-6-phosphate
GFAT L-glutamine-D-fructose-6-phosphate amidotransferases
Gln L-glutamine
Glu L-glutamic acid
Gly L-glycine
hr hour(s)
His L-histidine
Ile L-isoleucine
IPTG isopropyl-β-D thio-galacopyranose
kb kilobase pair(s)
kD kiloDalton
l litres
LA Luria agar
LB Luria Broth
Leu L-leucine
Lys L-lysine
M Molar
Met L-methionine
min minute(s)
ml millilitres
NAcG-6-P N-acetyl glucosamine-6-phosphate
ORF(s) open reading frames(s)
orC chromosomal origin of replication.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>L-phenylalanine</td>
</tr>
<tr>
<td>pho</td>
<td>phosphate</td>
</tr>
<tr>
<td>Pi</td>
<td>inorganic phosphate</td>
</tr>
<tr>
<td>Pro</td>
<td>L-proline</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>Ser</td>
<td>L-serine</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>Thr</td>
<td>L-threonine</td>
</tr>
<tr>
<td>Trp</td>
<td>L-tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>L-tyrosine</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>Tn</td>
<td>Transposon</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
<tr>
<td>UDP-glc</td>
<td>Uridyldiphosphogluicosamine</td>
</tr>
<tr>
<td>Val</td>
<td>L-valine</td>
</tr>
<tr>
<td>v/m</td>
<td>volume per mass</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-galactoside</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1. Source and media of iron-sulphur oxidizing bacteria used............ 44
Table 2.2. Constructs and subclones of p818.1... 45
Table 3.1. Constructs and subclones used to study the complementation of \textit{E.coli} CGSC 5392.... 79
ABSTRACT

A Tn7-like element was found in a region downstream of a cosmid (p818.1) isolated from a genomic library of *Thiobacillus ferrooxidans* ATCC 33020. A probe made from the Tn7-like element hybridized to restriction fragments of identical size from both cosmid p818.1 and *T. ferrooxidans* chromosomal DNA. The same probe hybridized to restricted chromosomal DNA from two other *T. ferrooxidans* strains (ATCC 23270 and 19859). There were no positive signals when an attempt was made to hybridize the probe to chromosomal DNA from two *Thiobacillus thiooxidans* strains (ATCC 19733 and DSM504) and a *Leptospirillum ferrooxidans* strain DSM 2705.

A 3.5 kb BamHI-BamHI fragment was subcloned from p818.1 downstream the *T. ferrooxidans* *unc* operon and sequenced in both directions. One partial open reading frame (ORF1) and two complete open reading frames (ORF2 and ORF3) were found. On the basis of high homology to previously published sequences, ORF1 was found to be the C-terminus of the *T. ferrooxidans glmU* gene encoding the enzyme GlcNAc 1-P uridyltransferase (EC 1.7.7.23). The ORF2 was identified as the *T. ferrooxidans glmS* gene encoding the amidotransferase, glucosamine synthetase (EC 2.6.1.16). The third open reading frame (ORF3) was found to have very good amino acid sequence homology to TnsA of transposon Tn7. Inverted repeats very similar to the imperfect inverted repeat sequences of Tn7 were found upstream of ORF3. The cloned *T. ferrooxidans glmS* gene was successfully used to complement an *E.coli glmS* mutant CGSC 5392 when placed behind a vector promoter, but was otherwise not expressed in *E.coli*.
Subcloning and single strand sequencing of DNA fragments covering a region of about 7 kb beyond the 3.5 kb BamHI-BamHI fragment were carried out and the sequences searched against the GenBank and EMBL databases. Sequences homologous to the TnsBCD proteins of Tn7 were found. The TnsD-like protein of the Tn7-like element (registered as Tn5468) was found to be shuffled, truncated and rearranged. Homologous sequence to the TnsE and the antibiotic resistance markers of Tn7 were not found. Instead, single strand sequencing of a further 3.5 kb revealed sequences which suggested the \textit{T. ferrooxidans spo} operon had been encountered. High amino acid sequence homology to two of the four genes of the \textit{spo} operon from \textit{E.coli} and \textit{H.influenzae} namely, \textit{spoT}, encoding guanosine-3',5'-bis (diphosphate) 3'-pyrophosphohydrolase, (EC 3.1.7.2) and \textit{recG} encoding ATP-dependent DNA helicase, RecG (EC 3.6.1.) was found. This suggests that Tn5468 is incomplete and appears to terminate with the reshuffled TnsD-like protein. The orientation of the \textit{spoT} and \textit{recG} genes with respect to each other was found to be different in \textit{T. ferrooxidans} compared to those of \textit{E.coli} and \textit{H.influenzae}. 
CHAPTER 1.

1.1. Bioleaching

1.1.2. Organism-substrate interaction 1
1.1.3. Leaching reactions 2
1.1.4. Industrial application 3

1.2. *Thiobacillus ferrooxidans* 4

1.3. Region around the *E.coli* unc operon.

1.3.1. Region immediately left of the unc operon 5
1.3.2. The unc operon 8
    1.3.2.1. F₀ subunit 8
    1.3.2.2. F₁ subunit 10
1.3.3. Region proximate right of the unc operon (*EcoURF-1*) 13
    1.3.3.1. Metabolic link between *glmU* and *glmS* 14
1.3.4. Glucosamine synthetase (*GlmS*) 15
1.3.5. The pho genes 18
    1.3.5.1. Phosphate uptake in *E.coli* 18
    1.3.5.2. The Pst system 19
    1.3.5.3 Pst operon 20
    1.3.5.4. Modulation of Pho uptake 21
1.3.6. Transposable elements 25
    1.3.6.1. Transposons and Insertion sequences (IS) 27
    1.3.6.2. Tn7 27
    1.3.6.2.1. Insertion of Tn7 into *E.coli* chromosome 29
    1.3.6.2.2 The Tn7 transposition mechanism 32
1.3.7. Region downstream of *E.coli* and *T.ferrooxidans* unc operons.
CHAPTER ONE

INTRODUCTION

1.1. BIOLEACHING.

The elements iron and sulphur circulate in the biosphere through specific paths from the environment to organism and back to the environment. Certain paths involve only microorganisms and it is here that biological reactions of relevance in leaching of metals from mineral ores occur (Sand et al., 1993; Liu et al., 1993). These organisms have evolved an unusual mode of existence and it is known that their oxidative reactions have assisted mankind over the centuries. Of major importance are the biological oxidation of iron, elemental sulphur and mineral sulphides. Metals can be dissolved from insoluble minerals directly by the metabolism of these microorganisms or indirectly by the products of their metabolism. Many metals may be leached from the corresponding sulphides and it is this process that has been utilized in the commercial leaching operations using microorganisms

1.1.2 Organism-Substrate interaction.

Some microorganisms are capable of direct oxidative attack on mineral sulphides. Scanning electron micrographs have revealed that numerous bacteria attach themselves to the surface of sulphide minerals in solution when supplemented with nutrients (Benneth and Tributsch, 1978). Direct observation has indicated that bacteria dissolve a sulphide surface of the crystal by means of cell contact (Buckley and Woods, 1987). Microbial cells have also been shown to attach to metal hydroxides (Kennedy et al., 1976). Silverman (1967), concluded that at least two roles were performed by the bacteria in the solubilization of minerals. One role involved the ferric-ferrous cycle (indirect mechanism), whereas the other involved the physical
contact of the microorganism with the insoluble sulfide crystals and was independent of the ferric-ferrous cycle. Insoluble sulphide minerals can be degraded by microorganisms in the absence of ferric iron under conditions that preclude any likely involvement of a ferrous-ferric cycle (Lizama and Sackey, 1993). Although many aspects of the direct attack by bacteria on mineral sulphides remain unknown, it is apparent that specific iron and sulphide oxidizers must play a part (Mustin et al., 1993; Suzuki et al., 1994). Microbial involvement is influenced by the chemical nature of both the aqueous and solid crystal phases (Mustin et al., 1993). The extent of surface corrosion varies from crystal to crystal and is related to the orientation of the mineral (Benneth and Tributsch, 1978; Claassen, 1993).

1.1.3. Leaching reactions.

Attachment of the leaching bacteria to surfaces of pyrite (FeS₂) and chalcopyrite (CuFeS₂) is followed by the following reactions.

For pyrite:

\[
\text{FeS}_2 + \frac{31}{2}\text{O}_2 + \text{H}_2\text{O} \rightarrow \text{FeSO}_4 + \text{H}_2\text{SO}_4
\]

\[
2\text{FeSO}_4 + \frac{1}{2}\text{O}_2 \text{ (bacteria)} \rightarrow \text{Fe(SO)}_4_3 + \text{H}_2\text{O}
\]

\[
\text{FeS}_2 + \text{Fe}_2(\text{SO})_4_3 \rightarrow 3\text{FeSO}_4 + 2\text{S}
\]

\[
2\text{S} + 3\text{O}_2 + 2\text{H}_2\text{O} \text{ (bacteria)} \rightarrow 2\text{H}_2\text{S}
\]

For chalcopyrite:

\[
2\text{CuFeS}_2 + \frac{1}{2}\text{O}_2 + \text{H}_2\text{SO}_4 \text{ (bacteria)} \rightarrow 2\text{CuSO}_4 + \text{Fe(SO)}_4_3 + \text{H}_2\text{O}
\]

\[
\text{CuFeS}_2 + 2\text{Fe}_2(\text{SO})_4_3 \rightarrow \text{CuSO}_4 + 5\text{FeSO}_4 + 2\text{S}
\]
Although the catalytic role of bacteria in these reaction is generally accepted, surface attachment is not obligatory for the leaching of pyrite or chalcopyrite; the presence of sufficient numbers of bacteria in the solutions in juxtaposition to the reacting surface is adequate to indirectly support the leaching process (Lungren and Silver, 1980)

1.1.4. Industrial application.

The ability of microorganisms to attack and dissolve mineral deposits and use certain reduced sulphur compounds as energy sources has had a tremendous impact on their application in industry. The greatest interest in bioleaching lies in the mining industries where microbial processes have been developed to assist in the production of copper, uranium, and more recently, gold from refractory ores. In the latter process, iron and sulphur-oxidizing acidophilic bacteria are able to oxidize certain sulphidic ores containing encapsulated particles of elemental gold. Pyrite and arsenopyrites are the prominent minerals in the refractory sulphidic gold deposits which are readily bio-oxidized. This results in improved accessibility of gold to complexation by leaching agents such as cyanide. Bio-oxidation of gold ores may be a less costly and less polluting alternative to other oxidative pretreatments such as roasting and pressure oxidation (Olson, 1994). In many places, rich surface ore deposits have been exhausted and therefore bioleaching presents the only option for the extraction of gold from the lower grade ores (Olson, 1994).

Aside from the mining industries, there is also considerable interest in using microorganisms for biological desulphurization of coal (Andrews et al., 1991; Karavaiko et al., 1988). This is due to the large sulphur content of some coals which cannot be burnt unless the unacceptable levels of sulphur are released as sulphur dioxide. Recently, the use of
bioleaching has been proposed for the decontamination of solid wastes or solids (Couillard & Mercier, 1992; van der Steen et al., 1992; Tichy et al., 1993). The most important mesophiles involved are *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans* and *Leptospirillum ferrooxidans*.

1.2. *Thiobacillus ferrooxidans*.

Much interest has been shown in *Thiobacillus ferrooxidans* because of its use in the industrial mineral processing and because of its unusual physiology. It is an autotrophic, chemolithotrophic, gram-negative bacterium that obtains its energy by oxidising Fe$^{2+}$ to Fe$^{3+}$ or reduced sulphur compounds to sulphuric acid. It is acidophilic (pH 2.5-3.5) and strongly aerobic with oxygen usually acting as the final electron acceptor. However under anaerobic conditions, ferric iron can replace oxygen as electron acceptor for the oxidation of elemental sulphur (Brock and Gustafson, 1976; Corbett and Ingledew, 1987). At pH 2, the free energy change of the reaction:

\[
S + 6Fe^{2+} + 4H_2O \rightarrow H_2SO_4 + 6Fe^{3+} + 6H^+
\]

is negative; \(\Delta G = -314\) kJ/mol (Brock and Gustafson, 1976). *T. ferrooxidans* is also capable of fixing atmospheric nitrogen as most of the strains have genes for nitrogen fixation (Pretorius et al., 1986). The bacterium is ubiquitous in the environment and may be readily isolated from soil samples collected from the vicinity of pyritic ore deposits or from sites of acid mine drainage that are frequently associated with coal waste or mine dumps.
Most isolates of *T. ferroxidans* have remarkably modest nutritional requirements. Aeration of acidified water is sufficient to support growth at the expense of pyrite. The pyrite provides the energy source and trace elements; the air provides the carbon, nitrogen and acidified water provides the growth environment. However, for very effective growth, certain nutrients, for example ammonium sulfate \((\text{NH}_4)_2\text{SO}_4\) and potassium hydrogen phosphate \((\text{K}_2\text{HPO}_4)\) have to be added to the medium. Some of the unique features of *T. ferroxidans* are its inherent resistance to high concentrations of metallic and other ions and its adaptability when faced with adverse growth conditions (Rawlings and Woods, 1991).

Since a major part of this study will concern a comparison of the genomes of *T. ferroxidans* and *E. coli* in the vicinity of the *atp* operon, the position and function of the genes in this region of the *E. coli* chromosome will be reviewed.

### 1.3. REGION AROUND THE *E. COLI* UNC OPERON

#### 1.3.1. Region immediately left of the unc operon.

The *Escherichia coli* unc operon, encoding the eight subunits of ATP synthase, is found near min 83 in the 100 min linkage map close to the single origin of bidirectional DNA replication, *oriC* (Bachmann, 1983). The region between *oriC* and *unc* is especially interesting because of its proximity to the origin of replication (Walker *et al.*, 1984). Earlier it had been suggested that an outer membrane protein binding at or near the origin of replication might be encoded in this region of the chromosome, or alternatively that the DNA segment to which the outer membrane protein is thought to bind might lie between *oriC* and *unc* (Wolf-Waltz & Norquist, 1979; Wolf-Watz and Masters, 1979). The phenotypic marker *het* (structural gene
Fig. 1a. *E. coli* chromosome in the vicinity of the unc operon. Genes on the right of oriC are transcribed from left to right.
for DNA-binding protein) has been used for this region (Wolf-Waltz & Norquist, 1979). Two DNA segments have been proposed to bind to the membrane protein; one overlaps oriC, the other lies within unc operon (Wolf-Watz, 1984).

A second phenotypic trait gid (glucose-inhibited division) has also been associated with the region between oriC and unc (Fig. 1a; Walker et al., 1984). This phenotype was designated following construction of strains carrying a deletion of oriC and part of gid and with an insertion of transposon Tn10 in asnA. This oriC deletion strain can be maintained by replication of an integrated F-plasmid (Walker et al., 1984). Various oriC minichromosomes were integrated into the oriC deletion strain by homologous recombination and it was observed that integrated minichromosomes carrying an intact gidA gene displayed a 30% higher growth rate on glucose media compared with ones in which gidA was partly or completely absent (von Meyenburg and Hansen, 1980). A protein of 70 kDa has been associated with gidA (von Meyenburg and Hansen, 1980). Insertion of transposon Tn10 in gidA also influences expression of a 25 kDa protein, the gene for which maps between gidA and unc. Therefore its been proposed that the 70 kDa and 25 kDa proteins are co-transcribed; gidB has been used to designate the gene for the 25 kDa protein (von Meyenburg et al., 1982).

Comparison of the region around oriC of B. subtilis and P. putida to E.coli revealed that, this region has been conserved in the replication origin of the bacterial chromosomes of both gram-positive and gram-negative eubacteria (Ogasawara and Yoshikawa, 1992). Detailed analysis of this region of E.coli and B.subtilis showed this conserved region to be limited to about nine genes covering a 10 kb fragment because of translocation and inversion event that
occurred in \textit{E.coli} chromosome (Ogasawara & Yoshikawa, 1992). This comparison also indicated that translocation of \textit{oriC} together with the \textit{gid} and \textit{unc} operons may have occurred during the evolution of the \textit{E.coli} chromosome (Ogasawara & Yoshikawa, 1992).

\subsection*{1.3.2. The \textit{unc} operon}

The ATP synthase (\textit{F}_{0}\textit{F}_{1}-\text{ATPase}), a key enzyme in energy converting reactions, couples the synthesis or hydrolysis of ATP to the translocation of protons (H\textsuperscript{+}) from across the membrane. It uses a protomotive force, generated across the membrane by electron flow to drive the synthesis of ATP from ADP and inorganic phosphate (Mitchell, 1966). The enzyme complex, which is present in procaryotic and eukaryotic organisms consists of a globular domain \textit{F}_{1} and an intrinsic membrane domain \textit{F}_{0}, linked by a slender stalk about 45Å long (Fillingame, 1990). Each sector of \textit{F}_{0}\textit{F}_{1} is composed of multiple subunits in unusual stoichiometric ratios, that is $\alpha_{i}\beta_{j}\tau_{i}\delta_{i}\epsilon_{i}$ for \textit{F}_{1} and $a_{i}b_{2}c_{10}$ for \textit{F}_{0} of the \textit{E.coli} enzyme (Fillingame, 1992). These eight subunits of the \textit{E.coli} \textit{F}_{1}\textit{F}_{0} complex are coded by the genes of the \textit{unc} operon (Walker \textit{et al.}, 1984).

\subsubsection*{1.3.2.1. \textit{F}_{0} subunit}

The three subunits of \textit{F}_{0} part of the gene has molecular masses of 30.276, 17.265, and 82.88 kDa (Walker \textit{et al.}, 1984) and a stoichiometry of 1:2:10±1 (Foster and Fillingame, 1982; Hermolin and Fillingame, 1989) respectively. Analyses of deletion mutants and reconstitution experiments with subcomplexes of all the three subunits of \textit{F}_{0} clearly demonstrated that, the presence of all the three subunits is indispensable for the formation of a complex active in proton translocation and \textit{F}_{1} binding (Friedl \textit{et al.}, 1983; Schneider and Altendorf, 1985). For
subunit \( a \), which is a very hydrophobic protein, a secondary structure with 5-8 membrane-spanning helices has been predicted (Fillingame, 1990; Lewis et al., 1990; Bjobæk et al., 1990; Vik and Dao 1992), but convincing evidence in favour of this secondary structures is still lacking (Birkenhager et al., 1995).

Subunit \( b \) is a hydrophilic protein anchored in the membrane by its apolar N-terminal region. Studies with proteases and subunit-\( b \)-specific antibodies revealed that, the hydrophilic antibody-binding part is exposed to the cytoplasm (Deckers-Hebestriet et al., 1992). Selective proteolysis of subunit \( b \) resulted in an impairment of \( F_{i} \) binding, whereas the proton translation remained unaffected (Hermolin et al., 1983; Perlin and Senior, 1985; Steffens et al., 1987; Deckers-Hebestriet et al., 1992). These studies and analyses of cells carrying amber mutations within the \( uncF \) gene indicated that, subunit \( b \) is also necessary for correct assembly of \( F_{o} \) (Steffens et al., 1987; Takeyama et al., 1988).

Subunit \( c \) exists as a hairpin-like structure with two hydrophobic membrane-spanning stretches connected by a hydrophilic loop region which is exposed to cytoplasm (Girvin and Fillingame, 1993; Fraga et al., 1994). Subunit \( c \) plays a key role in both \( H^{+} \) transport and the coupling of \( H^{+} \) transport with ATP synthesis (Fillingame, 1990). Several pieces of evidence have revealed that the conserved acidic residue within the C-terminal hydrophobic stretch Asp61, plays an important role in proton translocation process (Miller et al., 1990; Zhang and Fillingame, 1994). The number of \( c \) units present per \( F_{o} \) complex has been determined to be ±10 (Girvin and Fillingame, 1993). However, from the considerations of the mechanism, it is believed that 9 or 12 copies of subunit \( c \) are present for each \( F_{o} \).
which are arranged in three units of subunit c trimer or tetramers (Schneider and Altendorf, 1987; Fillingame 1992). Each unit is in close contact to a catalytic αβ pair of F₁ complex (Fillingame, 1992). Due to a high sequence similarity of the proteolipids from FₐF₁ ATPases, vacuolar H⁺-ATPases and gap junctions, a number of 12 copies of subunit c/F₀ must be favoured by analogy to the stoichiometry of the proteolipids in the gap junctions as revealed by electron microscopic analysis (Finbow et al., 1992; Holzenburg et al., 1993).

For the spatial arrangement of the three subunits in the F₀ complex, two different models have been proposed. Cox et al. (1986), have suggested that the $a_1b_1$ moiety is surrounded by a ring of c subunits. In the second model, $a_1b_2$ moiety is located outside the subunit c oligomer interacting only with one side of this oligomeric structure (Hoppe and Sebald, 1986; Schneider and Altendorf, 1987; Fillingame, 1992). However, Birkenhager et al. (1995), using transmission electron microscopy imaging (ESI) have proved that subunits $a$ and $b$ are located outside the subunit c oligomer (Hoppe and Sebald, 1986; Schneider and Altendorf, 1987; Fillingame, 1992).

1.3.2.2. F₁ subunit.

The F₁ domain is an approximate sphere 90-100Å in diameter and contains the catalytic binding sites for the substrates ADP and inorganic phosphate (Abrahams et al., 1994). The energy released by proton flux through F₀ is relayed to the catalytic sites in the F₁ domain, probably by conformational changes through the stalk (Abrahams et al., 1994). About three protons flow through the membrane per ATP synthesized; disruption of the stalk releases the water soluble enzyme, F₁-ATPase (Walker et al., 1994). This F₁ sector is the catalytic part
Fig. 1b. Schematic model of *E. coli* F₁ derived from projection views of unstained molecule interpreted in the framework of the three dimensional stain excluding outline. The directions of view (arrows) which produce the bilobed and trilobed projections are indicated. The peripheral subunits are presumed to be the α and β subunits, and the smaller density interior to them probably consists of at least parts of the α, β and/or ε subunits (Gogol et al., 1989).
of the complex; there are three catalytic sites per molecule which have been localised to \( \beta \) subunit whereas the function of nucleotides in the \( \alpha \)-subunits which do not exchange during catalysis is obscure (Vignais and Lunard, 1985).

According to the binding exchange mechanism of ATP synthesis (Cross et al., 1995), the structures of the three catalytic sites are always different, but each passes through a cycle of \"open\", \"loose\" and \"tight\" states. The mechanism suggests that F\(_1\)-ATPase is an inherent asymmetric assembly, as clearly indicated by subunit stoichiometry, electron microscopy (Boekemia et al., 1986 and Wilkens et al., 1994), and low resolution X-ray crystallography (Abrahams et al., 1993). The structure of F\(_1\) isolated from variety of sources has been studied by using both biochemical and biophysical techniques (Amzel and Pedersen, 1983). Electron microscopy has been particularly useful in defining the gross features of the protein complex (Brink et al., 1988). Molecules of F\(_1\) in negatively stained preparations are usually found in one predominant orientation which shows a hexagonal arrangement of six equal-sized lobes presumably representing the three \( \alpha \) and three \( \beta \) subunits (Akey et al., 1983; Tiedge et al., 1983; Tsuprun et al., 1984; Boekemia et al., 1988).

A seventh density, either centrally or asymmetrically located has been observed (Boekemia et al., 1986). Image analysis has revealed that six elongated protein densities (the \( \alpha \) and \( \beta \) subunits, each \( \approx 90 \, \text{Å} \times 30 \, \text{Å} \) in size) compromise its hexagonanally modulated periphery (Gogol et al., 1989). At the centre of F\(_1\) is an aqueous cavity which extends nearly or entirely through the length of the complex; a compact protein density located at one end of the hexagonal barrier and closely associated with one of the peripheral subunits, partially obstructs the central cavity (Gogol et al., 1989).
1.3.3. Region proximate right of unc operon (EcoURF-1)

DNA sequencing around the E.coli unc operon has previously shown that the glmS (encoding glucosamine synthetase) gene was preceded by an open reading frame of unknown function, named EcoURF-1, which theoretically encodes a polypeptide of 456 amino acids with a molecular weight of 49,130 (Walker et al., 1984). The short intergenic distance between EcoURF-1 and glmS and the absence of an obvious promoter consensus sequence upstream of glmS suggested that these genes were co-transcribed (Plumbridge et al., 1993; Walker et al., 1984). Mengin-Lecreulx et al. (1993), using a thermosensitive mutant in which the synthesis of EcoURF-1 product was impaired, have established that the EcoURF-1 gene is an essential gene encoding the GlcNAc-1-P uridylyltransferase activity. The N-acetylglucosamine-1-phosphate (GlcNAc-1-P) uridylyltransferase activity (also named UDP-GlcNAc pyrophosphorylase), which synthesizes UDP-GlcNAc from GlcNAc-1-P and UTP (see Fig. 1c), has previously been partially purified and characterized for Bacillus licheniformis and Staphylococcus aureus (Anderson et al., 1993; Strominger and Smith, 1959). Mengin-Lecreulx and Heijenoort (1993), have proposed the use of glmU (for glucosamine uridylyltransferase) as the name for this E.coli gene according to the nomenclature previously adopted for the glmS gene encoding glucosamine-6-phosphate synthetase (Walker et al., 1984; Wu and Wu, 1971).
Fructose-6-Phosphate

\[ \text{glucosamine synthetase (glmS)} \]

\[ \text{glucosamine-6-phosphate} \]

\[ \text{glucosamine-1-phosphate} \]

\[ \text{N-acetylglucosamine-1-P} \]

UTP \[ \text{(EcoURF-1) now called glmU} \]

UDP-N-acetylglucosamine

\[ \text{lipopolysaccharide} \]

\[ \text{peptidoglycan} \]

Fig. 1c. Biosynthesis and cellular utilization of UDP-Glc in *E.coli*.

1.3.3.1. **Metabolic link between glmU and glmS.**

The amino sugars D-glucosamine (GlcN) and N-acetyl-D-glucosamine (GlcNAc) are essential components of the peptidoglycan of bacterial cell walls and of the lipopolysaccharide of the outer membrane in gram-negative bacteria including *E.coli* (Höltje and Schwarz, 1985; Park, 1987). When present in the environment, both compounds are taken up and used for cell wall and lipid A (an essential component of outer membrane lipopolysaccharide) synthesis (Dobrogózz, 1968). In the absence of amino sugars in the environment, the bacteria must
synthesize glucosamine-6-phosphate from fructose-6-phosphate and glutamine via the enzyme glucosamine-6-phosphate synthase (L-glutamine: D-fructose-6-phosphate amidotransferase), the product of glmS gene. Glucosamine-6-phosphate (GlcNH₂-6-P) then undergoes sequential transformations involving the product of glmU (see Fig. 1c) leading to the formation of UDP-N-acetyl glucosamine, the major intermediate in the biosynthesis of all amino sugar containing macromolecules in both prokaryotic and eukaryotic cells (Badet et al., 1988). It is tempting to postulate that the regulation of glmU (situated at the branch point shown systematically in Fig. 1c) is a site of regulation considering that most of glucosamine in the E.coli envelope is found in its peptidoglycan and lipopolysaccharide component (Park, 1987; Raetz, 1987). Genes and enzymes involved in steps located downstream from UDP-GlcNAc in these different pathways have in most cases been identified and studied in detail (Doublet et al., 1992; Doublet et al., 1993; Kuhn et al., 1988; Miyakawa et al., 1972; Park, 1987; Raetz, 1987).

1.3.4. Glucosamine synthetase (glmS).

Glucosamine synthetase (2-amino-2-deoxy-D-glucose-6-phosphate ketol isomerase amino transferase EC 5.3.1.19), (formerly L-glutamine: D-fructose-6-phosphate amidotransferase, EC 2.6.1.16) transfers the amide group of glutamine to fructose-6-phosphate in an irreversible manner (Winterburn and Phelps, 1973). This enzyme is a member of glutamine amidotransferases (a family of enzymes that utilize the amide of glutamine for the biosynthesis of several amino acids including arginine, asparagine, histidine, and tryptophan as well as the amino sugar glucosamine 6-phosphate). Glucosamine synthetase is one of the key enzymes vital for the normal growth and development of cells. The amino sugars are also sources of carbon and nitrogen to the bacteria. For example, GlcNAc allows E.coli to growth at rates
comparable to those glucose (Plumbridge et al., 1993). The alignment of the first 194 amino acids of amidophosphoribosyl-transferase with glucosamine synthetase from *E. coli* produced 52 identical and 51 similar amino acids for an overall conservation of 53% (Mei and Zalkin, 1990). Glucosamine synthetase is unique among this group in that, it is the only one transferring the amide nitrogen to a keto group without the participation of a cofactor (Badet et al., 1986).

It is a dimer of identical, 68 kDa subunits showing classical properties of amidotransferases (Badet et al., 1987; Kucharczyk et al., 1990). The purified enzyme is stable, (could be stored on at -20°C for several months); does not exhibit lability; does not display any absorbance in the region of 300-500 nm and is colourless at a concentration of 5 mg/ml suggesting it is not an iron containing protein (Badet et al., 1986). Again, no hydrolytic activity could be detected by standard spectrophotometric assay and in contrast to mammalian glucosamine synthetase (Bates and Handshumacher, 1968; Winterburn and Phelps, 1973), UDP-GlcNAc does not affect *E. coli* glucosamine synthetase activity as shown by Kornfield (1961), in crude extracts from *E. coli* and *B. subtilis* (Badet et al., 1986).

Glucosamine synthetase is subject to weak product inhibition at millimolar concentrations of GlcN-6-P but it is not subject to allosteric regulation (Vermoote, 1988), unlike the equivalent eukaryotic enzymes which are allosterically inhibited by UDP-GlcNAc (Frisa et al., 1982; MacKnight et al., 1990). The intracellular concentration of GlmS protein is lowered about three fold by growth on the amino sugars glucosamine and N-acetylglucosamine; this regulation occurs at the level of transcription (Plumbridge et al., 1993). It is also subject to
a control mechanism which causes its expression to be reduced when the level of the nag (coding for proteins involved in the uptake and metabolism of N-acetylglucosamine) regulon genes are derepressed (Badet et al., 1993). Chimara et al. (1984), have also showed that anticapsin (the C-terminal epoxyamino acid of the antibiotic tetaine; also produced independently by a strain of Streptomyces griseoplanus), inactivates the glucosamine synthetase from E.coli, Pseudomonas aeruginosa, Arthrobacter aurenscens and Bacillus thuringiensis.

Studies performed earlier with other amidotransferases revealed that the sulfhydryl group of the active centre plays a vital role in the catalysis of transfer of the τ-amino group from glutamine to the acceptor substrate (Buchanan, 1973). The indispensability of the sulfhydryl group of the glmS gene product in glucosamine-6-phosphate synthesis suggests that anticapsin inactivates the glutamine binding site, presumably by covalent modification of cystein residue (Chmara et al., 1984). GlcNH2-6 synthetase has also been found to exhibit strong sensitivity to pyridoxal 5′-phosphate addition; the inhibition being competitive with respect to the substrate fructose-6-phosphate (Golinelli-Pimpaneaux and Badet, 1991). This inhibition by pyridoxal 5′-phosphate is irreversible and is thought to result from Schiff base formation with an active site lysine residue (Badet et al., 1993). In E.coli, the pst (phosphate specific transport) genes are found immediately downstream of the glmS gene therefore the pst genes will be the reviewed in this text after the glmS gene.
1.3.5. The *pho* genes.

Unless otherwise stated, information on *pho* genes has been based on the review of Rao and Torriano (1990).

1.3.5.1. Phosphate uptake in *E.coli*

Phosphate is an integral part of the globular cellular metabolism since it is indispensable for DNA and RNA synthesis, energy supply and membrane transport. Phosphate is utilized by the cell as phosphate ions, which are neither reduced nor oxidized during assimilation. However, a wide range of available phosphates that occur in nature cannot be metabolized by *E.coli* unless they are first degraded into Pi (inorganic phosphate). These phosphorylated compounds must first cross the outer membrane (OM), before they are hydrolysed to release Pi in the periplasm. The Pi is captured by binding proteins and finally transported across the inner membrane (IM) into the cytoplasm. In *E.coli*, two systems for inorganic phosphate (Pi) transport (Pst and Pit) have been reported (Willsky and Malamy, 1974, 1980; Sprague et al., 1975; Rosenberg et al., 1987).

*E.coli* normally transports inorganic phosphate (Pi) by the low-affinity transport system, Pit. When the level of external Pi is lower than 20 μM or when the only source of phosphate is organic phosphate other than glycerol phosphate and hexose phosphate, another transport system Pst, is induced. The latter is typical of a class of inducible high-affinity transport systems which are sensitive to osmotic shock and include periplasmic binding proteins (Medveczky and Rosenberg, 1970; Boos, 1974). Another protein, an outer membrane porin PhoE with a $K_m$ of about 1 μM is also induced; this outer membrane protein allows the
intake of organic phosphates which are degraded to Pi by phosphatases in the periplasm (Medveczky and Rosenberg, 1970).

1.3.5.2. The Pst system.

A comparison of the kinetic parameters shows the transport constant (Kt) for the Pst system (about 0.25 μM) to be two orders of magnitude lower than that of the Pit system (about 20 μM, Rosenberg, 1987). This makes the Pst system highly efficient and hence at low Pi concentrations, this is the main system for Pi uptake. The pst genes together with phoU gene, form an operon (Fig. 1d) that maps at about 83.5 min on the E.coli chromosome. Surin et al. (1987), established a definitive order in the confusing picture available on the genes of Pst region and their sequence on the E.coli map. The sequence is: bgI... phoU, pstB, psTA (formerly phoT), pstC (phoW), pstS (phoS)... glmS. All genes are transcribed counter clockwise on the E.coli chromosome and constitute an operon. The nuceotide sequences of all the five genes have been determined, and the amino acid sequences of the corresponding protein has been deduced (Surin et al., 1987). The gene products of the Pst system which have been isolated in pure forms are PiBP (phosphate binding protein) and PhoU (Surin et al., 1986; Rosenberg, 1987). The Pst system has two functions in E.coli; the transport of Pi and the negative regulation of the phosphate regulon (a complex of twenty proteins mostly related to organic phosphate transport).
Table 1: Pst genes and their role in phosphate uptake.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Genes</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>PstA</td>
<td><em>pstA</em>(phoT)</td>
<td>Cytoplasmic membrane</td>
</tr>
<tr>
<td>PstB</td>
<td><em>pstB</em></td>
<td>Energy coupling peripheral membrane</td>
</tr>
<tr>
<td>PstC</td>
<td><em>pstC</em>(phoW)</td>
<td>Cytoplasmic membrane protein</td>
</tr>
<tr>
<td>PiBP</td>
<td><em>pstS</em>(phoS)</td>
<td>Periplasmic Pi binding protein</td>
</tr>
</tbody>
</table>

1.3.5.3. **Pst operon**

*PhoS* (*pstS*) and *pstA* (phoT) mutations were initially isolated as alkaline phosphatase-constitutive mutants (Echols *et al.*, 1961), and Pst mutants were isolated either as arsenate-resistant (Bennet and Malamy, 1970) or as organic phosphate autotrophs (Sprague *et al.*, 1975). Regardless of the selection criteria, all of these mutants were defective in incorporation of Pi on a pit-background and synthesized alkaline phosphatase, coded for by the *phoA* gene, in high organic phosphate media. The activity of the Pst system depends on the presence of the PiBP which has a *K_m* of approximately 1 μM. This protein, encoded by the *pstS* gene has been purified and crystallized but no high resolution crystallographic data are yet available. The PiBP encoded by *pstS* consists of 346 amino acids, which is the precursor form of the phosphate binding protein (Surin *et al.*, 1984).
The mature phosphate binding protein (PiBP) has 321 amino acids and a molecular weight of 34,427. The 25 additional amino acids present in the pre-phosphate binding protein constitute a typical signal peptide (Michaelis and Rosenberg, 1970), with a positively charged N-terminus followed by a chain of 20 hydrophobic amino acid residues (Surin et al., 1984). The site of cleavage of the signal peptide to form the mature phosphate binding protein lies on the carboxyl side of the alanine residue preceding the N-terminal glutamate of the mature phosphate binding protein (Surin et al., 1984). After the passage of the organic phosphates through the outer membrane, Pi is cleaved off by phosphatase in the periplasm, and the Pi-binding protein captures the free Pi produced in the periplasm and directs it to the transmembrane channel of the cytoplasmic membrane. The channel consists of two proteins, PstA (phoT product) and PstC (phoW product), which have six and five transmembrane helices respectively. On the cytoplasmic side of the membrane, the channel is linked to PstB protein which carries a nucleotide (probably ATP)-binding site. PstB probably provides the energy required by the channel to free Pi.

1.3.5.4. Modulation of Pho uptake.

The expression of genes in 0.5-1.0% of the E.coli chromosome (4.7×10⁶ bp) is regulated by the level of external inorganic phosphate based on the proteins detected in two-dimensional gel electrophoresis system developed by F. Nieldhart (personal communication to Rao and Torriani). However Wanner and McSharry (1982), could detect only twenty promoters that were activated by Pi starvation (phosphate-starvation-induced, or Psi). This increased level of expression represents a survival mechanism for the cell and some of these genes constitute the Pho regulon (Fig. 1d). The products of the Pho regulon are proteins of the outer
Fig. 1d. Utilization of organic phosphate by cells of *E.coli* starved of inorganic phosphate (Pi). The organic phosphates permeate the outer membrane (OM) and are hydrolized to Pi by the phosphatases of the periplasm. The Pi produced is captured by the (PiBP) Pi-binding protein (gene *pstS*) and is actively transported through the inner membrane (IM) via the phosphate-specific transport (Pst) system. The *phoU* gene product may act as an effector of the Pi starvation signal and is directly or indirectly responsible for the synthesis of cytoplasmic polyphosphates. More important for the phosphate starved cells is the fact that *phoU* may direct the synthesis of positive co-factors (X) necessary for the activation of the positive regulatory genes *phoR* and *phoB*. The PhoR membrane protein will activate PhoB. The PhoB protein will recognize the 'Pho box' a consensus sequence present in the genes of the pho operon (Surin *et al*., 1987).
Fig. 1c. A model for Pst-dependent Pi transport modified from the one used (Treptow and Shuman, 1988) to explain the mechanism of maltose transport. (a) The PiBP has captured Pi (●); it adapts itself on the periplasmic domain of the two transmembrane proteins (PstC and PstA). The Pst protein is represented as being bound to PstA, but its specific position is not known. It has a nucleotide binding domain (black). (b) When ATP activates PstB, the PiBP is releases Pi to the PstA + PstB channel. (c) The energy released by the hydrolysis of ATP → ADP + Pi is utilized to free the transported Pi. The cycle is closed by the formation of the original structure (a). A, B, and C are PstA, PstB and PstC respectively.
membrane (porin PhoE), the periplasm (alkaline phosphatase, Pi-binding proteins, glycerol 3-phosphate binding protein), and the cytoplasmic membranes (PhoR, PstC, pstA, UgpA, UgpC). Genes coding for these proteins are positively regulated by the products of three genes: phoR and phoB, which are themselves repressed by Pi and phoM which is not. It was first established genetically and then in vitro that the PhoB is required to induce alkaline phosphatase production. The most recent results have established that PhoB is required to bind a specific DNA site upstream the genes of the pho regulon the 'Pho-box' (Nakata et al., 1987) as has been demonstrated directly for the pstS gene (Makino et al., 1988). Gene phoR is Pi-regulated and with phoB constitutes an operon. The present knowledge of the sequence and functions of the these two genes leads to the conclusion that PhoR is a membrane protein (Makino et al., 1985) that fulfils a modulatory function involved in Pi signal transduction. Furthermore, the homology of this operon with a number of two component regulatory systems (Ronson et al., 1987) suggests that PhoR activates PhoB by phosphorylation. This is now supported by the experiments of Yamada et al. (1990), which proved that a truncated PhoR (C-terminal region) is autophosphorylated and transphosphorylates PhoB (Makino et al., 1990).

It has been clearly established that the expression of the genes of the Pst system for Pi transport repressed by Pi. Any mutation in one of these five genes results in the constitutive expression of the Pho regulon. This implies that, the intact Pst structure exerts a negative effect on the expression of the Pho regulon when Pi levels are high. Thus the basal level of these proteins is involved in sensing Pi from the medium, but the transport and flow of Pi through Pst is not involved in the repression of the Pho regulon. If cells are starved for
Pi, the \textit{pst} genes are expressed at high levels and the Pho regulon is induced via \textit{phoR} and \textit{PhoB}. If Pi is added to these cells, \textit{pst} gene expression stops within five to ten minutes, probably because the positive regulators PhoR and PhoB are rapidly inactivated (dephosphorylated) and the Pst proteins may regain their function. Another gene of the Pst, \textit{phoU}, produces a protein involved in the negative regulation of the pho regulon, but the mechanism of this function has not been explained.

Transposons and Tn7 is reviewed because of Tn7-like segment which was encountered downstream the \textit{glmS} gene.

1.3.6. \textbf{Transposable elements.}

Transposons are precise DNA segments which can translocate from place to place in genome or from one replicon to another within a cell. This process does not involve the homologous or general recombination systems of the host, but requires one (or a few gene products) encoded by the element. In prokaryotes, the study of transposable elements dates from the discovery in the late 1960s of a new class of highly polar mutations (Malamy, 1966; Saedler and Starlinger, 1967; Saedler \textit{et al.}, 1968; Shapiro and Adhya, 1969), and from the identification of these mutations as insertions (Starlinger \textit{et al.}, 1968; Shapiro, 1969; Michealis \textit{et al.}, 1969; Malamy, 1970). Hybridization studies showed that the inserted DNA belonged to only a few classes and were called insertion sequences (\textit{IS}; Malamy \textit{et al.}, 1972; Hirsch \textit{et al.}, 1972; Kahn and Schaefer, 1995). Besides their presence on the chromosome, these \textit{IS} were discovered on bacteriophages (Brachet \textit{et al.}, 1970) and on plasmids such as fertility factor, F (Ohtsubo and Davidson, 1975; Hu \textit{et al.}, 1975). It became clear that the insertion sequences could only transpose as discrete units and could be integrated by mechanisms essentially
independent of DNA sequence homology. Furthermore, it was appreciated that the
determinants for antibiotic resistance found on R factors were themselves carried by
transposable elements with properties closely paralleling those of insertion sequences (Hedges
and Jacob, 1974; Gottesman and Rosner, 1975; Kopecko and Cohen 1975; Heffron, et al.,
1975; Berg et al., 1975).
In addition to the central phenomenon of transposition, that is, the appearance of a defined
length of DNA (the transposable element) in the midst of sequences where it had not
previously been detected, transposable elements typically display a variety of other properties.
They can fuse unrelated DNA molecules, mediate the formation of deletions and inversions
nearby; they can be excised; and they can contain transcriptional start and stop signals (Galas
and Chandler, 1989; Starlinger and Saedler, 1977; Sekine et al., 1996). They have been shown
in Psuedomonas cepacia to promote the recruitment of foreign genes creating new metabolic
pathways and to be able increase the expression of neighbouring genes (Scordilis et al., 1987).
They have also been found to generate miniplasmids as well as minicircles, consisting of the
entire insertion sequence and one of the flanking sequences in the parental plasmid (Sekine
et al., 1996). The frequency of transposition may, in certain cases, be correlated with the
environmental stress (Cullis, 1990). Prokaryotic transposable elements exhibit close functional
parallels. They also share important properties at the DNA sequence level. Transposable
insertion sequences, in general may promote genetic and phenotypic diversity of
microorganisms and could play an important role in gene evolution (Holmes et al., 1994).
Partial or complete sequence information is now available for most of the known insertion
sequences and for several transposons.
1.3.6.1. **Transposons and Insertional sequences (IS).**

Transposons were originally distinguished from insertional sequences because transposons carry detectable genes, often conferring antibiotic resistance. Transposons often terminate in long (800-1500 bp) inverted or direct repeats and often these repeat segments are themselves IS or IS-like elements (Calos and Miller, 1980; Boursaux-Eude *et al.*, 1995). Many transposons thus represent a segment of DNA that is mobile as a result of being flanked by IS units. For example, Tn9 and Tn1681 are flanked by copies of ISI, which accounts for the mobilization of the intervening genetic material (Calos and Miller, 1980). It is quite probable that the long inverted repeats of elements such as Tn5, Tn10 and Tn903 are also insertion sequences or are derived from them. Virtually all the insertion sequences and transposons characterized at the sequence level have a terminal inverted repeat. The only exception to date is bacteriophage Mu, where the situation is more complex; the ends share short regions of homology but do not form a convincing inverted repeat (Allet, 1979; Kahmann and Kamp, 1979; Radstrom *et al.*, 1994).

1.3.6.2. **Tn7**

Tn7 is relatively large, about 14 kb (Fig. 1f). It encodes several antibiotic resistance determinants in addition to its transposition functions: a novel dihydrofolate reductase that provides resistance to the anti-folate agent trimethoprim (Fling and Richards, 1983; Simonson *et al.*, 1983), an adenylyl transferase that provides resistance to the aminoglycosides streptomycin and spectinomycin (Fling *et al.*, 1985) and a transacetylase that provides resistance to streptothricin (Sundstrom *et al.*, 1991). Tn1825 and Tn1826 are Tn7-related transposons that apparently encode similar transposition functions but differ in their drug
Fig. 1f. Tn7. Shown are the Tn7-encoded transposition genes *tnsABCDE* and the sizes of their protein products. The *tns* genes are all similarly oriented as indicated by the arrows. Also shown are all the Tn7-encoded drug resistance genes *dhfr* (trimethoprin), *sat* (streptothricin), and *aadA* (spectinomycin/spectomycin). A pseudo-integrase gene (*p-int*) is shown that may mediate rearrangement among drug resistance cassettes in the Tn7 family of transposons.
resistance determinants (Tietze et al., 1987). These drug resistance genes appear to be encoded in genetic cassettes whose rearrangement can be mediated by an element-encoded recombinase (Ouellette and Roy, 1987; Sundstrom and Skold, 1990; Sundstrom et al., 1991). In Tn7, this recombinase gene appears to be interrupted by a stop codon and is, therefore an inactive pseudogene (Sundstrom et al., 1991). It should be noted that the transposition of the intact Tn7 does not require this recombinase (Waddell and Craig, 1988). Tn7 also encodes an elaborate array of transposition genes, tnsABCDE (Fig. 1f). These five tns genes mediate two distinct, but overlapping, recombination pathways (Rogers et al., 1986; Waddell and Craig, 1988; Kubo and Craig, 1990).

1.3.6.2.1. Insertion of Tn7 into E.coli chromosome.

Unlike other transposons, Tn7 is very site and orientation specific. When Tn7 transposes to the E.coli chromosome, it usually inserts in a specific site about 84 min of the 100 min chromosome map called attTn7 (Bath and Datta, 1976; Lichtenstein and Brenner, 1981). The specific point of insertion lies between the phoS and the glmS genes (Lichtenstein and Brenner, 1982; Walker et al., 1986) Fig 1g. In fact, the point of insertion in E.coli is within a region that produces the transcriptional terminator of the glmS gene while the sequence critical for attTn7 activity (called glmS box) encodes the carboxyl terminal 12 amino acids of the glucosamine synthetase enzyme (Waddell and Craig, 1989; Qadri et al. 1989; Walker et al., 1986).

Tn7 will also transpose to regions of DNA with sequence related to attTn7, pseudo attTn7 sites. TnsABC + tnsD promote high-frequency insertion into attTn7 and low frequency insertion into pseudo-attTn7 sites; a different ensemble of tns genes, tnsABC + tnsE mediates
Fig. 1g. The *E. coli* attTn7 region. (A) Physical map of attTn7 region. The upper box is Tn7 (not to scale) showing its orientation upon insertion into attTn7 (Lichenstein and Brenner, 1982). The next line is attTn7 region of the *E. coli* chromosome numbered as originally described by McKown et al. (1988). The middle base pair of the 5-bp *E. coli* sequence usually duplicated upon Tn7 insertion is designated 0, sequence towards *phoS* (leftward) are designated - and sequence towards *glmS* (rightward) are designated +.

(B). Nucleotide sequence of the attTn7 region (Orle et al., 1988). The solid bar indicates the region containing the nucleotides essential for attTn7 activity (Qadri et al., 1989).
low-frequency insertion into sites unrelated to attTn7 (Craig, 1991). Thus, tnsABC provide functions common to all Tn7 transposition events, whereas tnsD and tnsE are alternative target site-determining genes. The tnsD pathway chooses a limited number of target sites that are highly related in nucleotide sequence whereas the tnsE-dependent target sites appear to be unrelated in sequence to each other (or to the tnsD sites) and thus reflect an apparent random insertion pathway (Kubo and Craig, 1990).

As mentioned earlier, a notable feature of Tn7 is its high frequency of insertion into attTn7. For example, examination of the chromosomal DNA in cells containing plasmids bearing Tn7 reveals that up to 10% of the attTn7 sites are occupied by Tn7 in the absence of any selection for Tn7 insertion (Lichenstein and Brenner, 1981; Hauer and Shapiro, 1984). Tn7 insertion into attTn7 has no obvious deleterious effect on E.coli growth. The frequency of Tn7 insertion into sites other than attTn7 is about 100 to 1000-fold lower than insertion into attTn7. Non-attTn7 may result from either tnsE-mediated insertion into random sites or tnsD-mediated insertion into pseudo-attTn7 sites (Rogers et al., 1986; Waddell and Graig, 1988; Kubo and Craig, 1990). The nucleotide sequences of the tns genes have been determined (Smith and Jones, 1986; Flores et al., 1990; Orle and Craig 1990).

Inspection of the tns sequences has not, in general been informative about the functions and activities of the Tns proteins. Perhaps the most notable sequence similarity between a Tns protein and another protein (Flores et al., 1990) is a modest one between TnsC, an ATP-dependent DNA-binding protein that participates directly in transposition (Craig and Gamas, 1992) and MalT (Richet and Raibaud, 1989) which also binds ATP and DNA in its role as a transcriptional activator of the maltose operons of E.coli. Site-specific insertion of Tn7 into
the chromosomes of other bacteria has been observed. These organisms include *Agrobacterium tumefaciens* (Hernalsteens et al., 1980), *Pseudomonas aeruginosa* (Caruso and Shapiro, 1982), *Vibrio* species (Thomson et al., 1981), *Caulobacter crescentus* (Ely, 1982), *Rhodopseudomonas capsulata* (Youvan et al., 1982), *Rhizobium meliloti* (Bolton et al., 1984), *Xanthomonas campestris pv. campestris* (Turner et al., 1984), and *Pseudomonas fluorescens* (Barry, 1986). The ability of Tn7 to insert at a specific site in the chromosomes of many different bacteria probably reflects the conservation of *glmS* in prokaryotes (Qadri et al., 1989).

1.3.6.2.2. **The Tn7 transposition mechanism.**

The development of a cell-free system for Tn7 transposition to *attTn7* (Bainton et al., 1991) has provided a molecular view of this recombination reaction (Craig, 1991). Tn7 moves in *vitro* in an intermolecular reaction from a donor DNA to an *attTn7* target in a non-replicative reaction, that is Tn7 is not copied by DNA replication during transposition (Craig, 1991). Examination of Tn7 transposition to *attTn7* in *vivo* supports the view that this reaction is non-replicative (Orle et al., 1991). The DNA breaking and joining reactions that underlie Tn7 transposition are distinctive (Bainton et al., 1991) but are related to those used by other mobile DNAs (Craig, 1991). Prior to the target insertion in *vitro*, Tn7 is completely disconnected from the donor backbone by double-strand breaks at the transposon termini, forming an excised transposon which is a recombination intermediate (Fig. 1h; Craig, 1991). It is notable that no breaks in the donor molecule are observed in the absence of *attTn7*; thus the transposon excision is provoked by recognition of *attTn7* (Craig, 1991). The failure to observe recombination intermediates or products in the absence of *attTn7* suggests the
transposon ends flanked by donor DNA and the target DNA containing attTn7, are associated prior to the initiation of the recombination by strand cleavage (Craig, 1991). As neither

Fig. 1h. The Tn7 transposition pathway. Shown are a donor molecule containing Tn7 and a target molecule containing attTn7. Translocation of Tn7 from a donor to a target proceeds via excision of the transposon from the donor backbone via double-strand breaks and its subsequent insertion into a specific position in attTn7 to form a simple transposition product (Craig, 1991).
recombination intermediates nor products are observed in the absence of any single recombination protein, or in the absence of attTn7, it is believed that the substrate DNAs assemble into a nucleoprotein complex with the multiple transposition proteins in which recombination occurs in a highly concerted fashion (Bainton et al., 1991). The hypothesis that recognition of attTn7 provokes the initiation of Tn7 transposition is also supported by in vivo studies (Craig, 1995).

A novel aspect of Tn7 transposition is that this reaction involves staggered DNA breaks both at the transposition termini and the target site (Fig. 1h; Bainton et al., 1991). Staggered breaks at the transposon ends clearly expose the precise 3' transposon termini but leave several nucleotides (at least three) of donor backbone sequences attached to each 5' transposon end (Craig, 1991). The 3' transposon strands are then joined to 5' target ends that have been exposed by double-strand break that generates 5' overlapping ends 5 bp in length (Craig, 1991). Repair of these gaps, presumably by the host repair machinery, converts these gaps to duplex DNA and removes the donor nucleotides attached to the 5' transposition strands (Craig, 1991). The polarity of Tn7 transposition (that the 3' transposition ends join covalently to the 5' target ends), is the same as that used by the bacterial elements bacteriophage Mu (Mizuuchi, 1984) and Tn10 (Benjamin and Kleckner, 1989), by retroviruses (Fujiwara and Mizuuchi, 1988) and the yeast Ty retrotransposon (Eichinger and Boeke, 1990).

One especially intriguing feature of Tn7 transposition is that it displays the phenomenon of target immunity, that is the presence of a copy of Tn7 in a target DNA specifically reduces the frequency of insertion of a second copy of the transposon into the target DNA (Hauer and
Shapiro, 1984; Arciszewska et al., 1989). It should be emphasized that the target immunity is a cis-acting phenomenon, that is transposition into DNAs other than that containing the transposon is unaffected and thus immunity reflects the influence of the transposon on the target DNA in which it resides (Craig, 1991). The bacterial transposon Tn3 (Lee et al., 1983) and bacteriophage Mu (Adzumah and Mizuuchi, 1988) also display such target immunity. Tn7 immunity effect is observed over relatively large (>100 kb) molecules (Hauer and Shapiro, 1984; Arciszewska et al., 1989).

1.3.7. Region downstream of E.coli and T.ferrooxidans unc operons.

While examining the atp downstream region of T. ferrooxidans strain ATCC 33020, it was discovered that the genes occur in the same order as E.coli, that is atp_urf glmS Tn7-like transposon (Rawlings, unpublished information). The atp gene cluster from T.ferrooxidans has already been used to complement E.coli F1 unc mutants for growth on minimal media plus succinate (Brown et al., 1994). The second open reading frame (ORF2) in between the two mercury ion resistance genes (merR1 and merR2) in T.ferrooxidans strain E-15 has been found to have high sequence homology with tnsA of transposon Tn7 (Kusano et al., 1991). Since Tn7 was first isolated as part of a conjugative R-plasmid which had spread rapidly through the populations of enteric bacteria as a consequence of the heavy use of antibiotics (Barth et al., 1976), it was not expected to be present in an autotrophic chemolithotroph like T.ferrooxidans. That in itself raises the question of how similar the transposon is to Tn7 and what marker genes it may carry. There is also the question of whether this Tn7-like transposon is found in other strains of T.ferrooxidans and other species of bacteria which grow in its environment.
The objectives of this project were 1); to determine whether the region downstream of the cloned \textit{atp} genes is natural unrearranged \textit{T. ferrooxidans} DNA. 2); to determine whether a Tn7-like transposon is present in other strains of \textit{T. ferrooxidans} as well as metabolically related species like \textit{Leptospirillum ferrooxidans} and \textit{Thiobacillus thioxidans}. 3); to clone various pieces of DNA downstream of the Tn7-like transposon and carry out single strand sequencing to find out how much of Tn7-like transposon is present in \textit{T. ferrooxidans} strain ATCC 33020. 4); to determine whether the antibiotic resistance markers of \textit{Trp}^\prime, \textit{Str}^\prime/\textit{Spr}^\prime and streptothricin which are present on Tn7 are also present in the Tn7-like transposon. 5) whether in \textit{T. ferrooxidans}, the Tn7-like region is followed by the \textit{pho} genes as is the case in \textit{E.coli}. 6) to completely sequence the \textit{glmS} gene and test whether it will be able to complement an \textit{E.coli glmS} mutant.
CHAPTER 2

2.1. Summary 38
2.2. Introduction 38
2.3. Materials and Methods
   2.3.1. Bacterial strains and plasmids 40
   2.3.2. Media 40
   2.3.3. Chromosomal DNA preparation 41
   2.3.4. Restriction enzyme digest 41
   2.3.5. Agarose gel electrophoresis 41
   2.3.6. Preparation of probe 42
   2.3.7. Southern Hybridization 42
   2.3.8. Hybridization 43
   2.3.9. DIG detection 43
2.4. Results and discussion 48
   2.4.1. Restriction mapping and subcloning of *Thermoderovorans* cosmid p818.1 48
   2.4.2. Hybridization of p818.52 to various restriction fragments of cosmid p818.1 and *Thermoderovorans* 48
   2.4.3. Hybridization of p818.52 to chromosomal DNA of *Thermoderovorans*, *Thiothermoderovorans* and *L.thermoderovorans* 50

LIST OF FIGURES

Fig. 2.1. Restriction map of cosmid p818.1 and subclones 46
Fig. 2.2. Restriction map of p818.20 and subclones 47
Fig. 2.3. Restriction digest and hybridization results of p818.52 to cosmid p818.1 and *Thermoderovorans* chromosomal DNA 49
Fig. 2.4. Restriction digest and results of hybridization of p818.52 to *Thermoderovorans*, *Thiothermoderovorans* and *L.thermoderovorans* 51
CHAPTER TWO.

MAPPING AND SUBCLONING OF A 45 KB \textit{T.ferrooxidans'} COSMID (p818.1).

2.1. SUMMARY.

Cosmid p818.1, thought to extend approximately 35 kb downstream of \textit{T.ferrooxidans} \textit{atp} operon was physically mapped. Southern hybridization was then used to show that p818.1 was unrearranged DNA that originated from the \textit{T.ferrooxidans} chromosome. Subclone p818.52 which contained an insert which fell entirely within the Tn7-like element (Chapter 4), was used to make probe to show that, a Tn7-like element is present in three \textit{T.ferrooxidans} strains but not in two \textit{T.thiooxidans} strains or the \textit{L.ferrooxidans} type strain.

2.2. Introduction.

Cosmid p818.1, a 45 kb fragment of \textit{T.ferrooxidans} ATCC 33020 chromosome cloned into pHHC79 vector had previously been isolated by its ability to complement \textit{E.coli} \textit{atp} \textit{F}_{1} mutants (Brown \textit{et al.}, 1994) but had not been studied further. Two other plasmids p\textit{Tfatp1} and p\textit{Tfatp2} from \textit{T.ferrooxidans} chromosome were also able to complement \textit{E.coli} \textit{unc} \textit{F}_{1} mutants. Of these two, p\textit{Tfatp1} complemented only two \textit{unc} \textit{F}_{1} mutants (AN818B' and AN802\varepsilon') whereas p\textit{Tfatp2} complemented all four \textit{E.coli} \textit{F}_{1} mutants tested (Brown \textit{et al.}, 1994). Computer analysis of the primary sequence data of p\textit{Tfatp2}, which has been completely sequenced, revealed the presence of five open reading frames (ORFs) homologous to all the \textit{F}_{1} subunits as well as an unidentified reading frame (URF) of 42 amino acids with
50% and 63% sequence homology to URF downstream of *E.coli* unc (Walker et al., 1984) and *Vibrio alginolyticus* (Krumholz et al., 1989; Brown et al., 1994).

This chapter reports on the mapping of cosmid p818.1 and an investigation to determine whether the cosmid p818.1 is natural and unrearranged *T.ferrooxidans* chromosomal DNA. Furthermore, it reports on a study carried out to determine whether the Tn7-like element was found in other *T.ferrooxidans* strains as well as *T.thiooxidans* and *L.ferrooxidans*. 
2.3. **Materials and Methods.**

Details of solutions and buffers can be found in Appendix 2.

2.3.1. **Bacterial strains and plasmids.**

*T. ferrooxidans* strains ATCC 33020, ATCC 19859, and ATCC 23270, *T. thiooxidans* strains ATCC 19377 and DSM 504, *L. ferrooxidans* strains DSM 2705 as well as cosmid p818.1, plasmids pTfatp1 and pTfatp2 were provided by Prof. Douglas Rawlings. The medium in which they were grown before chromosomal DNA extraction and the geographical location of the original deposit of the bacteria is shown in Table 2.1. *E.coli* JM105 was used as the recipient in cloning experiments and pBluescript SK or pBluescript KSII (Stratagene, San Diego, USA) as the cloning vectors.

2.3.2. **Media**

Iron and tetrathionate medium was made from mineral salts solution (g/l): (NH₄)₂SO₄, 3.0; KCl, 0.1; K₂HPO₄, 0.5 and Ca(NO₃)₂, 0.01 adjusted to pH 2.5 with H₂SO₄ and autoclaved. Trace elements solution (mg/l): FeCl₃, 6H₂O, 11.0; CuSO₄.5H₂O, 0.5; HBO₃, 2.0; Na₂MoO₄.2H₂O, 0.8; CoCl₂.6H₂O, 0.6; and ZnSO₄.7H₂O were filter sterilized. Trace elements (1 ml) solution was added to 100 ml mineral salts solution and to this was added either 50 mM K₂S₂O₇ or 100 mM FeSO₄ and pH adjusted such that the final pH was 2.5 in the case of the tetrathionate medium and 1.6 in the case of iron medium.
2.3.3. **Chromosomal DNA preparation.**

Cells (10μl) were harvested by centrifugation. Washed three times in water adjusted to pH 1.8 using H₂SO₄ and resuspended in 500 μl TE buffer (pH 7.6). SDS (15 μl of a 20% solution) and proteinase K (3 μl of 20 mg/l) were added, mixed and allowed to incubate at 37 °C until cells had lysed and the solution cleared. Proteins and other debris were removed by extracting 3 times with a 25:24:1 solution of phenol/chloroform/isoamyl alcohol. The DNA was precipitated with ethanol, washed in 70% ethanol and resuspended in TE buffer (pH 8.0).

2.3.4. **Restriction enzyme digests.**

Restriction enzymes with their buffers were obtained commercially and were used in accordance with the specifications of the manufacturers. Plasmid p818.52 (5 μg) was restricted with *KpnI* and *SalI* restriction enzymes in a total reaction volume of 50 μl. Chromosomal DNAs of *T.ferrooxidans* strain ATCC 33020 and cosmid p818.1 were digested separately with *BamHI*, *HindIII* and *BgII*. *L.ferrooxidans* strain DSM 2705, *T.ferrooxidans* strains ATCC 33020, ATCC 19859 and ATCC 23270 together with *T.thiooxidans* strains ATCC 19377 and DSM 504 were all digested with *BgII*. Chromosomal DNA (10 μg) and p818.1 (5 μg) were digested in 50 and 80 μl respectively in each case. The standard molecular genetics techniques compiled by Maniatis *et al.* (1982), were followed in the restriction digests.

2.3.5. **Agarose gel electrophoresis.**

Agarose (0.8%) in Tris borate buffer (TBE) pH 8.0 with 1 μl of ethidium bromide (10 mg/ml) per 100 ml was used to separate the chromosomal DNAs, p818.1 and p818.52 (used as a control). The gel was run for 6 hours at 100 V. Half of the probe (p818.52) was run
separately on a 0.6 % low melting point agarose, electro-eluted and resuspended in 50 µl of H₂O.

2.3.6. Preparation of probe.
Labelling of probes, hybridization and detection were done with the digoxygenin-dUTP non-radioactive DNA labelling and detection kit (Boehringer Mannheim). DNA (probe) was denatured by boiling for 10 mins and then snap-cooled in a beaker of ice and ethanol. Hexanucleotide primer mix (2 µl of 10X), 2 µl of 10X dNTP's labelling mix, 5 µl of water and 1 µl Klenow enzyme were added and incubated at 37 °C for 1 hr. The reaction was stopped with 2 µl of 0.2 M EDTA. The DNA was then precipitated with 2.5 µl of 4 M LiCl and 75 µl of ethanol after it had been held at -20 °C for 5 mins. Finally, it was washed with ethanol (70%) and resuspended in 50 µl of H₂O.

2.3.7. Southern hybridization.
The agarose gel with the embedded fractionated DNA was denatured by washing in two volumes of 0.25 M HCl (fresh preparation) for approximately 15 mins at room temp (until stop buffer turned yellow) followed by rinsing with tap water. DNA in the gel was denatured using two volumes of 0.4 N NaOH for 10 mins until stop buffer turned blue again and capillary blotted overnight onto HyBond N+ membrane (Amersham) according to method of Sambrook et al. (1989). The membrane was removed the next day, air-dried and used for pre-hybridization.
2.3.8. **Hybridization.**

The blotted N+ HyBond was pre-hybridized in 50 ml of prehybridization fluid (see Appendix 2) for 6 hrs at 65 °C in a covered box. This was followed by another hybridization in fresh fluid (same as above) to which the probe (boiled for 10 mins and snap-cooled) had been added at 65 °C overnight according to method of Grunstein and Hodgness (1975). Following hybridization, the membrane was washed twice by incubating it in 100 ml wash buffer A (Appendix 2) for 10 mins at room temperature. This was followed by incubation at 65 °C in 100 ml of wash buffer B (Appendix 2) for 15 mins.

2.3.9. **DIG detection.**

All DIG detection reactions were carried out at room temp. The membrane treated in section 2.3.8. was washed in kit wash buffer (Appendix 2) for 5 mins, equilibrated in 50 ml of buffer 2 (Appendix 2) for 30 mins and incubated in anti-DIG buffer for 30 mins. It was then washed twice (15 mins each) in 100 ml wash buffer, after which the wash buffer was drained and incubated for 10 mins with a mixture of 50 μl of Lumigen (AMPDD) in DEA buffer. The DEA buffer was drained, the membrane sealed in bag (SaranWrap) and incubated at 37 °C for 15 mins before exposing the bag to a film (AGFA cuprix RP4) in dark room for 4 hours.
Table 2.1. Details of strains of bacteria used in this study.

<table>
<thead>
<tr>
<th>Bacterial Strain</th>
<th>Medium</th>
<th>Geographical location</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T. ferrooxidans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 22370</td>
<td>$S_4O_6^{2-}$</td>
<td>USA</td>
<td>K.Halberg</td>
</tr>
<tr>
<td>ATCC 19859</td>
<td>$S_4O_6^{2-}$</td>
<td>Canada</td>
<td>ATCC</td>
</tr>
<tr>
<td>ATCC 33020</td>
<td>$Fe^{2+}$</td>
<td>Japan</td>
<td>ATCC</td>
</tr>
<tr>
<td><strong>T. thiooxidans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 19377</td>
<td>$S_4O_6^{2-}$</td>
<td>Libya</td>
<td>K.Halberg</td>
</tr>
<tr>
<td>DSM 504</td>
<td>$S_4O_6^{2-}$</td>
<td>USA</td>
<td>K.Halberg</td>
</tr>
<tr>
<td><strong>L. ferrooxidans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM 2705</td>
<td>$Fe^{2+}$</td>
<td>Armenia</td>
<td>P.Norris</td>
</tr>
</tbody>
</table>
Table 2.2. Constructs and subclones of p818.1.

<table>
<thead>
<tr>
<th>Construct/subclone</th>
<th>Cloning sites</th>
<th>Approx size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p818.20</td>
<td>ApaI-EcoRI</td>
<td>11.0</td>
</tr>
<tr>
<td>p818.1ΔE</td>
<td>EcoRI</td>
<td>34.0</td>
</tr>
<tr>
<td>p818.30</td>
<td>BamHI</td>
<td>2.7</td>
</tr>
<tr>
<td>p818.41</td>
<td>BamHI-KpnI</td>
<td>0.8</td>
</tr>
<tr>
<td>p818.38</td>
<td>SalI-HindIII</td>
<td>1.0</td>
</tr>
<tr>
<td>p818.40</td>
<td>HindIII-BglII</td>
<td>3.4</td>
</tr>
<tr>
<td>p818.52</td>
<td>KpnI-SalI</td>
<td>1.7</td>
</tr>
<tr>
<td>p818.50</td>
<td>KpnI-BglII</td>
<td>2.6</td>
</tr>
</tbody>
</table>

All these subclones and constructs were cloned into pBluescript KSII.
Fig. 2.1. Restriction map of cosmid p818.1. Also shown are the subclones p818.20 (ApaI-EcoRI) and p818.1ΔE (EcoRI-EcoRI) as well as plasmids pTflatp1 and pTflatp2 which complemented *E.coli* unc F<sub>1</sub> mutants (Brown et al., 1994). Apart from p818.1 which was cloned into pHC79, all the other fragments were cloned into vector Bluescript KS.
Fig. 2.2. Subclones made from p818.20 including p818.52, the KpnI-SalI fragment used to prepare the probe. All the subclones were cloned into vector Bluescript KS.
2.4. Results and discussion.

2.4.1. Restriction mapping and subcloning of *T. ferrooxidans* cosmid p818.1

*T. ferrooxidans* cosmid p818.1 subclones, their cloning sites and their sizes are given in Table 2.2. Restriction endonuclease mapping of the constructs carried out and the resulting plasmid maps are given in Figs. 2.1 and 2.2. In the case of cosmid p818.1, there were too many restriction endonuclease sites so this construct was mapped for relatively few enzymes. The most commonly occurring recognition sites were for the *Bam*HI and *Bgl*II restriction enzymes. Cosmid p818.1 contained two *Eco*RI sites which enabled it to be subcloned as two pieces namely p818.20 (covering an *Apa*I-*Eco*RI sites -approx 11 kb) and p818.1ΔE being the *Eco*RI-*Eco*RI piece (approx 34 kb) adjacent to p818.20 (Fig 2.1.). The 11 kb *Apa*I-*Eco*RI p818.1 construct was further subcloned to produce plasmids p818.30, p818.40, p818.50, p818.38, p818.41 and p818.52 (Fig. 2.2). Some of these subclones were extensively mapped although not all the sites are shown in Fig. 2.2. The exact positions of the ends of *T. ferrooxidans* plasmids *pTfatpl* and *pTfatp2* on the cosmid p818.1 were also identified (Fig. 2.1).

2.4.2 Hybridization of p818.52 to various restriction fragments of cosmid p818.1 and *T. ferrooxidans*.

In order to confirm the origin of p818.1 and to demonstrate that the cosmid insert was natural unarranged DNA, digests of cosmid p818.1 and *T. ferrooxidans* chromosomal DNA were probed with p818.52. The sizes of the bands which gave a positive hybridization signal were identical for each of the three different restriction enzyme digests of p818.1 and chromosomal
Fig. 2.3. Hybridization of cosmid p818.1 and *T. ferrooxidans* (ATCC 33020) chromosomal DNA by the *KpnI-SalI* fragment of p818.52 (probe). (a) Autoradiographic image of the restriction digests. Lane x contains the probe, lane 1, 3 and 5 contain p818.1 restricted with *BamHI, HindIII* and *BgIII* respectively. Lanes 2, 4 and 6 contain *T. ferrooxidans* chromosomal DNA also restricted with same enzymes in similar order. (b) The sizes of restricted p818.1 and *T. ferrooxidans* chromosome hybridized by the probe. The lanes correspond to those in Fig 2.3a. *λ* DNA digested with *PstI* served as the molecular weight marker.
DNA (Fig. 2.3). The BamHI digests gave signals at 2.6 and 2.8 kb (lanes 1 and 2), HindIII at 10 kb (lanes 3 and 4) and BglII at 4.6 kb (lanes 5 and 6). The additional weak signals in the p818.1 lanes was because the purified KpnI-SalI probe from p818.52 had a small quantity of contaminating vector DNA (pBluecript KSII) which has regions of homology to the cosmid vector pHC79. The observation that the band sizes of hybridizing fragments were the same for both p818.1 and the T. ferrooxidans ATCC 33020 chromosome for the same restriction digest confirms that, the region extending from BglII site at 4.5 kb to the HindIII site at 16 kb on cosm id p818.1 represents native unrearranged chromosomal DNA from T. ferro oxidans ATCC 33020. Since there were no hybridization signals in addition to those predicted when comparing the map with the probe in lanes 2, 4 and 6, one can deduce that there are no multiple copies of the probe region (a Tn7-like segment) in the chromosome of T. ferro oxidans ATCC 33020.

2.4.3. Hybridization of chromosomal DNA of T. ferrooxidans, T. thiooxidans and L. ferrooxidans.

The insert of plasmid p818.52 was found to fall entirely within a Tn7-like transposon present on the chromosome of T. ferrooxidans 33020 (see Chapter 4). A Southern blot hybridization experiment was carried out to determine whether this Tn7-like element is present on other strains of T. ferrooxidans and/or strains of T. thiooxidans and L. ferrooxidans. The results of this experiment are shown in Fig. 2.4. Lanes D, E and F of Fig. 2.4 represent T. thiooxidans strains ATCC 19377, DSM 504 and L. ferrooxidans strain DSM 2705 respectively, all digested to completion with BglII restriction enzyme. A positive hybridization signal was obtained for each of the three T. ferrooxidans strains; ATCC 33020 (lane A), ATCC 19859 (lane B) and
Fig. 2.4. (a) Autoradiographic image of chromosomal DNA of *T. ferrooxidans* strains ATCC 33020, 19859, 23270 (lanes A, B, C), *T. thiooxidans* strains ATCC 19377 and DSM 504 (D and E) and *L. ferrooxidans* strain DSM 2705 (lane F) all restricted with BgII. λ *PstI* molecular weight marker was used for sizing. (b) Hybridization of the 1.7 kb *KpnI-SalI* piece of p818.52 (probe) to the chromosomal DNA of the organisms mentioned in Fig 2.4a. The lanes in Fig 2.4a correspond to those in Fig. 2.4b.
ATCC 23270 (lane C) examined (Fig. 2.4). The same size BgIII fragments (4.6 kb) were hybridized by the probe as shown in lanes A, B and C. This result indicates that, although the three *T.ferrooxidans* were isolated from different countries and grown on two different media, they all possess a single Tn7-like transposon in apparently the same location in their chromosomes. In contrast, no hybridization signal was obtained for *T.thiooxidans* strains ATCC 19377 and DSM 504 or *L.ferrooxidans* DSM 2705 (lanes D, E and F of Fig. 2.4).

*T.ferrooxidans* and *T.thiooxidans* have been found to be very closely related based on 16S rRNA sequences (Lane *et al.*, 1992). Since all *T.ferrooxidans* and no *T.thiooxidans* strains examined have the Tn7-like transposon, it implies either *T.ferrooxidans* and *T.thiooxidans* diverged before *T.ferrooxidans* acquired the Tn7-like transposon or *T.thiooxidans* does not have an attTn7 attachment site capable of hosting the Tn7-like element. Though these strains of *T.ferrooxidans* were isolated from geographical locations as far apart as the USA and Japan, it is difficult to estimate when the *T.ferrooxidans* strains acquired the Tn7-like transposon as bacteria get around the world pretty quickly and transposons are horizontally transmitted. It remains to be established whether the absence of this Tn7-like segment is a general property of all *L.ferrooxidans* and *T.thiooxidans* strains and also whether all *T.ferrooxidans* strains harbour this Tn7-like element in their chromosomes.
CHAPTER 3

3.1. Summary

3.2. Introduction

3.3. Materials and methods
  3.3.1. Bacterial strains and plasmid vectors
  3.3.2. Media and solutions
  3.3.3. Plasmid DNA extraction
  3.3.4. Agarose gel electrophoresis
  3.3.5. Competent cell preparation
  3.3.6. Transformation of DNA into cells
  3.3.7. Recombinant DNA techniques
  3.3.8. ExonucleaseIII shortening
  3.3.9. DNA sequencing
  3.3.10. Complementation studies

3.4. Results and discussion
  3.4.1. DNA sequence analysis
  3.4.2. Analysis of partial ORF-1
  3.4.3. Glucosamine synthetase gene
  3.4.4. Complementation of E.coli by T.ferrooxidans glmS gene.

LIST OF FIGURES

Fig. 3.1 Map of cosmid p818.1 and p818.20
Fig. 3.2 Plasmidmap of p818.16r
Fig. 3.3 Plasmidmap of p818.16f
Fig. 3.4 Restriction digest of p818.18r and p818.16f with EcoRI-Stul
Fig. 3.5 DNA sequence of the 3.5 kb BamHI-BamHI
Fig. 3.6 Codon preference and bias plot of the DNA sequence of p818.16
Fig. 3.7 Alignment of the C-terminal glmU sequences of E.coli and B.subtilis
  homology to that of T.ferrooxidans
Fig. 3.8 Alignment of amino acid sequences of organisms with high GFAT
  sequence homology to T.ferrooxidans
Fig. 3.9 Phylogenetic relationship between organisms with high glmS
  sequence homology to T.ferrooxidans
CHAPTER 3.

*T. FERROOXIDANS GLUCOSAMINE SYNTHETASE GENE.*

3.1. Summary.

A 3.5 kb *BamHI-BamHI* fragment of p818.20 was cloned into pUCBM20 and pUCBM21 to produce constructs p818.16f and p818.16r respectively. This fragment was completely sequenced from both directions and shown to cover the entire glmS gene (1.84 kb) Fig. 3.1 and 3.4. The derived amino acid sequence of the *T. ferrooxidans* glucosamine synthetase was compared to similar enzymes of other organisms and found to have very high sequence homology. Highest homology was to the glucosamine synthetase of the best studied eubacterium, *E. coli* (Fig. 3.8). Both constructs p818.16f and p818.16r containing the entire glmS gene of *T. ferrooxidans* also complemented an *E. coli* glmS mutant for growth on medium lacking N-acetyl glucosamine.

3.2. Introduction

Cell wall biosynthesis is vital to every microorganism. In most procaryotes, this process starts with amino sugars which are made from fructose-6-phosphate by the transfer of amide group of glutamine to a hexose sugar to form an amino sugar. This reaction is catalysed by glucosamine synthetase, the product of the glmS gene. The first part of the catalysis attributed to the 40-residue N-terminal glutamine-binding domain (Denisot *et al*., 1991), involves the participation of Cys1 to generate a glutamyl thiol ester and nascent ammonia (Buchanan, 1973). The 368-residue C-terminal domain is responsible for the second part of the reaction; formation of glucosamine 6-phosphate. This step has been shown to require the abstraction
of the C1pro-R hydrogen of a putative fructoseamine 6-phosphate to form a cis-enolamine intermediate which upon reprotonation to the face of C2, gives rise to the product (Golinelli-Pimpanneau et al., 1989).

The bacterial enzyme comprises two domains that can be separated by limited chymotryptic proteolysis (Denisot et al., 1991). The glutamine binding domain encompassing residues 1 to 240 has the same capacity to hydrolyse glutamine (and the corresponding p-nitroanilide derivative) into glutamate as the native protein. The amino acid sequence of the glutamine binding domain is highly conserved among members of the F-type group of amidotransferases. Enzymes in this subfamily include amidophosphoribosyl transferase (Tso et al., 1982), asparagine synthetase (Andrulis et al., 1987), glucosamine-6-P synthetase (Walker et al., 1984) and the NodM protein of *Rhizobium leguminosarum* (Surin and Downie, 1988). The 368-residue carboxyl-terminal domain retains the ability to bind fructose-6-phosphate.

The complete double-stranded DNA sequence of *T. ferrooxidans* glmS gene, a third of the glmU gene (preceding glmS) and sequences downstream of glmS gene are reported in this chapter. This chapter also contains a comparison of the sequenced *T. ferrooxidans* glmS gene product to other purF-type amidophosphoribosyl transferases and a complementation study of *T. ferrooxidans* glmS gene in *E.coli* glmS mutant CGSC 5392.
3.3. Materials and Methods.

3.3.1. Bacterial strains and plasmid vectors.

_E. coli_ strain JM109 (endA1 gyrA96 thi hsdR17(rk mK+) relA1 supE44 lac-proAB) (F' traD36 proAB lacI977 Z M15) was used for transformation. _E. coli_ glmS mutant CGSC 5392 was used for the complementation studies. Details of their phenotypes are listed in Appendix 3. Plasmid vectors pUCBM20 and pUCBM21 (Boehringer-Mannheim) were used for cloning of the constructs.

3.3.2. Media and Solutions.

Luria agar (LA) and Luria Broth media (LB) and Luria Broth supplemented with N-acetyl glucosamine (200 μg/ml) were used throughout in this chapter. Luria agar + ampicillin (100 μg/ml) was used to select exonuclease III shortened clones whilst Luria agar + X-gal (50 μl of X-gal + 25 μl IPTG per 20 ml of LA) was used to select for constructs (blue/white selection). Details of media can be found in Appendix 1. X-gal and ampicillin preparation details can be found in Appendix 2.

3.3.3. Plasmid DNA extraction

Pure plasmid DNA extractions were carried out using the Nucleobond kit according to the method developed by MACHEREY-NAGEL for routine separation of nucleic acid; details in Appendix 4. DNA was usually dissolved in 100 μl of TE buffer. Minipreparations of plasmid DNA were carried according to the protocol compiled by Sambrook _et al._ (1989). Miniprepped DNA pellets were usually dissolved in 20 μl of TE buffer and 2 μl
used in restriction endonuclease digests in a total volume of 20 μl.

3.3.4. Agarose gel Electrophoresis.
Agarose gels (0.8%) were used to check all restriction endonuclease reactions. For DNA fragments which were ligated into plasmid vector, low melting point agarose gels (0.6%) were used to separate the DNA fragments of interest.

3.3.5. Competent cell preparation.
*E.coli* strains JM109 and CGSC 5392 competent were prepared by inoculating single colonies into 5 ml LB and shaken vigorously at 37 °C for 2-3 hours. These starter cultures were then inoculated into 100 ml prewarmed LB and shaken at 37 °C until their respective OD₅₅₀'s reached 0.35 units. Each culture was separately transferred into a 2 x 35 ml capped centrifuge tubes, chilled on ice for 15 mins and centrifuged at 2500 rpm for 5 mins at 4°C. Supernatants were discarded and cells were resuspended by gentle vortexing in 10.5 ml ice cold TFB-1 (Appendix 2). After 90 mins on ice, cells were again centrifuged at 2500 rpm for 5 mins at 4°C. Supernatant was again discarded and the cells resuspended gently in 9 ml ice-cold TFB-2 (Appendix 2). The cell suspension was then aliquoted (200 μl) into Eppendorf microfuge tubes and stored at -70 °C until used.

3.3.6. Transformation of DNA into cells.
*E.coli* JM109 and CGSC 5392 cells (200 μl aliquots) were taken from -70 °C and put on ice until cells thawed. Plasmid DNA suitably diluted in TE either from shortening or ligation reactions were mixed with the competent cell suspension and kept on ice for 15 minutes. This
was followed by 5 minutes heat shock (37 °C) and replacement on the ice for another 15
minutes. LB was added (1.0 ml per tube) and incubated for 45 mins before being plated.

3.3.7. **Recombinant DNA techniques.**

General techniques as described by Sambrook *et al.* (1989) were followed. Plasmid constructs
p818.16f and p818.16r were made by ligating the 3.5 kb *BamHI*-*BamHI* into *BamHI* sites in
pUCBM20 and pUCBM21 respectively. Constructs were selected on X-gal plates, mini-
prepared and digested with various restriction enzymes to confirm that the correct construct
had been obtained.

3.3.8. **Exonuclease III shortenings.**

*ApaI* restriction digestion was used to protect both vectors (pUCBM20 and pUCBM21). *MluI*
restriction digestion provided the susceptible site for exonuclease III shortening. Exonuclease
III shortening was done using the protocol from Heinikoff (1984), see Appendix 4. DNA (8-
10 μg) was used in shortening reactions whilst approximately 1 μg of DNA was restricted for
cloning in each case.

3.3.9. **DNA sequencing.**

Ordered deletions of p818.16f and p818.16r (from exonuclease III shortenings) were used as
templates for DNA sequencing. Nucleotide sequence determination was by the dideoxy-chain
termination method (Sanger *et al.*, 1977) using a Sequitherm reaction kit (Epicentre
Technologies) and Alf Express Automatic Sequencer (Pharmacia Biotech). DNA sequence
data were analyzed by using the Genetics Computer Group Inc. software package (Version
Fig. 3.1. Map of cosmid p818.1 and construct p818.20 depicting where p818.20 maps unto p818.1. Below p818.20 are the constructs p818.16f and p818.16r, the 3.5 kb BamHI-BamHI fragment of p818.20 cloned into pUCBM20 and pUCBM21 respectively.
Fig. 3.2. Plasmidmap of p818.16r showing the positions of some of the commonly used restriction enzyme sites. *ApaI* restriction site was used to protect the vector while *MluI* acted as the susceptible site. Also shown is the insert of about 3.5 kb covering part of the *glmU* gene, the complete *glmS* gene and ORF3. The cloning vector is pUCBM21.
Fig. 3.3. Plasmid map of p818.16f showing the cloning orientation of the commonly used restriction enzyme sites of the pUCBM20 vector. *ApaI* restriction site was used to protect the vector while *MluI* acted as the susceptible site. Also shown is the insert of about 3.5 kb covering part of the *T. ferrooxidans glmU* gene, the complete *glmS* gene and ORF3.
Fig 3.4. p818.16r and p818.16f restricted with EcoRI and StuI restriction enzymes. Since the EcoRI site is at opposite ends of these vectors, the StuI-EcoRI digest gave different fragment sizes as illustrated in the diagram beneath. Both constructs are in the same orientation in the vectors. λ PstI molecular weight marker (extreme right lane) was used to determine the size of the bands. In lane x is an EcoRI digest of p818.16.
8.0) and the BLAST subroutines (NCIB, New Bethesda, Maryland, USA).

3.3.10. Complementation studies.

Competent *E.coli glmS* mutant (CGSC 5392) cells were transformed with plasmid constructs p818.16f and p818.16r. Transformants were selected on LA + amp. Transformations of the *E.coli glmS* mutant CGSC 5392 with p818.1 as well as p*Tfatp1* and p*Tfatp2* were also carried out. Controls were set by plating transformed pUCBM20 and pUCBM21 transformed cells as well as untransformed cells on Luria agar plate. Prior to these experiments, the *glmS* mutants were plated on LA supplemented with N-acetyl glucosamine (200 μg/ml) to serve as control.
3.4. Results and discussion

3.4.1. DNA sequence analysis.

Fig. 3.5 shows the entire DNA sequence of the 3.5 kb BamHI-BamHI fragment cloned into pUCBM20 and pUCBM21. The restriction endonuclease sites derived from the sequence data agreed with the restriction maps obtained previously (Fig. 3.1). Analysis of the sequence revealed two complete open reading frames (ORFs) and one partial ORF (Fig. 3.5. and 3.6). Translation of the first partial ORF and the complete ORFs produced peptide sequences with strong homology to the products of the glmU and glmS genes of E.coli and the tnsA gene of Tn7 respectively. Immediately downstream of the complete ORF is a region which resembles the inverted repeat sequences of transposon Tn7. The Tn7-like sequences will be discussed in Chapter 4.

3.4.2. Analysis of partial ORF-1.

This ORF was identified on the basis of its extensive protein sequence homology with the uridylytransferases of E.coli and B.subtilis (Fig. 3.6). There are two stop codons (547 and 577) before the glmS initiation codon ATG (bold and underlined in Fig. 3.5). Detailed analysis of the DNA sequences of the glmU ORF revealed the same peculiar six residue periodicity built around many glycine residues as reported by Ullrich and van Putten, (1995). In the 560 bp C-terminal sequence presented in Fig. 3.7, there are several (L/I/V)G pairs and a large number of tandem hexapeptide repeats containing the consensus sequence (L/I/V)(G/X)X4 which appear to be characteristic of a number of bacterial acetyl- and acyltransferases (Vaara, 1992).
Fig. 3.5. The complete nucleotide sequence of the 3.5 kb \textit{BamHI-BamHI} fragment of p818.16. The sequence includes part of the \textit{T.ferrooxidans glmU} gene (ORF1), the entire \textit{T.ferrooxidans glmS} gene (ORF2) and ORF3 which had high sequence homology to TnsA and inverted repeats of transposon Tn7 (Chapter 4). The deduced amino acid sequence for the open reading frames is shown below the coding sequence. Among the features highlighted are the initiation codons (bold and underlined) of the \textit{glmS} gene (ATG) and ORF3 (TTG), good Shine Dalgarno sequences (bold) immediately upstream the initiation codons of ORF2 and ORF3, and the stop codons (bold) of \textit{glmU} (ORF1) \textit{glmS} (ORF2), and \textit{tnsA} genes. Also shown are the restriction endonuclease sites for some of the enzymes which were used to map p818.16.
B	m
H
I partial ORF1

1	GGATCCCAACGTGTTGTTTCGCGAGGTGCATCTGGGGGCATCGGTCGCCGCTGGGAGC
   AspProAsnValLeuPheValGlyGluValHisLeuGlyHisArgValArgValGlyAla

61	TGGGGCCTTTTGACAGGATCGGAGATTGAGATGAGATCTACCTACCTACG
   GlyAlaValLeuGlnAspAlaArgIleGlyAspAspValGluIleLeuProThrSerHis

121	TATCGGAGGCGCCAGATGCGGCGGCGGCAGGATAGCCTTTTGCGGCGATTCGGC
   IleGluGlyAlaGlnIleGlyAlaGlyAlaArgIleGlyProPheAlaArgIleArgPro

181	CGGACCGAGATAGCCGCGACATATTGCTACATCAGGAGGATGGCGGAGATACCG
   GlyGlyGlyArgHisIleGlyGluArgHisAsnThrValGluValLysAlaAlaLysIle

241	CGGCGCAGGCAGCCAACCCGCAACCTGACATTTACGTTGAGGAGCGAGGATGGCA
   GlyGlySerLysAlaAsnHisLeuSerThrValGluGlnAlaIleGlyThrGlu

301	GGTCGAATAGGCGGCGGCTGGAGATGTCGTTACCTCAGGACACAACTACG
   ValAsnValGlyAlaGluThrIleThrCysAsnThrAspGlyAlaAsnLysHisArgThr

361	CACTGACGGCAATGCTCATCGCTCGGCACAGCCGAGATAGGCTACATTCTG
   IleIleGlyAsnAspValPheIleGlySerSerGlnLeuAlaValProAsnAla

421	CGGCGACGAGCGACGATTGCTGACGCTGAGCTGCTGCGGACGCTGCTGGAGG
   GlyGlyGlyAspValPheIleIleGlyAlaSerGluValIleGluAlaGlu

481	GCTGAGGTTGAGTCGCTGAGCCCACTGACAGAAGCTTACCTCAGGAGG
   LeuThrLeuSerArgSerGlnArgThrIleProHisTrpGlnArgArgArgAsp

541	GAAACTTACCGGATGAGCCGAAAGTGGCTTCGATAACGGGAGAGA
   LysLysEnd.. ORF2  MetCys

B	q
I

601	CGGTATTGTCGGTGGGGTGAGTAAAACAGATCTGGTCCCGATGATTCTGGAGGGGTTGCA
   GlyIleValGlyGlyValSerLysThrAspLeuValProMetIleLeuGlyLeuGin

661	GCCCTGAGATCCTGCTGAGTCGACTTCGCTCCGCGAGATTTGGGGCGGCGACTA
   ArgLeuGluGlyAspArgGluAspSerAspAlaGluLeuIleGlyAspAspAlaAsp

721	TTTTCGCGGTCCGCGGCGGCGGCCTCGGCGAGCTCGCAGCGCCCGCTTGCTGAC
   LeuLeuArgValArgSerValGlyArgValAlaGluThrAlaValAlaValGluAlaGlu

781	TGCGCTTCGAGGTTCCGCGCCACAGCCCTGCCGGCCACCCGCGGCGGCGCCGGG
   GlyLeuGlnGlyValGlyIleGlyThrArgTrpAlaThrHisGlyThrGlu

841	CGACTGAAATCGGCACTCCAGTATTGCTGAGAATGTCGCTGCTGCTGCTG
   GluCysAsnAlaHisProMetIleSerHisGluGlnIleLeuAlaAsnGlyIle

901	CTCCTTACCGACGAGGCTGCTGAGGAGGACTTGCCAGCAGGAGGCTGACG
   IleIleGluAsnPheHisAlaLeuArgAlaHisLeuGluGlyAlaGlyThrThrThrSer

961	CGAGACCTACCGAGGTTCCGATCGCACTGTTGACACATTGCGCATGGCGCATG
   ThrGluAspThrGluValAlaAlaThrArgValAspLeuAspArgGlyAlaAspAlaAspVal

1021	CTGGTTCCGGGCACCCGGGCGGCTGTTGCGCCTATGGCATGCGGG
   LeuPheAlaAlaThrArgArgAlaValGlyAspLeuArgGlyAlaAspAlaAspAlaVal

1081	GATCTCCGGCGGAGATCCGCGGTCGCTGCCAGGATGCGGCCCTCTGCTGCT
   IleSerSerGlyAspProGluThrValCysValAlaArgMetGlyCysProLeuLeuLeu
67

1141 GGGCGTTCGCCGATGATGGCGCATTACGCTCTGGACGTGGCGGCCCTGCTGCCGGTGAC 1200 GlyValAlaAspAspGlyHisTyrPheAlaSerAspValAlaLeuLeuProValThr
1201 CCGCCCGCTTGTATCGTGAACGGCGAGCTGGCCACAGCTGGCCACACGCCAGCCTGGCG 1260 ArgArgValLeuTyrLeuGluAspGlyAspValAslMetLeuGlnArgGlnThrLeuArg
1261 GATTACCGATCAAGGCCGAGCTGCCGAGGCGAAGAAACTGAGACGGCCACTGAGTCG 1320 IleThrAspGlnAlaGlyAlaSerArgGlnArgGluHisTrpSerGlnLeuSer Ala
1321 GGGGAGTGCCTGCCCCCGTCTCCGCTCTGCGACAGACACCGCACTGCGACAGCCGCG 1380 AlaAlaValAlaGluLeuProTyrArgHisPheMetGlnLysGluIleHisGluGlnPro
1381 CCGCGGTTGGCAGATACCTGGAGATTGGCGCTGACATCGGTGATCGGCTGATCCGGA 1440 ArgAlaValAlaAspThrLeuGluGlyAlaLeuAsnSerGlnLeuLeuAspLeuAspLeu
1441 CTGGGAGAGCTTTCTGCCCCCTTTCCGACAGCTGGCCACAGCTGGCCACAGCTGGCC 1500 TrpGlyAspGlyAlaAlaAlaMetPheArgAspValAspArgValLeuPheLeuAlaSer
1501 CGCCACTACGACACTGCGACATTGGCGACATTGGCGACATTGGCGACATTGGCGAC 1560 GlyThrSerHisTyrAlaLeuValTyrGlyArgTrpValGluSerIleValGlyIle
1561 TCCGCGGCAAGGCCAGCTGGGGCACTGCAATTCGCTACCGGAGACTCCATCCGACCGCG 1620 ProAlaAlaGluLeuHisTyrArgTyrArgAspSerAspIleProAspAspProArg
1621 CCAGGTCTTCGAGATCCTGGACCTCTGCGATCCGACCTCGGACCTCGGGGCGGCA 1680 GlnLeuValValThrLeuSerGlnSerGlyGluThrLeuAspThrPheAlaLeuArg
1681 CCGCAGGTTTTGCGCTGCTGGCCACAGCTGGCCACAGCTGGCCACAGCTGGCCAC 1740 ArgAlaLysAspLeuGluHisThrArgThrLeuAlaIleCysAsnValAlaGluSerAla
1741 CAATTCGCGGCTTTGCTGCCACAGCTGGCCACAGCTGGCCACAGCTGGCCACAGCTGGCC 1800 IleProArgAlaSerAlaLeuPheArgPheLeuThrArgAlaGlyProGluIleGlyValAla
1801 CTGGACACAAAGGCTTCTACCAACAGTCTTGCGGAGCTCTATCTGCTGCTCTGCTGCC 1860 SerThrTyrAspGlnLeuAlaLeuLeuLeuLeuThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu
1861 CAAGGCGCCAGGCGCTCTGCGGACAGCTGGGCACTGCAACATCGGACAGCTGGAAGCTGGCC 1920 LysAlaGluGlyAlaSerGluArgCysAlaAlaGlyGlySerProGlyArgLeuArgGln
1921 ACTGCGGCAAGCCATGCTCCGCACTGCGCTACCTGCAACCTGCGGATACGGTTTGCGGCC 1980 LeuProGlySerValGlnHisAlaLeuAsnLeuGluProGlnIleGlnGlyTrpAla
1981 ACGTGGTCCAGCAAGGCAACTGCGCCTTTTTCTGGGGCGCCGCCTGACACTACCACCATTGC 2040 ArgPheArgAlaSerLysThrAlaPheLeuPheGlyArgGlyHisProTyrIleAla
2041 GCTGGAGGCGGCGTTCAAGCTTAAACCTATCTATCCACCCGACAGCTTACCTTGAC 2100 LeuGluGlyAlaLeuLysLeuGluIleSerTyrIleHisAlaGluTyrProAla
2101 GGGGAGTGGAGTGCGGCGTGCCGCCCTGGGCTGGGCGCCACAGCTGGGCTGGGTGTA 2160 GlyGluLeuLeuHisProLeuAlaLeuValAspArgAspMetProValValIle
2161 CGGCGGGAAATGCGCCGCTCCCTGCGAAACAGCTGGGCAGAAGCTGAGACCCACGTCAACGCTGGC 2220 AlaProAsnArgAlaGluLeuLeuGluIleLeuAlaAsnMetGlnGluHisValArgAla
2221 TGCAGGTGACTCAGTGGTTGCGGATCACGCCCCCATTTAACCCGAGCGGCGGCGG 2280 GlyGluLeuTyrValPheAlaAspSerAspSerHisPheAsnAlaSerAlaGlyVal
2281 GCATGTGATCGTGTGCTGCACGGCGGTCTGCTCTCCCCATCGTCCATGCTATCCCTTG
2340 HisValMetArgLeuProArghisAlaGlyLeuLeuSerProIleValHisAlaIlePro -
2341 GGTGACGGTGCTGCTCATATGCCGCGGCGGGGTAAGGGGACGGATGGGTCGACGGCGG
2400 ValGlnLeuAlaTyrHisAlaAlaLeuValLysGlyThrAspValAspArgProArg -
2401 TAACCTCGGCAAAGCGGCGGCTGACGATGTTGTTGCGCAGGGGTGTTGTT -
2460 AsnLeuAlaLysSerValThrValGluEnd
End of gIms gene.

2461 AACGGCAATAGATGATCATCATGCTGGCACAAGTAGGTTGCTCCCACTCAGGACACGGCGG
2520 ArgGlu

2581 AATGATCAGCTAGGAAGGATGCCATCGGACGCCAACGCTGAGGTGCTGCAGACAGCGG
2640 LeuAlaArgGlnArgTyrGlyValAspGluAspArg -

2641 GTGCACGCTTCCAAAAGGAGGGAGCTTGTCAGGGCGTGGCGCAGACTACCCACCCCTGG
2700 ValAlaArgPheGlnLysGluGlyArgGlyGlnGlyArgAlaAspThrHisProTrp -
2701 CTACCACTCAAGCCGCTCCCTCCCAAGGCGGCTCCGACGATCCAGGCGATCGACGAGG
2760 LeuThrIleGlnAspValProSerGlnGlyArgSerHisArgLeuLysGlyIleThr -
2761 GGGCGAGTCGCTCAGCTCATATTGAGCGCAGACATCTCTCTGATATGGG
2820 GlyArgValHisLeuHisLeuSerAspIleGluArgPheThrLeuPheAspTrp -
2821 GCAGACGCTCCGAGTCAACCCGCGGACACACTTCGCTGAACTCCGACATCTCGGCG
2880 AlaAspAlaValThrAspIleArgGluGlnPheProLeuAsnArgAspIleThrArg -
2881 ATTCGCGATGACCTGGGTGCTCATCATCCCCGGATGTTGCACGGCGGACCCCTCTTCGTC
2940 ValIleAspAspLeuGlyValIleHisProArgAspValGlyGlyThrThrValValLeu -
2941 ATGACCACGAGACTTTTCTTGTTAGACAGATCAGCTGCTATGCTGGTGAACCTGGGCGG
3000 MetThrThrAspPheLeuValAspThrIleHisGlyArgMetGlnValAlaArg -
3001 GCGGAAAAACCCGCGTGAACAATCTGAAACCCGCGGCGTGGTTGGCAAAACTGAGATGGA
3060 AlaValLysProAlaGluLeuGluLeuGluProArgValGluValValGluIleGlu -
3061 CGCGTTATGGGGCGCCAGTAGGCGAGATTGGGGCGGCCGCTGACGGCGGACATCCCG
3120 ArgArgThrTrpAlaGlnGluLeuAspValGlyValThrValGluArgAspIlePro -
3121 AAAGCGATGGTTCGCAATATCGCCTGCTGCTGCTCAGTAATGCGCTGCTGCTGCTGCT
3180 LysAlaMetValArgAsnIleAlaTrpValHisSerTyrAlaValIleAspGlnMetSer -
3181 CGACCCCTACGACGCTACTACGTGAGAAAAGCAAAGCGCTGGTTTACAGAAACTTCCGCTG
3240 GlnProTyrAspGlyTyrAspGlyAlaArgLysLeuValLeuArgGluLeuGluProSer -
3241 CACCCCGGCGCTACCTGGCGCGCAATTCTGGCGCCGATCTGCGAGTTTCTATGCTGCT
3300 HisProGlyProThrLeuArgGlnPheCysAlaAspMetAspLeuGlnPheSerMetSer -
3301 GCTGGTGACTGTCCTCCTCCTAATTGCCACCTGCTGCAGGCAAGGCTAAGCTGCTGCT
3360 AlaGlyAspCysLeuLeuLeuIleArgHisLeuLeuAlaThrLysAlaAsnValCysPro -

3361 ATGGACGGGCTTACGAGCAGCTCCAGCTTTTTCGTCAGTTTGAGTGCGACGGAAGGTTTAA
3420 MetAspGlyProThrAspAspSerLysLeuLeuArgGlnPheArgValAlaGluGlyGlu -
3421 TCCAGGGCCAGCGGATGAGGATCTGTCGTTCAACCAATTGCTGGAATACCCCGATG 3480
SerArgArgAlaSerGlyEnd

B
a
m
H
i

3481 AAGGAAGATCGAAAGGATCC 3501
Fig. 3.6. Codon preference and bias plot of nucleotide sequence of the 3.5 kb p818.16 BamHI-BamHI fragment. The codon preference plot was generated using the an established codon usage table for *T. ferrooxidans*. The partial open reading frame (ORF1) and the two complete open reading frames (ORF2 and ORF3) are shown as open rectangles with rare codons shown as bars beneath them.
Fig. 3.7. Alignment of C-terminal 120 UDP-N-acetylglucosamine pyrophosphorylase (GlmU) of *E. coli* (E_c) and *B. subtilis* (B_s) to GlmU from *T. ferrooxidans*. Amino acids are identified by their single letter codes and the sequences are read from the amino to the carboxyl terminal. The consensus amino acids to *T. ferrooxidans* glmU are in bold.

<table>
<thead>
<tr>
<th>251</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_f</td>
<td>DP NVLFVGEVHL GHRVRVGAGA</td>
</tr>
<tr>
<td>E_c</td>
<td>DT NVIEGNVIL GHRVKI GTGC</td>
</tr>
<tr>
<td>B_s</td>
<td>YP GTVI KGEVQI GEDTII GPHT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>301</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_f</td>
<td>VLDARIGDD VEILPYSHIE GAQIGAGARI GPFA RIRPGT E.IGERHIGN</td>
</tr>
<tr>
<td>E_c</td>
<td>VIKNSVIGDD CEISPYTVVE DANLA AACTI GPFA RIRPGA ELLEGARVGN</td>
</tr>
<tr>
<td>B_s</td>
<td>EIMN SAI GSR TVIKQ.SVVN HSK VGD NVNI GPFA HIRPDS VIGNEV KIGN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>351</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_f</td>
<td>YYEVKAAKIG AGSKANHLSY LGDAEIGTVQ NVGAGTTICN YDGANKHR TI</td>
</tr>
<tr>
<td>E_c</td>
<td>FVEMKRAR LG KGSA GHLTY LGDAEIGDNV MIGAGTTICN YDGANK FKTI</td>
</tr>
<tr>
<td>B_s</td>
<td>FVEIKK1QFG DRSKASHLSY VGD AE VGDV N LGCGSITVN YDGK NKLTK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>401</th>
<th>450</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_f</td>
<td>IGDVFIGSD SQLVAPV NIG DGATIGAGST ITKEVFPPGL TLSPRPT QTI</td>
</tr>
<tr>
<td>E_c</td>
<td>IGDDVFVGSD TQLVAPVT VG KGATIAAGST VTRNVGENAL AISRP PQTQ K</td>
</tr>
<tr>
<td>B_s</td>
<td>IEDGAFICGN SNLVA PVT VG EGAYVAAGST VTED FP G KAI AIA RQVN K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>451</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_f</td>
</tr>
<tr>
<td>E_c</td>
</tr>
<tr>
<td>B_s</td>
</tr>
</tbody>
</table>
3.4.3. Glucosamine synthetase gene.

A second complete open reading frame (ORF-2) of 1.83 kb is shown in Fig. 3.5. A BLAST search of the GenBank and EMBL data bases indicated that ORF-2 was homologous to the glmS gene product of *E. coli* (47.8 % identical amino acids in the predicted protein sequences), *Haemophilus influenza* (51.9%), *Bacillus subtilis* (39.1%), *Saccharomyces cerevisiae* (39.4 %) and *Mycobacterium leprae* (42.2 %). An interesting observation was that, the *T. ferrooxidans* glmS gene had comparatively high homology sequences to the *Rhizobium leguminosarum* and *Rhizobium meliloti* nodulation protein M; the amino acid identity to each was 44% and 43.6% respectively. A consensus Shine Dalgarno sequence (AGGAG) upstream of the start codon (ATG) is highlighted in Fig. 3.5. No *E. coli* σ70-type promoter consensus sequence was detected in the 200 bp preceding the start codon. Based on the short intergenic distance between the two genes and the absence of any promoter consensus sequence, Plumbridge et al. (1993) suggested that the *E. coli* glmU and glmS genes were co-transcribed. In the case of the glmU-glmS region of the *T. ferrooxidans*, the situation appears to be similar. The sequences obtained have very high homology to six other amidotransferases (Fig. 3.8) which are designated purF-type (named after the purF-encoded glutamine phosphoribosyl pyrophosphate amidotransferase) Zalkin and Weng, (1987).

The glutamine amide transfer domain of approximately 194 amino acid residues is at the NH2 terminus of the protein chain. Zalkin and Mei (1989), using site-directed mutagenesis to replace several of the 9 invariant amino acids in the glutamine amide transfer domain of glutamine phosphoribosylpyrophosphated amidotransferase indicated in their report that, a Cys1-His101-Asp29 catalytic triad is involved in the glutamine amide transfer function of these
Fig.3.8. Comparison of the amino acid sequence alignment of glucosamine-6-phosphate amidotransferases (GFAT) of *Rhizobium meliloti* (R_m), *Rhizobium leguminosarum* (R_l), *E.coli* (E_c), *H.influenzae* (Hae_in), *Mycobacterium leprae* (M_l), *B.subtilis* (B_s) and *S.cerevisiae* (S_c) to that of *T.ferrooxidans* T_f. Amino acids are identified by their single letter codes. The asterisks (*) represent homologous amino acids of *T.ferrooxidans* glmS to at least two of the other organisms. Consensus amino acids to all eight organisms are highlighted (bold and underlined).
enzymes. His$^{101}$, which has been suggested by the authors to increase the nucleophilicity of Cys1 is believed to form a glutamine enzyme covalent intermediate. In these enzymes, the essential cysteine required for the formation of the covalent glutaminyl intermediate is the NH$_2$-terminal residue in the mature enzyme. The translated sequence of the \textit{T.ferrooxidans glmS} gene revealed a triad of Cys$^1$-His$^{102}$-Asp$^{29}$ which corresponds closely to the catalytic triad of the glutamine amide transfer domain suggested by Zalkin and Mei, (1990). Lys603, which is conserved in the C-terminal sequences of \textit{E.coli}, \textit{Saccharomyces cerevisiae}, and nodulation protein NodM from \textit{Rhizobium} (Surin and Downie, 1988) is also conserved in the same position in the \textit{glmS} of \textit{T.ferrooxidans} (seen as Lys721 in Fig. 3.8). In fact, the last 12 amino acid of all the amidotransferases (Fig. 3.8) are highly conserved. GlmS has been found to be inactivated by iodoacetamide (Badet \textit{et al.}, 1987), the glutamidine analogue 6-diazo-5-oxonorleucine (Badet \textit{et al.}, 1987), and N$^3$-fumaroyl-L-2,3-diaminopropionate derivatives (Kucharczyk \textit{et al.}, 1990), thus offering a potential target for antibacterial and antifungal agents (Andruszkiewicz \textit{et al.}, 1990; Badet-Denisot \textit{et al.}, 1993). Whether that is also the case for the \textit{T.ferrooxidans glmS} gene is worth investigating considering the high amino acid sequence homology to the other \textit{glmS} genes and the need to control \textit{T.ferrooxidans} growth to reduce pollution arising from acid mine drainage.

3.4.4. Complementation of \textit{E.coli} CGSC 5392 by \textit{T.ferrooxidans ATCC 33020 glmS} gene.

The complementation of an \textit{E.coli} \textit{glmS} mutant for growth on LA to which no N-acetyl glucosamine had been added is given in Table 3.1. As indicated earlier, the complete \textit{T.ferrooxidans glmS} gene was cloned as a 3.5 kb \textit{BamHI}-\textit{BamHI} fragment of p818.20 into pUCBM20 and pUCBM21 (p818.16f and p818.16r). In both constructs, the \textit{glmS} gene is
oriented in the 5'-3' direction with respect to the lacZ promoter of the cloning vectors. Complementation was detected by the growth of large colonies on Luria agar plates compared to E.coli mutant CGSC 5392 transformed with vector. Untransformed mutant cells produced large colonies only when grown on Luria agar supplemented with N-acetyl glucosamine (200 μg/ml). No complementation was observed for the E.coli glmS mutant transformed with pTfatpl and pTfatp2 as neither construct contained the entire glmS gene. Attempts to complement the E.coli glmS mutant with p818.1 and p818.20 were also not successful even though they contained the entire glmS gene. The reason for non-complementation of p818.1 and p818.20 could be that, the natural promoter of the T.ferrooxidans glmS gene is not expressed in E.coli.
Table 3.1. Complementation of *E. coli* glmS mutant CGSC 5392 by various constructs containing DNA fragments downstream the *T. ferrooxidans* ATCC 33020 unc operon.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>p818.16f</td>
<td>+</td>
</tr>
<tr>
<td>p818.16r</td>
<td>+</td>
</tr>
<tr>
<td>p818.1</td>
<td>-</td>
</tr>
<tr>
<td>p818.20</td>
<td>-</td>
</tr>
<tr>
<td>pTfatp1</td>
<td>-</td>
</tr>
<tr>
<td>pTfatp2</td>
<td>-</td>
</tr>
<tr>
<td>pUCBM20</td>
<td>-</td>
</tr>
<tr>
<td>pUCBM21</td>
<td>-</td>
</tr>
<tr>
<td>+ complementation</td>
<td>non-complementation</td>
</tr>
</tbody>
</table>


Fig 3.9. Deduced phylogenetic relationships between acetyl/acyltransferases (amidotransferases) which had high sequence homology to the T.ferrooxidans glmS gene; *Rhizobium melioti* (Rhime), *R.leguminosarum* (Rhilv), *E.coli* (Ec), *H.influenzae* (Haein) *T.ferrooxidans* (Tf), *S.cerevisiae* (Sacc), *M.lepra* (Myle), and *B.subtilis* (Bacsu).
CHAPTER 4.

4.1. Summary 83

4.2. Introduction 84

4.3. Materials and methods 86
   4.3.1. Bacterial strains and plasmids 86
   4.3.2. DNA constructs, subclones and shortenings 86
      4.3.2.1. Construct p818.52 86
      4.3.2.2. Construct p818.09 87
      4.3.2.3. Plasmid p818.09 and its shortenings 87
      4.3.2.4. p818.11 87

4.4. Results and discussion 88
   4.4.1. Analysis of the sequences downstream the gliS gene 88
   4.4.2. TnsBC homology 96
   4.4.3. Homology to the TnsD protein 99
   4.4.4. Analysis of DNA downstream of the region with TnsD homology 118

LIST OF FIGURES

Fig. 4.1. Alignment of Tn7-like sequence of *T. ferrooxidans* to *E.coli* Tn7 87
Fig. 4.2. DNA sequences at the proximal end of Tn5468 and *E.coli* gliS gene 88
Fig. 4.3. Comparison of the inverted repeats of Tn7 and Tn5468 89
Fig. 4.4. BLAST results of the sequence downstream of *T. ferrooxidans* gliS 90
Fig. 4.5. Alignment of the amino acid sequence of Tn5468 TnsA-like protein to TnsA of Tn7. 92
Fig. 4.6. Restriction map of the region of cosmid p818.1 with amino acid sequence homology to TnsABC and part of TnsD of Tn7.

Fig. 4.7. Restriction map of cosmid p818.1ΔE with its subclones and shortenings.

Fig. 4.8. BLAST results and DNA sequence of p818.52r.

Fig. 4.9. BLAST results and DNA sequence of p818.52f.

Fig. 4.10. BLAST results and DNA sequence of p818.09.

Fig. 4.11. Diagrammatic representation of the rearrangement of Tn5468 TnsD-like protein.

Fig. 4.12. Restriction digests on p818.1ΔE.

Fig. 4.13. BLAST results and DNA sequence of p818.10EM.

Fig. 4.14. BLAST results on joined sequences of p818.104 and p818.103.

Fig. 4.15. BLAST results and DNA sequence of p818.102.

Fig. 4.16. BLAST results and nucleotide sequence of p818.101.

Fig. 4.17. BLAST results and DNA sequence of the combined sequence of p818.11f and p818.10r.

Fig. 4.18. Results of the BLAST search on p818.11r and the DNA sequence obtained from p818.11r.

Fig. 4.19. Comparison of the spo operons of *E.coli*, *H.influenzae* and probable structure of *T.ferrooxidans* spo operon.
CHAPTER FOUR.

TN7-LIKE TRANSPOSON OF T.FERROOXIDANS.

4.1. Summary

Various constructs and subclones including exonuclease III shortenings were produced to gain access to DNA fragments downstream of the T.ferrooxidans glmS gene (Chapter 3) and sequenced. Homology searches were performed against sequences in GenBank and EMBL databases in an attempt to find out how much of a Tn7-like transposon and its antibiotic markers were present in the T.ferrooxidans chromosome (downstream the glmS gene). Sequences with high homology to the TnsA, TnsB, TnsC and TnsD proteins of Tn7 were found covering a region of about 7 kb from the T.ferrooxidans glmS gene.

Further sequencing (± 4.5 kb) beyond where TnsD protein homology had been found, revealed no sequences homologous to TnsE protein of Tn7 nor to any of the antibiotic resistance markers associated with Tn7. However, DNA sequences very homologous to E.coli ATP-dependent DNA helicase (RecG) protein (EC 3.6.1.) and guanosine-3’5-bis (diphosphate) 3’-pyrophosphohydrolase (EC 3.1.7.2) stringent response protein of E.coli, H.influenzae and Vibrio sp. were found about 1.5 kb and 4 kb respectively downstream of where homology to the TnsD protein had been detected.
4.2. Transposable insertion sequences in *Thiobacillus ferrooxidans*.

The presence of two families (family 1 and 2) of repetitive DNA sequences in the genome of *T. ferrooxidans* has been described previously (Yates and Holmes, 1987). One member of the family was shown to be a 1.5 kb insertion sequence (IST2) containing open reading frames (ORFs) (Yates *et al.*, 1988). Sequence comparisons have shown that the putative transposase encoded by IST2 has homology with the proteins encoded by IS256 and ISRm3 present in *Staphylococcus aureus* and *Rhizobium meliloti* respectively (Wheatcroft and Laberge, 1991). Restriction enzyme analysis and Southern hybridization of the genome of *T. ferrooxidans* is consistent with the concept that IST2 can transpose within *T. ferrooxidans* (Holmes and Haq, 1990). Additionally, it has been suggested that transposition of family 1 insertion sequences (IST1) might be involved in the phenotypic switching between iron and sulphur oxidizing modes of growth, including the reversible loss of the capacity of *T. ferrooxidans* to oxidise iron (Schrader and Holmes, 1988).

The DNA sequence of IST2 has been determined and exhibits structural features of a typical insertion sequence such as target-site duplications, ORFs and imperfectly matched inverted repeats (Yates *et al.*, 1988). A transposon-like element Tn5467 has been detected in *T. ferrooxidans* plasmid pTF-FC2 (Rawlings *et al.*, 1995). This transposon-like element is bordered by 38 bp inverted repeat sequences which has sequence identity in 37 of 38 and in 38 of 38 to the *tnpA* distal and *tnpA* proximal inverted repeats of Tn21 respectively. Additionally Kusano *et al.* (1991), showed that of the five potential ORFs containing *merR* genes in *T. ferrooxidans* strain E-15, ORFs 1 to 3 had significant homology to TnsA from transposon Tn7.
Analysis of the sequence at the terminus of the 3.5 kb BamHI-BamHI fragment of p818.20 revealed an ORF with very high homology to TnsA protein of the transposon Tn7 (Fig. 4.4). Further studies were carried out to determine how much of the Tn7 transposition genes were present in the region further downstream of the *T.ferroxidans glmS* gene. Tn7 possesses trimethoprim, streptomycin, spectinomycin and streptothricin antibiotic resistance markers. Since *T.ferroxidans* is not exposed to a hospital environment, it was of particular interest to find out whether similar genes were present in the Tn7-like transposon. In the *E.coli* chromosome, the Tn7 insertion occurs between the *glmS* and the *pho* genes (*pstS, pstC, pstA, pstB* and *phoU*). It was also of interest to determine whether *atp-glm-pst* operon order holds for the *T.ferroxidans* chromosome. It was these questions that motivated the study of this region of the chromosome.
4.3. **Materials and Methods.**

4.3.1. **Bacterial strains and plasmids.**

Bacterial strains and plasmids used were the same as in Chapter 3. Media and solutions (as in Chapter 3) can be found in Appendix 2. Plasmid DNA preparations, agarose gel electrophoresis, competent cell preparations, transformations and recombinant DNA techniques were all carried out as in Chapter 3. Probe preparation, Southern blotting, hybridization and DIG detection methods were as in Chapter 2. The same procedures of nucleotide sequencing and DNA sequence analysis described in Chapters 2 and 3 were used.

4.3.2. **DNA constructs, subclones and shortenings.**

(see Figs. 4.5 and 4.6)

4.3.2.1. **Construct p818.52**

Plasmid p818.52 is a *KpnI-SalI* subclone of p818.20 in the Bluescript vector KS (1.7 kb) and was described in Chapter 2 where it was used to prepare the probe for Southern hybridization. DNA sequencing was from both the *KpnI* (p818.52r) and from the *SalI* restriction sites (p818.52f).

4.3.2.2. **Construct p818.09.**

Construct p818.09 was made by *EcoRI-HindIII* digestion of p818.20. The resulting fragment (about 4.2 kb) was then ligated to a Bluescript vector KS+ (restricted with the same enzymes). DNA sequencing was carried out from the *EcoRI* end of the construct.
4.3.2.3. **Plasmid p818.10 and its shortenings.**

The 4.0 kb *Eco*RI-*Cla*I fragment of p818.1ΔE cloned into Bluescript was called p818.10. Exonuclease III shortening of the fragment was based on the method of Heinikoff (1984). The vector was protected with *Apa*I, and *Cla*I was used as the susceptible site for exonuclease III. Another construct p818.10EM was made by digesting p818.10 and pUCBM20 with *Mlu*I and *Eco*RI and ligating the approximately 1 kb fragment to the vector.

4.3.2.4. **p818.11**

A 2.5 kb *ApaI-ClaI* restriction digest of p818.1ΔE was cloned into Bluescript vector KS+ and sequenced from both ends.
4.4. Results and discussion.

4.4.1. Analysis of the sequences downstream the glmS gene.

A comparison of the nucleotide sequences of the glmS gene termini (approximately 170 bp downstream) between *T. ferrooxidans* and *E. coli* is shown in Fig. 4.1. This region includes the *att*Tn7 site of Tn7 insertion within the chromosome of *E. coli* and the imperfect inverted repeat sequences of Tn7 (Fig. 4.3). There was marked homology at both the nucleotide and amino acid sequence levels within the glmS gene but this homology decreased substantially beyond the stop codon. The "GCGGG" which has been associated with target duplication at the insertion at *att*Tn7 by Tn7 in *E. coli* (Lichtenstein and Brenner, 1982) appears as "CCGGGG" in *T. ferrooxidans* (bold and underlined in Figs.4.1 and 4.2). The two sequences "CCGGGG" and "GCGGG" are almost equidistant from their respective glmS translational stop codons. The alignment of Tn7 and the Tn7-like transposon indicated that although there is some homology between the transposons in the regions which includes the imperfect inverted repeats sequences, the homology appears to be random thereafter (Fig. 4.1). The inverted repeats of Tn7 are very similar to the inverted repeats of the Tn7-like element of *T. ferrooxidans* with gAcaAtagAgt*tcat*c**aa (Fig. 4.3) being conserved in the eight repeats (four each). The transposon was registered with Esther Lederberg of Stanford University, California, as Tn5468.

A search on the databases (GenBank and EMBL) with the nucleotide sequence downstream of those in Fig. 4.1 showed high homology to the TnsA protein of Tn7 (Fig. 4.4). Comparison of the predicted amino acid sequence of the TnsA-like protein of Tn5468 to the predicted
Fig. 4.1

Alignment of the DNA sequence of the Tn7-like element in *T. ferrooxidans* to the *E. coli* Tn7 sequence at the site of insertion into the transcriptional termination site of *glmS*. The intermittent homology shown below occurs within the imperfect inverted repeats (underlined) of both the Tn7-like element and transposon Tn7 (Fig. 4.3). Homology becomes random just before the end of their corresponding imperfect repeats.

<table>
<thead>
<tr>
<th>T</th>
<th>V</th>
<th>E</th>
<th>*</th>
<th>*</th>
</tr>
</thead>
</table>
| 10 | 20 | 30 | 40 | 50 | 60

**Ec. Tn7**

ACGTTGAGTAAATGGATGCCCTGCCTAACGGGGTGTGGGCCGACAAATGAATCTTA

**Tf.7-**

ACGTTGAGTAAATGGGCCGTTCCGGTTAACGACCTGATCGATTA

(like)

| 10 | 20 | 30 | 40 | 50 | 60 |

**Ec. Tn7**

AACTGAACAAATGATCTCAAACTATGACAAATTAAAGGCTTTAAGCAATGACAGG

**Tf.7-**

TTCTGACAATAGATCTTCTCCGAAATAAAGTACTTGATTCCTCAACCAAATAGAGAT

(like)

| 70 | 80 | 90 | 100 | 110 | 120 |

**Ec. Tn7**

AAACTGAACTCACTGCTCTGTTGGTTGGGGCGGGTTTGGGCTAAGG

**Tf.7-**

CATCCGCTTGGCTCCGGAGGATGTGGAGTTCTGGCCCTAATGACAGCTAGGAGA

(like)

| 130 | 140 | 150 | 160 | 170 | 180 |
Fig. 4.2. DNA sequences at the 3' end of the *E. coli* glmS gene and the tnsA proximal of transposon Tn7 compared to the DNA sequences at the 3' end of the *T. ferrooxidans* glmS gene and the proximal end of the Tn7-like transposon. (a) DNA sequence determined in the *E. coli* strain GD92::Tn7 in which Tn7 has been inserted into the glmS transcriptional terminator (Walker et al., 1986). Nucleotide 38 onwards are the sequence of the left end of Tn7. (b) DNA sequence determined in *T. ferrooxidans* strain ATCC 33020 with the insertion of the Tn7-like transposon at the transcriptional termination site of the glmS gene. Also shown are the 22 bp imperfect repeats of the Tn7-like transposon as well as the region where homology to the tnsA protein begins. A good Shine Dalgarno sequence is shown immediately upstream of what appears to be the TTG initiation codon for the transposon.

(a)

(b)
Fig. 4.3. Inverted repeats: (a) Alignment of the sequences of the four tnsA proximal inverted repeats of Tn7. Consensus bases among the four different repeats are highlighted. (b) Alignment of the nucleotide sequences of four proximal repeats of the Tn7-like element from T. ferrooxidans. (c) Nucleotide sequence alignment of the consensus bases of Tn7 repeats with that of the Tn7-like element. The nucleotide bases marked with asterisks have equal presence (2 out of four) with the nucleotide beneath it.

(a) TN7

REPEAT 1: GACAATAAAG TCTTAAACTG AA
REPEAT 2: AACAAAATAG ATCTAAACTA TG
REPEAT 3: GACAATAAAG TCTTAAACTA GA
REPEAT 4: GACAATAAAG TCTTAAACTA GA
CONSENSUS NUCLEOTIDES gACAAtAaAG tctTAAACTa *a

(b) TN7-LIKE

REPEAT 1: GACAATAGAG TATCATTCTG GA
REPEAT 2: GACAATAGAG TTTCATCCCCG AA
REPEAT 3: AACAAATAAG TATCATCCTC AA
REPEAT 4: AACAAATAAG TATCATCCTG GC
CONSENSUS (all four) gACAAATAgAG TaTCATcCtgc aa

(c)

Transposon Tn7 GAcAAttAgAg tttcatTccc ga
Tn7-like element gACAAATAgAG TaTCATcCtgc aa
gACAAAtAgAg t*tcat*c** aa
**Fig. 4.4. Results of the BLAST search obtained using the DNA sequence downstream of the glmS gene from 2420-3502. Consensus amino acids are in bold.**

<table>
<thead>
<tr>
<th>Probability</th>
<th>Sequences producing High-scoring Segment Pairs:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading High</strong></td>
<td><strong>Frame Score P(N)</strong></td>
</tr>
<tr>
<td></td>
<td>sp</td>
</tr>
<tr>
<td></td>
<td>piri</td>
</tr>
<tr>
<td></td>
<td>piri</td>
</tr>
<tr>
<td></td>
<td>piri</td>
</tr>
</tbody>
</table>

**>sp|P13988|TNSA_ECOLI TRANSPOSASE FOR TRANSPOSON TN7. >piri|S12637|S12637 tnsA protein - Escherichia coli transposon Tn7 >gp|X17693|ISTN7TNS 1**

Transposon Tn7 transposition genes tnsA, B, C, and E. [Escherichia coli] 

Length = 273

**Plus Strand HSPs:**

Score = 302 (138.9 bits), Expect = 1.1e-58, Sum P(3) = 1.1e-58

Identities = 58/102 (56%), Positives = 71/102 (69%), Frame = +3

Query: 189 LARQRYGVEDDRVARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGIKTGRVHLLLS +A+E ++AR. KEGRGQG Y PWLTVPSQGRSHRLKGIKTGRVHLLLS 368

Sbjct: 1 MAKANSSFSEVQIAARRIKEGRGQGHGKDYPWLTQEVPSQGRSHRLKGIKTGRVHLLLS 60

Score = 369 DIERDIFYLFDIADVTIDREQPFLNLRDIADDGVTIVHP 494

D+E +F +W +V DIREQPFL NR+IA D G+ HP

Sbjct: 61 DLELAVFILSEWESSVLIDIREQPFLPFDTRQIAIDSIGKHP 102

Score = 148 (68.1 bits), Expect = 1.1e-58, Sum P(3) = 1.1e-58

Identities = 27/61 (44%), Positives = 39/61 (63%), Frame = +3

Query: 558 DGRMQLARAVPKAEELEKPRVVEKLEIERRYWAQGQVDQVVTIDIPKAMVRNIAWV 737

DG Q A VKPA L+ R +E:KLE+ER.R.YW Q+ + W + T+++I + NI W++

Sbjct: 121 DGPFQPFAQIPQAIKVPAAALQDERTLEKLEIERRYWQQKIPWFIFDTDEINPVKVENEWLY 180

Score = 738 S 740

S

Sbjct: 181 S 181

Score = 44 (20.2 bits), Expect = 1.1e-58, Sum P(3) = 1.1e-58

Identities = 9/13 (69%), Positives = 10/13 (76%), Frame = +3

Query: 510 GTPLVMTTDDLVD 548

G VM+TDLVD

Sbjct: 106 GVDQVMSTDLVD 118

**>piri|S18587|S18587 hypothetical protein 2 - Thiobacillus ferrooxidans**

>gp|X57326|TFMERRCQ_4 T.ferrooxidans merR and merC genes. [Thiobacillus ferrooxidans]

Length = 138

**Plus Strand HSPs:**

Score = 190 (87.4 bits), Expect = 8.8e-24, Sum P(2) = 8.8e-24

Identities = 36/59 (61%), Positives = 43/59 (72%), Frame = +3

Query: 225 VARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGIKTGRVHLLLSDIERDIFYLFD 401

AR KEGRGQG Y PWLT+DVPS+G S R+KG KTGRVHLLLS +E F +D

Sbjct: 13 IARRKKEGRGQGSEKSYPKWLPVRDVPSGLSVRKGRKTGRVHLLSQLELESYFLMLD 71
Score = 54 (24.8 bits), Expect = 8.8e-24, Sum P(2) = 8.8e-24
Identities = 11/13 (84%), Positives = 11/13 (84%), Frame = +3

Query: 405 ADAVTDIREQFPL 443
   A VTDIREQFPL
Sbjct: 75 AGCVTDIREQFPL 87
Fig. 4.5. (a) Alignment of the amino acid sequence of Tn5468 (which had homology to TnsA of Tn7, Fig. 4.4) to the amino acid sequence of ORF2 (between the merR genes of T.ferrooxidans strain E-15). ORF2 had been found to have significant homology to TnsA of Tn7 (Kusano et al., 1991). (b) Alignment of the amino acids translation of Tn5468 (above) to TnsA of Tn7. (c) Alignment of the amino acid sequence of TnsA of Tn7 to the hypothetical ORF2 protein of T.ferrooxidans strain E-15.

(a)
Percent Similarity: 60.000  Percent Identity: 44.444

| Tf  | Tn5468 | 1 | LARQRYGVDEDVRADFQKEGRGGQQRGADYHPWLTIQDVSQGRSHRLKG | 50 |
| Tf  | E-15   | 1 | MSKGRSRRSASEGARIKEGRGGQGSEKYPWLTVROVSRSGVKIKG | 50 |
| Tf  | Tn5468 | 51 | KTGRVHHLSDIHEREDFYLFD..WADAVTIREFQFPLNR.DITRRIADDL | 97 |
| Tf  | E-15   | 51 | KTGRVHHLSDLSEYFLMLDDIRAQCVTIDEQFPLPIETTLEIADIR | 100 |
| Tf  | Tn5468 | 98 | GVIHPVDGSGTPLVMTDFTLIDGMRVQARVKPAELEKPVVE | 147 |
| Tf  | E-15   | 101 | CAGAHRLVDDDGSCV......VDLNAHRKVAMRRRRTASGDQ......138 |

(b)
Percent Similarity: 59.542  Percent Identity: 42.366

| Tf  | Tn5468 | 1 | LARQRYGVDEDVRADFQKEGRGGQQRGADYHPWLTIQDVSQGRSHRLKG | 50 |
| Tf  | Tn5468 | 51 | KTGRVHHLSDIHEREDFYLFD..WADAVTIREFQFPLNR.DITRRIADDL | 97 |
| Tf  | Tn5468 | 101 | HPRDVGSGETPLVMTDFTLIDGMRVQARVKPAELEKPVVEKLE | 150 |
| Tf  | Tn5468 | 151 | IERRYWQQGVVDVVTEDIPKAMVRNIAWHYVAVIDQMSQPDGYD | 200 |
| Tf  | Tn5468 | 201 | ELKARVLRELPSHPGTLQFCADMDLQFSMSAGDCLLIRRHAL....246 |
| Tf  | Tn5468 | 247 | .........KANVCPMDGPTDDSKLIRQFRVAEGES | 273 |
| Tf  | Tn5468 | 243 | FNIYKSFRAK.DLCISQVVNMEELYVAN..273 |
| TnsA | Tn7   | 1 | MAKANSSFSEVQIARRIKEGRGGQSHGKDYIPWLTVQEFVSPSRSRIYSH | 50 |
| TnsA | Tn7   | 101 | HP..VIRGDQVMSTDFTLVD.CKDGFEQFQAIVKPAALQIERTLEKLE | 147 |
| TnsA | Tn7   | 151 | LERRYWQQGIPFWFIDKEINPVVKENVLYVS.KTEVEAELAQLS | 196 |
| TnsA | Tn7   | 197 | PLAHI......LQEGDNEINVCKVQDIAYDELKTLSEIRALTANGFIK | 242 |
| TnsA | Tn7   | 243 | FNIYKSFRAK.DLCISQVVNMEELYVAN..273 |
| TnsA Tn7 | 1 | MAKANSSFSEVQIARRIKEGRGQGHGKDYIPWLTVQEVFPSGRS | 50 |
| Tf E-15 | 1 | MSKGSRRSESGAIARLKEGRGQGSEKSYKPFWLT | 50 |
| TnsA Tn7 | 51 | KTGRVHLLSDLELAVFLS | 96 |
| Tf E-15 | 51 | KTGRVHLLSQLELYFLMLDD | 100 |
| TnsA Tn7 | 97 | SGIKHPVIRGVDQVMSTDLVCDKGD | 134 |
| Tf E-15 | 101 | CAGAHRLVDD | 138 |
product of ORF2 (hypothetical protein 2, only 138 amino acids) of *T. ferrooxidans* E-15 (Kusano *et al.*, 1991) showed 60% similarity and 44.4% identical amino acids (Fig. 4.5a). The homology obtained from the amino acid sequence comparison of Tn5468 to TnsA of Tn7 (59.5% identity and 42.4% similarity) was lower (Fig. 4.5b) than the homology between *T. ferrooxidans* strain E-15 and TnsA of Tn7 (66.7% similarity and 47.3% identity, Fig. 4.5c). However, the lower percentage (59% similarity and 42.4% identity) can be explained in that, the entire TnsA of Tn7 (273 amino acids) is compared to the TnsA-like protein of Tn5468 unlike the other two. Homology of the Tn7-like transposon to TnsA of Tn7 appears to begin with the codon TTG (Fig. 4.2) instead of the normal methionine initiation codon ATG. There is an ATG codon two bases upstream of what appears to be the TTG initiation codon but it is out of frame with the rest of the amino acid sequence. This is near a consensus ribosome binding site AGAGG, at 5 bp from the apparent TTG initiation codon. From these results, it was apparent that a Tn7-like transposon had been inserted in the translational terminator region of the *T. ferrooxidans glmS* gene. The question to answer was how much of this Tn7-like element was present and whether there were any genetic markers linked to it.

**4.4.2. TnsBC homology**

Single strand sequencing from both the *KpnI* and *SalI* ends of construct p818.52 (Fig. 4.6) was carried out. A search for sequences with homology to each end of p818.52 was performed using the NCBI BLAST subroutine and the results are shown in Figs. 4.8 and 4.9 respectively. The TnsB protein of Tn7 consists of 703 amino acids and aside from being required for transposition, it is believed to play a role in sequence recognition of the host DNA. It is also required for homology specific binding to the 22 bp repeats at the termini of
Fig. 4.6. Restriction map of p818.1 and construct p818.20 showing the various restriction sites on the fragments. The subclones p818.16, p818.52 and p818.09 and the regions which revealed high sequence homology to TnsA, TnsB, TnsC and TnsD are shown below construct p818.20.
Fig. 4.7. Restriction map of cosmid p818.1ΔE showing the constructs, shortenings and subclones used to search for sequence homology to the TnsD and TnsE proteins of Tn7. Also shown is construct p818.11 which showed good amino acid sequence homology to *E.coli* and *H.influenzae* SpoT and RecG proteins at the ends.
Tn7 (Orle and Craig, 1991). Homology of the sequence from the KpnI site of p818.52 to TnsB begins with the third amino acid (Y) corresponding to amino acid 270 of TnsB. The amino acid sequences are highly conserved up to amino acid 182 for Tn5468. Homology to the rest of the sequence is lower but clearly well above background (Fig. 4.8). The region of TnsB to which homology was found falls approximately in the middle of the TnsB of Tn7 with about 270 amino acids of TnsB protein upstream and 320 amino acids downstream. Another interesting observation was the comparatively high homology of the p818.52r sequence to the Tra3 transposase of Tn552 (Fig. 4.8).

The TnsC protein of Tn7 binds non-specifically to the DNA in the presence of ATP and is required for transposition (Orle and Craig, 1991). It is 555 amino acids long. Homology to TnsC from Tn7 was found in a single frame which extended for 81 amino acids along the sequence obtained from p818.52. The region of homology corresponded to amino acids from 163 to 232 of TnsC. Analysis of the size of p818.52 (1.7 kb) with respect to the regions of homology to TnsB and TnsC of Tn7 suggests that the size of the Tn5468 TnsB and C like ORFs correspond to those found in Tn7. On average, about 40-50% identical and 60-65% similar amino acids were revealed in the BLAST search (Fig. 4.9).

4.4.3. Homology to TnsD protein.

The location of construct p818.09 and p818.10 is shown in Figs. 4.6 and 4.7 respectively. The DNA sequences obtained from the EcoRI sites of these constructs were joined (after inverting and complementing the sequence of p818.09). In this way, 799 bp of sequence which had been determined from one or other strand was obtained. The results of the BLAST search
Fig. 4.8. (a) The results of the BLAST search of the DNA sequence obtained from p818.52r (from the KpnI site) showing clear homology to the TnsB protein of Tn7. The amino acid sequence also showed high homology to Tra3 transposase of transposon Tn552. (b) Nucleotide sequence of p818.52 and the product of the open reading frame.

(a)

<table>
<thead>
<tr>
<th>Reading</th>
<th>High</th>
<th>Proba</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>Score</td>
<td>P(N)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>+3</td>
<td>316</td>
<td>3.1e-37</td>
</tr>
<tr>
<td>+1</td>
<td>46</td>
<td>0.00038</td>
</tr>
<tr>
<td>+3</td>
<td>69</td>
<td>0.028</td>
</tr>
<tr>
<td>-3</td>
<td>37</td>
<td>0.060</td>
</tr>
<tr>
<td>+1</td>
<td>45</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Plus Strand HSPs:

Score = 316 (145.4 bits), Expect = 3.1e-37, P = 3.1e-37
Identities = 59/114 (51%), Positives = 77/114 (67%), Frame = +3

Query: 3 YQIDATIADVYLVSRYDRTKIVGRPVLYIVIDVFSRMITGCVYVGFEGSPWSVGAMMALNSNT 182
Y+IDATIAD+YLV +DR KI+GRP LIVIDVFSRMITG Y+GFE PS+V AM A N
Sbjct: 270 YEIDATIADYLVHIDRQKI+GRPTLYIVIDVFSRMITGFIYFGENPYSYLVAMQAFVMA 329

Query: 183 ATEKVEYCRQFGVEIGAADWPCRCPCTLSRPRGIEAIGSALRPFINNFQVRVEN 344
++K C Q ++EI ++DWC P + E + ++F VRVE+
Sbjct: 330 CSDKTAICAQHDIEISSSDWPCVGLPVDLLADRGELMSHVQEAIVSSFVNVRVES 383

sp|P13989|TNSB_ECOLI TRANSPOSON TN7 TRANSPOSITION PROTEIN TNSB.
Length = 702

Plus Strand HSPs:

Score = 69 (31.7 bits), Expect = 0.028, Sum P(2) = 0.028
Identities = 14/48 (29%), Positives = 24/48 (50%), Frame = +3

Query: 51 DRTKIVGRPVLYIVIDVFSRMITGCVYVGFEGSPWSVGAMMALNSNTATEK 194
D+ + RP L I+D +SR I G ++ P+ + + L K
Sbjct: 176 DQKGNINRPWLTIIMDDYSAIAGYFISFDAPNAQTALTHQAIWNK 223

Score = 36 (16.6 bits), Expect = 0.028, Sum P(2) = 0.028
Identities = 5/15 (33%), Positives = 11/15 (73%), Frame = +3

Query: 3 YQIDATIADVYLVSRYDRTKIVGRPVLYIVIDVFSRMITGCVYVGFEGSPWSVGAMMALNSNT 47
+Q D T+ D+Y++ +
Sbjct: 163 WQADHTLLLDIYLDQ 177
GGTACCAGATCGACGCCACCACGCAGGTGTACTGCTGATCCGATATGCCTAAGCGTACAA
10  20  30  40  50  60
Y Q I D A T I A D V Y L V S R Y D R T K -

AGATCGTCCGGACGCCCGGGTGCTCTATATCGTACATCGACGTTCCAGCCGATGATCACC
70  80  90  100 110 120
I V G R P V L Y I V I D V F S R M I T G -

GCGTGATATGGGCTTGAGAGCCTTTCTGCCGGTGGCGGCGATGATGCGCCCTGTCGAACA
130  140  150  160 170 180
V Y V G F E G P S W V G A M M A L S N T -

CCGCCACGGAAAGGTTGGAATATTGCCGCCAGTTGCCGCTCAGATCAGCGGCAGGGCGGACCT
190  200  210  220 230 240
A T E K V E Y C R Q F G V E I G A A D W -

GCCGTGCAGGCGCTGCCGAGCGTTTCGCCGACGCGGAGAGAAGGATGCCGCGCAGCC
250  260  270  280 290 300
P C R P C F T L S R R P G R E I A G S A -

CATTGAGACCCCTTTATCAACACTTCCAGTCGGGTGGAAAC
310  320  330  340
L R P F I N N F Q V R V E N -
Fig. 4.9. (a) Results of the BLAST search on the inverted and complemented nucleotide sequence of p818.52f (from the Sall site). Results indicate high amino acid sequence homology to the TnsC protein of Tn7 (protein E is the same as TnsC protein). (b) The nucleotide sequence and open reading frame of p818.52f.

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Reading</th>
<th>High</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>Score</td>
<td>P(N)</td>
</tr>
</tbody>
</table>

piR|B25543|QQECE7 hypothetical protein E - Escher... +1 176 1.5e-20
gp|X04492|INN7EUR_1 Transposon Tn7 fragment E with ... +1 176 1.5e-20
sp|P05846|TNSC_EODLI TRANSPOSON Tn7 TRANSPOSITION PR... +1 176 6.4e-20
gp|U41011|CEL2024_7 D2024.8 gene product [Caenorhab... +3 59 1.7e-06

piR|B25543|QQECE7 hypothetical protein E - Escherichia coli transposon Tn7 (fragment)
Length = 294

Plus Strand HSPs:

Score = 176 (81.0 bits), Expect = 1.5e-20, Sum P(2) = 1.5e-20
Identities = 29/70 (41%), Positives = 51/70 (72%), Frame = +1

Query: 34 SLDOVWLKDPSYKSPQLCISFFQEMDNLGTPRYRAGRSRS5DEMVMQAMHAN 213
       ++QQVH+D + GS K++C++FF +D LG+ Y RYG R ++ M+ M+Q+AN
Sbjct: 163 NVEQVVVYKIDCHNGLKEICLNNFFRADIRALGSNYERYGLKRKHEITMLALMSQIAN 222

Query: 214 LQSLGVLIVD 243
       +LG+L+D
Sbjct: 223 AHALGLLVID 232

Score = 40 (18.4 bits), Expect = 1.5e-20, Sum P(2) = 1.5e-20
Identities = 6/9 (66%), Positives = 8/9 (88%), Frame = +1

Query:  YPQVHHAEE 27
       YPQV++H E
Sbjct: 153 YPQVHYHRE 161

(b)

1 TATCCGCAAGTCGTTCCACCATTGCCGAGCCTTCATCAGTCTGCAAGTGGTCTGGCTGAAG 60
10 20 30 40 50 60
Y P Q V V H H A E P F S L D Q V V W L K -
61
CTGGATAGCCCTACAAAGGATCCCGCAACACAGATCTGCATCGTTTTCAGGAGATG 120
70 80 90 100 110 120
L D S P Y K G S P K Q L C I S F F Q E M -
121
GACATCATATGACCGACCCGTACCCGGCGTACCGGAGCCGGAGTCTGAGTCAC 180
130 140 150 160 170 180
D N L L G T P Y R A R Y G A S R S S L D -
181
GAGATGATGTCGCAGATGCGCAAACTGCGATCTGCTGCTCACGTCACGTCGTC 240
190 200 210 220 230 240
E M M V Q M A Q M A N L Q S L G V L I V -
241
GAC. 244
D -
using the GenBank and EMBL databases of the joined DNA sequences and open reading frames are shown in Fig. 4.10a and 4.10b respectively. Homology of the TnsD-like of protein of Tn5468 to TnsD of Tn7 (amino acid 68) begins at nucleotide 38 of the 799 bp sequence and continues downstream along with TnsD of Tn7 (A, B, C, E, Fig. 4.11). However, homology to a region of Tn5468 TnsD marked D (Fig. 4.11, 384-488 bp) was not to the predicted region of the TnsD of Tn7, but to a region further towards the C-terminus (279-313, Fig. 4.11). Thus it appears there has been a rearrangement of this region of Tn5468.

Because of what appears to be a rearrangement of the Tn5468 insD gene, it was necessary to show that this did not occur during the construction of p818.1ΔE or p818.10. The 12 kb fragment of cosmid p818.1 from the BgIII site to the HindIII site (Fig.2.1) had previously been shown to be natural unrearranged DNA from the genome of T. ferrooxidans ATCC 33020 (Chapter 2). Plasmid p818.1ΔE had been shown to have restriction fragments which corresponded exactly with those predicted from the map of p818.1 (Fig. 4.12) including the EcoRI-ClaI p818.10 construct (lane 7) shown in Fig. 4.12. A subclone p818.01EM containing 1 kb EcoRI-MluI fragment of p818.10 (Fig. 4.7) cloned into pUCBM20 was sequenced from the MluI site. A BLAST search using this sequence produced no significant matches to either TnsD or to any other sequences in the Genbank or EMBL databases (Fig. 4.13). The end of the nucleotide sequence generated from the MluI site is about 550 bp from the end of the sequence generated from the p818.10 EcoRI site.

In order to determine whether any homology to Tn7 could be detected further downstream, construct p818.10 was shortened from ClaI site (Fig. 4.7). Four shortenings namely, p818.101,
Fig. 4.10. Results of BLAST obtained from the inverted complemented sequence of p818.09 joined to the forward sequence of p818.10f showing clear homology to TnsD of Tn7 (b) The sequence from p818.09r joined to p818.10 and the predicted translation products in all three forward reading frames.

(a)

Probability

Sequences producing High-scoring Segment Pairs: Frame Score P(N)

sp|P13991|TNSD_ECOLI TRANSPONSON TN7 TRANSPOSITION PR... +2 86 8.1e-13
gp|D13972|AQUAQO1_1 ORF 1 [Plasmid pAQ1] +3 48 0.046
sp|P42148|HSP1 Plain SPERM PROTAMINE (CYSTEINE-RICH ... -3 44 0.26

>sp|P13991|TNSD_ECOLI TRANSPONSON TN7 TRANSPOSITION PROTEIN TNSD.
pir|S12640|S12640 tnsD protein - Escherichia coli transposon Tn7
>gp|X17693|ISTN7NS_4 Transposon Tn7 transposition genes tnsA, B, C, D and E. [Escherichia coli]
Length = 508

Plus Strand HSPs:

Score = 86 (39.6 bits), Expect = 8.1e-13, Sum P(5) = 8.1e-13
Identities = 14/26 (53%), Positives = 18/26 (69%), Frame = +2

Query: 236 LSRYGEGYWRRSHQLPGVLVCPDHGA 313
L+RYGE +W+R LF + CP HGA
Sbjct: 132 LNRYGAEAFQRDWYLPALPYCPKNGA 157

Score = 55 (25.3 bits), Expect = 8.1e-13, Sum P(5) = 8.1e-13
Identities = 14/48 (29%), Positives = 23/48 (47%), Frame = +2

Query: 38 ERLTRDFTLIRLITFAEPKSVGSVLASLADGPADAVHVRGLGAASAI 181
++L + TL L+ F K + + AVH+ LG+AAS +
Sbjct: 68 QQLIYEHYTLFLYAFPVGKERRDEAIRLMEYQAQGAVHLMGLGVAASRV 115

Score = 49 (22.5 bits), Expect = 8.1e-13, Sum P(5) = 8.1e-13
Identities = 9/14 (64%), Positives = 11/14 (78%), Frame = +2

Query: 671 VVRKHRKAFHPLRH 712
+ RKHRKAF L+H
Sbjct: 274 IFRKHRKAFSYLQH 287

Score = 40 (18.4 bits), Expect = 6.5e-09, Sum P(4) = 6.5e-09
Identities = 10/35 (29%), Positives = 18/35 (51%), Frame = +3

Query: 384 QKNCSPVRHSLGLRVAAPSDAVARDCQADAALDH 488
++K S ++HS++ + P V +QL +AL +H
Sbjct: 279 RKAFSYLOQHSLVWQALLPKLTVIALQQASALTEH 313

Score = 38 (17.5 bits), Expect = 8.1e-13, Sum P(5) = 8.1e-13
Identities = 7/23 (30%), Positives = 10/23 (43%), Frame = +3

Query: 501 PSFRDWSAYRSAVIARGFGK 569
PS W+ +Y+ G K K
Sbjct: 213 PSLEQWTLFYQRLAQDLGLTKSK 235

Score = 37 (17.0 bits), Expect = 8.1e-13, Sum P(5) = 8.1e-13
Identities = 6/12 (50%), Positives = 8/12 (66%), Frame = +2

Query: 188 SRHTLRYCPICL 223
S + RCYP C+
Sbjct: 117 SDNRFRYCPDCV 128
p818.102, p818.103 and p818.104 were selected (Fig. 4.7) for further studies. The sequences obtained from the smallest fragment (p818.104) about 1.15 kb from the EcoRI end overlapped with that of p818.103 (about 1.34 kb from the same end). These two sequences were joined and used in a BLAST homology search. The results are shown in Fig. 4.14. Homology to TnsD protein (amino acids 338-442) was detected with the translation product from one of these frames. Other proteins with similar levels of homology to that obtained for the TnsD protein were found but these were in different frames or the opposite strand. Homology to the TnsD protein appeared to be above background. The BLAST search of sequences derived from p818.101 and p818.102 failed to reveal anything of significance as far as TnsD or TnsE of Tn7 was concerned (Fig. 4.15 and 4.16 respectively).

A clearer picture of the rearrangement of the TnsD-like protein of Tn5468 emerged from the BLAST search of the combined sequence of p818.104 and p818.103. Regions of homology of the TnsD-like protein of Tn5468 to TnsD of Tn7 (depicted with the letters A-G, Fig. 4.11) correspond with increasing distance downstream. There are minor intermittent breaks in homology. As discussed earlier, the region marked "D" (Fig. 4.11) between nucleotides 384 and 488 of the TnsD-like protein showed homology to a region further downstream on TnsD of Tn7. Regions D and E of the Tn5468 TnsD-like protein were homologous to an overlapping region of TnsD of Tn7. The largest gap in homology was found between nucleotides 712 and 1212 of the Tn5468 TnsD-like protein (Fig. 4.11). Together, these results suggest that the TnsD-like protein of Tn5468 has been rearranged, duplicated, truncated and shuffled.
Fig. 4.11. Rearrangement of the TnsD-like ORF of Tn5468. Regions on TnsD-like protein of Tn5468 identified by the letters A-G are matched unto the corresponding areas on TnsD of Tn7. At the bottom is the map of Tn5468 which indicates the areas where homology to TnsD of Tn7 was obtained. (Note that the numbers on the line representing TnsD of Tn7 are amino acids, whereas the numbers on the TnsD-like protein of Tn5468 represent distances in base pairs).
Fig. 4.12. Restriction digests carried out on p818.1ΔE with the following restriction enzymes; 1 HindIII, 2 EcoRI-ApaI, 3 HindIII-ApaI, 4. ClaI, 5. HindIII-EcoRI, 6. ApaI-ClaI, 7. EcoRI-ClaI, and 8. ApaI. The smaller fragment in lane 7 (arrowed) is the 4.0 kb insert in the p818.10 construct. λ PstI marker (first lane) was used for sizing the DNA fragments.
(a) BLAST results of the nucleotide sequence obtained from p818.10EM. (b) The inverted complemented nucleotide sequence of p818.10EM.

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Frame</th>
<th>Score</th>
<th>Expect</th>
<th>Sum P(2)</th>
<th>Proba</th>
</tr>
</thead>
<tbody>
<tr>
<td>pir</td>
<td>S40626</td>
<td>S40626</td>
<td>androgen receptor - mouse</td>
<td>+2</td>
<td>45</td>
<td>0.16</td>
</tr>
<tr>
<td>pir</td>
<td>S39636</td>
<td>S39636</td>
<td>cydD protein - Escherichia coli (...-1</td>
<td>40</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>gp</td>
<td>D50624</td>
<td>SIMVBRA1_4</td>
<td>NusG like protein [Streptomyces v...+1</td>
<td>63</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>pir</td>
<td>S40626</td>
<td>S40626</td>
<td>androgen receptor - mouse</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Length = 48

Score = 45 (20.7 bits), Expect = 0.17, Sum P(2) = 0.16
Identities = 10/25 (40%), Positives = 13/25 (52%), Frame = +2

Query: 329 GTAQEMV*SACGLGDIGS*HGDP TA 403
G+A + C GD+GS HG A
Sbjct: 24 GSAWAAAAAACRYGDLSHLHGSVA 48

Score = 33 (15.2 bits), Expect = 0.17, Sum P(2) = 0.16
Identities = 5/9 (55%), Positives = 6/9 (66%), Frame = +2

Query: 29 NPAESGEAW 55
NP + G AW
Sbjct: 19 NPLDYGSW 27

(b)

1 TGGACTTGG TCCCTACTG GATGGTCAA TA CCAGAA AGTGCTGAA Ag
51 CCTGGGACTC AGGGCGGTAG CAAATATCT TGCATGTAag CCAAGGGACGG
101 TGAATTCGA CGCTCCCGA CGCGGGTTGA ATGTGCCCTG GAAGCCCTTA
151 GCGGCAGAC ATTCGCGAAC ATGGGCGCCCGA GCCGCGGATG CACATACGAT
201 AGCTGTTGTG GACATSCAG GBAATTACGC TGATTTATCA
251 CTGGCACTCC TGCTACCCAA AGAACACTCA TGGTTGTACC
301 GGGAGTTGGTC GGAGACCAAT TCACCACG GACTGCAACAA GAAATGGCTT
351 GATCAGCGTG TGGACTGGGA GACATTGGAT CGTAGCATGG CGAtCCAACT
401 GCCTTGGGCA GCGCGGGAAA tcAtctTCGG GAGGTTCCGC CACACCGGt
Fig. 4.14. (a) BLAST check on the sequences obtained from p818.104 and p818.103 joined. Homology to TnsD was detected. b) The nucleotide sequence and frame in which homology to TnsD was found.

(a)

<table>
<thead>
<tr>
<th>Sequences producing High-scoring Segment Pairs:</th>
<th>Reading Frame</th>
<th>Score</th>
<th>Prob P(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pir</td>
<td>A47283</td>
<td>A47283</td>
<td>calphotin - Drosophila</td>
</tr>
<tr>
<td>sp</td>
<td>P13991</td>
<td>TNSD ECOLI</td>
<td>TRANSPOSON TN7 TRANSPOSITION PROTEIN TNSD.</td>
</tr>
<tr>
<td>gp</td>
<td>D49399</td>
<td>RABA_TIC</td>
<td>alpha 2 type I collagen</td>
</tr>
<tr>
<td>pir</td>
<td>A44950</td>
<td>A44950</td>
<td>merozoite surface antigen 2 - P</td>
</tr>
</tbody>
</table>

>sp|P13991|TNSD ECOLI | TRANSPOSON TN7 TRANSPOSITION PROTEIN TNSD. | +1 |
| pir|S12640|S12640 | tnsD protein - Escherichia coli transposon Tn7 | +1 |
| gp|X17693|ISTN7TNS_4 | Transposon Tn7 transposition genes tnsA, B, C, D and E. [Escherichia coli] | +1 |

Length = 508

Plus Strand HSPs:

Score = 56 (25.6 bits), Expect = 0.33, Sum P(2) = 0.28
Identities = 14/54 (25%), Positives = 24/54 (44%), Frame = +1

Query: 319 RVDW|DR|QL + ++ + RT + L ++ +++ KLP
Sbjct: 389 RVDW|NQ|DR|AVR|QL|R|L|K|P|G

Score = 50 (23.0 bits), Expect = 0.33, Sum P(2) = 0.28
Identities = 11/42 (26%), Positives = 19/42 (45%), Frame = +1

Query: 169 WL|DMQQ|NYAD|LS|RR|ALL|LP|KE|SH|WLYR|HT|GSGWS|NIHQ|R
Sbjct: 338 WQQL|VH|K|Q|G|I|A|R|Q|S|LE|GG|V|L|A|R|Y|W|L|R|D|W|L|H|W|N|Q|H

(b)

GAGGAAATCGGGAAGGCTGCGATCAGGGCTGACCTATAATCATTCATTGCTTACCTGATACCAGGG
E E I G R P G I G R * P I I I A L I P G
ACGGTGAGATTGCAAGAGCTCAGCAGCAGGGTGGATGAAATGGCCCTGCCAGCCATTGCCAAGGCA
130 140 150 160 170 180
R H S R T L V P D R D A I R I R W L D M
CAGCAGAATTACGCGTATTATCAGCTGCAGCAGCTGACCTCCCGGCTGACCTGAGCTGAGCTGAGCTGAGCTG
190 200 210 220 230 240
Q Q N Y A D L S R R R L A L L L P K E H
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Amino Acid</th>
<th>Start</th>
<th>Stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCATGGTTGTACC3CCATACAGGGAGTGCTGGCTGGGACATTCACCAACGGCACTGCACA</td>
<td>SWL YRHTGSNGWSNHQQRHCT</td>
<td>250</td>
<td>300</td>
</tr>
<tr>
<td>AGAAATGGTCCTGATCCACGCTGTGGACTGGAAGACATTGGATCGTACGTGAGCAGATCCAA</td>
<td>RNGLIQRVWDWEKLDRSMAIQ</td>
<td>310</td>
<td>360</td>
</tr>
<tr>
<td>CTGGTAAGGCGCGCGGAAATCCTCAGGGGAGGTTCCGCCACACGGTTACCCAGGGCG</td>
<td>LRAAREIITREVVPPQVRVTQA</td>
<td>370</td>
<td>420</td>
</tr>
<tr>
<td>GCTTTGGACGCCAAGTTGGGCGGCCACTCCGGATG1CAGAAACACAGGCAAAACTGCCT</td>
<td>ALERKSLGRPLRMSEQQAKLP</td>
<td>430</td>
<td>480</td>
</tr>
<tr>
<td>AAAAGTATCGGTGTACTTGCGGA</td>
<td>KSIGVLAG</td>
<td>490</td>
<td>500</td>
</tr>
</tbody>
</table>
Fig. 4.15. (a) Results of the BLAST search on p818.102 (b) DNA sequence of p818.102.

(a)

**Reading High Probability**
Sequences producing High-scoring Segment Pairs:  
Frame Score $P(N)$

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score</th>
<th>Expect</th>
<th>$P$</th>
<th>Identities</th>
<th>Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>P29424</td>
<td>TX25 PHONI NEUROTOXIN TX2-5. &gt;pir</td>
<td>S29215</td>
<td>S2...</td>
<td>+3</td>
</tr>
<tr>
<td>sp</td>
<td>P28880</td>
<td>C XO A_CONST OMEGA-CONOTOXIN SVIA. &gt;pir</td>
<td>B4437...</td>
<td>+3</td>
<td>55</td>
</tr>
<tr>
<td>gp</td>
<td>X77576</td>
<td>BNACCG8_1 acetyl-CoA carboxylase [Brassica...</td>
<td>+2</td>
<td>39</td>
<td>0.999</td>
</tr>
<tr>
<td>gp</td>
<td>X90568</td>
<td>HSTITIN2B_1 titin gene product [Homo sapiens]</td>
<td>+2</td>
<td>43</td>
<td>0.9995</td>
</tr>
</tbody>
</table>

>sp|P29424|TX25 PHONI NEUROTOXIN TX2-5. >pir|S29215|S29215 neurotoxin Tx2 - spider (Phoneutria nigriventer)
Length = 49

Plus Strand HSPs:
Score = 57 (26.2 bits), Expect = 4.7, $P = 0.99$
Identities = 12/30 (40%), Positives = 14/30 (46%), Frame = +3

Query: 105 EKGSXVRKSPAPCRQANLPCGAYLLGCSRC 194
E+G V P CRQ N AY L +C
Sbjct: 19 ERGECVCGPC1CRQGFLIAAYKLASCKC 48

sp|P28880|C XO A_CONST OMEGA-CONOTOXIN SVIA. >pir|B44379|B44379
omega-conotoxin
SVIA - cone shell (Conus striatus)
Length = 24

Plus Strand HSPs:
Score = 55 (25.3 bits), Expect = 6.1, $P = 1.0$
Identities = 8/19 (42%), Positives = 11/19 (57%), Frame = +3

Query: 141 CRQANLPCGAYLLGCSRCW 197
CR + PCG + C RC+
Sbjct: 1 CRSSGSPCGVTS1CCGRCY 19

(b)

1 GGAGTCTCTG TGATGTTTAA ATTGAGTGAG AAAAGACGCT ATTCGGCCGA
51 AGTGCCTGAA AATTTGACTC TACACTGGCT GGCAGAATGT CAAACAAGAC
101 GATGAAAAAA GGTTGCGAGG TCCGTAAGTC ACCCGCCCGC TGTCGTCAG
151 CGAATTTGCC ATGTGGCGCG TACCTATTTG GCTGTTCCCG GTGCTGGCAC
201 TCGCTATAG CTTCTCCGAC GTTAGATTGC TCCCAATAAC CTAGCGCGCAA
251 TATTCGAAAG TGATTATTCC CAATTTGCG GGAAGGAGAG GAAATCAACT
301 CGTCCGACAT TCCGGCCGAA TTATATTCTGT TTGGAAGCAA AGCACAGGGG
Fig. 4.16. (a) BLAST results of the nucleotide sequence obtained from p818.101. (b) The nucleotide sequence obtained from p818.101.

(a)

Sequences producing High-scoring Segment Pairs:  

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Reading Frame</th>
<th>Score</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>P02241</td>
<td>HCYD EURCA HEMOCYANIN D CHAIN.</td>
<td>+2</td>
</tr>
<tr>
<td>sp</td>
<td>P03108</td>
<td>VL2 CRPVK PROBABLE L2 PROTEIN.</td>
<td>+3</td>
</tr>
<tr>
<td>sp</td>
<td>Q09816</td>
<td>YAC(2) SCHPO HYPOTHETICAL 33.9 KD PROTEIN C16C</td>
<td>+2</td>
</tr>
<tr>
<td>sp</td>
<td>P06278</td>
<td>AMY BACLI ALPHA-AMYLASE PRECURSOR (EC 3.2.1</td>
<td>+2</td>
</tr>
<tr>
<td>sp</td>
<td>P26805</td>
<td>GAG MLVFP GAG POLYPROTEIN (CORE POLYPROTEIN</td>
<td>+3</td>
</tr>
</tbody>
</table>

Length = 627  
Plus Strand HSPs:  
Score = 47 (21.6 bits), Expect = 0.039, Sum P(3) = 0.038  
Identities = 6/12 (50%), Positives = 9/12 (75%), Frame = +2  

Query:  140 HFHWYPQHPCFY 175  
Sbjct:  170 HWHWHLVYPFAFY 181  
Score = 38 (17.5 bits), Expect = 0.039, Sum P(3) = 0.038  
Identities = 12/36 (33%), Positives = 15/36 (41%), Frame = +2  

Query:  41 TPWAGDRAETQKRL*AVGFFHPFDIFSFPFH 148  
Sbjct:  17 TSLSPDPLPAERDPRKGFVFLRGSTFGFHEEH 52  
Score = 32 (14.7 bits), Expect = 0.039, Sum P(3) = 0.038  
Identities = 7/21 (33%), Positives = 11/21 (52%), Frame = +2  

Query:  188 DPYFFVPPADVFYPAKSGSGQ 250  
Sbjct:  548 DFYLFLMLDYLEEDSVQGAGE 568  

(b)

1 CGTCCGTACC TTCACCTTTC TCGACCGCAA GCAGCCTGAA ACTCCATGGG  
51 CAGGAGATCG TCTGTCAGAG ACTCAGGGGA AACCCTCCTT TTAGCCGGTG  
101 GGGTTTTTCC ACTTCCCCGA TATTTTCGGG TCTTTCCCTC ATTTTCACTG  
151 GTACCCCGCAG CACCCCCGG TATTTTTCGG TCTTCCCTC ATTTTCACTG  
201 TTGTACCCGG CCGACACTTC GTTTACC CGG AAAAAAGTG G TCTGGTCA  
251 TATCGCGACGG GTGTTGAGAA CACCG
Fig. 4.17. (a) BLAST results obtained from combined sequence of p818.11f (inverted and complemented) and p818.10r. Very good sequence homology to E.coli and H.influenzae ATP-dependent DNA helicase RecG (EC 3.6.1.) was found. (b) The combined sequence and open reading frame which was homologous to RecG.

(a) Reading High Prob

Sequences producing High-scoring Segment Pairs: Frame Score P(N)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Frame</th>
<th>Score</th>
<th>Expect</th>
<th>Sum P(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>P43809</td>
<td>RECG_HAEIN ATP-DEPENDENT DNA HELICASE RECG. ...</td>
<td>+3</td>
<td>158</td>
<td>1.3e-14</td>
</tr>
<tr>
<td>gi</td>
<td>290502</td>
<td>(L10328) DNA recombinase [Escheri...</td>
<td>+3</td>
<td>156</td>
<td>2.6e-14</td>
</tr>
<tr>
<td>sp</td>
<td>P24230</td>
<td>RECG_ECOLI ATP-DEPENDENT DNA HELICASE RECG. ...</td>
<td>+3</td>
<td>156</td>
<td>2.6e-14</td>
</tr>
<tr>
<td>gi</td>
<td>1001580</td>
<td>(D64000) hypothetical protein [Sy...</td>
<td>+3</td>
<td>127</td>
<td>5.6e-10</td>
</tr>
<tr>
<td>gi</td>
<td>1150620</td>
<td>(249988) MmsA [Streptococcus pnu...</td>
<td>+3</td>
<td>132</td>
<td>8.0e-10</td>
</tr>
</tbody>
</table>

Length = 693

Plus Strand HSPs:

Score = 158 (72.7 bits), Expect = 1.3e-14, Sum P(2) = 1.3e-14
Identities = 34/76 (44%), Positives = 50/76 (65%), Frame = +3

Query: 3 EILAEQLPLAFQWLEPAGPGGLSGQGFTRARRETALAGGLAGSLVIGTLSFQQGV 182
Sbjct: 327 EILAEQHANNFRWRKPGIEVGLAGKVRGSQAELEIKTGAVQMVGVTHALFQEVEV 386

Score = 50 (23.0 bits), Expect = 1.3e-14, Sum P(2) = 1.3e-14
Identities = 9/16 (56%), Positives = 12/16 (75%), Frame = +3

Query: 261 RRGAMPHLLVMTASI 308
Sbjct: 416 KAGYPHQILMTAPI 431

Length = 693
Plus Strand HSPs:

Score = 156 (71.8 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
Identities = 33/76 (43%), Positives = 46/76 (60%), Frame = +3

Query: 3 EILAEQLPLAFQQWLEPAGGGGLGGQPFTRARRRETAEQLAGSLRLVIGTQSLFQGGV 182

Sbjct: 327 EILAEQHANNRFNWAPLGLIEVGWLAKQKKARLAQQAEEIASGQVQM1VTHAIFQEQV 386

Query: 183 VFACLGLVIIIIDEQHRF 230

F L LVIIDEQHRF

Sbjct: 387 QFNGLALVIIIIDEQHRF 402

Score = 50 (23.0 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
Identities = 9/16 (56%), Positives = 13/16 (81%), Frame = +3

Query: 261 RRGAMPHLVLVMTPASI 308

++G PH L+MTA+PI

Sbjct: 416 QQGFHPHQLIMTAPI 431

>gi|1001580 (D64000) hypothetical protein [Synechocystis sp.]
Length = 831

Plus Strand HSPs:

Score = 127 (58.4 bits), Expect = 5.6e-10, Sum P(2) = 5.6e-10
Identities = 33/76 (43%), Positives = 40/76 (52%), Frame = +3

Query: 3 EILAEQLPLAFQQWLEPAGGGGLGGQPFTRARRRETAEQLAGSLRLVIGTQSLFQGGV 182

Sbjct: 464 EVLAEQHNYQKLVSFWNLTLPLVEQSLTVQALQKSLVLTHALIQETV 523

Query: 183 VFACLGLVIIIIDEQHRF 230

F L LVIIDEQHRF

Sbjct: 524 NFQLVGLVIIIIDEQHRF 539

Score = 49 (22.5 bits), Expect = 5.6e-10, Sum P(2) = 5.6e-10
Identities = 9/15 (60%), Positives = 12/15 (80%), Frame = +3

Query: 264 RRGAMPHLVLVMTPASI 308

++G PH L+MTA+PI

Sbjct: 550 KGNAPHLVLSMTAPI 564

(b)

CCGAGATTTCTCGGAGCAGCTCCCATTGCGTCCAGCAATGGTGAACCAGCGCTGGGC
10 20 30 40 50 60
E I L A E Q L P L A F Q Q W L E P A G P -

CTGGAGGTCGCTACCTGAGCTCGTACCCCTTCGAGCTGGCCGACCCGACCCGACCGGCAGACCGGAAA
70 80 90 100 110 120
G G G L S G G Q P F T R A R R E T A E T -

CGCTTGCGCTCGAGCCAGCTTCGTTGGAACCCAGCCTGCTGTTCCAGCAAGGG
130 140 150 160 170 180
L A G G S L R L V I G T Q S L F Q Q G V -
TGGTGGTGCATGTCTCGGACTGGTCATCATCGACGAGCAACACCGCTTTTGGCCGTGGA
190 200 210 220 230 240
V F A C L G L V I I D E Q H R F W P W S -
GCAGCGCCGTCAATTGCTGGAAGAGGGGCGCCATGCCAACCTGTGTAATGACCGCTA
250 260 270 280 290 300
S A V N C W R R G A M P H L V M T A S -
GCCGATCATGGTGAGGACGCGCATTACGTCAGGCCATTTCTGTGCGGTAGGACACCCA
310 320 330 340 350 360
P I M V E D G I T S G I L L C G R T P N -
ATCGATTCTGCTCAAGCTGGTAGGCTCGGCCGTTGCTCTCCGGGCGTGAATGACAGCT
370 380 390 400 410 420
R F I P Q A G I D A V A F P G V D K D Y -
ATGCCGCCCCGTGAAAGGCGCGCAAGCCGGCGCATGCACCCGCTGCTAAACGACCCAG
430 440 450 460 470 480
A A R E A A K R G R W T P L L N D A G -
GCATACGTGCGCTGGCTTTGGTGACCAAGCTGCCCTTGCTGTCAGGCCAACAACACTGCTA
490 500 510 520 530 540
I R R S W L G R A S I A F V Q R N T A I -
TCGGAGGGCAACTTGAAGAGGCCTCGCCGCGGCGAGAATCTCCAGTTAATCCCGCGGA
550 560 570 580 590 600
G G Q L E E G G R A R T S Q L I P A S -
GCCATCCGCAACGGATCTGACGCTCAAGCTGTACTCCGACATCTGCTCTCAAGCTG
610 620 630 640 650 660
H P R T D R Q R L I H R D Y L L S S T D -
ACATTGAGCTTTCGATCTATAGAGGACGACTAGAATATTACCTCAGGCAGATTGCAATTG
670 680 690 700 710 720
I E L S I Y E D R L E I T S R Q I O M V -
TATTACGCGGACGTGATGCTGCGGCGTGTGACACGAAACCAGCTCATCAAGGATGTTAT
730 740 750 760 770 780
L R A D R D L A G R H E T S S S R M L C -
GCGCAGCATCC
791
A A S -
4.4.4. Analysis of DNA downstream of region with TnsD homology.

The location of the of p818.11 Apal-ClaI construct is shown in Fig. 4.7. The single strand sequence from the ClaI site was joined to p818.1or and searched using BLAST against the GenBank and EMBL databases. Good homology to E.coli and H.influenzae ATP-dependent DNA helicase recombinase proteins (EC 3.6.1.) was obtained (Fig. 4.17). The BLAST search with the sequence from the ApaI end showed high sequence homology to the stringent response protein guanosine-3',5'-bis (diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2) of E.coli, H.influenzae and S.coelicolor (Fig. 4.18). In both E.coli and H.influenzae, the two proteins RecG helicase recombinase and ppGpp stringent response protein constitute part of the spo operon.

4.4.5. Spo operon.

A brief description will be given on the spo operons of E.coli and H.influenzae which appear to differ in their arrangements. In E.coli, spoT gene encodes guanosine-3',5-bis pyrophosphohydrolase (ppGpp) which is synthesized during stringent response to amino acid starvation. It is also known to be responsible for cellular ppGpp degradation (Gentry and Cashel, 1995). The RecG protein is required for normal levels of recombination and DNA repair. RecG protein is a junction specific DNA helicase that acts post-synaptically to drive branch migration of Holliday junction intermediate made by RecA during the strand exchange stage of recombination (Whitby and Lloyd, 1995).

The spoS (also called rpoZ) encodes the omega subunit of RNA polymerase which is found associated with core and holoenzyme of RNA polymerase. The physiological function of the omega subunit is unknown. Nevertheless it binds stoichiometrically to RNA polymerase,
Fig. 4.18. (A) Results of the BLAST search of the nucleotide sequence of p818.11r (from the ApaI restriction site) inverted and complemented. Results revealed high protein sequence homology to stringent response protein: guanosine-3', 5'-bis(diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2) [ppGpp]-ase] (penta-phosphate guanosine -3'-pyrophosphorylase) of *E.coli*, *M.leprae*, *S.coelicolor* and *H.influenzae* (b). The nucleotide sequence of p818.11r and the open reading frame with homology to the SpoT protein.

(a)

<table>
<thead>
<tr>
<th>Sequences producing High-scoring Segment Pairs</th>
<th>Reading Frame Score</th>
<th>High Sum Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>*sp</td>
<td>P17580</td>
<td>SPOT_ECOLI GUANOSINE-3',5'-BIS(DIPHOSPHATE)</td>
</tr>
<tr>
<td>*sp</td>
<td>P43811</td>
<td>SPOT_HAEIN GUANOSINE-3',5'-BIS(DIPHOSPHATE)</td>
</tr>
<tr>
<td>*gp</td>
<td>U22374</td>
<td>VSU22374_1 carS gene product [Vibrio sp. S14]</td>
</tr>
<tr>
<td>*gp</td>
<td>U29580</td>
<td>ECU29580_9 GTP pyrophosphokinase [Escherichia...</td>
</tr>
<tr>
<td>*gp</td>
<td>U13769</td>
<td>VSU13769_2 ppGpp synthetase I [Vibrio sp. ]</td>
</tr>
<tr>
<td>*gp</td>
<td>X87267</td>
<td>SCAPTUREL2 putative ppGpp synthetase [Stre...</td>
</tr>
<tr>
<td>*sp</td>
<td>P15855</td>
<td>REL_HAEIN GTP PYROPHOSPHOKINASE (EC 2.7.6...</td>
</tr>
<tr>
<td>*pir</td>
<td>S39975</td>
<td>S39975 stringent response-like protein</td>
</tr>
<tr>
<td>*sp</td>
<td>P44644</td>
<td>REL_HAEIN GTP PYROPHOSPHOKINASE (EC 2.7.6...</td>
</tr>
</tbody>
</table>

Length = 702

Plus Strand HSPs:

Score = 277 (127.4 bits), Expect = 2.5e-31, P = 2.5e-31
Identities = 53/82 (64%), Positives = 62/82 (75%), Frame = +2

Query: 14 QASGREKHYVRMRKQMGKYGAFGDIHDLHAFRITIVADVTWCYRVLGKHSLSYRPFPGRF 193 + SGREKH+Y K M K F I D+++AFR+IVD D TCTYRVLG +HSLY+P QGR
Sbjct: 232 RVSGREKHLYSYIYCKMVLKEQRFS1MD1YAFRVIVNDSDTCYRVLGQMSLYKPRPGRV 291

Query: 194 KIDYIAIPKNSNYQSLHTVLAPG 259 KIDYIAIPK+NGYQSLHT + GP
Sbjct: 292 KIDYIAIPKANGYQSLHTSMIGP 313

>sp|P43811|SPOT_HAEIN GUANOSINE-3',5'-BIS(DIPHOSPHATE) 3'-PYROPHOSPHOHYDROLASE (EC 3.1.7.2) [ppGpp]-ase (penta-phosphate guanosine-3'-pyrophosphohydrolase) homolog - Haemophilus influenzae (strain Rd KW20) >gp|U00087|HIU00087_15 penta-phosphate guanosine-3'-pyrophosphohydrolase [Haemophilus influenzae] >gp|I44642|HEH1741_1 penta-phosphate guanosine-3'-pyrophosphohydrolase [Haemophilus influenzae] >gp|U32847|HIU32847_2 penta-phosphate guanosine-3'-pyrophosphohydrolase [Haemophilus influenzae]

Length = 677
Plus Strand HSPs:

Score = 268 (123.3 bits), Expect = 4.5e-30, P = 4.5e-30
Identities = 49/83 (59%), Positives = 64/83 (77%), Frame = +2

Query: 11 AQASGREKHYVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVTDCYRVLGLVHSLYRIPGR 190
       +GREKH+Y+ +KM+ K F I D++AFR+IV++VD CYRVLG +H+LY+P PGR
Sbjct: 204 ARVWGREKHLKYQKMRLKQDQEFHSHIDYAFRIVKVNVDCCYRVLQGMHNLKPRPGR 263

Query: 191 FKDYSIAIPKSNGYQSLHTVLGP 259
       KDYIA+PK+NGYQSL T + GP
Sbjct: 264 VKDYIAVPKANGYQSLQTSMIGP 286

>gp|U22374|VSU22374_1 csrS gene product [Vibrio sp. S14]
   Length = 119

   Plus Strand HSPs:

Score = 239 (109.9 bits), Expect = 5.7e-28, P = 5.7e-28
Identities = 44/68 (64%), Positives = 54/68 (79%), Frame = +2

Query: 56 KMQKKGYAFGDIHDLHAFRIIVADVTDCYRVLGLVHSLYRIPGRFKDYIAIPKSNGYQ 235
       KMQKKGYAFGDIHDLHAFRIIVADVTDCYRVLGLVHSLYRIPGRFKDYIAIPKSNGYQ 235
Sbjct: 3 KMKNKQERPHSIDYAFRIVLVSDDTCYRVLQGMHNLKPRPFSMMDYIAIPKSNGYQ 62

Query: 236 LHTVLGP 259
       L T L GP
Sbjct: 63 LTTSLVGP 70

>gp|U29580|ECU2958_9 GTP pyrophosphokinase [Escherichia coli]
   Length = 744

   Plus Strand HSPs:

Score = 239 (109.9 bits), Expect = 5.1e-26, P = 5.1e-26
Identities = 43/83 (51%), Positives = 56/83 (67%), Frame = +2

Query: 11 AQASGREKHYVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVTDCYRVLGLVHSLYRIPGR 190
       +GR KH+Y RKMQKK AF ++ D+ A RI+ ++ CY LG+LY+P PGR
Sbjct: 247 AEYVGRPKHYISIRKMQKKNLAFDELFDVRAVRVAERLQDCYAAAGLGVHTHRHLFDE 306

Query: 191 FKDYSIAIPKSNGYQSLHTVLGP 259
       F D+APK NGYQSLHTVLGP
Sbjct: 306 FDDYVANPKNGYQSIHTVLGP 329

>gp|U13769|VSU13769_2 ppGpp synthetase I [Vibrio sp.]
   Length = 744

   Plus Strand HSPs:

Score = 239 (109.9 bits), Expect = 5.1e-26, P = 5.1e-26
Identities = 42/83 (50%), Positives = 56/83 (67%), Frame = +2

Query: 11 AQASGREKHYVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVTDCYRVLGLVHSLYRIPGR 190
       +GR KH+Y RKMQKK F ++ D++ A RI+ ++ CY LG+LY+P PGR
Sbjct: 246 AEYVGRPKHYISIRKMQKKSLFDELFDVRAVRVAERLQDCYAAAGLGVHTHRHLFDE 305

Query: 191 FKDYSIAIPKSNGYQSLHTVLGP 259
       F D+APK NGYQSLHTVLGP
Sbjct: 306 FDDYVANPKNGYQSIHTVLGP 329
SCAPTRELA_2 putative ppGpp synthetase [Streptomyces coelicolor]

Length = 847

Plus Strand HSPs:
Score = 233 (107.2 bits), Expect = 3.6e-25, P = 3.6e-25
Identities = 44/86 (51%), Positives = 56/86 (65%), Frame = +2

Query: 2 RFAAQASGREKHVYRSIKMKQKGYAFGDIHDLHAFRIVYADVFYRTLGLVHSLYRPI 181
   R A +GR K H Y +K M +G F +I+DL R++V V CY LG VH+ +P+
Sbjct: 330 RIKATVTGRFHYSSYQMMIVRGRFDYAEYDLYTGIRVLVDVDRDCAALGTVHAKNPV 389

Query: 182 PGRFKDYIAIPKNSGYQSLHTVLAGP 259
   PGRFKDYIA+PK N YQSLHT+GP
Sbjct: 390 PGRFKDIAMPKFNYQSLHTTVISP 415

(b)
cross-links specifically to $\beta'$ subunit and is immunologically conserved among bacteria (Gentry et al., 1991).

*SpoU* is the only functionally uncharacterized ORF in the *spo* operon. Its been reported to have high amino acid similarity to RNA methylase encoded by *tsr* gene of *Streptomyces azureus* (Koonin and Rudd, 1993). Expression of *tsr* gene prevents the antibiotic thiopeptin from inhibiting ppGpp synthesis during nutritional shift-down in *S.lividans* (Ochi, 1989). Thus putative SpoU rRNA methylase may be involved in stringent starvation response and functionally connected to SpoT (Koonin and Rudd, 1993).

The *recG, spoT, spoU* and *spoS* genes form the *spo* operon in both *E.coli* and *H.influenzae* (Fig. 4.19). The arrangement of genes in the *spo* operon between these two organisms differ with respect to where the *spoU* gene is located in the operon. In *E.coli*, *spoU* is found between the *spoT* and the *recG* genes whereas in *H.influenzae*, it is found between the *spoS* and the *spoT* genes (Fig. 4.19). In the case *T.ferrooxidans*, the *spoT* and *recG* genes are physically linked but are arranged in opposite orientations to each other. Transcription of the *T.ferrooxidans recG* and *spoT* genes appear to be divergent from a common region about 1.0-1.3 kb from the *ApaI* restriction site (Fig. 4.11). From the limited amount of sequence information available, it is not possible to predict whether *spoU* or *spoS* genes are present and which of the genes constitute an operon.
Fig. 4.19. Comparison of the spo operons of (1) E.coli and (2) H.influenzae and the probable structure of the spo operon of T.ferrooxidans (3). Note the size of the spoU gene in E.coli as compared to that of H.influenzae. Areas with bars indicate the respective regions on the both operons where good homology to T.ferrooxidans was observed. Arrows indicate the direction of transcription of the genes in the operon.
CHAPTER FIVE

GENERAL CONCLUSIONS
CHAPTER 5.

GENERAL CONCLUSIONS.

The objective of this project was to study the region beyond the *T.ferrooxidans* (strain ATCC 33020) *unc* operon. The study was initiated after random sequencing of a *T.ferrooxidans* cosmid (p818.1) harbouring the *unc* operon (which had already been shown to complement *E.coli* *unc* mutants) revealed a putative transposase with very high amino acid sequence homology to Tn7. Secondly, nucleotide sequences (about 150 bp) immediately downstream the *T.ferrooxidans unc* operon had been found to have high amino acid sequence homology to the *E.coli glmU* gene.

A piece of DNA (p818.52) within the Tn7-like transposon covering a region of 1.7 kb was used to prepare a probe which was hybridized to cosmid p818.1 and to chromosomal DNA from *T.ferrooxidans* (Fig. 2.4). This result confirmed the authenticity of a region of more than 12 kb from the cosmid as being representative of unrearranged *T.ferrooxidans* chromosome. The results from the hybridization experiment showed that, only a single copy of this transposon (Tn5468) exists in the *T.ferrooxidans* chromosome. This chromosomal region of *T.ferrooxidans* strain ATCC 33020 from which the probe was prepared also hybridized to two other *T.ferrooxidans* strains namely ATCC 19859 and ATCC 23270 (Fig. 2.6). The results revealed that, not only is this region with the Tn7-like transposon native to *T.ferrooxidans* strain ATCC 33020, but appears to be widely distributed amongst *T.ferrooxidans* strains. *T.ferrooxidans* strain ATCC 33020 was isolated from a uranium mine in Japan, strain ATCC
23270 from coal drainage water in Pittsburgh, USA and strain ATCC 19859 from Canada and therefore the Tn7-like transposon has a wide geographical distribution amongst *T. ferrooxidans* strains.

On other hand, hybridization of this region to two *T. thiooxidans* strains ATCC 19377 and DSM 504 and *L. ferrooxidans* strain DSM 2705 proved negative (Fig. 2.6). Thus all three *T. ferrooxidans* strains tested carry a transposon homologous to Tn7 while the other three isolates of gram-negative bacteria which were tested and which grow in the same environment do not. The fact that all the bands of the *T. ferrooxidans* strains which hybridized to the Tn5468 probe were of the same size implies that, this chromosomal region is highly conserved within the *T. ferrooxidans* strains studied.

A 3.5 kb *BamHI-BamHI* fragment of *T. ferrooxidans* DNA downstream of the *unc* operon was cloned into pUCBM20 and pUCBM21 vectors (p818.16f and p818.16r). This piece of DNA has been sequenced from both directions and found to have one partial open reading frame (ORF1) and two complete open reading frames (ORF2 and ORF3). The partial open reading frame (ORF-1) was found to have very high amino acid sequence homology to *E. coli* and *B. subtilis* *glmU* gene products and represents about 150 amino acids from the C-terminus of the *T. ferrooxidans* GlmU protein. The second complete open reading frame (ORF-2) has been shown to be the *T. ferrooxidans* glucosamine synthetase gene. It has high amino acid sequence homology to six *purF*-type amidotransferases. The constructs p818.16f and p818.16r complemented the *E. coli* *glmS* mutant (CGSC #5392) as it enabled the mutant to grow as large colonies when N-acetyl glucosamine was added to the media. The larger fragment p818.20 (10.2 kb), containing the entire gene failed to complement the *E. coli* *glmS* mutant.
This was probably due to a lack of expression in the absence of a suitable vector promoter. The third open reading frame (ORF-3) had shown high amino acid sequence homology to the TnsA protein of Tn7 (Fig. 4.3). The region beyond the BamHI-BamHI 3.5 kb construct covering about 10 kb of the *T. ferrooxidans* chromosome has been studied by subcloning and single strand sequencing. Homology to the TnsB, TnsC and TnsD in addition to the already mentioned TnsA protein of Tn7 was found within this region. The TnsD-like ORF appears to have undergone some rearrangement and is almost certainly no longer functional.

Homology to the TnsE and the antibiotic resistance genes of Tn7 was not found though a fragment size of about 4 kb beyond where homology to TnsD had been found was searched. The search for the TnsE and the antibiotic resistance markers was aborted when it became apparent that, the genes *spoT*, encoding guanosine-3',5'-bis (diphosphate) 3'-pyrophosphohydrolase, (EC 3.1.7.2) and *recG* encoding ATP-dependent DNA helicase, RecG (EC 3.6.1.) had been encountered about 1.5 and 3.5 kb respectively beyond where final homology to TnsD had been found. These genes were identified by their high sequence homology to *E.coli* and *H.influenzae* SpoT and RecG proteins (Fig 4.17 and 4.18). Though these genes are transcribed in the same direction and form part of the *spo* operon in *E.coli* and *H.influenzae*, in *T.ferrooxidans* the *spoT* and the *recG* genes appear to be transcribed in the opposite directions from a common region (Fig. 4.19).

The transposon had been registered as Tn5468 (before it was known that it is probably non-functional). The degree of similarity between Tn7 and Tn5468 suggests that they originated from a common ancestor. Since *T.ferrooxidans* only grows in an acidic inorganic
environment, one would predict that Tn5468 has evolved in a very different environment to Tn7. For example, Tn7 has acquired antibiotic resistance determinants as accessory genes which are presumably advantageous to its *E.coli* host. One would not expect the same antibiotic resistance genes to confer a selective advantage to *T.ferrooxidans*, which is not exposed to a medical environment. It was disappointing that the TnsD-like protein at distal end of Tn5468 appears to have been truncated as it might have provided an insight into the structure of the common ancestor of Tn7 and Tn5468.

The region beyond the *T.ferrooxidans* unc operon has been studied and found to have genetic arrangement very similar to that an *E.coli* strain which has had a Tn7 insertion at attTn7. The arrangement of genes in such an *E.coli* strain is unc_glmU_glmS_Tn7, which is identical to that of *T.ferrooxidans*, unc_glmU_glmS_(Tn7-like). This arrangement must be a carry over from a common ancestor before the organisms became exposed to different environmental conditions and differences in their chromosomes were magnified. Based on 16Sr RNA sequence data, *T.ferrooxidans* and *T.thiooxidans* are phylogenetically very closely related whereas *L.ferrooxidans* is very distantly related (Lane et al. 1992). Presumably *T.ferrooxidans* and *T.thiooxidans* originated from a common ancestor. If Tn5468 had been inserted into the common ancestor before *T.ferrooxidans* and *T.thiooxidans* diverged, one might have expected Tn5468 to be present in the *T.thiooxidans* strains examined. A plausible reason for the absence of Tn5468 in *T.thiooxidans* could be that these two organisms diverged before *T.ferrooxidans* acquired Tn5468. Nevertheless, the Tn7-like transposon must have become inserted into *T.ferrooxidans* a long time ago because the three strains with Tn7-like transposon were isolated from geographical locations as far apart as the USA, Japan and Canada.
APPENDIX
APPENDIX 1

MEDIA

Luria agar (LA)

Tryptone 10 g
Yeast extract 5 g
NaCl 5 g
Agar 15 g
Distilled water 1000 ml
Autoclaved.

Luria broth (LB)

Tryptone 10 g
Yeast extract 5 g
NaCl 5 g
Distilled water 1000 ml
Autoclaved.

N-acetyl glucoammine + Luria agar

Luria agar + N-acetyl glucosamine solution (200μg/ml). N-acetyl glucosamine added after autoclaved LA has been melted and temp of melted LA had fallen below 45 °C. N-acetyl glucosamine solution was filter sterilized.
APPENDIX 2.

BUFFERS AND SOLUTIONS

Plasmid extraction solutions (Ish-Horowicz & Burk, 1981)

**Solution I**

50 mM glucose
25 mM Tris-HCl
10 mM EDTA

Dissolve Tris-HCl in water, and adjust pH to 8.0 with concentrated HCl. Add glucose and EDTA. Dissolve and make up to volume with water. Sterilize by autoclaving, and store at room temp.

**Solution II**

Prepare stock solutions of SDS (25% m/v) and NaOH (10 N). Autoclave separately, and store at 4 °C. Prepared weekly by adding 4 ml SDS to 2 ml NaOH. Make up to 100 ml with water.

**Solution III**

3 M K\textsubscript{2}OAc
2 M HOAc

Dissolve 29.4 g potassium acetate in 70 ml H\textsubscript{2}O. Adjust pH to 4.8 with glacial acetic acid. Sterilize by autoclaving. Store on the shelf.
TBE buffer (pH 8.0)

Tris 89 mM
Glacial acetic acid 89 mM
EDTA 2.5 mM

TE buffer (pH 8.0)

Tris 10 mM
EDTA 1 mM

Dissolve in water and adjust pH to 8.0 with conc. HCl.

TE buffer (pH 7.5)

Tris 10 mM
EDTA 1 mM

Dissolve in H₂O, and adjust pH to 7.5 with concentrated HCl. Autoclave.

Salt saturated isopropanol

Dissolve 29.22 g NaCl in 10 ml 10 X TE buffer; make up to 100 ml with water. Add 200 ml isopropanol. Mix well, and allow to stand overnight at room temperature.
Plasmid extraction buffer solutions (Nucleobond Kit).

S1.

50 mM Tris/HCl, 10 mM EDTA, 100 μg RNase /ml, pH 8.0

Store at 4 °C

S2.

200 mM NaOH, 1% SDS. Storage: room temp.

S3.

2.60 M KAc, pH 5.2. Store at room temp.

N2.

100 mM Tris, 15% ethanol and 900 mM KCl adjusted with H₃PO₄ to pH 6.3; store at room temp.

N3.

100 mM Tris, 15% ethanol and 1150 mM KCl adjusted with H₃PO₄ to pH 6.3; store at room temp.

N5.

100 mM Tris, 15% ethanol and 1000 mM KCl adjusted with H₃PO₄ to pH 8.5; store at room temp.

Competent cells preparation buffers.

Stock solutions:

i) TFB-1 (100 mM RbCl, 50 mM MnCl₂, 30 mM KOAc, 10 mM CaCl₂, 15% glycerol).

For 50 ml:

1 mM RbCl (Sigma) 5.0 ml
MnCl$_2$.4H$_2$O (Sigma, tetrahydrate) 0.495 g
KOAc (BDH) Analar 0.147 g
750 mM CaCl$_2$.2H$_2$O (Sigma) 0.67 ml
50% glycerol (BDH analar) 15.0 ml
Adjust pH to 5.8 with glacial acetic acid, make up volume with H$_2$O and filter sterilize.

ii) TFB-2. (10mM MOPS pH 7.0 (with NaOH) 5.0 ml
1 M RbCl$_2$ (Sigma) 0.5 ml
750 mM CaCl$_2$ (Sigma, dihydrate) 5.0 ml
50% glycerol (BDH Analar) 15.0 ml
Make up volume with dH$_2$O and filter sterilize.

Southern blotting and Hybridization solutions.

20 x SSC
175.3g NaCl + 88.29 sodium citrate in 800 ml H$_2$O
pH adjusted to 7.0 with 10 N NaOH. Distilled water added to 1000 ml.

Prehybridization solutions.

stock solution

<table>
<thead>
<tr>
<th>stock solution</th>
<th>final</th>
<th>volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x SSC</td>
<td>20 X</td>
<td>25.0 ml</td>
</tr>
<tr>
<td>5 % Elite Milk powder</td>
<td>-</td>
<td>4.0 g</td>
</tr>
<tr>
<td>0.1 N-laurylsarcosine</td>
<td>10 %</td>
<td>1.0 ml</td>
</tr>
<tr>
<td>0.02 % SDS</td>
<td>10 %</td>
<td>0.2 ml</td>
</tr>
</tbody>
</table>
Water - 69.0 ml

Microwave to about 65 °C and stir to dissolve. Make fresh.

Buffer 1:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maleic acid</td>
<td>11.6 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>8.79 g</td>
</tr>
<tr>
<td>H₂O</td>
<td>to 1 litre.</td>
</tr>
</tbody>
</table>

pH to 7.5 with conc. NaOH (± 20 ml). Autoclave.

Wash buffer:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tween 20</td>
<td>1.0 ml</td>
</tr>
<tr>
<td>Buffer 1</td>
<td>199.0 ml</td>
</tr>
</tbody>
</table>

Buffer 2:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elite milk powder</td>
<td>8.0 g</td>
</tr>
<tr>
<td>Buffer 1</td>
<td>190.0 ml</td>
</tr>
</tbody>
</table>

DEA Buffer:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diethanolamine</td>
<td>193.0 μl</td>
</tr>
<tr>
<td>H₂O</td>
<td>19.7 ml</td>
</tr>
<tr>
<td>1 M MgCl₂</td>
<td>100.0 μl</td>
</tr>
</tbody>
</table>

pH to 10 with conc. HCl (± 15 μl).
Washes

<table>
<thead>
<tr>
<th>20 x SSC</th>
<th>10 % SDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (2 x SSC, 0.1 % SDS)</td>
<td>10.0 ml</td>
</tr>
<tr>
<td>B (0.1 x SSC, 0.1 % SDS)</td>
<td>0.5 ml</td>
</tr>
</tbody>
</table>

Both A and B made up to a total of 100 ml with water.

Stop Buffer

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromophenol Blue</td>
<td>0.25 g</td>
</tr>
<tr>
<td>Xylene cyanol</td>
<td>0.25 g</td>
</tr>
<tr>
<td>Ficoll type 400</td>
<td>15.0 g</td>
</tr>
</tbody>
</table>

Dissolve in 100ml water. Keep at room temp.

Ethidium bromide

A stock solution of 10 mg/ml was prepared by dissolving 0.1 g in 10 ml of water. Shaken vigorously to dissolve and stored at room temp in a light proof bottle.

Ampicillin (100 mg/ml)

Dissolve 2 g in 20 ml water. Filter sterilize and store aliquots at 4 ºC.

Photography of gels

Photos of gels were taken on a short wave Ultra violet (UV) transilluminator using Polaroid camera. Long wave transilluminator was used for DNA elution and excision.
Preparation of agarose gels

Unless otherwise stated, gels were prepared by dissolving 0.8% (m/v) type II low endo-osmotic agarose (sigma) in 1 x TBE buffer. Solution is heated in microwave oven until the agarose is completely dissolved. It is then cooled to 45 °C and prepared into a gel mold.

IPTG (Isopropyl-β-D-Thio-galactopyranoside)

Prepare a 100 mM "stock solution" Filter sterilize.

X-gal (5-Bromo-4-chloro-3 indolyl-β-galactoside).

Dissolve 25 mg in 1.25 ml dimethylformamide. To prepare, dissolve 500 μl X-gal, 125 μl IPTG and 500 μl ampicillin (10 mg/ml) in 500 ml sterilized LA of about 45 °C in temp.
APPENDIX 3
GENOTYPE AND PHENOTYPE OF STRAINS OF BACTERIA USED.

*E.coli* JM105

**GENOTYPE:** endA1, thi, rpsL, sbcB15, hsdR4, Δ(lac-proAB), (F, traD36, proAB, lacIqZΔM15).

**PHENOTYPE:** thi-, strR, lac-, pro-.


*E.coli* JM109

**GENOTYPES:** endA1, recA1, gyrA96, thi, hsdR17(rK,mK), relA1, supE44, Δ(lac-proAB), (F,traD36, proAB, lacIqZΔM15).

**PHENOTYPE:** thi-, nalR, lac-, pro-

Reference: Promega catalog.

*E.coli* CGSC5392

Thr-1, ara-14, leuB6, DE (gpt-proA), lacY1, tsx-33, qsr'-0, glnV44 (AS), galK2, LAM-, rac-0, hisG4 (OC), rfbD1, mgl51, rspL31 (Str), kdgK51, xylA5, mtl-1, glms1, argE3 (OC), thi-1.

Reference: CGSC Webserver.
APPENDIX 4.

STANDARD TECNIQUES

Plasmid DNA extraction method (Nucleobond Kit)

Innoculate from single colony into 100 ml LB plus antibiotic.

Grow O/N at 37 °C with shaking.

Harvest cells, Room temp.

Resuspend cell pellet in 4 ml S1.

Add 4 ml S2, mix by inversion; keep at room temp for 5 mins.

Add 4 ml S3. Mix by shaking to homogenous suspension.

Spin at 15 K for 40 mins at 4 °C.

Carefully remove supernatant to fresh tube.

Equilibrate Nucleobond column with 2 ml N2

Load supernatant in 2 to 4 ml amounts.

Wash column with 2 X 4 ml of N3. Elute the DNA with 1.5 ml of N5 (discard the first bed volume of about 8 to 10 drops). To the eluent (two tubes of about 700 μl each), add 0.7 volumes of plain isopropanol.

Spin at 4 °C. Wash with 70 % Ethanol. Resuspended pellet in approx 100 μl of TE and scan to get the concentration.
SEQUITHERM CYCLE SEQUENCING

Alf-express Cy5 end labelled promer method.

Use only DNA transformed into end- *E.coli* strain.

The label is sensitive to light, do all steps with fluorescent lights off.

3-5 kb  3μg  
3-7 kb  4μg  
7-10 kb  6μg

Thaw all reagents from kit at RT; mix well before use and keep on ice.

1) Label 200 PCR tubes on side or little cap flap.

(The heated lid removes markings from the top of the tubes)

2) Add 3 μl of termination mixes to labelled tubes.

3). On ice with fluorescent lights off, using 1.2 ml Eppendorf tubes, Make DNA up to 12.5 μl with MilliQ water.

Add 1 μl of Primer

Add 2.5 μl of 10X sequencing buffer.

Add 1 μl Sequitherm DNA polymerase.

Mix well, spin. Aliquot 3.8 μl from the eppendorf to each termination tube. Spin down.

Push caps on properly.
Hybaid thermal Cycler.

Program

- 93 °C for 5 mins  1 cycle
- 93 °C for 30 secs
- 55 °C for 30 secs
- 70 °C for 60 secs  30 cycle  93 °C- 30s: 55 °C-30s:
- 70 °C for 5 mins. 1 cycle

Primer must be min 20 bp long and min. 50% GC content if the annealing step is to be omitted. Incubate at 95 °C for 5 mins to denature before running. Spin down. Load 3 µl.

SEQUITHERM CYCLE SEQUENCING

Ordinary method.

Use only DNA transformed into end- E.coli strain.

- 3-5 kb  3µl
- 3-7 kb  4µl
- 7-10 kb  6µl

Thaw all reagents from kit at RT; mix well before use and keep on ice.

1) Label 200 PCR tubes on side or little cap flap.

(The heated lid removes markings from the top of the tubes)

2) Add 3 µl of termination mixes to labelled tubes.

3). On ice using 1.2 ml Eppendorf tubes, make DNA up to 12.5 µl with MilliQ water.
Add 1 μl of Primer

Add 2.5 μl of 10X sequencing buffer.

Add 1 μl Sequitherm DNA polymerase.

Mix well, spin. Aliquot 3.8 μl from the eppendorf to each termination tube. Spin down.

Push caps on properly.

Hybaid thermal Cycler.

Program

93 °C for 5 mins 1 cycle
93 °C for 30 secs
55 °C for 30 secs
70 °C for 60 secs 30 cycles
93 °C- 30s: 55 °C-30s: 70 °C for 5 mins. 1 cycle

Primer must be minimum of 20 bp long and min. 50% GC content if the annealing step is to be omitted. Incubate at 95 °C for 5 mins to denature before running.

Spin down. Load 3 μl for short and medium gels; 2μl for long gel runs.
REFERENCES.


Fraga, D., Hermolin, J., Oldenburg, M., Miller, M.J. and Fillingame, R. H. 1994. Arginine 41 of subunit c of *Escherichia coli* H+-ATP synthase is essential in binding and coupling of F\textsubscript{i} to F\textsubscript{o}, J. Biol. Chem. **269**:7532-7537.


Girvin, M. E. and Fillingame, R. H. 1993. Helical structure and folding of subunit c of F\textsubscript{i}F\textsubscript{o} ATP synthase: \textsuperscript{1}H NMR resonance assignments and NOE analysis, Biochemistry **32**:12167-12177.


Hoppe, J. and Sebald, W. 1986. Topological studies suggest that the pathway of the protons through F0 is provided by amino acid residues accessible from the lipid phase, Biochimie. (Paris) 68:427-434.


Schneider, E and Altendorf, K. 1985. All the three subunits are required for the reconstitution of an active proton channel (F_0), of Escherichia coli ATP synthase (F_1F_0), EMBO J. 4:515-518.


