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.

TABLE OF CONTENTS

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)
oriC	chromosomal origin of replication.

Phe	L-phenylalanine
pho	phosphate
Pi	inorganic phosphate
Pro	L-proline
RNA	ribonucleic acid
rpm	revolutions per minute.
SDS	Sodium Dodecyl Sulphate
Ser	L-serine
T	thymine
Thr	L-threonine
Trp	L-tryptophan
Tyr	L-tyrosine
TE	Tris-EDTA buffer
Tn	Transposon
UV	Ultra-violet
UDP-glc	Uridyldiphosphoglucosamine
Val	L-valine
v/m	volume per mass
v/v	volume per volume
X-gal	5-bromo-4-chloro-3-indolyl-galactoside.

LIST OF TABLES

ABSTRACT

A Tn*7*-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 Tn*7*-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 *Bam*HI-*Bam*HI 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 Tn*7*. Inverted repeats very similar to the imperfect inverted repeat sequences of Tn*7* 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 *Bam*HI-*Bam*HI fragment were carried out and the sequences searched against the GenBank and EMBL databases. Sequences homologous to the TnsBCD proteins of Tn*7* were found. The TnsD-like protein of the Tn*7*-like element (registered as Tn*5468*) was found to be shuffled, truncated and rearranged. Homologous sequence to the TnsE and the antibiotic resistance markers of Tn*7* were not found. Instead, single strand sequencing of a further 3.5 kb revealed sequences which suggested the *T.ferrooxidans spo* operon had been encountered. High amino acid sequence homology to two of the four genes of the *spo* operon from *E.coli* and *H.influenzae* namely, *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.) was found. This suggests that Tn*5468* is incomplete and appears to terminate with the reshuffled TnsD-like protein. The orientation of the *spoT* and *recG* genes with respect to each other was found to be different in *T.ferrooxidans* compared to those of *E.coli* and *H.influenzae.*

CHAPTER 1.

LIST OF FIGURES.

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 indrectly 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 microrganism with the insoluble sulfide crystals and was independent of the the ferric-ferrous cycle. Insoluble sulphide minerals can be degraded by microrganisms 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 (FeS_2) and chalcopyrite ($CuFeS_2$) is followed by the following reactions.

For pyrite :

$$FeS_2 + 31/2O_2 + H_2O \rightarrow FeSO_4 + H_2SO_4$$

$$2FeSO_4 + 1/2O_2 \text{ (bacteria)} \rightarrow Fe(SO_4)_3 + H_2O$$

$$FeS_2 + Fe_2(SO_4)_3 \rightarrow 3FeSO_4 + 2S$$

$$2S + 3O_2 + 2H_2O \text{ (bacteria)} \rightarrow 2H_2S$$

For chalcopyrite:

$$2CuFeS_2 + 1/2O_2 + H_2SO_4 \text{ (bacteria)} \rightarrow 2CuSO_4 + Fe(SO_4)_3 + H_2O$$

$$CuFeS_2 + 2Fe_2(SO_4)_3 \rightarrow CuSO_4 + 5FeSO_4 + 2S$$

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 *Thiobaccillus ferrooxidans* because of its use in the industrial mineral processing and because of its unusual physiology. It is an autotrophic, chemolithotropic, 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 $(NH_4)_2SO_4$ and potassium hydrogen phosphate (K_2HPO_4) have to be added to the medium. Some of the unique features of *T.ferrooxidans* 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. ferrooxidans* 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

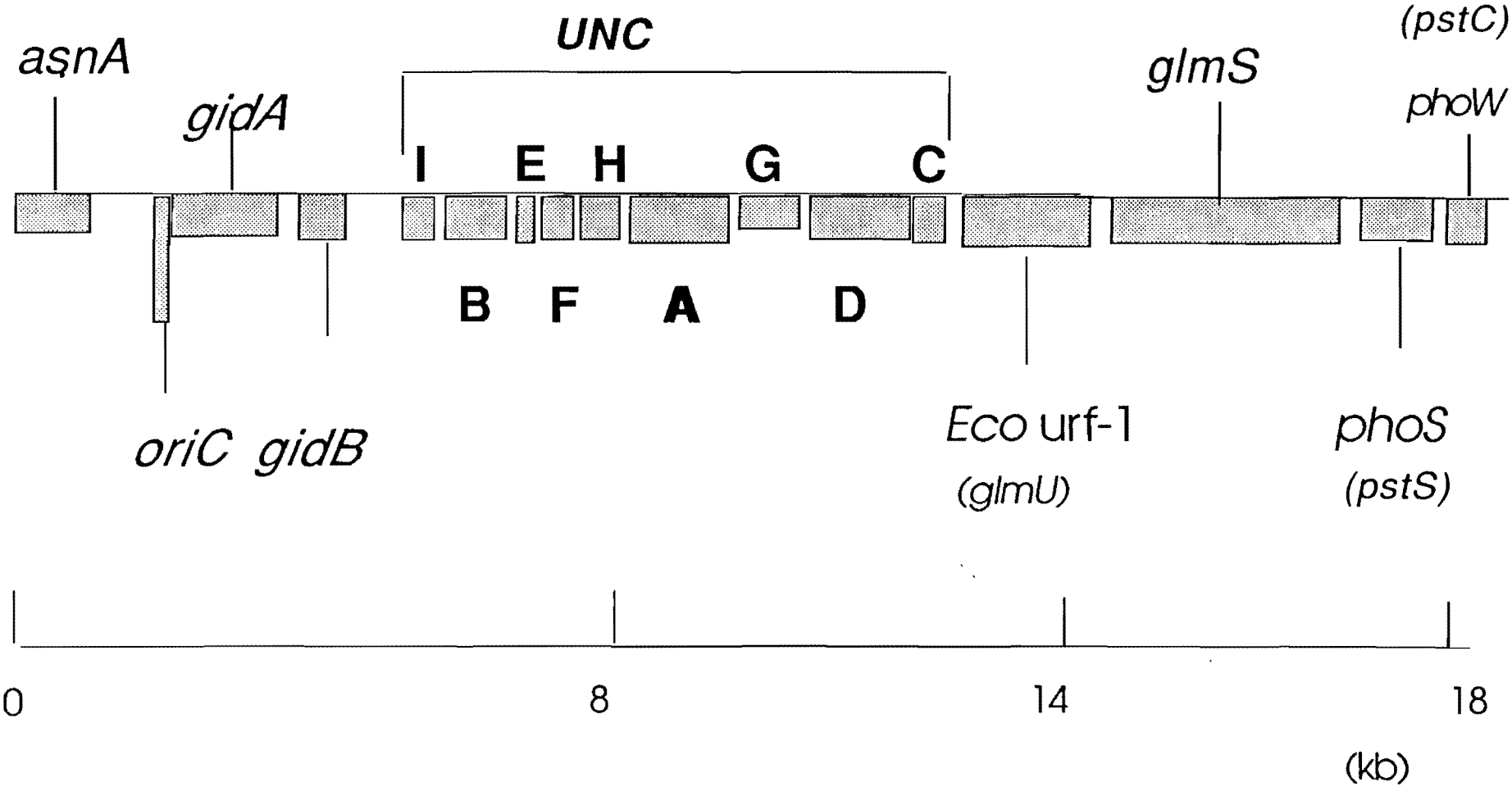


Fig. 1a. *E.coli* chromosome in the vicinity of the *unc* operon. Genes on the right of *oriC* are transcribed from left to right.

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 Tn*10* 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 Tn*10* 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

occured in *E.coli* chromosome (Ogasawara & Yoshikawa, 1992). This comparison also indicated that translocation of *oriC* together with the *gid* and *unc* operons may have occured during the evolution of the *E.coli* chromosome (Ogasawara & Yoshikawa, 1992).

1.3.2. The *unc* operon.

The ATP synthase (F_0F_1-ATPase), a key enzyme in energy converting reactions, couples the synthesis or hydrolysis of ATP to the translocation of protons (H^+) from across the 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 F_1 and an intrinsic membrane domain F_0, linked by a slender stalk about 45Å long (Fillingame, 1990). Each sector of F_0F_1 is composed of multiple subunits in unusual stoichiometric ratios, that is $\alpha_3\beta_3\tau_1\delta_1\varepsilon_1$ for F_1 and $a_1b_2c_{10}$ for F_0 of the *E.coli* enzyme (Fillingame, 1992). These eight subunits of the *E.coli* F_1F_0 complex are coded by the genes of the *unc* operon (Walker *et al.*, 1984).

1.3.2.1. F_0 subunit.

The three subunits of F_0 part of the gene has molecular masses of 30.276, 17.265, and 82.88 kDa (Walker *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 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 F_1 binding (Friedl *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; Bjobaek *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_1 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_0 (Steffens *at 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_0 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_0,

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 F_1 complex (Fillingame, 1992). Due to a high sequence similariity of the proteolipids from F_0F_1 ATPases, vacuolar H^+-ATPases and gap junctions, a number of 12 copies of subunit *c*/F_0 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 F_0 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; Filliingame, 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. F_1 subunit.

The F_1 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 F_0 is relayed to the catalytic sites in the F_1 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, F_1-ATPase (Walker *et al.*, 1994). This F_1 sector is the catalytic part

Fig. 1b.

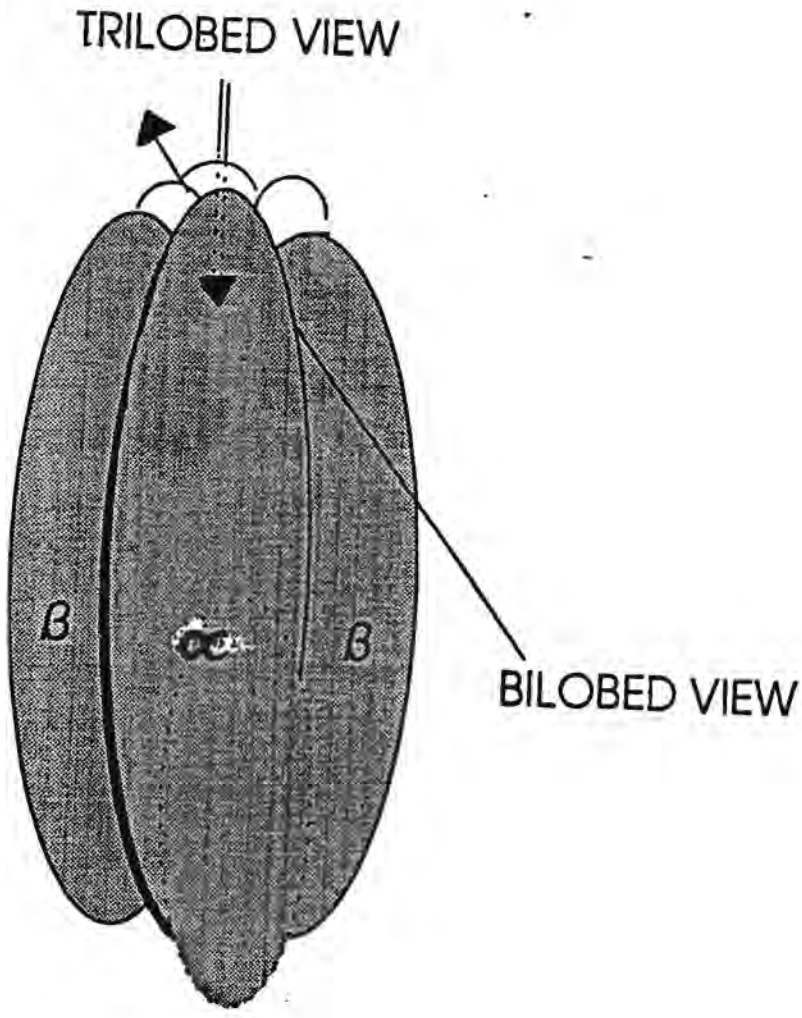


Fig. 1b. Schematic model of *E.coli* F_1 derived from projection views of unstained molecule interpreted in the framework of the three dimensional stain excluding outline. The directions of view (arrows) which produce the bilobed and trilobed projections are indicated. The peripheral subunits are presumed to be the α and ß subunits, and the smaller density interior to them probably consists of at least parts of the α, ß and/or ε subunits (Gogol *et al.*, 1989).

of the complex; there are three catalytic sites per molecule which have been localised to β subunit whereas the function of nucleotides in the α-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 α and three β 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 α and β subnits, each ≈ 90 Å X ≈30Å 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 (*Eco*URF-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 *Eco*URF-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 *Eco*URF-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 *Eco*URF-1 product was impaired, have established that the *Eco*URF-1 gene is an essential gene encoding the GlcNAc-1-P uridyltransferase activity. The *N*-acetylglucosamine-1-phosphate (GlcNAc-1-P) uridyltransferase 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 uridyltransferase) 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).

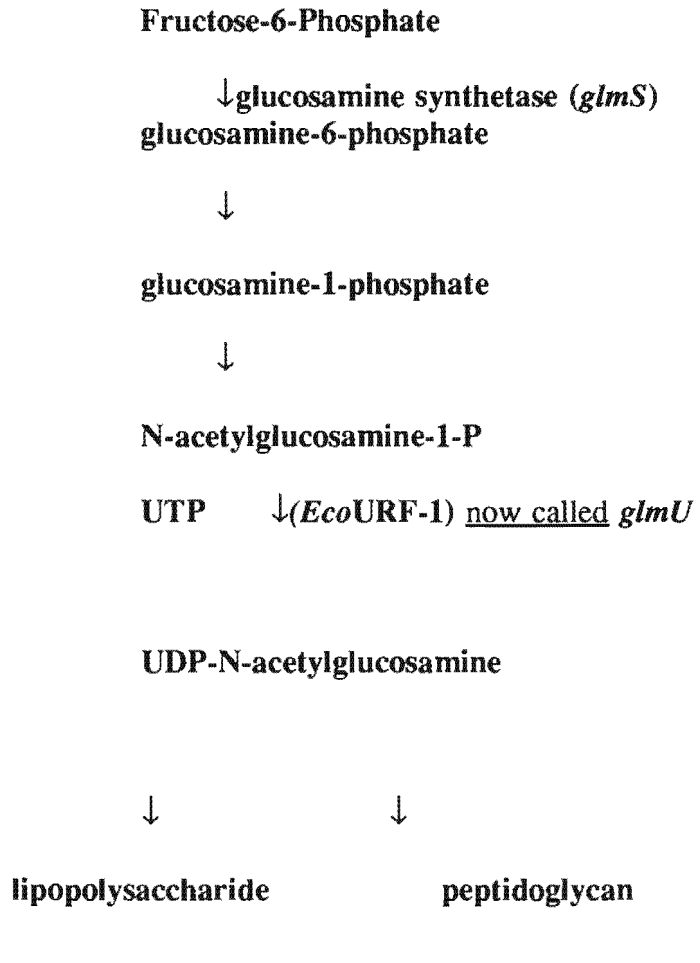


Fig. 1c. Biosynthesis and cellular utilization of UDP-Glc in *E.coli*.

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 (Dobrogozz, 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 ($GlcNH_2$-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 downtream 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-phospate 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 amidotranferases (a family of enzymes that utilize the amide of glutamine for the biosynthesis of serveral amino acids including arginine, asparagine, histidine, and trytophan as well as the amino sugar glucosamine 6-phospate). 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 (Badget *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 indispensabilty 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; Rosenburg *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 Rosenburg, 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 Rosenburg, 1970).

1.3.5.2. The Pst system.

A comparison of the kinetic parameters shows the transport constant (K_t) for the Pst system (about 0.25 μM) to be two orders of magnitude lower than that of the Pit system (about 20 μM, Rosenburg, 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: *bgl*... *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; Rosenburg, 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.

Proteins	Genes	Function
PstA	*pstA*(*phoT*)	Cytoplasmic membrane
PstB	*pstB*	Energy coupling peripheral membrane
PstC	*pstC*(*phoW*)	Cytoplasmic membrane protein
PiBP	*pstS*(*phoS*)	Periplasmic Pi binding proteun

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 Rosenburg, 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 ($4{,}7 \times 10^6$ 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

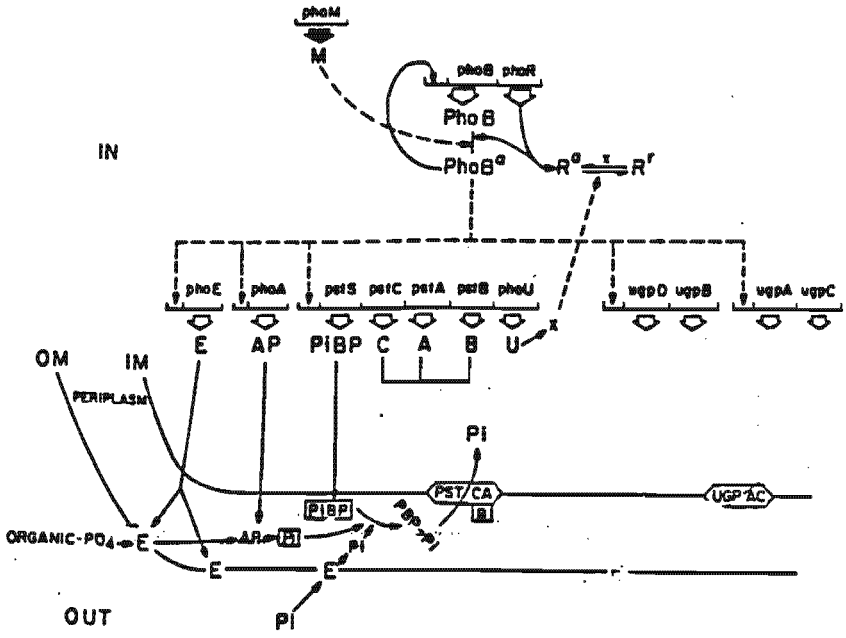


Fig. 1d. Utilization of organic phosphate by cells of *E.coli* starved of inorganic phosphate (Pi). The organic phosphates permeate the outer membrane (OM) and are hydrolized to Pi by the phosphatases of the periplasm. The Pi produced is captured by the (PiBP) Pi-binding protein (gene *pstS*) and is 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).

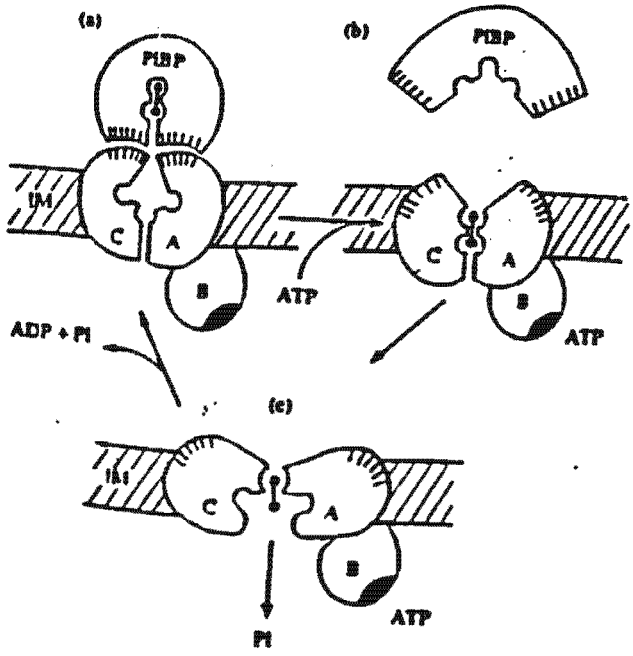


Fig. 1e. A model for Pst-dependent Pi transport modified from the one used (Treptow and Shuman, 1988) to explain the mechanism of maltose transport. (a) The PiBP has captured Pi (●—●); it adapts itself on the periplasmic domain of the two transmembrane proteins (PstC and PstA). The Pst protein is represented 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 cytoplsmic 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 establised 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 *pst* genes are expressed at high levels and the Pho regulon is induced via *phoR* and *PhoB*. If Pi is added to these cells, *pst* gene expresssion 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, *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 Tn*7* is reviewed because of Tn*7*-like segment which was encountered downstream the *glmS* gene.

1.3.6. 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 *et al.*, 1968; Shapiro and Adhya, 1969), and from the identification of these mutations as insertions (Starlinger *et al.*, 1968; Shapiro, 1969; Michealis *et al.*, 1969; Malamy, 1970). Hybridization studies showed that the inserted DNA belonged to only a few classes and were called insertion sequences (*IS*; Malamy *et al.*, 1972; Hirsch *et al*, 1972; Kahn and Schaefer, 1995). Besides their presence on the chromosome, these *IS* were discovered on bacteriophages (Brachet *et al.*, 1970) and on plasmids such as fertility factor, F (Ohtsubo and Davidson, 1975; Hu *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).

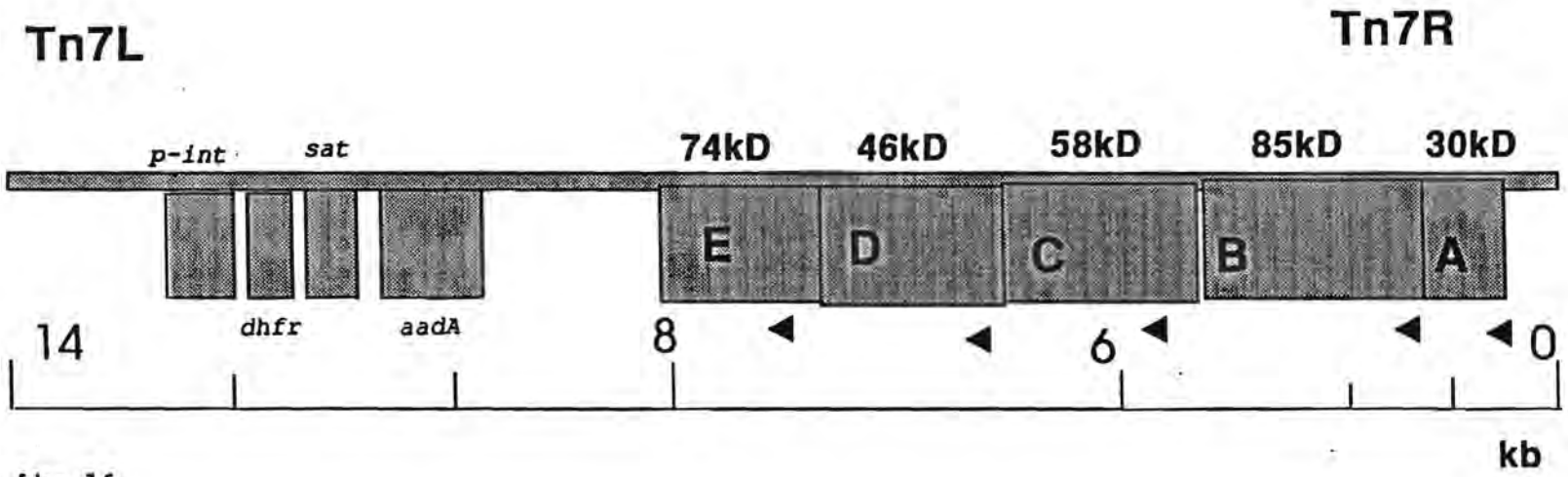
In addition to the central phenomenon of transposition, that is, the appearance of a defined length of DNA (the transposable element) in the midst of sequences where it had not previously been detected, transposable elements typically display a variety of other properties. They can fuse unrelated DNA 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, Tn*9* and Tn*1681* are flanked by copies of IS*1*, 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 Tn*5*, Tn*10* and Tn*903* 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. Tn*7*

Tn*7* 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 trasferase 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). Tn*1825* and Tn*1826* are Tn*7*-related transposons that apparently encode similar transposition functions but differ in their drug



Hg. 1f.

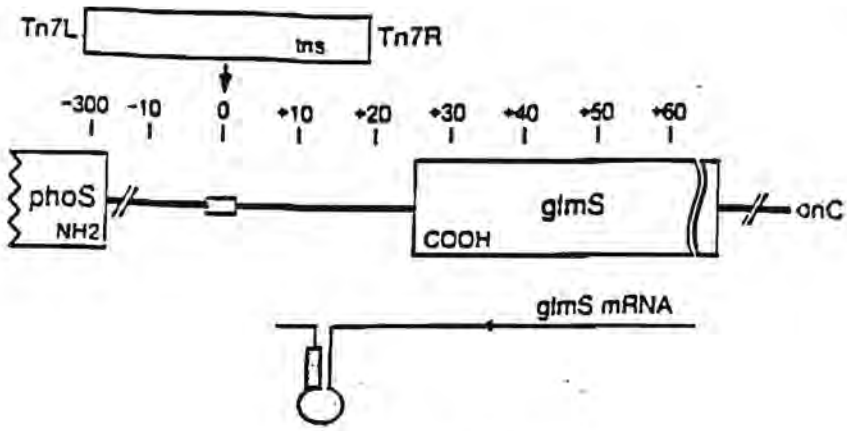
Fig. 1f. Tn7. Shown are the Tn7-encoded transposition genes *tnsABCDE* and the sizes of their protein products. The *tns* genes are all similarly oriented as indicated by tha arrows. Also shown are all the Tn7-encoded drug resistance genes *dhfr* (trimethoprin), *sat* (streptothricin), and *aadA* (spectinomycin/spectomycin). A pseudo-integrase gene (*p-int*) is shown that may mediate rearrangement among drug resistance casettes 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 Tn*7*, 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 Tn*7* does not require this recombinase (Waddell and Craig, 1988). Tn*7* also encodes an elaborate array of transposition genes, *tnsABCDE* (Fig. 1f). These five *tns* genes mediate two distict, but overlapping, recombination pathways (Rogers *et al.*, 1986; Waddell and Craig, 1988; Kubo and Craig, 1990).

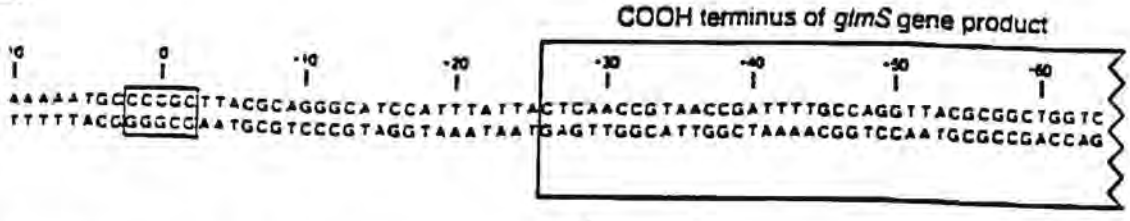
1.3.6.2.1. Insertion of Tn*7* into *E.coli* chromosome.

Unlike other transposons, Tn*7* is very site and orientation specific. When Tn*7* 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 termiminator of the *glmS* gene while the sequence critical for *att*Tn*7* 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).

Tn*7* will also transpose to regions of DNA with sequence related to *att*Tn*7*, pseudo *att*Tn*7* sites. *TnsABC* + *tnsD* promote high-frequency insertion into *att*Tn*7* and low frequency insertion into pseudo-*att*Tn*7* sites; a different ensemble of *tns* genes, *tnsABC* + *tnsE* mediates



B.



Region containing nucleotides required for attTn7 target activity.

Fig. 1g. The *E.coli attTn7* region. (A) Physical map of *attTn7* region. The upper box is Tn*7* (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 Tn*7* 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 *att*Tn*7* (Craig, 1991). Thus, *tnsABC* provide functions common to all Tn*7* 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 Tn*7* is its high frequency of insertion into *att*Tn*7*. For example, examination of the chromosomal DNA in cells containing plasmids bearing Tn*7* reveals that up to 10% of the *att*Tn*7* sites are occupied by Tn*7* in the absence of any selection for Tn*7* insertion (Lichenstein and Brenner, 1981; Hauer and Shapiro, 1984). Tn*7* insertion into *att*Tn*7* has no obvious deleterious effect on *E.coli* growth. The frequency of Tn*7* insertion into sites other than *att*Tn*7* is about 100 to 1000-fold lower than insertion into *att*Tn*7*. Non-*att*Tn*7* may result from either *tnsE*-mediated insertion into random sites or *tnsD*-mediated insertion into pseudo-*att*Tn*7* 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 *Mal*T (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. <u>The Tn7 transposition mechanism</u>.

The developement of a cell-free system for Tn7 transposition to *att*Tn7 (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 *att*Tn7 target in a non-replicative reaction, that is Tn7 is not copied by DNA replication during transposition (Craig 1991). Examination of Tn7 transposition to *att*Tn7 *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 *att*Tn7; thus the transposon excision is provoked by recognition of *att*Tn7 (Craig, 1991). The failure to observe recombination intermediates or products in the absence of *att*Tn7 suggests the

transposon ends flanked by donor DNA and the target DNA containing *att*Tn*7*, are associated prior to the initiation of the recombination by strand cleavage (Craig, 1991). As neither

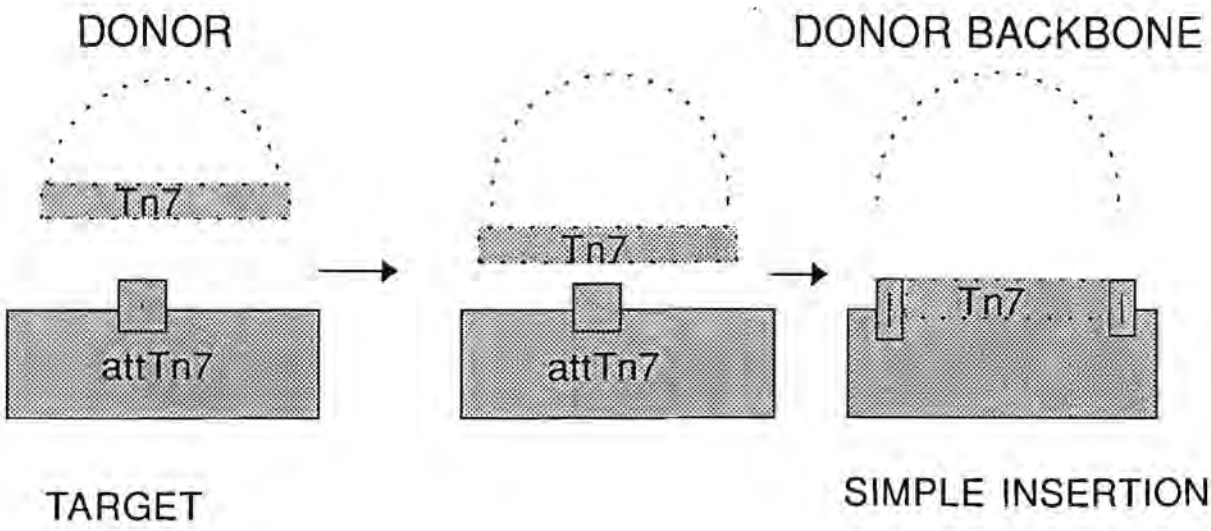


Fig. 1h. The Tn*7* transposition pathway. Shown are a donor molecule containing Tn*7* and a target molecule containing *att*Tn*7*. Translocation of Tn*7* from a donor to a target proceeds via excison of the transposon from the donor backbone via double-strand breaks and its subsequent insertion into a specific position in *att*Tn*7* 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 *att*Tn*7*, 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 *att*Tn*7* provokes the initiation of Tn*7* transposition is also supported by *in vivo* studies (Craig, 1995).

A novel aspect of Tn*7* 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 Tn*7* 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 Tn*10* (Benjamin and Kleckner, 1989), by retroviruses (Fujiwara and Mizuuchi, 1988) and the yeast Ty retrotransposon (Eichinger and Boeke, 1990).

One especially intriguing feature of Tn*7* transposition is that it displays the phenomenon of target immunity, that is the presence of a copy of Tn*7* 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 Tn*3* (Lee *et al.*, 1983) and bacteriophage Mu (Adzumah and Mizuuchi, 1988) also display such target immunity. Tn*7* 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*_Tn*7*-like transposon (Rawlings, unpublished information). The *atp* gene cluster from *T.ferrooxidans* has already been used to complement *E.coli* F_1 *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 (*mer*R1 and *mer*R2) in *T.ferrooxidans* strain E-15 has been found to have high seqeunce homology with *tnsA* of transposon Tn*7* (Kusano *et al.*, 1991). Since Tn*7* 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 chemolitroph like *T.ferrooxidans*. That in itself raises the question of how similar the transposon is to Tn*7* and what marker genes it may carry. There is also the question of whether this Tn*7*-like transposon is found in other strains of *T. ferroxidans* 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 *atp* genes is natural unrearranged *T. ferrooxidans* DNA. 2); to determine whether a Tn*7*-like transposon is present in other strains of *T.ferrooxidans* as well as metabolically related species like *Leptospirillum ferrooxidans* and *Thiobacillus thioxidans*. 3); to clone various pieces of DNA downstream of the Tn*7*-like transposon and carry out single strand sequencing to find out how much of Tn*7*-like transposon is present in *T.ferrooxidans* strain ATCC 33020. 4); to determine whether the antibiotic resistance markers of Trp^r, Str^r/Spr^r and streptothricin which are present on Tn*7* are also present in the Tn*7*-like transposon. 5) whether in *T.ferrooxidans,* the Tn*7*-like region is followed by the *pho* genes as is the case in *E.coli*. 6) to completely sequence the *glmS* gene and test whether it will be able to complement an *E.coli glmS* mutant.

CHAPTER 2

LIST OF FIGURES

CHAPTER TWO.

MAPPING AND SUBCLONING OF A 45 KB *T.FERROOXIDANS*'COSMID (p818.1).

2.1. SUMMARY.

Cosmid p818.1, thought to extend approximately 35 kb downstream of *T.ferrooxidans atp* operon was physically mapped. Southern hybridization was then used to show that p818.1 was unrearranged DNA that originated from the *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 *T.ferrooxidans* strains but not in two *T.thiooxidans* strains or the *L.ferrooxidans* type strain.

2.2. Introduction.

Cosmid p818.1, a 45 kb fragment of *T.ferrooxidans* ATCC 33020 chromosome cloned into pHC79 vector had previously been isolated by its ability to complement *E.coli atp* F_1 mutants (Brown *et al.*, 1994) but had not been studied further. Two other plasmids p*Tfatp1* and p*Tfatp2* from *T.ferrooxidans* chromosome were also able to complement *E.coli unc* F_1 mutants. Of these two, p*Tfatp1* complemented only two *unc* F_1 mutants (AN818ß$^-$ and AN802∈$^-$) whereas p*Tfatp2* complemented all four *E.coli* F_1 mutants tested (Brown *et al.*, 1994). Computer analysis of the primary sequence data of p*Tfatp2,* which has been completely sequenced, revealed the presence of five open reading frames (ORFs) homologous to all the 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 p*Tfatp1* and p*Tfatp2* 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); $(NH_4)_2SO_4$, 3.0; KCl, 0.1; K_2HPO_4, 0.5 and $Ca(NO_3)_2$, 0.01 adjusted to pH 2.5 with H_2SO_4 and autoclaved. Trace elements solution (mg/l); $FeCl_3.6H_2O$, 11.0; $CuSO_4.5H_2O$, 0.5; HBO_3, 2.O; $Na_2MoO_4.2H_2O$, 0.8; $CoCl_2.6H_2O$, 0.6; and $ZnSO_4.7H_2O$ were filter sterilized. Trace elements (1 ml) solution was added to 100 ml mineral salts solution and to this was added either 50 mM $K_2S_4O_6$ or 100 mM $FeSO_4$ 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 (10l) were harvested by centrifugation. Washed three times in water adjusted to pH 1.8 using H_2SO_4 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 *Kpn*I and *Sal*I 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 *Bam*HI, *Hin*dIII and *Bgl*II. *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 *Bgl*II. 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 H_2O.

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 H_2O.

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.

Bacterial Strain	Medium	Geographical location	Source
T.ferrooxidans			
ATCC 22370	$S_4O_6^{2-}$	USA	K.Halberg
ATCC 19859	$S_4O_6^{2-}$	Canada	ATCC
ATCC 33020	Fe^{2+}	Japan	ATCC
T.thiooxidans			
ATCC 19377	$S_4O_6^{2-}$	Libya	K.Halberg
DSM 504	$S_4O_6^{2-}$	USA	K.Halberg
L.ferrooxidans			
DSM 2705	Fe^{2+}	Armernia	P.Norris

Table 2.2. Constructs and subclones of p818.1.

Construct/subclone	Cloning sites	Approx size (kb)
p818.20	*Apa*I-*Eco*RI	11.0
p818.1ΔE	*Eco*RI	34.0
p818.30	*Bam*HI	2.7
p818.41	*Bam*HI-*Kpn*I	0.8
p818.38	*Sal*I-*Hin*dIII	1.0
p818.40	*Hin*dIII-*Bgl*II	3.4
p818.52	*Kpn*I-*Sal*I	1.7
p818.50	*Kpn*I-*Bgl*II	2.6

All these subclones and constructs were cloned into pBluescript KSII.

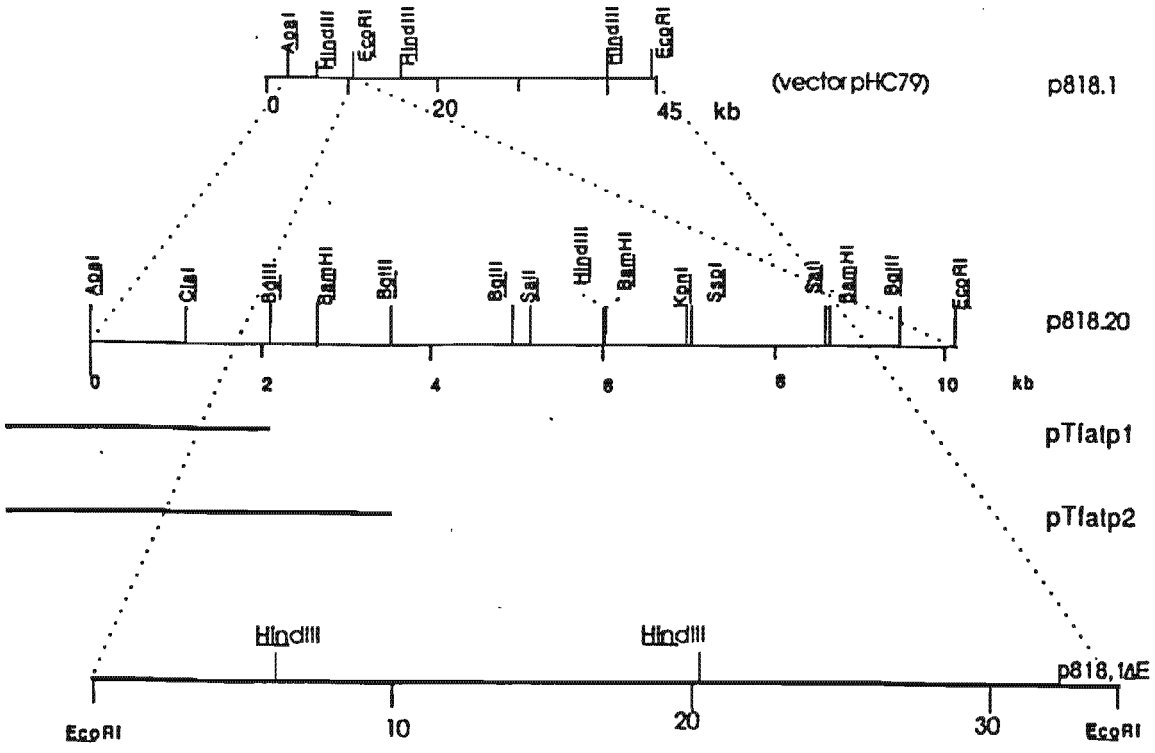


Fig. 2.1. Restriction map of cosmid p818.1. Also shown are the subclones p818.20 (*Apa*I-*Eco*RI) and p818.1ΔE (*Eco*RI-*Eco*RI) as well as plasmids p*Tfatp1* and p*Tfatp2* which complemented *E.coli unc* F_1 mutants (Brown *et al.*, 1994). Apart from p818.1 which was cloned into pHC79, all the other fragments were cloned into vector Bluescript KS.

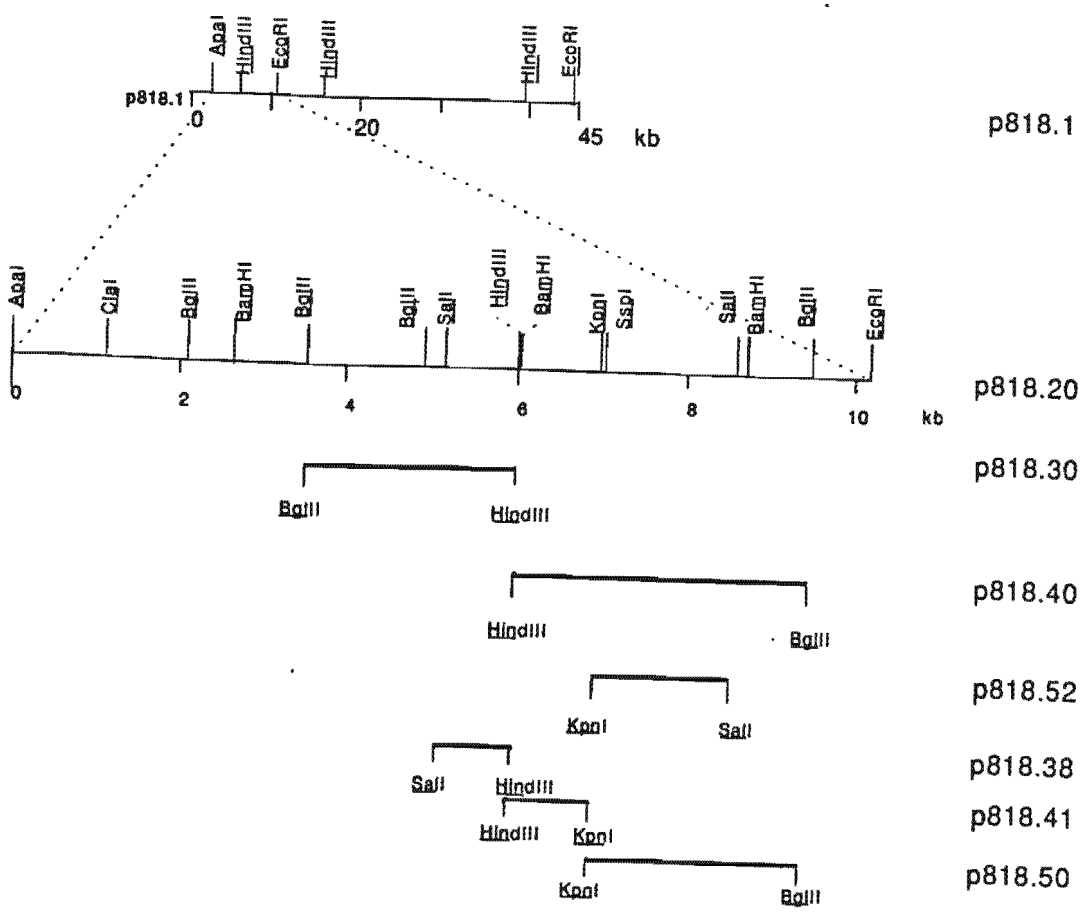


Fig. 2.2. Subclones made from p818.20 including p818.52, the *Kpn*I-*Sal*I fragment used to prepare the probe. All the subclones were cloned into vector Bluescript KS.

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 p*Tfatp1* and p*Tfatp2* 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 unrearranged 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

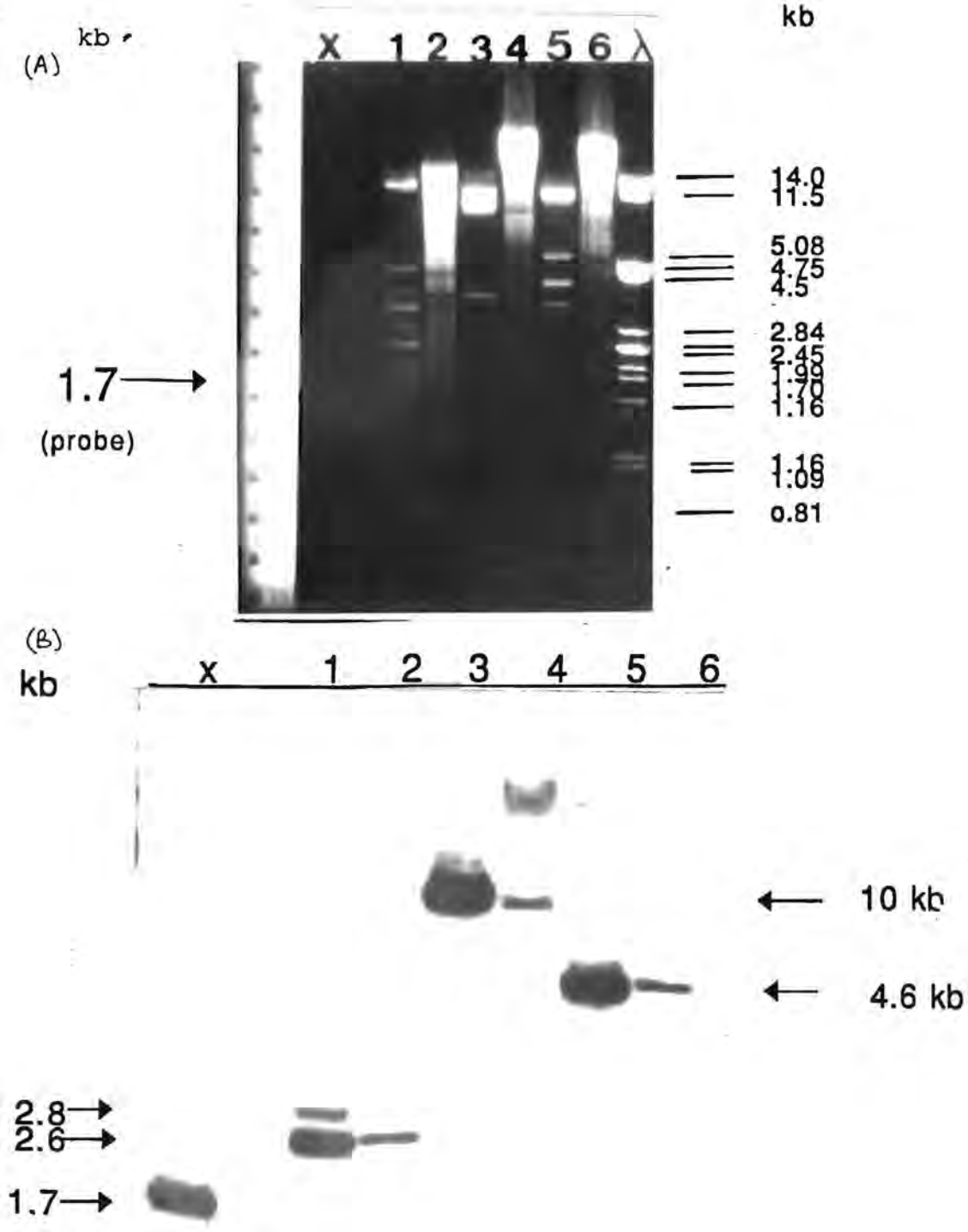


Fig. 2.3. Hybrdization of cosmid p818.1 and *T.ferrooxidans* (ATCC 33020) chromosomal DNA by the *KpnI-SalI* fragment of p818.52 (probe). (a) Autoradiographic image of the restriction digests. Lane x contains the probe , lane 1, 3 and 5 contain p818.1. restricted with *Bam*HI, *Hin*dIII and *Bgl*II respectively. 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 *Pst*I served as the molecular weight marker.

DNA (Fig. 2.3). The *Bam*HI digests gave signals at 2.6 and 2.8 kb (lanes 1 and 2), *Hin*dIII at 10 kb (lanes 3 and 4) and *Bgl*II at 4.6 kb (lanes 5 and 6). The additional weak signals in the p818.1 lanes was because the purified *Kpn*I-*Sal*I 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 *Bgl*II site at 4.5 kb to the *Hin*dIII site at 16 kb on cosmid p818.1 represents native unrearranged chromosomal DNA from *T.ferrooxidans* 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 Tn*7*-like segment) in the chromosome of *T.ferrooxidans* 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 Tn*7*-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 Tn*7*-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 *Bgl*II 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

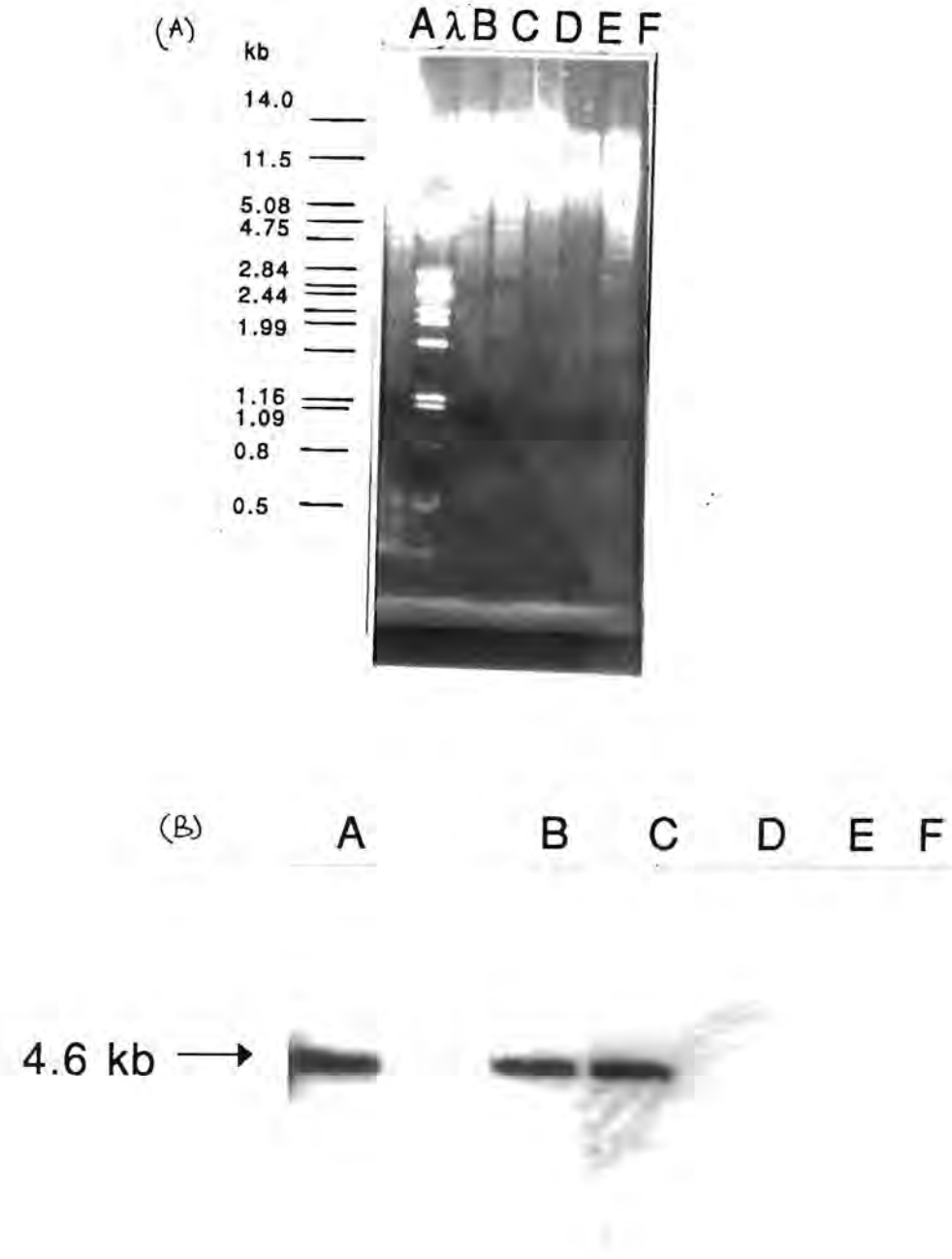


Fig. 2.4. (a) Autoradiographic image of chromosomal DNA of *T.ferrooxidans* strains ATCC 33020, 19859, 23270 (lanes A, B, C), *T.thiooxidans* strains ATCC 19377 and DSM 504 (D and E) and *L.ferrooxidans* strain DSM 2705 (lane F) all restricted with *Bgl*II. λ *Pst*I molecular weight marker was used for sizing. (b) Hybridization of the 1.7 kb *Kpn*I-*Sal*I 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 *Bgl*II 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 Tn*7*-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 Tn*7*-like transposon, it implies either *T.ferrooxidans* and *T.thiooxidans* diverged before *T.ferroxidans* acquired the Tn*7*-like transposon or *T.thiooxidans* does not have an *att*Tn*7* attachment site capable of hosting the Tn*7*-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 Tn*7*-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 Tn*7*-like segment is a general property of all *L.ferrooxidans* and *T.thiooxidans* strains and also whether all *T.ferrooxidans* strains harbour this Tn*7*-like element in their chromosomes.

CHAPTER 3

LIST OF FIGURES

CHAPTER 3.

T.FERROOXIDANS GLUCOSAMINE SYNTHETASE GENE.

3.1. Summary.

A 3.5 kb *Bam*HI-*Bam*HI 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 C1*pro*-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-Pimpaneau *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 amidophosphophoribosyl 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 *pur*F-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*(r^-_k m_k+) *relA1 supE44 lac-proAB*) (*F'traD36 proAB* lacI^{q}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 cells 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 OD_{550}'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 *Bam*HI-*Bam*HI into *Bam*HI 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.

*Apa*I restriction digestion was used to protect both vectors (pUCBM20 and pUCBM21). *Mlu*I 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

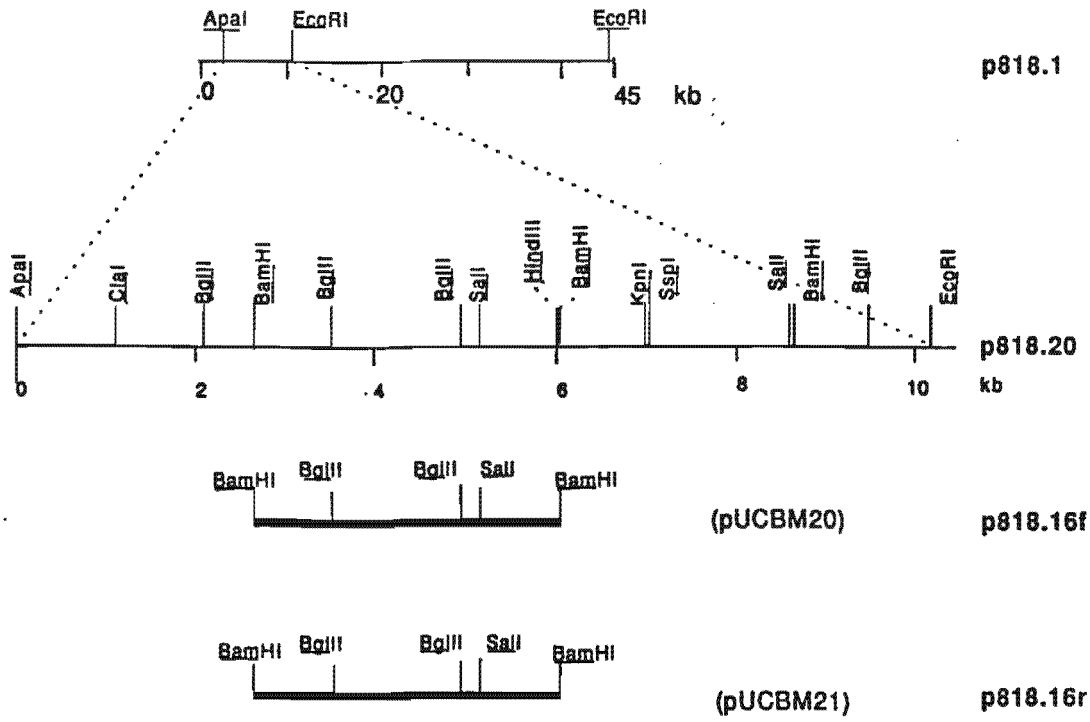


Fig.3.1. Map of cosmid p818.1 and construct p818.20 depicting where p818.20 maps unto p818.1. Below p818.20 are the constructs p818.16f and p818.16r, the 3.5 kb *Bam*HI-*Bam*HI fragment of p818.20 cloned into pUCBM20 and pUCBM21 respectively.

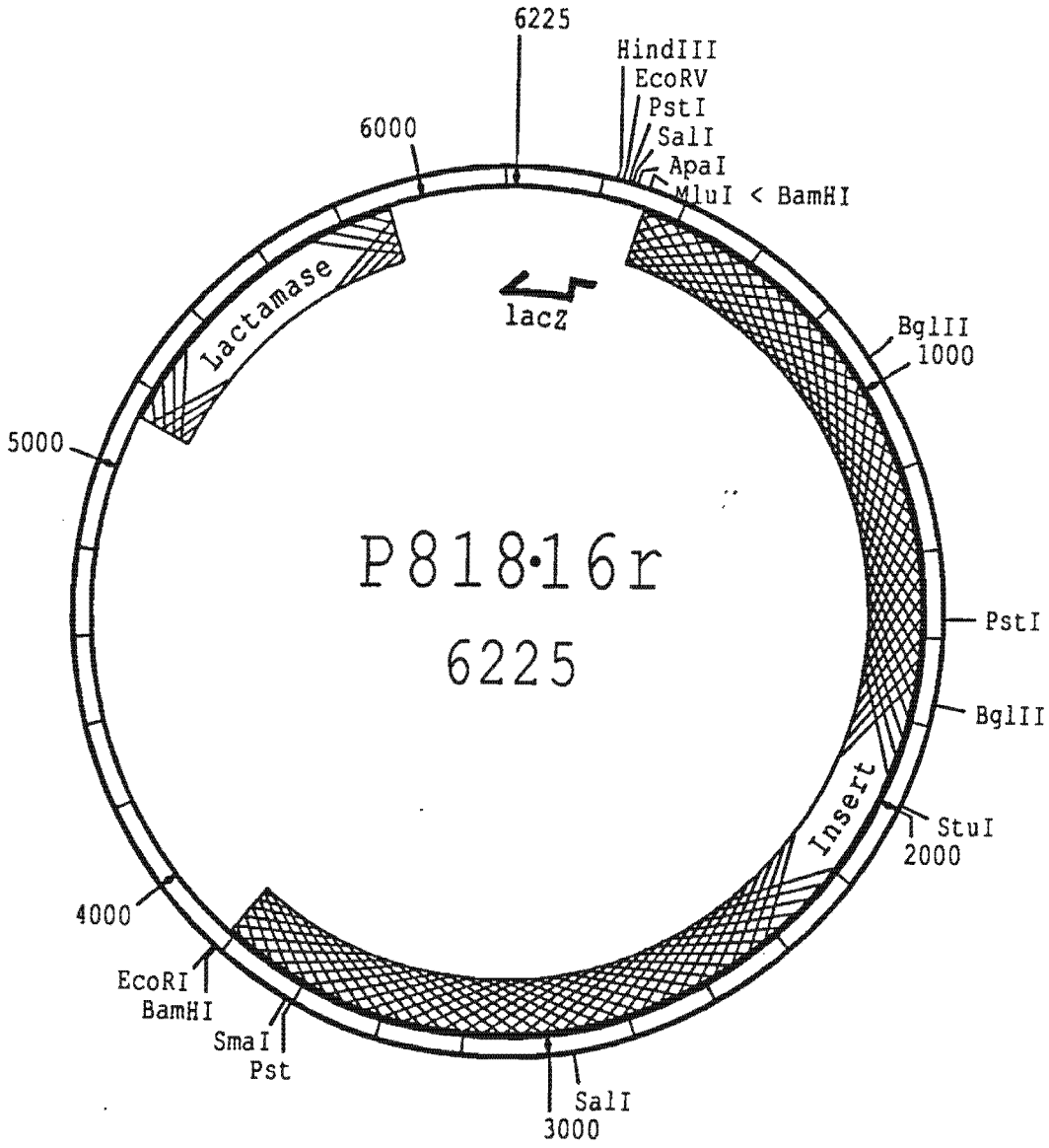


Fig. 3.2. Plasmidmap of p818.16r showing the positions of some of the commonly used restriction enzyme sites. *Apa*I restriction site was used to protect the vector while *Mlu*I acted as the susceptible site. Also shown is the insert of about 3.5 kb covering part of the *glmU* gene, the complete *glmS* gene and ORF3. The cloning vector is pUCBM21.

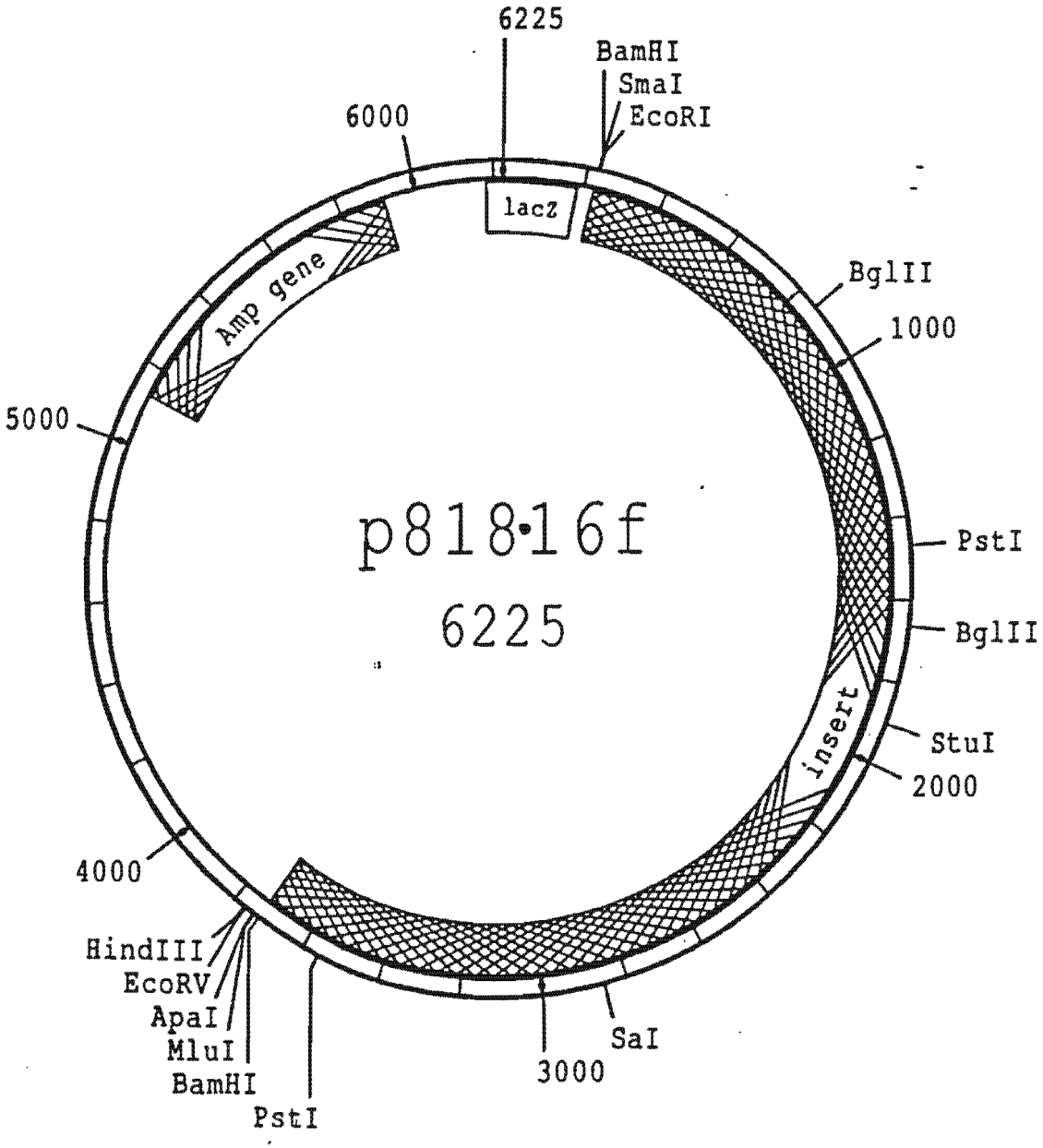


Fig. 3.3. Plasmid map of p818.16f showing the cloning orientation of the commonly used restriction enzyme sites of the pUCBM20 vector. *Apa*I restriction site was used to protect the vector while *Mlu*I acted as the susceptible site. Also shown is the insert of about 3.5 kb covering part of the *T.ferrooxidans glmU* gene, the complete *glmS* gene and ORF3.

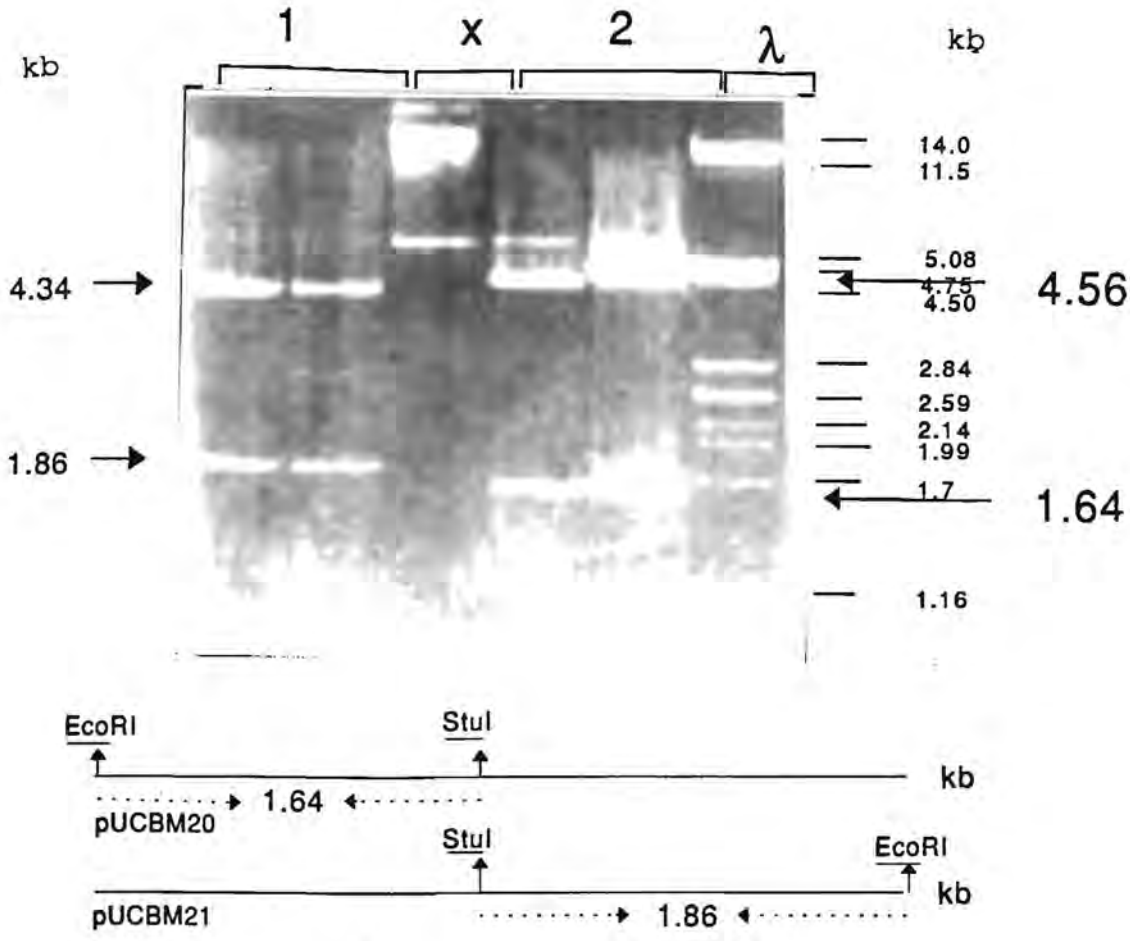


Fig 3.4. p818.16r and p818.16f restricted with *Eco*RI and *Stu*I restriction enzymes. Since the *Eco*RI site is at opposite ends of these vectors, the *Stu*I-*Eco*RI digest gave different fragment sizes as illustrated in the diagram beneath. Both constructs are in the same orientaion in the vectors. λ *Pst*I molecular weight marker (extreme right lane) was used to determine the size of the bands. In lane x is an *Eco*RI digest of p818.16.

8.0) and the BLAST subroutines (NCIB, New Bethesda, Maryland. USA).

3.3.10. Complementaton 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 pT*fatp1* and pT*fatp2* 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 *Bam*HI-*Bam*HI 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 Tn*7* respectively. Immediately downstream of the complete ORF is a region which resembles the inverted repeat sequences of transposon Tn*7*. The Tn*7*-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 uridyltransferases 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)X_4 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 *Bam*HI-*Bam*HI fragment of p818.16. The sequence includes part of the *T.ferrooxidans glmU* gene (ORF1), the entire *T.ferrooxidans glmS* gene (ORF2) and ORF3 which had high sequence homology to TnsA and inverted repeats of transposon Tn*7* (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 *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 *glmU* (ORF1) *glmS* (ORF2), and *tnsA* genes. Also shown are the restriction endonuclease sites for some of the enzymes which were used to map p818.16.

(Legend is on the previous page).

```
    B
    a
    m
    H
    I partial ORF1
  1 GGATCCCAACGTGTTGTTTGTCGGCGAGGTGCATCTGGGGCATCGCGTCCGCGTGGGAGC 60
     AspProAsnValLeuPheValGlyGluValHisLeuGlyHisArgValArgValGlyAla  -

 61 TGGCGCCGTTTTGCAGGATGCGCGGATTGGTGACGATGTGGAGATTCTACCCTACAGCCA 120
     GlyAlaValLeuGlnAspAlaArgIleGlyAspAspValGluIleLeuProTyrSerHis  -

121 TATCGAAGGCGCCCAGATTGGGGCCGGGGCGCGGATAGGACCTTTCGCGCGGATTCGCCC 180
     IleGluGlyAlaGlnIleGlyAlaGlyAlaArgIleGlyProPheAlaArgIleArgPro  -

181 CGGAACGGAGATCGGCGAACGTCATATCGGCAACTATGTCGAGGTGAAGGCGGCCAAAAT 240
     GlyThrGluIleGlyGluArgHisIleGlyAsnTyrValGluValLysAlaAlaLysIle  -

241 CGGCGCGGGCAGCAAGGCCAACCACCTGAGTTACCTTGGGGACGCGGAGATCGGTACCGG 300
     GlyAlaGlySerLysAlaAsnHisLeuSerTyrLeuGlyAspAlaGluIleGlyThrGly  -

301 GGTGAATGTGGGCGCCGGGACGATTACCTGCAATTATGACGGTGCGAACAAACATCGGAC 360
     ValAsnValGlyAlaGlyThrIleThrCysAsnTyrAspGlyAlaAsnLysHisArgThr  -

361 CATCATCGGCAATGACGTGTTCATCGGCTCCGACAGCCAGTTGGTGGCGCCAGTGAACAT 420
     IleIleGlyAsnAspValPheIleGlySerAspSerGlnLeuValAlaProValAsnIle  -

421 CGGCGACGGAGCGACCATCGGCGCGGGCAGCACCATTACCAAAGAGGTACCTCCAGGAGG 480
     GlyAspGlyAlaThrIleGlyAlaGlySerThrIleThrLysGluValProProGlyGly  -

481 GCTGACGCTGAGTCGCAGCCCGCAGCGTACCATTCCTCATTGGCAGCGGCCGCGGCGTGA 540
     LeuThrLeuSerArgSerProGlnArgThrIleProHisTrpGlnArgProArgArgAsp  -

541 CAAAAAGTAATTCCGGATGACGCGGAATGGTCTTCGATAACAGGAGAACAGAATCATGTG 600
     LysLysEnd..                                ORF2          MetCys  -

                                    B
                                    g
                                    l
                                    I
                                    I
601 CGGTATTGTCGGTGGGGTGAGTAAAACAGATCTGGTCCCGATGATTCTGGAGGGGTTGCA 660
     GlyIleValGlyGlyValSerLysThrAspLeuValProMetIleLeuGluGlyLeuGln  -

661 GCGCCTGGAGTATCGTGGCTACGACTCTGCCGGGCTGGCGATATTGGGGGCCGATGCGGA 720
    ArgLeuGluTyrArgGlyTyrAspSerAlaGlyLeuAlaIleLeuGlyAlaAspAlaAsp  -

721 TTTGCTGCGGGTGCGCAGCGTCGGGCGGGTCGCCGAGCTGACCGCCGCCGTTGTCGAGCG 780
     LeuLeuArgValArgSerValGlyArgValAlaGluLeuThrAlaAlaValValGluArg  -

781 TGGCTTGCAGGGTCAGGTGGGCATCGGCCACACGCGCTGGGCCACCCATGGCGGCGTCCG 840
     GlyLeuGlnGlyGlnValGlyIleGlyHisThrArgTrpAlaThrHisGlyGlyValArg  -

841 CGAATGCAATGCGCATCCCATGATCTCCCATGAACAGATCGCTGTGGTCCATAACGGCAT 900
     GluCysAsnAlaHisProMetIleSerHisGluGlnIleAlaValValHisAsnGlyIle  -

901 CATCGAGAACTTTCATGCCTTGCGCGCCCACCTGGAAGCAGCGGGGTACACCTTCACCTC 960
     IleGluAsnPheHisAlaLeuArgAlaHisLeuGluAlaAlaGlyTyrThrPheThrSer  -

961 CGAGACCGATACGGAGGTCATCGCGCATCTGGTGCACCATTATCGGCAGACCGCGCCGGA 1020
     GluThrAspThrGluValIleAlaHisLeuValHisHisTyrArgGlnThrAlaProAsp  -

1021 CCTGTTCGCGGCGACCCGCCGGGCAGTGGGCGATCTGCGCGGCGCCTATGCCATTGCGGT 1080
     LeuPheAlaAlaThrArgArgAlaValGlyAspLeuArgGlyAlaTyrAlaIleAlaVal  -

1081 GATCTCCAGCGGCGATCCGGAGACCGTGTGCGTGGCACGGATGGGCTGCCCGCTGCTGCT 1140
     IleSerSerGlyAspProGluThrValCysValAlaArgMetGlyCysProLeuLeuLeu  -
```

```
1141 GGGCGTTGCCGATGATGGGCATTACTTCGCCTCGGACGTGGCGGCCCTGCTGCCGGTGAC 1200
      GlyValAlaAspAspGlyHisTyrPheAlaSerAspValAlaAlaLeuLeuProValThr  -

1201 CCGCCGCGTGTTGTATCTCGAAGACGGCGATGTGGCCATGCTGCAGCGGCAGACCCTGCG 1260
      ArgArgValLeuTyrLeuGluAspGlyAspValAlaMetLeuGlnArgGlnThrLeuArg  -

1261 GATTACGGATCAGGCCGGAGCGTCGCGGCAGCGGGAAGAACACTGGAGCCAGCTCAGTGC 1320
      IleThrAspGlnAlaGlyAlaSerArgGlnArgGluGluHisTrpSerGlnLeuSerAla  -

1321 GGCGGCTGTCGATCTGGGGCCTTACCGCCACTTCATGCAGAAGGAAATCCACGAACAGCC 1380
      AlaAlaValAspLeuGlyProTyrArgHisPheMetGlnLysGluIleHisGluGlnPro  -

                                                                BglII
1381 CCGCGCGGTTGCCGATACCCTGGAAGGTGCGCTGAACAGTCAACTGGATCTGACAGATCT 1440
      ArgAlaValAlaAspThrLeuGluGlyAlaLeuAsnSerGlnLeuAspLeuThrAspLeu  -

1441 CTGGGGAGACGGTGCCGCGGCTATGTTTCGCGATGTGGACCGGGTGCTGTTCCTGGCCTC 1500
      TrpGlyAspGlyAlaAlaAlaMetPheArgAspValAspArgValLeuPheLeuAlaSer  -

1501 CGGCACTAGCCACTACGCGACATTGGTGGGACGCCAATGGGTGGAAAGCATTGTGGGGAT 1560
      GlyThrSerHisTyrAlaThrLeuValGlyArgGlnTrpValGluSerIleValGlyIle  -

1561 TCCGGCGCAGGCCGAGCTGGGGCACGAATATCGCTACCGGGACTCCATCCCCGACCCGCG 1620
      ProAlaGlnAlaGluLeuGlyHisGluTyrArgTyrArgAspSerIleProAspProArg  -

                                                            StuI
1621 CCAGTTGGTGGTGACCCTTTCGCAATCCGGCGAAACGCTGGATACTTTCGAGGCCTTGCG 1680
      GlnLeuValValThrLeuSerGlnSerGlyGluThrLeuAspThrPheGluAlaLeuArg  -

1681 CCGCGCCAAGGATCTCGGTCATACCCGCACGCTGGCCATCTGCAATGTTGCGGAGAGCGC 1740
      ArgAlaLysAspLeuGlyHisThrArgThrLeuAlaIleCysAsnValAlaGluSerAla  -

1741 CATTCCGCGGGCGTCGGCGTTGCGCTTCCTGACCCGGGCCGGCCCAGAGATCGGAGTGGC 1800
      IleProArgAlaSerAlaLeuArgPheLeuThrArgAlaGlyProGluIleGlyValAla  -

1801 CTCGACCAAGGCGTTCACCACCCAGTTGGCGGCGCTCTATCTGCTGGCTCTGTCCCTGGC 1860
       SerThrLysAlaPheThrThrGlnLeuAlaAlaLeuTyrLeuLeuAlaLeuSerLeuAla  -

1861 CAAGGCGCCAGGGGCATCTGAACGATGTGCAGCAGGCGGATCACCTGGACGCTTGCGGCA 1920
      LysAlaProGlyAlaSerGluArgCysAlaAlaGlyGlySerProGlyArgLeuArgGln  -

1921 ACTGCCGGGCAGTGTCCAGCATGCCCTGAACCTGGAGCCGCAGATTCAGGGTTGGGCGGC 1980
      LeuProGlySerValGlnHisAlaLeuAsnLeuGluProGlnIleGlnGlyTrpAlaAla  -

1981 ACGTTTTGCCAGCAAGGACCATGCGCTTTTTCTGGGGCGCGGCCTGCACTACCCCATTGC 2040
      ArgPheAlaSerLysAspHisAlaLeuPheLeuGlyArgGlyLeuHisTyrProIleAla  -

2041 GCTGGAGGGCGCGCTGAAGCTCAAGGAAATCTCCTATATCCACGCCGAGGCTTATCCCGC 2100
      LeuGluGlyAlaLeuLysLeuLysGluIleSerTyrIleHisAlaGluAlaTyrProAla  -

2101 GGGCGAATTGAAGCATGGCCCCTTGGCCCTGGTGGACCGCGACATGCCCGTGGTGGTGAT 2160
      GlyGluLeuLysHisGlyProLeuAlaLeuValAspArgAspMetProValValValIle  -

2161 CGCGCCCAATGACCGCCTCCTCGAAAAGCTGGCCGCCAACATGCAGGAAGTCCACGCCCG 2220
      AlaProAsnAspArgLeuLeuGluLysLeuAlaAlaAsnMetGlnGluValHisAlaArg  -

2221 TGGCGGTGAGCTCTATGTTTTTGCCGATTCGGACAGCCACTTTAACGCCAGTGCGGGCGT 2280
      GlyGlyGluLeuTyrValPheAlaAspSerAspSerHisPheAsnAlaSerAlaGlyVal  -
```

```
2281 GCATGTGATGCGTTTGCCCCGTCACGCCGGTCTGCTCTCCCCCATCGTCCATGCTATCCC 2340
     HisValMetArgLeuProArgHisAlaGlyLeuLeuSerProIleValHisAlaIlePro  -

2341 GGTGCAGTTGCTGGCCTATCATGCGGCGCTGGTGAAGGGCACCGATGTGGATCGACCGCG 2400
      ValGlnLeuLeuAlaTyrHisAlaAlaLeuValLysGlyThrAspValAspArgProArg  -

2401 TAACCTCGCGAAGAGCGTGACGGTGGAGTAATGGGGCGGTTCGGTGTGCCGGGTGTTGTT 2460
      AsnLeuAlaLysSerValThrValGluEnd
                       End of glmS gene.

2461 AACGGACAATAGAGTATCATTCTGGACAATAGAGTTTCATCCCGAACAATAAAGTATCAT 2520

2521 CCTCAACAATAGAGTATCATCCTGGCCTTGCGCTCCGGAGGATGTGGAGTTAGCTTGCCT 2580
                                                        S
                                                        a
                                                        l
                                                        I

2581 AATGATCAGCTAGGAGAGGATGCCTTGGCACGCCAACGCTACGGTGTCGACGAAGACCGG 2640
                                LeuAlaArgGlnArgTyrGlyValAspGluAspArg   -
                                ORF3

2641 GTCGCACGCTTCCAAAAGGAGGGACGTGGTCAGGGGCGTGGCGCAGACTACCACCCCTGG 2700
     ValAlaArgPheGlnLysGluGlyArgGlyGlnGlyArgGlyAlaAspTyrHisProTrp  -

2701 CTTACCATCCAAGACGTGCCCTCCCAAGGGCGGTCCCACCGACTCAAGGGCATCAAGACC 2760
     LeuThrIleGlnAspValProSerGlnGlyArgSerHisArgLeuLysGlyIleLysThr  -

2761 GGGCGAGTGCATCACCTGCTCTCGGATATTGAGCGCGACATCTTCTACCTGTTCGATTGG 2820
     GlyArgValHisHisLeuLeuSerAspIleGluArgAspIlePheTyrLeuPheAspTrp  -

2821 GCAGACGCCGTCACGGACATCCGCGAACAGTTTCCGCTGAATCGCGACATCACTCGCCGT 2880
     AlaAspAlaValThrAspIleArgGluGlnPheProLeuAsnArgAspIleThrArgArg  -

2881 ATTGCGGATGATCTTGGGGTCATCCATCCCCGCGATGTTGGCAGCGGCACCCCTCTAGTC 2940
     IleAlaAspAspLeuGlyValIleHisProArgAspValGlySerGlyThrProLeuVal  -

2941 ATGACCACGGACTTTCTTGTGGACACGATCCATGATGGACGTATGGTGCAACTGGCGCGG 3000
     MetThrThrAspPheLeuValAspThrIleHisAspGlyArgMetValGlnLeuAlaArg  -

3001 GCGGTAAAACCGGCTGAAGAACTTGAAAAGCCGCGGGTGGTCGAAAAACTGGAGATTGAA 3060
     AlaValLysProAlaGluGluLeuGluLysProArgValValGluLysLeuGluIleGlu  -

3061 CGCCGTTATTGGGCGCAGCAAGGCGTGGATTGGGGCGTCGTCACCGAGCGGGACATCCCG 3120
     ArgArgTyrTrpAlaGlnGlnGlyValAspTrpGlyValValThrGluArgAspIlePro  -

3121 AAAGCGATGGTTCGCAATATCGCCTGGGTTCACAGTTATGCCGTAATTGACCAGATGAGC 3180
     LysAlaMetValArgAsnIleAlaTrpValHisSerTyrAlaValIleAspGlnMetSer  -

3181 CAGCCCTACGACGGCTACTACGATGAGAAAGCAAGGCTGGTGTTACGAGAACTTCCGTCG 3240
     GlnProTyrAspGlyTyrTyrAspGluLysAlaArgLeuValLeuArgGluLeuProSer  -

3241 CACCCGGGGCCTACCCTCCGGCAATTCTGCGCCGACATGGACCTGCAGTTTTCTATGTCT 3300
     HisProGlyProThrLeuArgGlnPheCysAlaAspMetAspLeuGlnPheSerMetSer  -

3301 GCTGGTGACTGTCTCCTCCTAATTCGCCACTTGCTGGCCACGAAGGCTAACGTCTGTCCT 3360
     AlaGlyAspCysLeuLeuLeuIleArgHisLeuLeuAlaThrLysAlaAsnValCysPro  -

                                H
                                i
                                n
                                d
                                I
                                I
                                I

3361 ATGGACGGGCCTACGGACGACTCCAAGCTTTTGCGTCAGTTTCGAGTGGCCGAAGGTGAA 3420
     MetAspGlyProThrAspAspSerLysLeuLeuArgGlnPheArgValAlaGluGlyGlu  -
```

```
3421 TCCAGGAGGGCCAGCGGATGAGGGATCTGTCGGTCAACCAATTGCTGGAATACCCCGATG 3480
     SerArgArgAlaSerGlyEnd                                         -

                      B
                      a
                      m
                      H
                      I
3481 AAGGGAAGATCGAAAGGATCC 3501
```

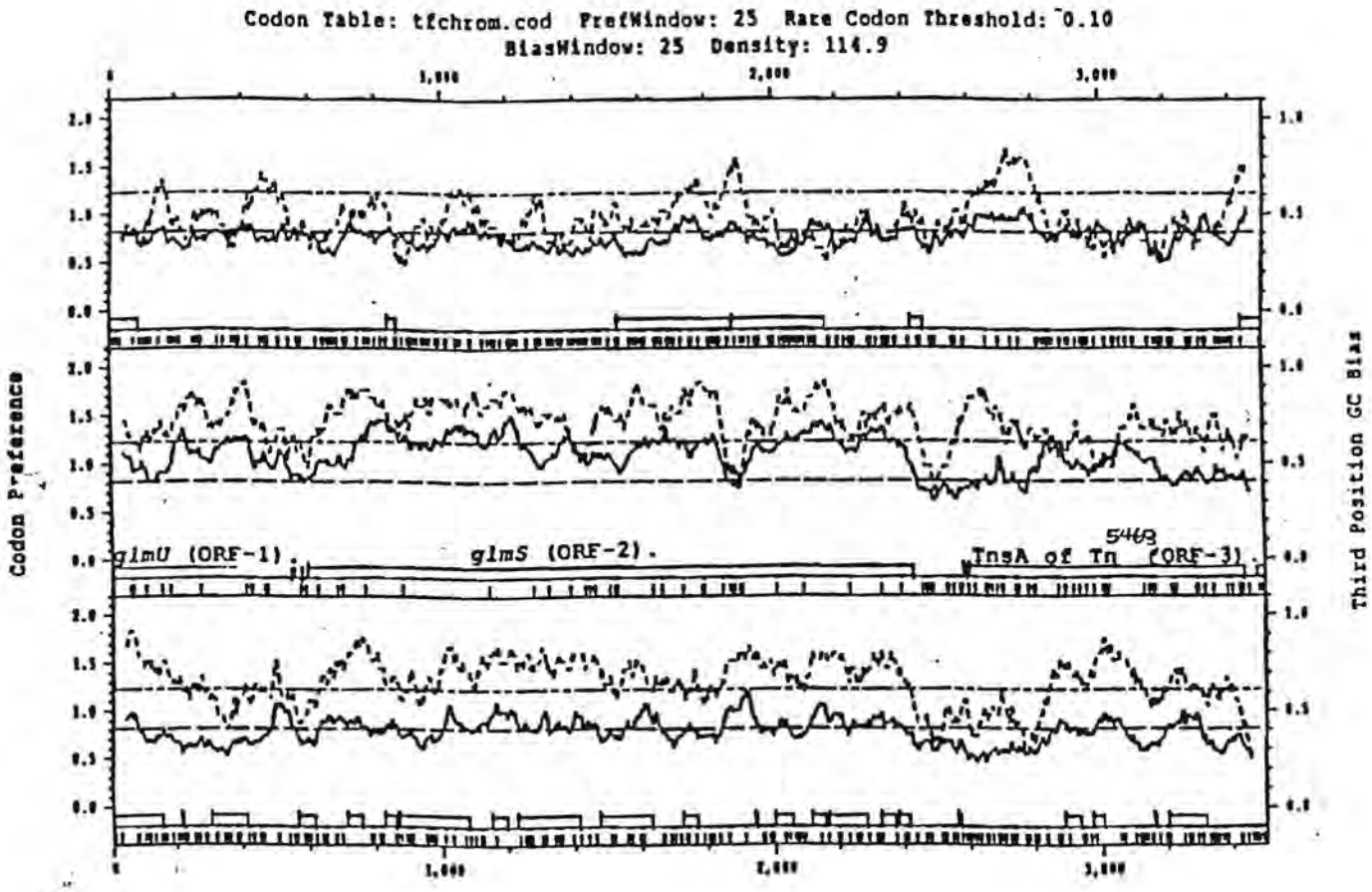


Fig. 3.6. Codon preference and bias plot of nucleotide sequence of the 3.5 kb p818.16 *Bam*HI-*Bam*HI fragment. The codon preference plot was generated using the an established codon usage table for *T.ferrooxidans*. The partial open reading frame (ORF1) and the two complete open reading frames (ORF2 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-acetylclucosamine 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.

```
     251                                                  300
T_f                                DP NVLFVGEVHL GHRVRVGAGA
E_c                                DT NVIIEGNVTL GHRVKIGTGC
B_s                                YP GTVIKGEVQI GEDTIIGPHT


     301                                                  350
T_f  VLQDARIGDD VEILPYSHIE GAQIGAGARI GPFARIRPGT E.IGERHIGN
E_c  VIKNSVIGDD CEISPYTVVE DANLAAACTI GPFARLRPGA ELLEGAHVGN
B_s  EIMNSAIGSR TVIKQ.SVVN HSKVGNDVNI GPFAHIRPDS VIGNEVKIGN


     351                                                  400
T_f  YVEVKAAKIG AGSKANHLSY LGDAEIGTGV NVGAGTITCN YDGANKHRTI
E_c  FVEMKKARLG KGSKAGHLTY LGDAEIGDNV NIGAGTITCN YDGANKFKTI
B_s  FVEIKKTQFG DRSKASHLSY VGDAEVGTDV NLGCGSITVN YDGKNKYLTK


     401                                                  450
T_f  IGNDVFIGSD SQLVAPVNIG DGATIGAGST ITKEVPPGGL TLSRSPQRTI
E_c  IGDDVFVGSD TQLVAPVTVG KGATIAAGTT VTRNVGENAL AISRVPQTQK
B_s  IEDGAFIGCN SNLVAPVTVG EGAYVAAGST VTEDVPGKAL AIARARQVNK


     451
T_f  PHWQRPRRDK
E_c  EGWRRPVKKK
B_s  DDYVKNIHKK
```

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* σ^{70}-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 *pur*F-type (named after the *pur*F-encoded glutamine phosphoribosyl pyrophosphate amidotransferase) Zalkin and Weng, (1987).

The glutamine amide transfer domain of approximately 194 amino acid residues is at the NH_2 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 Cys^{1}-His^{101}-Asp^{29} 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 amidotranferases (GFAT) of *Rhizobium meliloti* (R_m), *Rhizobium leguminnosarum* (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).

```
          1                                                  50
R_m       .*****i*.. ....*hkp*s *r*i*a*g** *******s** *t*f....**
R-l       .*****i*.. ....*hqp*s *r*v*a*ep* ********** *tmd....a*
E_c       ******a*.. ....a****a ***l***r** *********l ***d...a**
Hae_in    ******a*.. ....a****a ***in**h** ********** ****...kqn
T_f       ******g*.. ....skt*lv pmil***q** *********l **lg...a*a
          MCGIVG V          D       GL RL EYRGYDSAG  AI      D
M_l       .**l****.. ....***pac g*vm*a*r** *******s** *l**gsaks*
B_s       ********.. ....**l*ak ***lk**ek* ********** **a*....*q
S_c       .***f**cny lversrg**i *t*v***q** *******t** **dgd..ead

          51                                                 100
R_m       **h*r*ae** lg**ktrlk* ...*...**s *t******** ****a*t*c*
R_l       t*q*r*ae** lg**r*klk* ...*...**s *t******** ****a*t*r*
E_c       h*t*l*rl** *qm*aq*ae* ...h...**h *gt******* ******s*v*
Hae_in    **q*i*cl** *ka*d***s* ...k...**i *gt******* ******s*t*
T_f       **l*v*sv*r *ae*ta**v* ...r...g*q *q***g**** ****gvr*c*
          DL R R  G  V  L  AV E         L  G VGI HTRW ATHG   E N
M_l       n*tvr*ra*r ls**esvla* mv**...s*a *n**lg*i** ****r*t*r*
B_s       gihvfkek*r *ad*r*v*d. ...*...nve aka**g**** ******syl*
S_c       stfiykqi** *sa*k*e*tk qn*nrdvt*v shc******* ****r*eqv*

          101                                                150
R_m       ****.f.t.* g********* ***sk*k*a* aev*tk*qt* *********.
R_l       ****.f.t.* g********* ***a**k*** a*g*a**qt* *********.
E_c       ****.v.*.* h*v******* **h*p*r*** k*r**t*v** *********.
Hae_in    ****.s.*.g tf******** **h***r*l* ksr**v*l*q *********.
T_f       ***m.i.*h* q********* ***ha*rah* e*a**t*t** *********.
          AHP    S E  IAVVHNGII ENF  L   L  A GY F SE TDTEVIAHL.
M_l       ****.r*atg k********* ***ps*rh** eia*v**v** *****av**.
B_s       ****.qsalg rftl****** ***vq*kq*y .lqdv**k** ******vqv.
S_c       c**qrs*pe* qfv******* t**r**ktl* ink**k*e** ****c**k*y

          151                                                200
R_m       ..ltk**... *d**gc***m h*m**cve** ******fe** *a***v**n*
R_l       ..lak**... *d**grr**m h*m**rvk** ******fe** *s******t*
E_c       ..*nwel... kq*gt*r*** lr*ipq*r'* *gtv****rh ***l****s*
Hae_in    ..*ewem... **tds***** kk*v*q*t** *g*v****rh **hlv***s*
T_f       ..*hh**... q*ap***a*t rr*vgd*r** **i**is*g* ****cv**m*
            V  YR     T  DLF A    A   L GA YA AV  S D PETV  AR G
M_l       ..*aqaycag e*ag**vgs* l*v*rr*q*h *t*vfana** *g**v***rs
B_s       ..*eq*v... ng***te**f rkt*tl*k*s **i*l**n*n r****v*knk
S_c       lhlyntnlqn ghd***h*lt klv*le*e*s *g*lckschy *ne*i*t*k*
```

```
        201                                                        250
R_m     p*****h... .......... .......... .......... ..........
R_l     .**a**h... .......... .......... .......... ..........
E_c     ******l... .......... .......... .......... ..........
Hae_in  ******l... .......... .......... .......... ..........
T_f     c**ll*v... .......... .......... .......... ..........
M_l     t***l*i... .......... .......... .......... ..........
B_s     ***l**l... .......... .......... .......... ..........
S_c     ***l**vkse kklkvdfvdv efpeenagqp eiplksnnks fglgpkkare

        251                                                        300
R_m     .......... .......... .......... .....****m *lg***i**a
R_l     .......... .......... .......... .....*n**m *lg***i**a
E_c     .......... .......... .......... .....*m**n *i***ql***
Hae_in  .......... .......... .......... .....*i**n *l***ql***
T_f     .......... .......... .......... .....a*dgh *f***va***
                                                    D    Y ASD  ALL
M_l     .......... .......... .......... .....****m *vg**va**i
B_s     .......... .......... .......... .....**tfn vv****m***
S_c     feagsqnanl lpiaanefnl rhsqsrafls edgsptpv*f *vs***asvv

        301                                                        350
R_m     *f*nd*t**n ***w***gkt s*q**.*i** *..****pr* **lat*dl*g
R_l     *f*ne*t**i ***w***gkt g*h**.**** *..**e*prq **taa*fl**
E_c     ****rf**** *****e***r s*n**.*kt* a..e*k*qdi e*n*qyd*g*
Hae_in  s***rf**** *****e***r t*d**.*th* *..kak*ei* e*n**nd*a*
T_f     ****r*l*** *****mlq*q tlr*t.*qa* a..srq*ee* w*q*s*a***
        PVTR V YLE DGDVA   R     I  D  G       R  H  S L A AVD
M_l     eh**qa*e*g qdqav***ad gyr*s.**** *dda*nartf hi*w*la*a*
B_s     q**ney*e*m *k*mv***dd q*v*k.n*** d..****asy *a****sd**
S_c     kh*kk*l*** *d*l*h*ydg elh*hrsrre vgasm**siq tl***laq*m

        351                                                        400
R_m     **nhp***e* *******v** ra*ghy**.. vn*shvttt* t.did*ag**
R_l     **nh****e* *******v** ia*ghy**.. v**kscrsd* d.*id*ag**
E_c     **i******* ******n**k ***t*r*s*g q**lsel*pn *.del*sk**
Hae_in  **k******* ******t*li ***e*r**** n*iv*si*ng *.kgi*ek**
T_f     l*p******* **h***r*** d**e*al*sq ldlt*lw*dg *.*am*r***
        G YRHFMQK EI EQP AVA  TL G  N       D  G   A A  F DVD
M_l     **g*ey**l* **a******v d**l*hftgg r*vl*eqrl* d.qe.*r***
B-s     **t*p***l* *td***vvmr kiiqty..q* engklsvpgd i.*aava*a*
S-c     **p*d***** *******stf ***r*r*dy* nnkvilg*lk *wlpvvrrar

        401                                                        450
R_m     sla*s****a *l*****k** **r**rli** ********** *ip***r*aa
R_l     sla*sc***a *l*****k** **r**rl*** *a******** *ip***q*.a
E_c     h*q*l***** *ns***sr** ***l*g**c* ********** ksavrrn***
Hae_in  h*q******* *n****ar** ***l*g*sc* ********** kfvtr*n***
T_f     ***fl*s*** hy*t***rq* v**ivg**aq a*lgh***** *sipd*rq*v
        RVL  A GTS   A LVG  W  ES   IP    E   EYRYR D   P  L
M_l     k********a **s**lak*t i*h*trl*** **l******* *pv*drst*v
B_s     *****g**** *******kq* i*m**n**** *h*****s*n mpl**kkp**
S_c     *lim****** **sc*atrai **elsd***s **l****ld* kcpvfrddvc
```

```
          451                                                  500

R_m       l*i******* ****s***c* ah*.***g** v*tre*t*a* *a*a**pil*
R_l       l*i******* ****s***c* ah*.***g** v*are*t*a* ***a**pil*
E_c       *tl******* ****g**ls* ***y*gs*** ***p**s*v* ***la***n*
Hae_in    *tl******* *******la* *k*y*aa*t* ***a**s*v* ***la*****
T_f       *tl******l ***e***ra* ***ht*t*** ***ae*aip* a*alr*****
          V  SQSGET  DTF ALR  K DLG  R LAI CNV  S   R  S·  FLTRA
M_l       *ai******* ***e*v*ha* *qk.ak**** **tn**qip* *c*a**y***
B_s       **l******* *sr*v*vqv* a**h.ka*t* t**p**t*s* *a*yt**lh*
S_c       **v******* ***l**n*cl *r*a*.tvg* v*sv**sis* vthcgvhin*

          501                                                  550
R_m       ********** ***c***v** a*rigag*** ***td****v l*qs*at**g
R_l       *******r** ***c***v** a*ragag*** ***sg****a l**s*ae**a
E_c       *t******** ******tv*l **vanvsrlk *.ldasi*h* i*hg*q***s
Hae_in    *v******** *******a*l **vta*g*vk *h*s**k*r* i**a*qs**a
T_f       ********** *******a*y ****s*a**p *asercaagg spgr*rq**g
          GPEIGVASTK AFTTQLA L  LLAL L KA  G             L  LP
M_l       ********** t*la*v*any **g*a*aq** **kyp**v*r eyre*e***d
B_s)      ****a***** ***a*i*v** v**svaadkn *i...nigf* l**e*.giaa
S_c)      ********** ***s*yia*v ****s*sdd* .vsk**rri* i*qg*kli*g

          551                                                  600
R_m       *mr*****it ****l*s**. *shy*dv*** ***ts**l** **********
R_l       *mg****siq ****v*s**. *snc*dv*** ***ts**l** **********
E_c       r***m*s.q* kr**a**e*. *s**h***** ***dq***** **********
Hae_in    e**ka*a.** t***a**e*. *a**h***** ***af****v *as*******
T_f       s*qha**.** *q*qgw*ar. *as*d***** ***lh***** **********
           V   LN LE P I   A    F  K HALFL GRG  YPIAL EGALKLKEIS
M_l       l*ar*iag*g *.*ad**yr. *aqsttv*** **hvg***** ********la
B_s       namea*c*qk d*m*mi***y *tvs*n***i ***ld*f*cv **********
S_c       q*k***k.** *r*kk*cate *k*qksl*** ***yq*aa** *****i****

          601                                                  650
R_m       ********** ******i*** ********** *h**....** **t***mq**
R_l       **qpks**** q*****y*** ********** *h**....** **t*t*mq**
E_c       *****a**** ********** *ad******* **ne....** **lk***e**
Hae_in    *****a**** ********** *ad******* **ne....** **vk***e**
T_f       *****a*p** ********** *rd******* ****....** **laa*mq**
          YIHAE Y AG ELKHGPLALV D  MPVVVIA PNDR    LL EK   N  EV
M_l       *m******** ******i*** *********m *spkgsam*h a*ll***r**
B_s       **q*****g* *****ti*** *qgt**fal* tq*hvnl... .sirg**k**
S_c       *m*s**vl** *****v**** *******afg tr*s....** p*v**s*eq*

          651                                                  700
R_m       a****riil* t**tgas... m*k*pt**t* v****a**i* ****sl****
R_l       .****riil* t**kgaa... a*k**t**t* v****d**i* ****sl****
E-c       r****q*y*f **q*ag*... **s.*n**i* e**h*e**i* *if*******
Hae_in    r****q*y*f **k*ag*... tps.*g*ki* t**k*n**v* *if*******
T_f       h****e*y*f **s*sh*... nas.agv*vm r**rhaglls *i*ha**v**
           ARGG L V  AD D  F          H    LP        P V  IP QL
M_l       qis*avti** ***gd*t... *rlyad.*l* e**a*stllq **ls*****v
B_s       a***antc** slkgl*.... ....*addrf v****npal* ***sv*****
S_c       t**k*hpi** cn*nd*vwaq k*ksid*qtl ev*qtv*clq g**ni*****
```

```
          701
R_m       ****t*v*m* a********* ******
R_l       ****t*vfm* ********** ******
E_c       ********** ********** ******
Hae_in    *s*y****** ********** ******
T_f       ****a***** ****r***** ******
          LAYH ALVKG TDVD PRNLA KSVTVE
M_l       **as**qar* y***k***** ******
B_s       i**ya**hr* c***k***** ******
S_c       *s*wl*vn** i***f***** ******
```

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 *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 *E.coli*, *Saccharomyces cerevisiae*, and nodulation protein NodM from *Rhizobium* (Surin and Downie, 1988) is also conserved in the same position in the *glmS* of *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 *et al.*, 1987), the glutamide analogue 6-diazo-5-oxonorleucine (Badet *et al.*, 1987), and N^3 -fumaroyl-L-2,3-diaminopropionate derivatives (Kucharczyk *et al.*, 1990), thus offering a potential target for antibacterial and antifungal agents (Andruszkiewicz *et al.*, 1990; Badet-Denisot *et al.*, 1993). Whether that is also the case for the *T.ferrooxidans glmS* gene is worth investigating considering the high amino acid sequence homology to the other *glmS* genes and the need to control *T.ferrooxidans* growth to reduce pollution arising from acid mine drainage.

3.4.4. Complementation of *E.coli* CGSC 5392 by *T.ferrooxidans* ATCC 33020 *glmS* gene.

The complementation of an *E.coli 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 *T.ferrooxidans glmS* gene was cloned as a 3.5 kb *Bam*HI-*Bam*HI fragment of p818.20 into pUCBM20 and pUCBM21 (p818.16f and p818.16r). In both constructs, the *glmS* gene is

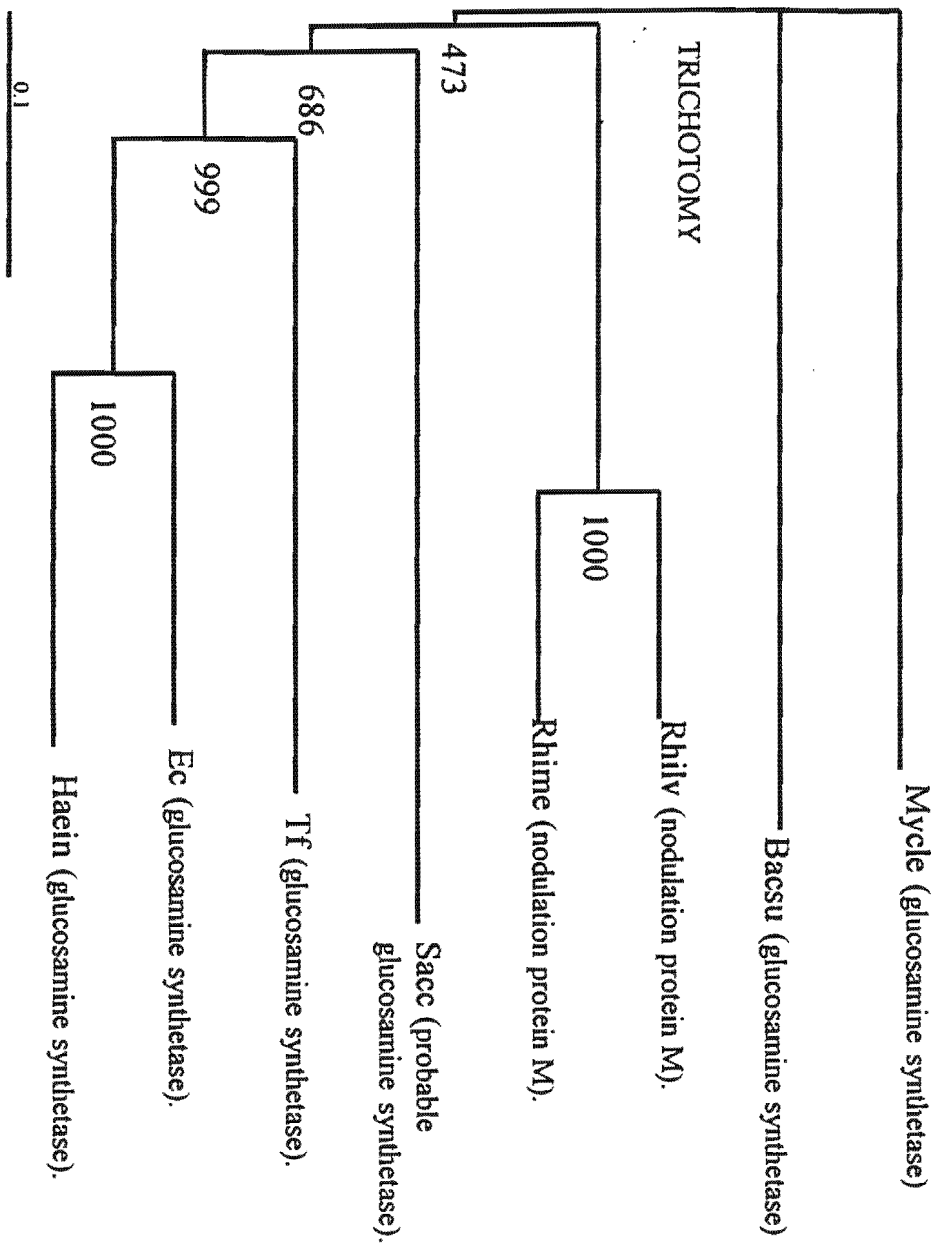
oriented in the 5'-3' direction with respect to the *lacZ* promoter of the cloning vectors. Complementation was detected by the growth of large colonies on Luria agar plates compared to *E.coli* mutant CGSC 5392 transformed with vector. Untransformed mutant cells produced large colonies only when grown on Luria agar supplemented with N-acetyl glucosamine (200 μg/ml). No complementation was observed for the *E.coli glmS* mutant transformed with pT*fatp1* and pT*fatp2* 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.

Construct	Results
p818.16f	+
p818.16r	+
p818.1	-
p818.20	-
pT*fatp1*	-
pT*fatp2*	-
pUCBM20	-
pUCBM21	-

+ complementation - non-complementation

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.leprae* (Mycle), and *B.subtilis* (Bacsu).



CHAPTER 4.

LIST OF FIGURES

CHAPTER FOUR.

TN*7*-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 Tn*7*-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 Tn*7* 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 Tn*7* nor to any of the antibiotic resistance markers associated with Tn*7*. 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 (IS*T2*) containing open reading frames (ORFs) (Yates *et al.*, 1988). Sequence comparisons have shown that, the putative transposase encoded by IS*T2* has homology with the proteins encoded by IS*256* and IS*Rm3* 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 IS*T2* can transpose within *T.ferrooxidans* (Holmes and Haq, 1990). Additionally, it has been suggested that transposition of family 1 insertion sequences (IS*T1*) 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 IS*T2* 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 Tn*5467* 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 Tn*21* 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 Tn*7*.

Analysis of the sequence at the terminus of the 3.5 kb *Bam*HI-*Bam*HI fragment of p818.20 revealed an ORF with very high homology to TnsA protein of the transposon Tn*7* (Fig. 4.4). Further studies were carried out to determine how much of the Tn*7* transposition genes were present in the region further downstream of the *T.ferroxidans glmS* gene. Tn*7* possesses trimethoprim, streptomycin, spectinomycin and streptothricin antibiotic resistance markers. Since *T.ferrooxidans* is not exposed to a hospital environment, it was of particular interest to find out whether similar genes were present in the Tn*7*-like transposon. In the *E.coli* chromosome, the Tn*7* 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.ferrooxidans* 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 *Kpn*I-*Sal*I 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 *Kpn*I (p818.52r) and from the *Sal*I restriction sites (p818.52f).

4.3.2.2. Construct p818.09.

Construct p818.09 was made by *Eco*RI-*Hin*dIII 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 *Eco*RI 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 *EcoR*I and ligating the approximately 1 kb fragment to the vector.

4.3.2.4. p818.11

A 2.5 kb *Apa*I-*Cla*I 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*Tn*7* site of Tn*7* insertion within the chromosome of *E.coli* and the imperfect inverted repeat sequences of Tn*7* (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*Tn*7* by Tn*7* in *E.coli* (Lichtenstein and Brenner, 1982) appears as "CCGGG" in *T.ferrooxidans* (bold and underlined in Figs.4.1 and 4.2). The two sequences "CCGGG" and "GCGGG" are almost equidistant from their respective *glmS* translational stop codons. The alignment of Tn*7* and the Tn*7*-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 Tn*7* are very similar to the inverted repeats of the Tn*7*-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 Tn*5468.*

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 Tn*7* (Fig. 4.4). Comparison of the predicted amino acid sequence of the TnsA-like protein of Tn*5468* to the predicted

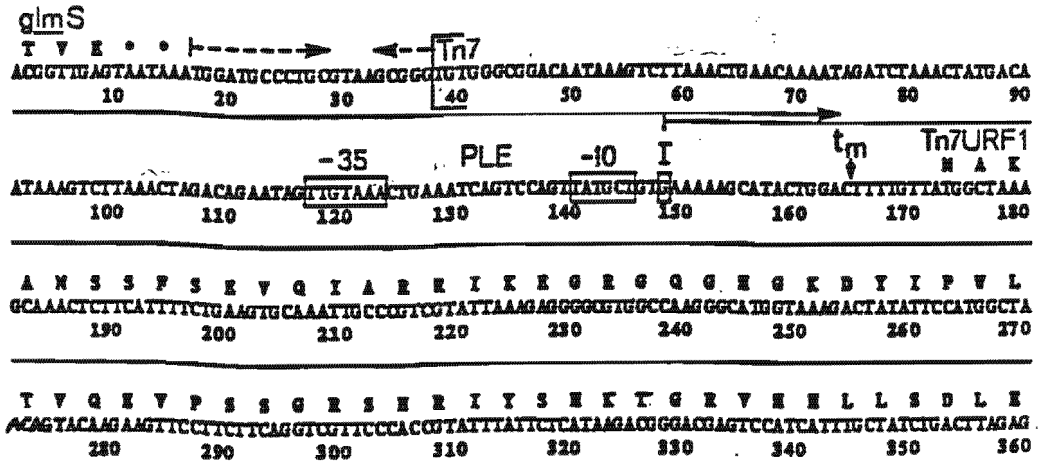
Fig. 4.1

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.

```
          T  V  E  * *
                  10        20        30        40        50        60
Ec.Tn7    ACGGTTGAGTAATAAATGGATGCCCTGCGTAACGGGGTGTGGGCGGACAATAAAGTCTTA
          ||||| |||||||     | | |  || |     | |||   |||||||||| ||| | |
Tf.7-     ACGGTGGAGTAATGGGGCGGTTCGGTGTGCCGGGTGTTGTTAACGGACAATAGAGTATCA
(like)            10        20        30        40        50        60

                  70        80        90       100       110       120
Ec.Tn7    AACTGAACAAAATAGATCTAAACTATGACAATAAAGTCTTAAACTAGACAGAATAGTTGT
            ||| |||| | || |   |  |       |  ||         ||  ||  |     | |
Tf.7-     TTCTGGACAATAGAGTTTCATCCCGAACAATAAAGTATCATTCCTCAACCAATAGAGTAT
(like)            70        80        90       100       110       120

                 130       140       150       160       170       180
Ec.Tn7    AAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGCTAAA
            | |         |  |      |       |               |           | |
Tf.7-     CATCCTGGCCTTGCGCTCCGGAGGATGTGGAGTTAGCTTGCCTAATGATCAGCTAGGAGA
(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)

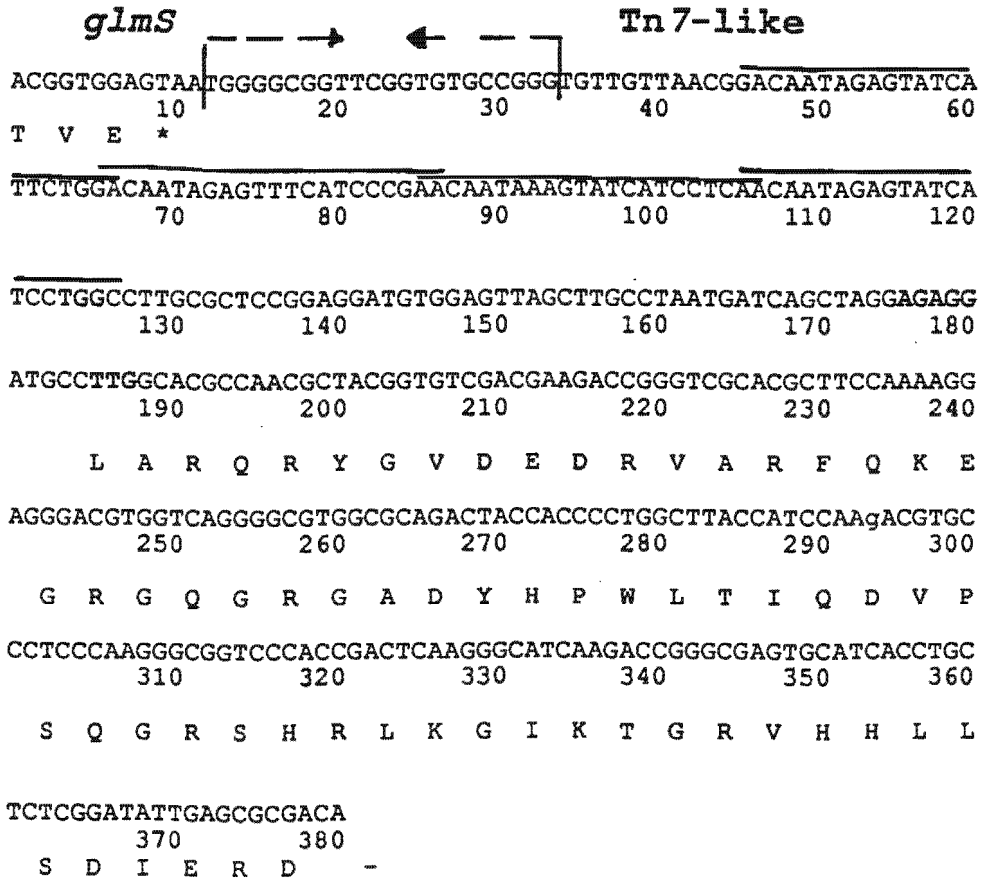


Fig. 4.3. Inverted repeats: (a) Alignment of the sequences of the four *tnsA* proximal inverted repeats of Tn7. Consensus bases among the four different repeats are highlighted. (b) Alignment of the nucleotide sequences of four proximal repeats of the Tn7-like element from *T.ferrooxidans*. (c) Nucleotide 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	g**ACAA**tAa**AG** tct**TAAACT**a *a

(b) TN7-LIKE

REPEAT 1	GACAATAGAG TATCATTCTG GA
REPEAT 2	GACAATAGAG TTTCATCCCG AA
REPEAT 3	AACAATAAAG TATCATCCTC AA
REPEAT 4	AACAATAGAG TATCATCCTG GC
CONSENSUS (all four)	**gACAATAgAG TaTCATcCtg aa** * * **a a**

(c)

Transposon Tn7	GAcaAtagAg tttcatTccc ga
Tn7-like element	gACAATAgAG TaTCATcCtg aa
	gAcaAtagAg 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.

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


sp|P13988|TNSA_ECOLI  TRANSPOSASE FOR TRANSPOSON TN7. ...  +3    302  1.1e-58
pir|JS0585|JS0585     transposition protein A - Escher...  +3    302  1.6e-58
pir|S03133|S03133     transposase - Escherichia coli t...  +3    302  3.8e-38
pir|S18587|S18587     hypothetical protein 2 - Thiobac...  +3    190  8.8e-24


>sp|P13988|TNSA_ECOLI TRANSPOSASE FOR TRANSPOSON TN7. >pir|S12637|S12637
tnsA protein - Escherichia coli transposon Tn7 >gp|X17693|ISTN7TNS_1
Transposon Tn7 transposition genes tnsA, B, C, D 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 LARQRYGVDEDRVARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGIKTGRVHHLLS 368
           +A+      E ++AR  KEGRGQG G DY PWLT+Q+VPS GRSHR+   KTGRVHHLLS
Sbjct:   1 MAKANSSFSEVQIARRIKEGRGQGHGKDYIPWLTVQEVPSSGRSHRIYSHKTGRVHHLLS 60

Query:   369 DIERDIFYLFDWADAVTDIREQFPLNRDITRRIADDLGVIHP 494
             D+E  +F   +W  +V DIREQFPL    TR+IA D G+ HP
Sbjct:    61 DLELAVFLSLEWESSVLDIREQFPLLPSDTRQIAIDSGIKHP 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 DGRMVQLARAVKPAEELEKPRVVEKLEIERRYWAQQGVDWGVVTERDIPKAMVRNIAWVH 737
           DG   Q A  VKPA  L+  R +EKLE+ERRYW Q+ + W + T+++I   +  NI W++
Sbjct: 121 DGPFEQFAIQVKPAAALQDERTLEKLELERRYWQQKQIPWFIFTDKEINPVVKENIEWLY 180

Query:   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 GTPLVMTTDFLVD 548
             G   VM+TDFLVD
Sbjct:   106 GVDQVMSTDFLVD 118


>pir|S18587|S18587 hypothetical protein 2 - Thiobacillus ferrooxidans
            >gp|X57326|TFMERRCG_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 VARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGIKTGRVHHLLSDIERDIFYLFD 401
           +AR  KEGRGQG    Y PWLT++DVPS+G S R+KG KTGRVHHLLS +E   F + D
Sbjct:  13 IARRLKEGRGQGSEKSYKPWLTVRDVPSRGLSVRIKGRKTGRVHHLLSQLELSYFLMLD 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 Tn*5468* (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 Tn*5468* (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 LARQRYGVDEDRVARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGI 50
                  : :     : : :||  ||||||| :  | ||||: |||| | | |:||
Tf  E-15        1 MSKGSRRSESGAIARRLKEGRGQGSEKSYKPWLTVRDVPSRGLSVRIKGR 50

Tf  Tn5468     51 KTGRVHHLLSDIERDIFYLFD..WADAVTDIREQFPLNR.DITRRIADDL 97
                  ||||||||||::|   | ::|  :|: ||||||||||   : |  |||
Tf  E-15       51 KTGRVHHLLSQLELSYFLMLDDIRAGCVTDIREQFPLTPIETTLEIADIR 100

Tf  Tn5468     98 GVIHPRDVGSGTPLVMTTDFLVDTIHDGRMVQLARAVKPAEELEKPRVVE 147
                     | |: : :       ||     | ||  |  : | : :
Tf  E-15      101 CAGAHRLVDDDGSVC......VDLNAHRRKVQAMRIRRTASGDQ..... 138
```

(b)

```
Percent Similarity: 59.542          Percent Identity: 42.366

Tf Tn5468       1 LARQRYGVDEDRVARFQKEGRGQGRGADYHPWLTIQDVPSQGRSHRLKGI 50
                  :|:   :  |  :||  |||||||:| || ||||:|:||| |||||: :
TnsA Tn7        1 MAKANSSFSEVQIARRIKEGRGQGHGKDYIPWLTVQEVPSSGRSHRIYSH 50

Tf Tn5468      51 KTGRVHHLLSDIERDIFYLFDWADAVTDIREQFPLNRDITRRIADDLGVI 100
                  ||||||||||:|  :|  ::|   | ||||||||    || || | |:
TnsA Tn7       51 KTGRVHHLLSDLELAVFLSLEWESSVLDIREQFPLLPSDTRQIAIDSGIK 100

Tf Tn5468     101 HPRDVGSGTPLVMTTDFLVDTIHDGRMVQLARAVKPAEELEKPRVVEKLE 150
                  ||  |  |   || ||||||   || : |:|  ||||  |:  | :||||
TnsA Tn7      101 HP..VIRGVDQVMSTDFLVD.CKDGPFEQFAIQVKPAAALQDERTLEKLE 147

Tf Tn5468     151 IERRYWAQQGVDWGVVTERDIPKAMVRNIAWVHSYAVIDQMSQPYDGYYD 200
                  :|||||  |  : | : |:::|   :  || |: |    :::|    :
TnsA Tn7      148 LERRYWQQKQIPWFIFTDKEINPVVKENIEWLYS.VKTEEVSAELLAQLS 196

Tf Tn5468     201 EKARLVLRELPSHPGPTLRQFCADMDLQFSMSAGDCLLLIRHLLAT.... 246
                    |::    |    :  :   | ::|: : :  |  |  || | |
TnsA Tn7      197 PLAHI....LQEKGDENIINVCKQVDIAYDLELGKTLSEIRALTANGFIK 242

Tf Tn5468     247 .......KANVCPMDGPTDDSKLLRQFRVAEGES 273
                        :|| |: |    :   : ::|
TnsA Tn7      243 FNIYKSFRANKCA.DLCISQVVNMEELRYVAN.. 273
```

(c)

Percent Similarity: 66.667 Percent Identity: 47.287

```
TnsA Tn7      1 MAKANSSFSEVQIARRIKEGRGQGHGKDYIPWLTVQEVPSSGRSHRIYSH 50
                | |:         ||||:|||||||| :| | ||||| :||| | | || ::
Tf E-15       1 MSKGSRRSESGAIARRLKEGRGQGSEKSYKPWLTVRDVPSRGLSVRIKGR 50

TnsA Tn7     51 KTGRVHHLLSDLELAVFLSLE..WESSVLDIREQFPLLPSDTR.QIA.ID 96
                ||||||||||:|||  || |:  : ::| |||||||| | :|  :|| |
Tf E-15      51 KTGRVHHLLSQLELSYFLMLDDIRAGCVTDIREQFPLTPIETTLEIADIR 100

TnsA Tn7     97 SGIKHPVIRGVDQVMSTDFLVDCKDGPFEQFAIQVKPA..... 134
                ::  | :: : |  ::.|:    :       |: ::
Tf E-15     101 CAGAHRLVDD.DGSVCVDLNAHRRKVQ....AMRIRRTASGDQ 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 Tn*5468* to TnsA of Tn*7* (59.5% identity and 42.4% similarity) was lower (Fig. 4.5b) than the homology between *T.ferrooxidans* strain E-15 and TnsA of Tn*7* (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 Tn*5468* unlike the other two. Homology of the Tn*7*-like transposon to TnsA of Tn*7* 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 Tn*7*-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 Tn*7*-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 *Kpn*I and *Sal*I 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 Tn*7* 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

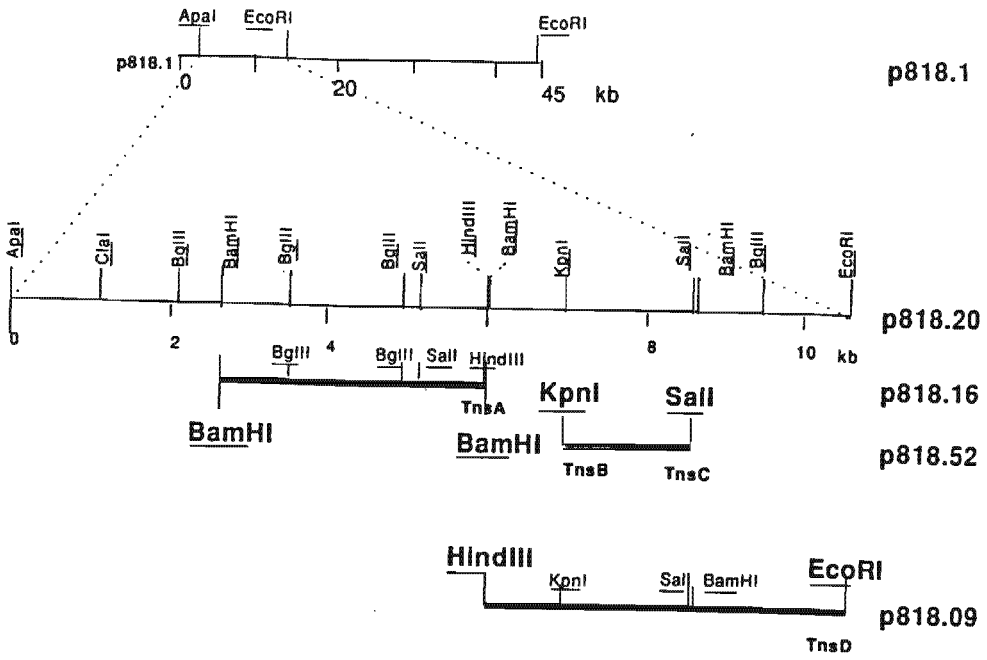


Fig. 4.6. Restriction map of p818.1 and construct p818.20 showing the various restriction sites on the fragments. The subclones p818.16, p818.52 and p818.09 and the regions which revealed high sequence homology to TnsA, TnsB, TnsC and TnsD are shown below construct p818.20.

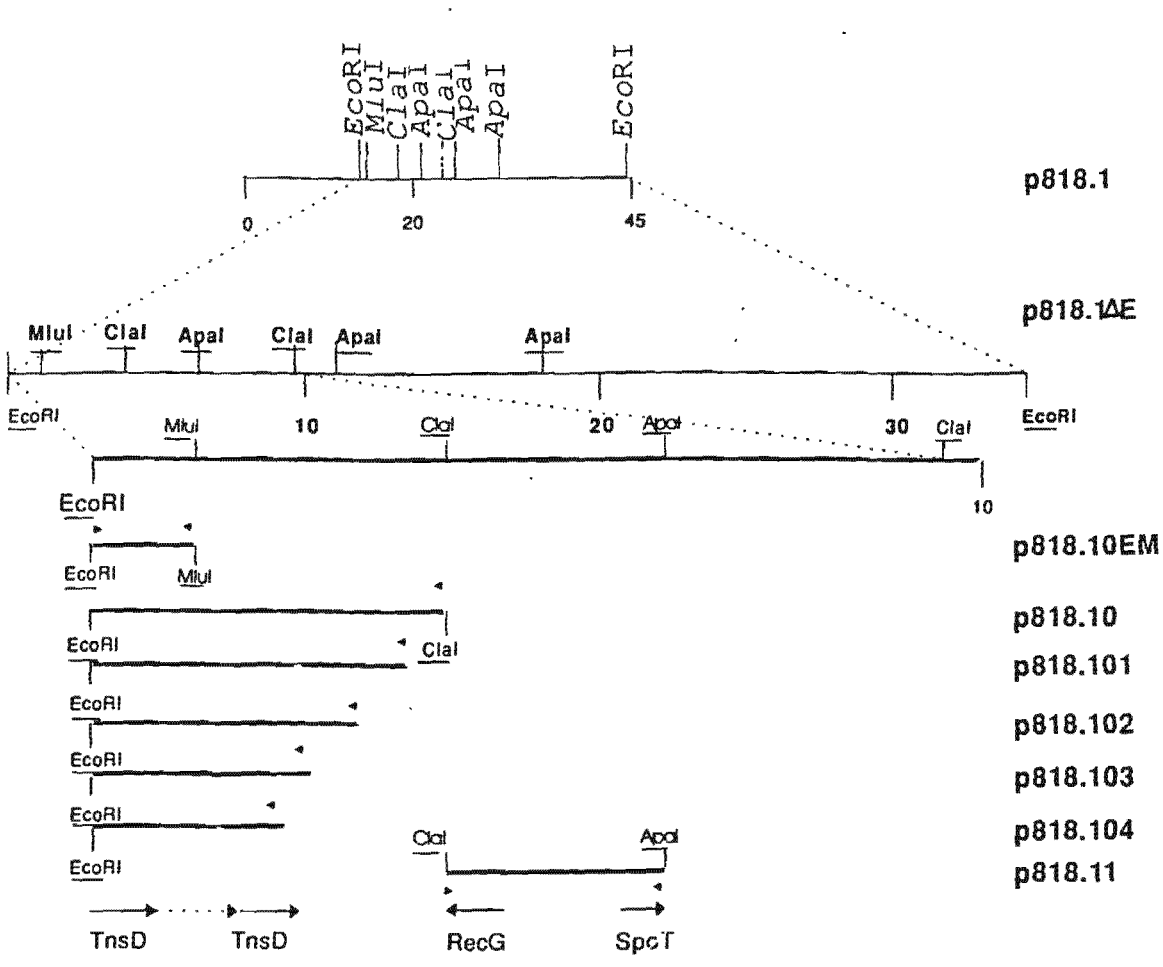


Fig. 4.7. Restriction map of cosmid p818.1ΔE showing the constructs, shortenings and subclones used to search for sequence homology to the TnsD and TnsE proteins of Tn7. Also shown is construct p818.11 which showed good amino acid sequence homology to *E.coli* and *H.influenzae* SpoT and RecG proteins at the ends.

Tn*7* (Orle and Craig, 1991). Homology of the sequence from the *Kpn*I 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 Tn*5468*. 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 Tn*7* 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 Tn*552* (Fig. 4.8).

The TnsC protein of Tn*7* 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. Homolgy to TnsC from Tn*7* 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 Tn*7* suggests that the size of the Tn*5468* TnsB and C like ORFs correspond to those found in Tn*7*. 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 *Eco*RI 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 *Kpn*I site) showing clear homology to the TnsB protein of Tn7. The amino acid sequence also showed high homology to Tra3 transposase of transposon Tn*552*. (b) Nucleotide sequence of p818.52 and the product of the open reading frame.

(a)

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Proba P(N)
sp|P13989|TNSB_ECOLI TRANSPOSON TN7 TRANSPOSITION PROT...	**+3**	**316**	**3.1e-37**
sp|P22567|ARGA_PSEAE AMINO-ACID ACETYLTRANSFERASE (EC ...	+1	46	0.00038
sp|P18416|TRA3_STAAU TRANSPOSASE FOR TRANSPOSON TN552 ...	**+3**	69	0.028
sp|P03181|YHL1_EBV HYPOTHETICAL BHLF1 PROTEIN.	-3	37	0.060
sp|P08392|ICP4_HSV11 TRANS-ACTING TRANSCRIPTIONAL PROT...	+1	45	0.082

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

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 YQIDATIADVYLVSRYDRTKIVGRPVLYIVIDVFSRMITGVYVGFEGPSWVGAMMALSNT 182
           Y+IDATIAD+YLV  +DR KI+GRP LYIVIDVFSRMITG Y+GFE PS+V AM A  N
Sbjct: 270 YEIDATIADIYLVDHHDRQKIIGRPTLYIVIDVFSRMITGFYIGFENPSYVVAMQAFVNA 329

Query: 183 ATEKVEYCRQFGVEIGAADWPCRPCPTLSRRPGREIAGSALRPFINNFQVRVEN 344
             ++K   C Q  +EI ++DWPC   P +      E+    +   +++F VRVE+
Sbjct: 330 CSDKTAICAQHDIEISSSDWPCVGLPDVLLADRGELMSHQVEALVSSFNVRVES 383
```

sp|P18416|TRA3_STAAU TRANSPOSASE FOR TRANSPOSON TN552 (ORF 480).
Length = 480

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 DRTKIVGRPVLYIVIDVFSRMITGVYVGFEGPSWVGAMMALSNTATEK 194
           D+   + RP L I++D +SR I G ++ F+ P+     + L      K
Sbjct: 176 DQKGNINRPWLTIIMDDYSRAIAGYFISFDAPNAQNTALTLHQAIWNK 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 YQIDATIADVYLVSR 47
           +Q D T+ D+Y++ +
Sbjct: 163 WQADHTLLDIYILDQ 177
```

(b)

```
GGTACCAGATCGACGCGACCATCGCGGACGTGTACCTGGTATCGCGATATGACCGTACAA
      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 -

AGATCGTCGGACGCCCGGTGCTCTATATCGTCATCGACGTGTTCAGCCGCATGATCACCG
      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 -

GCGTGTATGTGGGGTTCGAGGGACCTTCCTGGGTCGGGGCGATGATGGCCCTGTCCAACA
     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 -

CCGCCACGGAAAAGGTGGAATATTGCCGCCAGTTCGGCGTCGAGATCGGCGCGGCGGACT
     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 -

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

CATTGAGACCCTTTATCAACAACTTCCAGGTGCGGGTGGAAAAC
     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 *Sal*I 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.

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

pir|B25543|QQECE7       hypothetical protein E - Escher...  +1    176   1.5e-20
gp|X04492|ISTN7EUR_1    Transposon Tn7 fragment E with ...  +1    176   1.5e-20
sp|P05846|TNSC_ECOLI    TRANSPOSON TN7 TRANSPOSITION PR...  +1    176   6.4e-20
gp|U41011|CELD2024_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 SLDQVVWLKLDSPYKGSPKQLCISFFQEMDNLLGTPYRARYGASRSSLDEMMVQMAQMAN 213
           +++QVV+LK+D  + GS K++C++FF+ +D  LG+ Y  RYG  R  ++ M+  M+Q+AN
Sbjct: 163 NVEQVVYLKIDCSHNGSLKEICLNFFRALDRALGSNYERRYGLKRHGIETMLALMSQIAN 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:    1 YPQVVHHAE 27
             YPQV++H E
Sbjct:  153 YPQVIYHRE 161
```

(b)

```
   1 TATCCGCAAGTCGTCCACCATGCCGAGCCCTTCAGTCTTGACCAAGTGGTCTGGCTGAAG 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 CTGGATAGCCCCTACAAAGGATCCCCCAAACAACTCTGCATCAGCTTTTTCCAGGAGATG 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 GACAATCTATTGGGCACCCCGTACCGCGCCCGGTACGGAGCCAGCCGGAGTTCCCTGGAC 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 GAGATGATGGTCCAGATGGCGCAAATGGCCAACCTGCAGTCCCTCGGCGTACTGATCGTC 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 Tn*5468* to TnsD of Tn*7* (amino acid 68) begins at nucleotide 38 of the 799 bp sequence and continues downstream along with TnsD of Tn*7* (A, B, C, E, Fig. 4.11). However, homology to a region of Tn*5468* TnsD marked D (Fig. 4.11, 384-488 bp) was not to the predicted region of the TnsD of Tn*7*, 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 Tn*5468*.

Because of what appears to be a rearrangement of the Tn*5468 tnsD* 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 *Bgl*II site to the *Hin*dIII 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 *Eco*RI-*Cla*I p818.10 construct (lane 7) shown in Fig. 4.12. A subclone p818.01EM containing 1 kb *Eco*RI-*Mlu*I fragment of p818.10 (Fig. 4.7) cloned into pUCBM20 was sequenced from the *Mlu*I 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 *Mlu*I site is about 550 bp from the end of the sequence generated from the p818.10 *Eco*RI site.

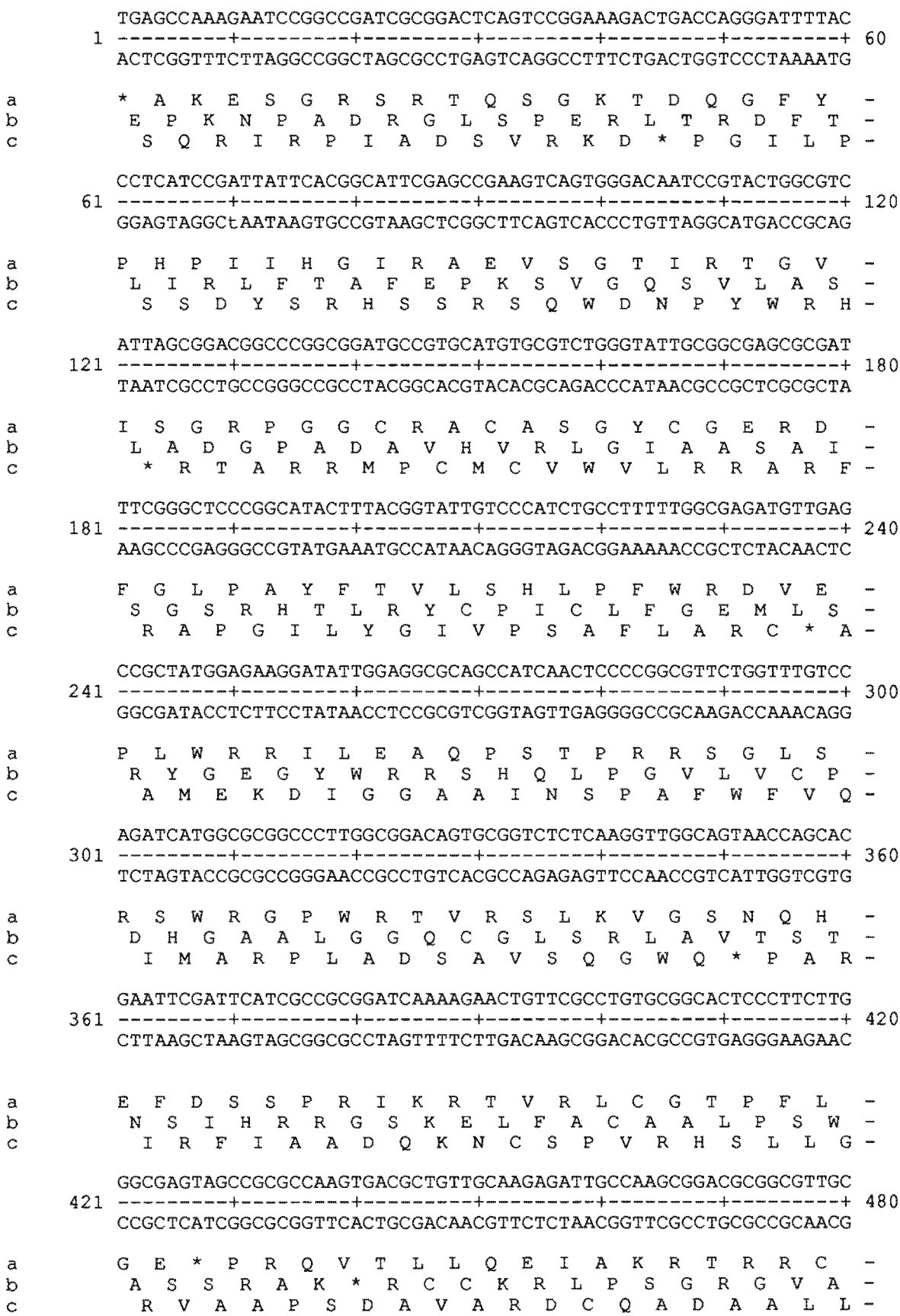
In other to determine whether any homology to Tn*7* could be detected further downstream, construct p818.10 was shortened from *Cla*I site (Fig. 4.7). Four shortenings namely, p818.101,

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)

```
                                                          Reading  High
Probability
Sequences producing High-scoring Segment Pairs:           Frame Score  P(N)
sp|P13991|TNSD_ECOLI   TRANSPOSON TN7 TRANSPOSITION PR... +2    86  8.1e-13
gp|D13972|AQUPAQ1_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 TRANSPOSON TN7 TRANSPOSITION PROTEIN TNSD.
pir|S12640|S12640 tnsD protein - Escherichia coli transposon Tn7
>gp|X17693|ISTN7TNS_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   LP +  CP HGA
Sbjct:   132 LNRYGEAFWQRDWYLPALPYCPKHGA 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 ERLTRDFTLIRLFTAFEPKSVGQSVLASLADGPADAVHVRLGIAASAI 181
             ++L  + TL  L+  F  K      +  +      AVH+ LG+AAS +
Sbjct:    68 QQLIYEHTLFPLYAPFVGKERRDEAIRLMEYQAQGAVHLMLGVAASRV 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 (28%), Positives = 18/35 (51%), Frame = +3

Query:   384 QKNCSPVRHSLLGRVAAPSDAVARDCQADAALLDH 488
             +K  S ++HS++ +   P   V    Q  +AL +H
Sbjct:   279 RKAFSYLQHSIVWQALLPKLTVIEALQQASALTEH 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 PSFRDWSAYYRSAVIARGFGKGK 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 +  RYCP C+
Sbjct:   117 SDNRFRYCPDCV 128
```

(b)


TGAGCCAAAGAATCCGGCCGATCGCGGACTCAGTCCGGAAAGACTGACCAGGGATTTTAC
1
60
ACTCGGTTTCTTAGGCCGGCTAGCGCCTGAGTCAGGCCTTTCTGACTGGTCCCTAAAATG
a * A K E S G R S R T Q S G K T D Q G F Y -
b E P K N P A D R G L S P E R L T R D F T -
c S Q R I R P I A D S V R K D * P G I L P -
CCTCATCCGATTATTCACGGCATTCGAGCCGAAGTCAGTGGGACAATCCGTACTGGCGTC
61
120
GGAGTAGGCtAATAAGTGCCGTAAGCTCGGCTTCAGTCACCCTGTTAGGCATGACCGCAG
a P H P I I H G I R A E V S G T I R T G V -
b L I R L F T A F E P K S V G Q S V L A S -
c S S D Y S R H S S R S Q W D N P Y W R H -
ATTAGCGGACGGCCCGGCGGATGCCGTGCATGTGCGTCTGGGTATTGCGGCGAGCGCGAT
121
180
TAATCGCCTGCCGGGCCGCCTACGGCACGTACACGCAGACCCATAACGCCGCTCGCGCTA
a I S G R P G G C R A C A S G Y C G E R D -
b L A D G P A D A V H V R L G I A A S A I -
c * R T A R R M P C M C V W V L R R A R F -
TTCGGGCTCCCGGCATACTTTACGGTATTGTCCCATCTGCCTTTTTGGCGAGATGTTGAG
181
240
AAGCCCGAGGGCCGTATGAAATGCCATAACAGGGTAGACGGAAAAACCGCTCTACAACTC
a F G L P A Y F T V L S H L P F W R D V E -
b S G S R H T L R Y C P I C L F G E M L S -
c R A P G I L Y G I V P S A F L A R C * A -
CCGCTATGGAGAAGGATATTGGAGGCGCAGCCATCAACTCCCCGGCGTTCTGGTTTGTCC
241
300
GGCGATACCTCTTCCTATAACCTCCGCGTCGGTAGTTGAGGGGCCGCAAGACCAAACAGG
a P L W R R I L E A Q P S T P R R S G L S -
b R Y G E G Y W R R S H Q L P G V L V C P -
c A M E K D I G G A A I N S P A F W F V Q -
AGATCATGGCGCGGCCCTTGGCGGACAGTGCGGTCTCTCAAGGTTGGCAGTAACCAGCAC
301
360
TCTAGTACCGCGCCGGGAACCGCCTGTCACGCCAGAGAGTTCCAACCGTCATTGGTCGTG
a R S W R G P W R T V R S L K V G S N Q H -
b D H G A A L G G Q C G L S R L A V T S T -
c I M A R P L A D S A V S Q G W Q * P A R -
GAATTCGATTCATCGCCGCGGATCAAAAGAACTGTTCGCCTGTGCGGCACTCCCTTCTTG
361
420
CTTAAGCTAAGTAGCGGCGCCTAGTTTTCTTGACAAGCGGACACGCCGTGAGGGAAGAAC
a E F D S S P R I K R T V R L C G T P F L -
b N S I H R R G S K E L F A C A A L P S W -
c I R F I A A D Q K N C S P V R H S L L G -
GGCGAGTAGCCGCGCCAAGTGACGCTGTTGCAAGAGATTGCCAAGCGGACGCGGCGTTGC
421
480
CCGCTCATCGGCGCGGTTCACTGCGACAACGTTCTCTAACGGTTCGCCTGCGCCGCAACG
a G E * P R Q V T L L Q E I A K R T R R C -
b A S S R A K * R C C K R L P S G R G V A -
c R V A A P S D A V A R D C Q A D A A L L -

```
    TCGATCATCCGCCAGCGCGCCCCAGCTTTCGCGATTGGAGTGCATATTATCGGAGCGCAG
481 ---------+---------+---------+---------+---------+---------+ 540
    AGCTAGTAGGCGGTCGCGCGGGGTCGAAAGCGCTAACCTCACGTATAATAGCCTCGCGTC

a       S  I  I  R  Q  R  A  P  A  F  A  I  G  V  H  I  I  G  A  Q  -
b        R  S  S  A  S  A  P  Q  L  S  R  L  E  C  I  L  S  E  R  S -
c         D  H  P  P  A  R  P  S  F  R  D  W  S  A  Y  Y  R  S  A  V -

    TGATTGCTCGCGGCTTCGGCAAAGGCAAGGAGAATGTCGCTCAGAACCTTCTCAGGGAAG
541 ---------+---------+---------+---------+---------+---------+ 600
    ACTAACGAGCGCCGAAGCCGTTTCCGTTCCTCTTACAGCGAGTCTTGGAAGAGTCCCTTC

a       *  L  L  A  A  S  A  K  A  R  R  M  S  L  R  T  F  S  G  K  -
b        D  C  S  R  L  R  Q  R  Q  G  E  C  R  S  E  P  S  Q  G  S -
c         I  A  R  G  F  G  K  G  K  E  N  V  A  Q  N  L  L  R  E  A -

    CAATTTCAAGCCTGTTTGAACCAGTAAGTCGACCTTATCCCCGGTGGTCTCGGCGACGAT
601 ---------+---------+---------+---------+---------+---------+ 660
    GTTAAAGTTCGGACAAACTTGGTCATTCAGCTGGAATAGGGGCCACCAGAGCCGCTGCTA

a       Q  F  Q  A  C  L  N  Q  *  V  D  L  I  P  G  G  L  G  D  D  -
b        N  F  K  P  V  *  T  S  K  S  T  L  S  P  V  V  S  A  T  I -
c         I  S  S  L  F  E  P  V  S  R  P  Y  P  R  W  S  R  R  R  L -

    TGGCCTACCAGTGGTACGGAAACACAGAAAGGCATTTCATCCGTTGCGGCATGTTCATTC
661 ---------+---------+---------+---------+---------+---------+ 720
    ACCGGATGGTCACCATGCCTTTGTGTCTTTCCGTAAAGTAGGCAACGCCGTACAAGTAAG


a       W  P  T  S  G  T  E  T  Q  K  G  I  S  S  V  A  A  C  S  F  -
b        G  L  P  V  V  R  K  H  R  K  A  F  H  P  L  R  H  V  H  S -
c         A  Y  Q  W  Y  G  N  T  E  R  H  F  I  R  C  G  M  F  I  Q -

    AGATTTTCTGACAAACGGTCGCCGCAACCGGACGGAAACCCCCTTCGGCAAGGGACCGGT
721 ---------+---------+---------+---------+---------+---------+ 780
    TCTAAAAGACTGTTTGCCAGCGGCGTTGGCCTGCCTTTGGGGGAAGCCGTTCCCTGGCCA

a       R  F  S  D  K  R  S  P  Q  P  D  G  N  P  L  R  Q  G  T  G  -
b        D  F  L  T  N  G  R  R  N  R  T  E  T  P  F  G  K  G  P  V -
c         I  F  *  Q  T  V  A  A  T  G  R  K  P  P  S  A  R  D  R  C -

    GCTCTTTAATTCGCTTGCA
781 ---------+--------- 799
    CGAGAAATTAAGCGAACGT

a       A  L  *  F  A  C    -
b        L  F  N  S  L  A    -
c         S  L  I  R  L       -
```

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 *Eco*RI 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 Tn*7* was concerned (Fig. 4.15 and 4.16 respectively).

A clearer picture of the rearrangement of the TnsD-like protein of Tn*5468* emerged from the BLAST search of the combined sequence of p818.104 and p818.103. Regions of homology of the TnsD-like protein of Tn*5468* to TnsD of Tn*7* (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 Tn*7*. Regions D and E of the Tn*5468* TnsD-like protein were homologous to an overlapping region of TnsD of Tn*7*. The largest gap in homology was found between nucleotides 712 and 1212 of the Tn*5468* TnsD-like protein (Fig. 4.11). Together, these results suggest that the TnsD-like protein of Tn*5468* has been rearranged, duplicated, truncated and shuffled.

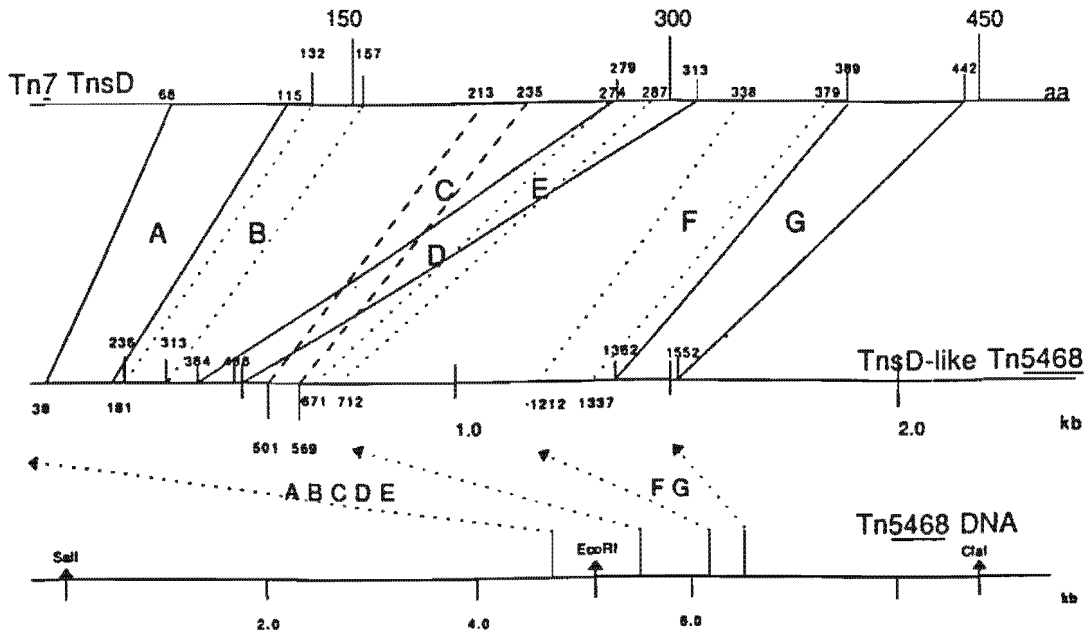


Fig. 4.11. Rearrangement of the TnsD-like ORF of Tn*5468*. Regions on TnsD-like protein of Tn*5468* identified by the letters A-G are matched unto the corresponding areas on TnsD of Tn*7*. At the bottom is the map of Tns*5468* which indicates the areas where homology to TnsD of Tn*7* was obtained. (Note that the numbers on the line representing TnsD of Tn*7* are amino acids, whereas the numbers on the TnsD-like protein of Tn*5468* represent distances in base pairs).

Fig. 4.12. Restriction digests carried out on p818.1ΔE with the following restriction enzymes; 1 *Hin*dIII, 2 *Eco*RI-*Apa*I, 3 *Hin*dIII-*Apa*I, 4. *Cla*I, 5. *Hin*dIII-*Eco*RI, 6. *Apa*I-*Cla*I, 7. *Eco*RI-*Cla*I, and 8. *Apa*I. 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.

Fig. 4.13. (a) BLAST results of the nucleotide sequence obtained from p818.10EM. (b) The inverted complemented nucleotide sequence of p818.10EM.

(a)

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Proba P(N)
pir|S40626|S40626 androgen receptor - mouse	+2	45	0.16
pir|S39636|S39636 cydD protein - Escherichia coli (...	-1	40	0.72
gp|D50624|STMVBRA1_4 NusG like protein [Streptomyces v...	+1	63	0.87

pir|S40626|S40626 androgen receptor - mouse
Length = 48

Plus Strand HSPs:

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*HGDPTA 403
             G+A     + C  GD+GS HG   A
Sbjct:    24 GSAWAAAAAQCRYGDLGSLHGGSVA 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 NPLDYGSAW 27
```

(b)

```
      1  TGGACTTTGG TCCCCTACTG GATGGTCAAA TCCGgCAGAA AGTGGCGAAg
     51  CCTGGGACTC AGGGCGGTAG CcAAATATCT TGCCATTGAg CCAGGGACGG
    101  TGAGATTGCA CGCCTCCCGA CGCGGGTTGA ATGTGCCCTG GAAGCCCTTA
    151  GCCGCACGAC ATTCCCGAAC ATTGGTGCCA GACCGGGATG CCATAAGAAT
    201  ACGGTGGTTG GACATGCAGC AGAATTACGC TGATTTATCA CGTCGGCtgA
    251  CTGGCACTCC TGCTACCCAA AGAACACTCA TGGTTGTACC GCCATTgACA
    301  GGGAGTGGCT GGAGCAACAT TCACcAACGG CACTGCACAA GAAATGGTCT
    351  GATCAGCGTG TGGACTGGGA GACATTGGAT CGTAGCATGG CGAtCCAACT
    401  GCGTAAGGCc GCGCGGGAAA tcAtctTCGG GAGGTTCCGC CACAACGCGt
```

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)

Sequences producing High-scoring Segment Pairs:		Reading Frame	High Score	Prob P(N)
pir|A47283|A47283	calphotin - Drosophila >gp|L050...	-2	57	0.011
sp|P13991|TNSD_ECOLI	**TRANSPOSON TN7 TRANSPOSITION PR...**	**+1**	**56**	**0.28**
gp|D49399|RABA2TIC_1	alpha 2 type I collagen [Orycto...	-3	42	0.32
pir|A44950|A44950	merozoite surface antigen 2 - P...	+3	74	0.33

>sp|P13991|TNSD_ECOLI TRANSPOSON TN7 TRANSPOSITION PROTEIN TNSD.
pir|S12640|S12640 tnsD protein - Escherichia coli transposon Tn7
gp|X17693|ISTN7TNS_4 Transposon Tn7 transposition genes tnsA, B, C, D and E. [Escherichia coli]
Length = 508

Plus Strand HSPs:

Score = 56 (25.8 bits), Expect = 0.33, Sum P(2) = 0.28
Identities = 14/54 (25%), Positives = 24/54 (44%), Frame = +1

```
Query:   319 RVDWETLDRSMAIQLRKAAREITREVPPQRVTQAALERKLGRPLRMSEQQAKLP 480
             RVDW   DR    QL +  + +   +   R T + L ++       +++   KLP
Sbjct:   389 RVDWNQRDRIAVRQLLRIIKRLDSSLDHPRATSSWLLKQTPNGTSLAKNLQKLP 442
```

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 WLDMQQNYADLSRRRLALLLPKEHSWLYRHTGSGWSNIHQRH 294
             W  +   Y  +   R +L     ++WLYRH      + +Q+H.
Sbjct:   338 WQQLVHKYQGIKAARQSLEGGVLYAWLYRHDRDWLVHWNQQH 379
```

(b)

```
GAGGAAATCGGAAGGCCTGGCATCGGACGGTGACCTATAATCATTGCCTTGATACCAGGG
        10        20        30        40        50        60
E  E  I  G  R  P  G  I  G  R  *  P  I  I  I  A  L  I  P  G

ACGGTGAGATTGCACGCCTCCCGACGCGGGTTGAATGTGCCCTGGAAGCCCTTGACCGCA
        70        80        90       100       110       120
T  V  R  L  H  A  S  R  R  G  L  N  V  P  W  K  P  L  T  A

CGACATTCCCGAACATTGGTGCCAGACCGGGATGCCATAAGAATACGGTGGTTGGACATG
       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

CAGCAGAATTACGCTGATTTATCACGTCGGCGACTGGCACTCCTGCTACCCAAAGAACAC
       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
```

```
TCATGGTTGTACCGCCATACAGGGAGTGGCTGGAGCAACATTCACCAACGGCACTGCACA
       250       260       270       280       290       300
S  W  L  Y  R  H  T  G  S  G  W  S  N  I  H  Q  R  H  C  T

AGAAATGGTCTGATCCAGCGTGTGGACTGGGAGACATTGGATCGTAGCATGGCGATCCAA
       310       320       330       340       350       360
R  N  G  L  I  Q  R  V  D  W  E  T  L  D  R  S  M  A  I  Q

CTGCGTAAGGCCGCGCGGGAAATCACTCGGGAGGTTCCGCCACAACGCGTTACCCAGGCG
       370       380       390       400       410       420
L  R  K  A  A  R  E  I  T  R  E  V  P  P  Q  R  V  T  Q  A

GCTTTGGAGCGCAAGTTGGGGCGGCCACTCCGGATGTCAGAACAACAGGCAAAACTGCCT
       430       440       450       460       470       480
A  L  E  R  K  L  G  R  P  L  R  M  S  E  Q  Q  A  K  L  P

AAAAGTATCGGTGTACTTGGCGA
       490       500
K  S  I  G  V  L  G
```

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)

sp|P29424|TX25_PHONI  NEUROTOXIN TX2-5. >pir|S29215|S2...  +3     57  0.991
sp|P28880|CXOA_CONST  OMEGA-CONOTOXIN SVIA. >pir|B4437...  +3     55  0.998
gp|X77576|BNACCG8_1   acetyl-CoA carboxylase [Brassica...  +2     39  0.999
gp|X90568|HSTITIN2B_1 titin gene product [Homo sapiens]    +2     43  0.9995


>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 ERGECVCGGPCICRQGNFLIAAYKLASCKC 48


sp|P28880|CXOA_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 CRSSGSPCGVTSICCGRCY 19
```

(b)

```
       1  GGAGTCCTtG TGATGTTTAA ATTGAGTGAG AAAAGAGCGT ATTCCGGCGA

      51  AGTGCCTGAA AATTTGACTC TACACTGGCT GGCCGAATGT CAAACAAGAC

     101  GATGGAAAAA GGGTCGCAGG TCCGTAAGTC ACCCGCGCCG TGTCGTCAGG

     151  CGAATTTGCC ATGTGGCGCG TACCTATTGG GCTGTTCCCG GTGCTGGCAC

     201  TCGGCTATAG CTTCTCCGAC GGTAGATTGC TCCCAATAAC CTACGGCGAA

     251  TATTCGGAAG TGATTATTCC CAATTCTGGC GGAAGGAGAG GAAATCAACT

     301  CGTCCGACAT TCCGCCGGAA TTATATTCGT 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)

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

sp|P02241|HCYD_EURCA HEMOCYANIN D CHAIN.                           +2    47  0.038
sp|P03108|VL2_CRPVK  PROBABLE L2 PROTEIN.                          +3    66  0.072
sp|Q09816|YAC2_SCHPO HYPOTHETICAL 33.9 KD PROTEIN C16C...          +2    39  0.69
sp|P06278|AMY_BACLI  ALPHA-AMYLASE PRECURSOR (EC 3.2.1...         +2    52  0.75
sp|P26805|GAG_MLVFP  GAG POLYPROTEIN (CORE POLYPROTEIN...         +3    53  0.77


sp|P02241|HCYD_EURCA HEMOCYANIN D CHAIN.
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
             H+HW+  +P FY
Sbjct:   170 HWHWHLVYPAFY 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 TPWAGDRRAETQGKRSL*AVGFFHFPDIFGSFPHFH 148
             T  + D   E +    L  VGF     +FGSF   H
Sbjct:    17 TSLSPDPLPEAERDPRLKGVGFLPRGTLFGSFHEEH 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 DPYFFVPPADFVYPAKSGSGQ 250
             D Y FV   D+   +  G+G+
Sbjct:   548 DFYLFVMLTDYEEDSVQGAGE 568
```

(b)

```
    1  CGTCCGTACC TTCACCTTTC TCGACCGCAA GCAGCCTGAA ACTCCATGGG

   51  CAGGAGATCG TCGTGCAGAG ACTCAGGGGA AACGCTCACT TTAGGCGGTG

  101  GGGTTTTTCC ACTTCCCCGA TATTTTCGGG TCTTTCCCTC ATTTTCACTG

  151  GTACCCGCAG CACCCTTGTT TCTACGCCTG ATAACACGAC CCGTACTTTT

  201  TTGTACCACC CGCAGACTTC GTTTACCCGG CAAAAAGTGG TTCTGGTCAA

  251  TATCGGCAGG GTGGTGAGAA 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)

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Prob P(N)
sp|P43809|RECG_HAEIN ATP-DEPENDENT DNA HELICASE RECG. ...	+3	158	1.3e-14
gi|290502 (L10328) DNA recombinase [Escheri...	+3	156	2.6e-14
sp|P24230|RECG_ECOLI ATP-DEPENDENT DNA HELICASE RECG. ...	+3	156	2.6e-14
gi|1001580 (D64000) hypothetical protein [Sy...	+3	127	5.6e-10
gi|1150620 (Z49988) MmsA [Streptococcus pneu...	+3	132	8.0e-10

>sp|P43809|RECG_HAEIN ATP-DEPENDENT DNA HELICASE RECG. >pir||E64139 DNA recombinase (recG) homolog - Haemophilus influenzae (strain Rd KW20) >gi|1008823 (L44641) DNA recombinase [Haemophilus influenzae] >gi|1205971 (U00087) DNA recombinase [Haemophilus influenzae] >gi|1221898 (U32794) DNA recombinase, helicase [Haemophilus influenzae]

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 EILAEQLPLAFQQWLEPAGPGGGLSGGQPFTRARRETAETLAGGSLRLVIGTQSLFQQGV 182
          EILAEQ    F++W +P G   G   G+   ++R+   E +  G++++V+GT +LFQ+ V
Sbjct:327 EILAEQHANNFRRWFKPFGIEVGWLAGKVKGKSRQAELEKIKTGAVQMVVGTHALFQEEV 386


Query:  183 VFACLGLVIIDEQHRF 230
            F+ L LVIIDEQHRF
Sbjct:  387 EFSDLALVIIDEQHRF 402
```

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 RRGAMPHLLVMTASPI 308
            + G  PH L+MTA+PI
Sbjct:  416 KAGFYPHQLIMTATPI 431
```

>sp|P24230|RECG_ECOLI ATP-DEPENDENT DNA HELICASE RECG. >pir||JH0265 RecG protein - Escherichia coli >pir||S18195 recG protein - Escherichia coli >gi|42669 (X59550) recG gene product [Escherichia coli] >gi|147545 (M64367) DNA recombinase [Escherichia coli]

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 EILAEQLPLAFQQWLEPAGPGGGLSGGQPFTRARRETAETLAGGSLRLVIGTQSLFQQGV 182
          E+LAEQ    F+ W  P G   G   G+   +AR    E +A G +++++GT ++FQ+ V
Sbjct:327 ELLAEQHANNFRNWFAPLGIEVGWLAGKQKGKARLAQQEAIASGQVQMIVGTHAIFQEQV 386


Query:  183 VFACLGLVIIDEQHRF 230
              F  L LVIIDEQHRF
Sbjct:  387 QFNGLALVIIDEQHRF 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 RRGAMPHLLVMTASPI 308
             ++G  PH L+MTA+PI
Sbjct:  416 QQGFHPHQLIMTATPI 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 EILAEQLPLAFQQWLEPAGPGGGLSGGQPFTRARRETAETLAGGSLRLVIGTQSLFQQGV 182
          E+LAEQ       W         L  G   T  RRE    L+ G L L++GT +L Q+ V
Sbjct:464 EVLAEQHYQKLVSWFNLLYLPVELLTGSTKTAKRREIHAQLSTGQLPLLVGTHALIQETV 523

Query:  183 VFACLGLVIIDEQHRF 230
             F  LGLV+IDEQHRF
Sbjct:  524 NFQRLGLVVIDEQHRF 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 RGAMPHLLVMTASPI 308
             +G  PH+L MTA+PI
Sbjct:  550 KGNAPHVLSMTATPI 564


(b)

          CCGAGATTCTCGCGGAGCAGCTCCCATTGGCGTTCCAGCAATGGCTGGAACCGGCTGGGC
                  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 -

          CTGGAGGTGGGCTATCTGGTGGGCAGCCGTTCACCCGTGCCCGTCGCGAGACGGCGGAAA
                  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 -

          CGCTTGCTGGTGGCAGCCTGAGGTTGGTAATCGGCACCCAGTCGCTGTTCCAGCAAGGGG
                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 -
```

```
TGGTGTTTGCATGTCTCGGACTGGTCATCATCGACGAGCAACACCGCTTTTGGCCGTGGA
      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 -

GCAGCGCCGTCAATTGCTGGAGAAGGGGCGCCATGCCCCACCTGCTGGTAATGACCGCTA
     250       260       270       280       290       300
  S   A   V   N   C   W   R   R   G   A   M   P   H   L   L   V   M   T   A   S -

GCCCGATCATGGTCGAGGACGGCATCACGTCAGGCATTCTTCTGTGCGGTAGGACACCCA

      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 -

ATCGATTCCTGCCTCAAGCTGGTATCGACGCCGTTGCCTTTCCGGGCGTAGATAAGGACT
      370       380       390       400       410       420
  R   F   L   P   Q   A   G   I   D   A   V   A   F   P   G   V   D   K   D   Y -

ATGCCGCCCGTGAAAGGGCCGCCAAGCGGGGCCGATGgACGCCGCTGCTAAACGACGCAG
      430       440       450       460       470       480
  A   A   R   E   R   A   A   K   R   G   R   W   T   P   L   L   N   D   A   G -

GCATACGTCGAAGCTGGCTTGGTCGAGCAAGCATCGCCTTCGTCCAGCGCAACACTGCTA
      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 -

TCGGAGGGCAACTTGAAGAAGGCGGTCGCGCCGCGAGAACCTCCCAGCTTATCCCCGCGA
      550       560       570       580       590       600
  G   G   Q   L   E   E   G   G   R   A   A   R   T   S   Q   L   I   P   A   S -

GCCATCCGCGAACGGATCGTCAACGCCTGATTCATCGAGACTATCTGCTCTCAAGCACTG
      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 -

ACATTGAGCTTTCGATCTATGAGGACCGACTAGAAATTACTTCCAGGCAGATTCAAATGG
     670       680       690       700       710       720
  I   E   L   S   I   Y   E   D   R   L   E   I   T   S   R   Q   I   Q   M   V -

TATTACGCGCCGATCGTGACCTGGCCGGTCGACACGAAACCAGCTCATCAAGGATGTTAT
      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 *Apa*I-*Cla*I construct is shown in Fig. 4.7. The single strand sequence from the *Cla*I site was joined to p818.10r and searched using BLAST against the GenBank and EMBL databases. Good homology to *E.coli* and *H.influezae* ATP-dependent DNA helicase recombinase proteins (EC 3.6.1.) was obtained (Fig. 4.17). The BLAST search with the sequence from the *Apa*I 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 *rpo*Z) 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 *Apa*I 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)

Sequences producing High-scoring Segment Pairs:		Reading Frame	High Score	Sum Prob P(N)
sp|P17580|SPOT_ECOLI	GUANOSINE-3',5'-BIS(DIPHOSPHATE...	+2	277	2.5e-31
sp|P43811|SPOT_HAEIN	GUANOSINE-3',5'-BIS(DIPHOSPHATE...	+2	268	4.5e-30
gp|U22374|VSU22374_1	csrS gene product [Vibrio sp. S14]	+2	239	5.7e-28
gp|U29580|ECU29580_9	GTP pyrophosphokinase [Escheric...	+2	239	5.1e-26
gp|U13769|VSU13769_2	ppGpp synthetase I [Vibrio sp.]	+2	239	5.1e-26
gp|X87267|SCAPTRELA_2	putative ppGpp synthetase [Stre...	+2	233	3.6e-25
sp|P11585|RELA_ECOLI	GTP PYROPHOSPHOKINASE (EC 2.7.6...	+2	230	9.0e-25
pir|S39975|S39975	stringent response-like protein...	+2	225	4.5e-24
sp|P44644|RELA_HAEIN	GTP PYROPHOSPHOKINASE (EC 2.7.6...	+2	221	1.6e-23

>sp|P17580|SPOT_ECOLI GUANOSINE-3',5'-BIS(DIPHOSPHATE) 3'-PYROPHOSPHOHYDROLASE (EC 3.1.7.2) ((PPGPP)ASE) (PENTA-PHOSPHATE GUANOSINE-3'-PYROPHOSPHOHYDROLASE). >pir|B30374|SHECGD guanosine 3',5'-bis(diphosphate) 3'-pyrophosphatase (EC 3.1.7.2) - Escherichia coli >gp|M24503|ECOSPOT_2 (p)ppGpp3'-pyrophosphohydrolase [Escherichia coli] >gp|L10328|ECOUW82_16(p)ppGpp 3'-pyrophosphohydrolase [Escherichia coli]

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 QASGREKHVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPIPGRF 193
           + SGREKH+Y    KM  K   F  I D++AFR+IV D DTCYRVLG +HSLY+P PGR
Sbjct: 232 RVSGREKHLYSIYCKMVLKEQRFHSIMDIYAFRVIVNDSDTCYRVLGQMHSLYKPRPGRV 291

Query:  194 KDYIAIPKSNGYQSLHTVLAGP 259
           KDYIAIPK+NGYQSLHT + GP
Sbjct: 292 KDYIAIPKANGYQSLHTSMIGP 313
```

>sp|P43811|SPOT_HAEIN GUANOSINE-3',5'-BIS(DIPHOSPHATE) 3'-PYROPHOSPHOHYDROLASE (EC 3.1.7.2) ((PPGPP)ASE) (PENTA-PHOSPHATE GUANOSINE-3'-PYROPHOSPHOHYDROLASE). >pir|F64139|F64139 penta-phosphate guanosine-3'-pyrophosphohydrolase (spoT) homolog - Haemophilus influenzae (strain Rd KW20) >gp|U00087|HIU00087_15 penta-phosphate guanosine-3'-pyrophosphohydrolase [Haemophilus influenzae] >gp|L44642|HEAHI1741_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 AQASGREKHVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPIPGR 190
           A+  GREKH+Y+  +KM+ K   F  I D++AFR+IV +VD CYRVLG +H+LY+P PGR
Sbjct: 204 ARVWGREKHLYKIYQKMRIKDQEFHSIMDIYAFRVIVKNVDDCYRVLGQMHNLYKPRPGR 263

Query:  191 FKDYIAIPKSNGYQSLHTVLAGP 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 KMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPIPGRFKDYIAIPKSNGYQS 235
           KM+ K   F  I D++AFR++V+D+DTCYRVLG VH+LY+P P R KDYIAIPK+NGYQS
Sbjct:   3 KMKNKEQRFHSIMDIYAFRVLVSDLDTCYRVLGQVHNLYKPRPSRMKDYIAIPKANGYQS 62

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


>gp|U29580|ECU29580_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 AQASGREKHVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPIPGR 190
           A+  GR KH+Y   RKMQKK  AF ++ D+ A RI+   +  CY  LG+VH+ YR +P
Sbjct: 247 AEVYGRPKHIYSIWRKMQKKNLAFDELFDVRAVRIVAERLQDCYAALGIVHTHYRHLPDE 306

Query:  191 FKDYIAIPKSNGYQSLHTVLAGP 259
            F DY+A PK NGYQS+HTV+ GP
Sbjct:  307 FDDYVANPKPNGYQSIHTVVLGP 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 AQASGREKHVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPIPGR 190
           A+  GR KH+Y   RKMQKK   F ++ D+ A RI+   +  CY  LG+VH+ YR +P
Sbjct: 246 AEVQGRPKHIYSIWRKMQKKSLEFDELFDVRAVRIVAEELQDCYAALGVVHTKYRHLPKE 305

Query:  191 FKDYIAIPKSNGYQSLHTVLAGP 259
            F DY+A PK NGYQS+HTV+ GP
Sbjct:  306 FDDYVANPKPNGYQSIHTVVLGP 328
```

```
>gp|X87267|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 RFAAQASGREKHVYRSIRKMQKKGYAFGDIHDLHAFRIIVADVDTCYRVLGLVHSLYRPI 181
           R  A  +GR KH Y   +KM  +G  F +I+DL   R++V  V  CY  LG VH+ + P+
Sbjct: 330 RIKATVTGRPKHYYSVYQKMIVRGRDFAEIYDLVGIRVLVDTVRDCYAALGTVHARWNPV 389

Query: 182 PGRFKDYIAIPKSNGYQSLHTVLAGP 259
           PGRFKDYIA+PK N YQSLHT + GP
Sbjct: 390 PGRFKDYIAMPKFNMYQSLHTTVIGP 415
```

(b)

```
ACGGTTCGCCGCCCAGGCTAGTGGCCGTGAGAAACACGTCTACAGATCTATCAGAAAAAT
        10        20        30        40        50        60
 R  F  A  A  Q  A  S  G  R  E  K  H  V  Y  R  S  I  R  K  M  -

GCAGAAGAAGGGGTATGCCTTCGGTGACATCCACGACCTCCACGCCTTCCGGATTATTGT
        70        80        90       100       110       120
 Q  K  K  G  Y  A  F  G  D  I  H  D  L  H  A  F  R  I  I  V  -

TGCTGATGTGGATACCTGCTATCGCGTTCTGGGTCTGGTTCACAGTCTGTATCGACCGAT
       130       140       150       160       170       180
 A  D  V  D  T  C  Y  R  V  L  G  L  V  H  S  L  Y  R  P  I  -

TCCCGGACGTTTCAAAGACTACATCGCCATTCCGAAATCCAATGGATATCAGTCTCTGCA
       190       200       210       220       230       240
 P  G  R  F  K  D  Y  I  A  I  P  K  S  N  G  Y  Q  S  L  H  -


CACCGTGCTGGCTGGGCCC
       250       259
 T  V  L  A  G  P  -
```

cross-links specifically to ß' 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 *Apa*I 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.

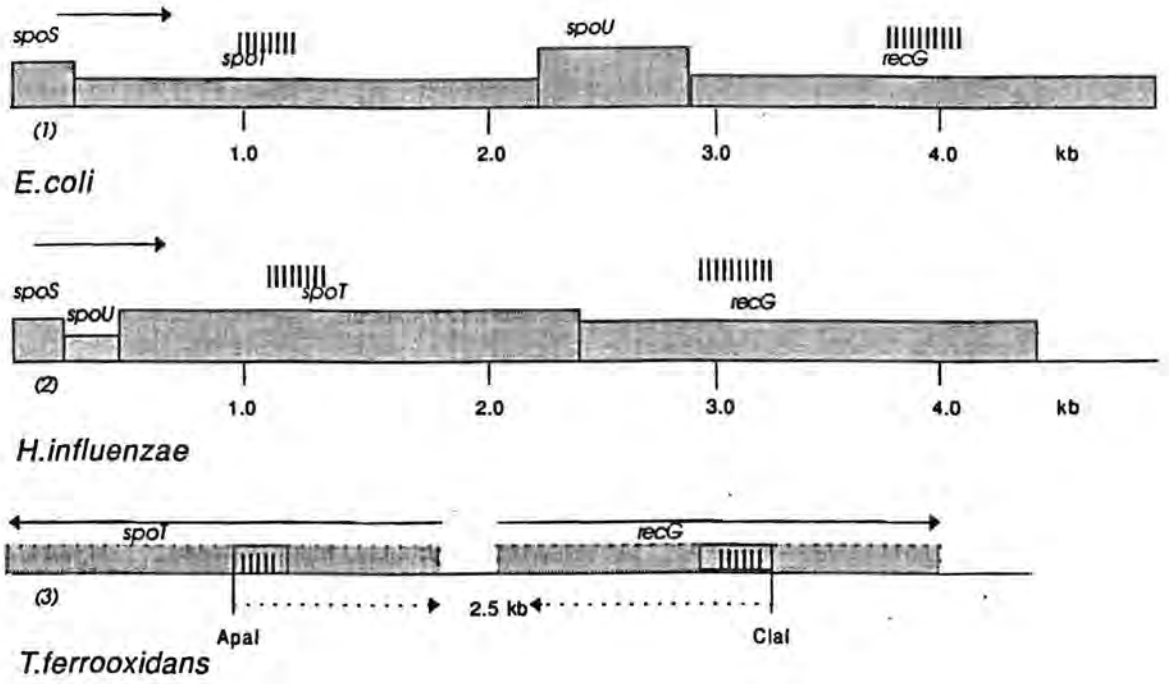


Fig. 4.19. Comparison of the *spo* operons of (1) *E.coli* and (2) *H.influenzae* and the probable structure of the *spo* operon of *T.ferrooxidans* (3). Note the size ot the *spoU* gene in *E.coli* as compared to that of *H.influenzae*. Areas with bars indicate the respective regions on the both operons 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 Tn*7*. 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 Tn*7*-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 (Tn*5468*) 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 Tn*7*-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 Tn*7*-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 Tn*7* 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 Tn*5468* probe were of the same size implies that, this chromosomal region is highly conserved within the *T.ferrooxidans* strains studied.

A 3.5 kb *Bam*HI-*Bam*HI 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 *pur*F-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 Tn*7* (Fig. 4.3). The region beyond the *Bam*HI-*Bam*HI 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 Tn*7* 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 Tn*7* 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 Tn*5468* (before it was known that it is probably non-functional). The degree of similarity between Tn*7* and Tn*5468* suggests that they originated from a common ancestor. Since *T.ferrooxidans* only grows in an acidic inorganic

environment, one would predict that Tn*5468* has evolved in a very different environment to Tn*7*. For example, Tn*7* 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 Tn*5468* appears to have been truncated as it might have provided an insight into the structure of the common ancestor of Tn*7* and Tn*5468*.

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 Tn*7* insertion at *att*Tn*7*. The arrangement of genes in such an *E.coli* strain is *unc_glmU_glmS*_Tn*7*, which is identical to that of *T.ferrooxidans*, *unc_glmU_glmS*_(Tn*7*-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 Tn*5468* had been inserted into the common ancestor before *T.ferrooxidans* and *T.thiooxidans* diverged, one might have expected Tn*5468* to be present in the *T.thiooxidans* strains examined. A plausible reason for the absence of Tn*5468* in *T.thioooxidans* could be that these two organisms diverged before *T.ferrooxidans* acquired Tn*5468*. Nevertheless, the Tn*7*-like transposon must have become inserted into *T.ferrooxidans* a long time ago because the three strains with Tn*7*-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 glucoamamine + 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 KoAc

2 M HoAc

Dissolve 29.4 g potassium acetate in 70 ml H_2O. 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 H_2O, 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, 10mM EDTA, 100 μg RNase /ml, pH 8.0

Store at 4 °C

S2.

200 mM NaOH, 1% SDS. Strorage: 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 H_3PO_4 to pH 6.3; store at room temp.

N3

100 mM Tris, 15% ethanol and 1150 mM KCl adjusted with H_3PO_4 to ph 6.3; store at room temp.

N5.

100 mM Tris, 15% ethanol and 1000 mM KCl adjusted with H_3PO_4 to pH 8.5; store at room temp.

Competent cells preparation buffers.

Stock solutions:

i) TFB-1 (100 mM RbCl, 50 mM $MnCl_2$), 30mM KOAc, 10mM $CaCl_2$, 15% glycerol).

For 50 ml:

1 mM RbCl (Sigma)	5.0 ml

$MnCl_2.4H_2O$ (Sigma, tetrahydrate) 0.495 g

KOAc (BDH) Analar) 0.147 g

750 mM $CaCl_2.2H_2O$ (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_2O 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_2O and filter sterilize.

Southern blotting and Hybridization solutions.

20 x SSC

175.3g NaCl + 88.29 sodium citrate in 800 ml H_2O

pH adjusted to 7.0 with 10 N NaOH. Distilled water added to 1000 ml.

Prehybridization solutions.

	stock solution	
5 x SSC	20 X	25.0 ml
5 % Elite Milk powder	-	4.0 g
0.1 N-laurylsarcosine	10 %	1.0 ml
0.02 % SDS	10 %	0.2 ml

Water - 69.0 ml

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

Buffer 1:

Maleic acid	11.6 g
NaCl	8.7 9
H_2O	to 1 litre.

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

Wash buffer:

Tween 20	1.0 ml
Buffer 1	199.0 ml

Buffer 2:

Elite milk powder	8.0 g
Buffer 1	190.0 ml

DEA Buffer:

Diethanolamine	193.0 µl
H_2O	19.7 ml
1 M $MgCl_2$	100.0 µl

pH to 10 with conc. HCl (± 15 µl).

Washes

	20 x SSC	10 % SDS
A (2 x SSC, 0.1 % SDS)	10.0 ml	1.0 ml
B (0.1 x SSC, 0.1 % SDS)	0.5 ml	1.0 ml

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

Stop Buffer

Bromophenol Blue	0.25 g
Xylene cyanol	0.25 g
Ficoll type 400	15.0 g

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 °C.

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-galactopyronoside)

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: *end*A1, *thi*, *rps*L, *sbc*B15, *hsd*R4, Δ(*lac-proAB*), (F', *tra*D36, *pro*AB, lacI^{q}ZΔM15).

PHENOTYPE: thi$^-$, **str^R**, lac$^-$, pro$^-$.

Reference: Gene, 33 (1985) 103-119.

E.coli JM109

GENOTYPES: *end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17(r_{K-},m_{K+}), *rel*A1, sup*E44*, Δ(*lac-proAB*), (F',*tra*D36, *proAB*, *lac*I^{q}ZΔM15).

PHENOTYPE: thi-, **nal^R**, lac$^-$, pro$^-$

Reference: Promega catalog.

***E.coli* CGSC5392**

Thr-1, *ara*-14, *leu*B6, DE (*gpt-pro*A), *lac*Y1, *tsx*-33, *qsr*'-0, *gln*V44 (AS), *gal*K2, LAM-, *rac*-0, *his*G4 (OC), *rfb*D1, *mgl*51, *rsp*L31 (Str), *kdg*K51, *xyl*A5, *mtl*-1, *glmS*1, *arg*E3 (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	3μ
3-7 kb	4μ
7-10 kb	6μ

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

REFERENCES.

Abrahams, J. P., Lutter, R., Todd, R. J., van Raaij, M. J., Leslie, A. G. W. and Walker, J. E. 1993. Inherent asymmetry of the structure of F_1-ATPase from bovine heart mitochondria at 65 Å resolution, EMBO J. **12**:1775-1780.

Abrahams, J. P., Leslie, A. G. W., Lutter, R. and Walker J. E. 1994. Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria, Nature **370**:621-628

Adzumah, K. and Mizuuchi, K. 1988. Target immunity of Mu transposition reflects a differential distribution of Mu B protein. Cell **53**:257-266.

Akey, C. W., Crepeau, R. H., Dunn, S. D., McCarty, R. E. and Edelstein, S. J. 1983. Electron microscopy of single molecules and crystals of F_1-ATPases. EMBO J. **2**:1409-1415.

Allet, B. 1979. Mu insertion duplicates a 5 bp sequence at the host inserted site. Cell **16**:123-129.

Amzel, L. M. and Pedersen, P. L. 1983. Proton ATP-ases: structure and mechanism. Ann. Rev. Biochem. **52**:801-824.

Andersson, L., Mac Neela, J. and Wolfenden, R. 1985. Use of secondary isotope effects and varying pH to investigate mode of binding of inhibitory amino aldehydes by leucine aminopeptidase. Biochem **24**:330-333.

Andrews, G. F., Dugan, R. P. and Stevens, C. J. 1992. Combining physical and bacterial treatment for removing pyritic sulfur from coal. In Processing and Utilization of High-sulfur coals IV. Dugan, P. R., D. Quigley and Y. Attia (eds), p. 515, Elsevier, New York.

Andrulis I. L., Chen, J. and Ray, P. N. 1987. Isolation of human cDNAs for asparagine synthetase and expression in Jansen rat sarcoama cells. Mol. Cell Biol. **7**:2435-2443.

Andruszkiewicz, R., Milewsky, S., Zieniawa, T. and Borowski, E. 1990. Anticandidal properties of N^3 -(4-methoxy-fumaroyl)-L-2, 3-diamino propanoic acid oligopeptides. J. Med. Chem. **33**:132-135.

Arciszewska, L. K., Drake, D. and Craig, N.L. 1989. Transposon Tn*7* cis-acting sequences in transposition and transposition immunity. J. Mol. Biol. **207**:35-42

Bachmann, B. J. 1983. Linkage map of *Escherichia coli* K-12 edition 7. Microbiol. Rev. **47**:180-230

Badet, B., Plumbridge, J. A., Cochet, D., Souza, J. M., Altamirano, M. M. and Calcagno, M. L. 1988. Cordinated regulation of amino sugar-synthesizing and degrading enzyme. J. Bacteriol. **175**:4951-4956

Badet, B., Vermoote, P. V. and Le Goffic, F. 1993. Glucosamine synthetase from *E.coli*; kinteic mechanism and inhibition by N^3-fumaroyl-2, 3-diaminopropionic derivatives. Biochem. **27**:2282-2287.

Badet, B., Vermoote, P., Haumont, P.Y., Lederer, F. and Le-Goffic, F. 1987. Glucosamine synthetase from *Escherichia coli*: purification, properties and glutamine-utilizing site location. Biochemistry **26**:1940-1948.

Badet, B., Inagaki, K., Soda, K. and Walsh, C. T. 1987. Time dependent inhibition of *Bacillus stearothermophilus* alanine racemase by (1-aminoethyl) phosphonate isomers by isomerization to non covalent enzyme-(1-aminoethyl) phosphonate complexes. Biochem. **25**:3275-3282.

Badet-Denisot, M. A. and Badet, B. 1992: Chemical modification of glucoseamine-6-phosphate synthetase by diethylpyrocarbonates: evidence of histidine requirement for enzymatic activity. Arch. Biochem Biophys, **292**:475-478.

Bainton, R., Gamas, P. and Craig, N. L. 1991. Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell **65**:805-816.

Barry, G. 1986. Permanent insertion of foreign genes into the chromosomes of soil bacteria. Bio/Technology **4**:446-449.

Barth, P. T., Datta, N., Hedges, R. W. and Grinter, N. S. 1976. Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. J. Bacteriol. **125**:800-810

Bates, C. J., Adams, W. R. and Handschumacher, R. R. 1968. Control of the formation of uridine diphospho-N-acetyl hexoseamine and glycoprotein synthesis in rat liver. J. Biol. Chem. **241**:1705-1712.

Benjamin, H. W. and Klecker, N. 1989. Intramolecular transposition of Tn*10*. Cell **59**:373-383.

Bennet, R. L. and Malamy, M. H. 1970. Arsenate resistant mutants of *Escgerichia coli* and phosphate transport. Biochem. Biophys. Res. Commun. **40**:496-503.

Benneth, J. C. and Tributsch, H. 1978. Bacterial leaching patterns on pyrite crystal surfaces. J. Bacteriol. **134**:310-317

Berg, D. E., Davies, J., Allet, B. and Rochaix, J. D. 1975. Transposition of R factor genes to bacteriophage lambda. Proc. Natl. Acad. Sci USA. **72**:3268-3632.

Berg, D. E. and Drummond, M.1978. Absence of DNA sequences homologous to transposable element Tn5 (Kan) in the chromosome of *E.coli* K-12. J. Bcateriol. **136**:419-422.

Birkenhager, R., Hoppert, M., Deckers-Hebestriet, G. Mayer, F. and Altendorf K. 1995. The F_0 complex of the *E.coli* ATP synthase: investigation by electron imaging and immunoelectron microscopy. Eur. J. Biochem. **230**:58-67

Bjϕrbaek, C., Foërsom, V. and Michelsen, O. 1990. The transmembrane topology of the *a* subunit from ATPase in *Escherichia coli* analysed by PhoA protein fusions, FEBS lett. **260**:31-34.

Boekemia, E. J., Berden, J.A. and van Heel, M.G. 1986. sructure of mitochondrial F_1-ATPase studied by electron microscopy and image processing. Biochim. Biophys. Acta. **851**:353-360.

Bolton, E., Glynn, P. and O'Gara, F. 1984. Site specific transposition of Tn*7* into a *Rhizobium meliloti* megaplasmid. Mol. Gen. Genet. **193**:153-157.

Boos, W. 1974. Bacterial transport. Annu. Rev. Biochem. **43**:123-146.

Boursaux-Eude, C., Girons, I. S. and Zuerner, R. 1995. IS1500, an IS*3*-like element from *Leptospira interrogans*. Microbiol. **141**:2165-2173.

Boyer, P. D. 1993. The binding change mechanism for ATP synthase:-some probabilities and possibilities. Biochim. Biophys. Acta. **1140**:215-250.

Brachet, P., Eisen, H. and Rambach, A. 1970. Mutations in coliphage lambda affecting the expression of replicative functions O and P. Mol. Gen. Genet. **108**:266-276.

Brink, J., Boekemia, E. J. and van Bruggen, E. F. 1987. The structure of NADH: ubiquitone oxidoreductase fron beef heart mitochondria. Crystals containing an octameric arrangement of iron-sulphur protein fragments. Eur. J. Biochem. **166**:287-294.

Brock T. D. and Gustafson, J. 1976. Ferric iron reduction by sulphur- and iron-oxidizing bacteria. Appl Environ Microbiol, **32**: 567-571.

Brown, L. D., Dennehy, M. E. and Rawlings, D. E. 1994. The F_1 genes of the F_1F_0 ATP synthase from the acidophilic bacterium *Thiobacillus ferrooxidans* complement *Escherichia coli* F_1 *unc* mutants. FEMS Microbiol Lett. **122**:19-25.

Buchanan, J. M. 1973. Adv. The Amidotransferases. Enzymol. **39**:91-183.

Buckley, A. N. and Woods, R. D. 1987. The surface oxidation of pyrite. Applied Surface Science, **27**:437-452.

Caruso, M. and Shapiro, J. 1982. Interactions of Tn*7* and temperate phage F116L of *Pseudomonas aeruginosa*. Mol. Gen. Genet. **188**:292-298.

Chmara, H., and H. Zähner. 1984. The inactivation of glucosamine synthetase from bacteria by anticapsin, the C-terminal epoxyamino acid of the antibiotic tetaine. Biochimica et Biophysica Acta. **787**:45-52.

Claassen, R. 1993. Mineralogical controls on the bacterial oxidation of refractory Barberton gold ores. FEMS Microbiol. Lett. **11**:197-206.

Cohen, S. N. 1976. Transposable genetic elements and plasmid evolution. Nature **263**:731-738.

Corbet, C. M. and W. J. Ingledew. 1987. Is $Fe^{3+/2+}$ cycling an intermediate in sulphur oxidation by Fe^{2+} -grown *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. **60**:29-33.

Couillard, D. and Mercier, G. 1992. Metallurgical residue for solubilization of metals from sewage sludge. J. Environ. Engng. **118**(5):808-813.

Cox, G. B., Fimmel, A. L. Gibson, F and Hatch, L. 1986. The mechanism of ATP synthase: a reassessment of the functions of the *b* and *a* subunits, Biochim. Biophys. Acta. **849**:62-69.

Craig, N. L. 1995. Unity in transposition. Science **270**:253-254.

Craig, N. and Gamas, P. 1992. Purification and characterization of TnsC, a Tn*7* transposition protein that binds ATP and DNA. Nucl. Acids Res. **20**:2525-2532.

Craig, N.L. 1991. Tn*7*: A target site-specific transposon. Mol. Microbiol. **5**:493-501.

Cross, R. L., Duncan, T. M., Bulygin, V. V., Zhou, Y. and Hutcheon, M. L. 1985. Rotation of the subunits during catalysis by *E.coli* F_1-ATPase. Proc. Natl. Acad Sci. **92**:10964-10968.

Cullis, C. A. 1990. DNA rearrangements in response to environmental stress. Advances in Genetics **28**:73-97.

Deckers-Hebestriet, G., Simoni, R. D. and Altendorf, K. 1992. Influence of subunit-specific antibodies on the activity of the F_o complex of the ATP synthase of *Escherichia coli*. Effect of subunit *b*-specific polyclonal antibodies, J. Biol. Chem. **267**:12364-12369.

Denisot, M. A., Le Goffic, F. and Badet, B. 1991. Glucosamine- 6P synthase yields two proteins on limited proteolysis. Arch. Biochem. Biophys. **288**:225-230

Dobrogosz, W. J. 1968. N-acetylglucosamine assimilation in *Escherichia coli* and its relation to catabolite repression. J. Bacteriol. **95**:585-591.

Doublet, P., J. van Heijenoort, and D. Mengin-Lecreulx. 1993. The *murI* gene of *Escherichia coli* is an essential gene that encodes glutamase racemase activity. J. Bacteriol. **175**:2970-2979.

Doublet, P., J. van Heijenoort, and D. Mengin-Lecreulx. 1992. Identification of the *E.coli murI* gene, which is required for the biosynthesis of D-glutamic acid, a specific component of bacterial peptidoglycan. J. Bacteriol. **174**:5772-5779.

Echols, H., Garren, A., Garen, S. and Torriani, A. 1961. Genetic control of repression of alkaline phosphatase in *E.coli*. J. Mol. Biol. **3**:425-438

Einchinger, D. J. and Boeke, J. D. 1990. A specific terminal structure is required for T*y1* transposition. Genes Develop. **4**:324-330.

Ely, B. 1982. Transposition of Tn*7* occurs at a single site on the *Caulobacter crescentus* chromosome. J. Bacteriol. **151**:1056-1058.

Fillingame, R. H. 1990. Molecular mechanisms of ATP synthesis by type H^+ -transporting ATP synthases, p. 345-391. *In* T. A. Krulwich (ed), The bacterial, vol. 12. Bacterial energetics. Academic Press Ltd., London.

Fillingame, R. H. 1992a. H^+ transport and coupling by the F_0 sector of ATP synthase: insights into the molecular mechanism of function, J. Bioenerg. Biomembr. **24**:485-491.

Fillingame, R. H. 1992b. Subunit *c* of F_1F_0 ATP synthase: structure and role in transmembrane energy transduction, Biochim. Biophys. Acta. **1101**:240-243.

Finbow, M. E., Eliopoulos, E. E., Jackson, P. J., Keen, J.N., Meagher, L., Thompson, P. and Findlay, J. B. C. 1992. Structure of a 16 kDa integral membrane protein that has identity to the putative proton channel of the vacuolar H^+ -ATPase, Protein Eng. **5**:7-15.

Fling M., Kopf, J. and Richards, C. 1985. Nucleotide sequence of the transposon Tn*7* gene encoding the aminoglycoside-modifying enzyme, 3'(9)-O-nucleotidyltransferase. Nucleic Acids Res. **13**:7095-7106.

Fling, M. and C Richards. 1983. The nucleotide sequence of the the trimethoprim-resistant dihydrofolate reductase gene harbored by Tn*7*. Nucleic Acids Res. **11**:5147-5158.

Flores, C. C., Qadri, M. I., Lichtenstein C. P. 1990. DNA sequence anlysis of five genes, *tnsA, tnsB, tnsC tnsD* and *tnsE* required for the Tn*7* transposition. Nucleic Acids Res. **18**:901-911.

Foster, D. L. and Fillingame, R. H. 1982. Stoichiometry of subunits in the H^+-ATPase complex of *Escherichia coli*, J. Biol. Chem. **257**:2009-2015.

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_1 to F_0, J. Biol. Chem. **269**:7532-7537.

Friedl, P., Hoppe, J., Gunsalus, R. P., Michelsen O., von Meyenburg, K. and Schairer, H. U. 1983. Membrane integration and function of the three F_0 subunits of the ATP-synthase of *Escherichia coli* K12, EMBO J. **2**:99-103.

Frisa, P. S. and Sonneborn, D. R. 1982. Developmentally regulated interconversion between end product inhibitable and non-inhibitable forms of a first pathway-specific enzyme activity can be mimicked in *vitro* by dephosphorylation reactions. Proc. Natl. Acad. Sci. USA **79**:6289-6293.

Fujiwara, T. and Mizuuchi, K. 1988. Retroviral DNA integration: Structure of an integration intermediate. Cell **54**:497-504.

Galas, D. J. and Chandler, M. 1989. Bacterial insertion sequences. In Mobile DNA. Berg, D. E., and Howe, M. M. (eds). Washington, D. C. American Society for Microbiology, pp. 109-162.

Gay, N. J., Tybulewicz, V. L.J., Walker J.E. 1986. Insertion of transposon Tn*7* into the *E.coli glmS* transcriptional terminator. Biochem. J. **234**:111-117.

Gentry, D. R. and Cashel, M. 1995. Cellular localization of the *Escherichia coli* SpoT protein. J. Bacteriol. **177**: 3890-3893.

Gentry, D. R., Xiao, H., Burgess, R. R. and Cashel, M. 1991. The omega subunit of *E.coli* K-12 RNA polymerase is not required for stringent RNA control in *vivo*. J. Bacteriol. **173**: 3901-3903.

Girvin, M. E. and Fillingame, R. H. 1993. Helical structure and folding of subunit *c* of F_1F_0 ATP synthase: ^1H NMR resonace assignments and NOE analysis, Biochemistry **32**:12167-12177.

Gogol, E. P., Lucken, U., Bork, T. and Capaldi, R. A. 1989. Molecular architecture of *Escherichia coli* F_1 Adenosinetriphosphatase. Biochemistry **28**:4709-4716.

Gogol, E. P., Aggeler, R., Sagermann, M. and Capaldi, R. A. 1989. Cryoelectronic Microscopy of Escherichia coli F_1 Adenosinetriphosphatase Decorated with Monoclonal Antibodies to Individual Subunits of the complex. Biochemistry **28**:4717-4724.

Golinelli-Pimpaneaux, B. and Badet, B. 1991. Possible involvement of Lys603 from *E.coli* glucosamoine-6-phosphate synthase in the binding of its substrate fructose-6-phosphate. Eur. J. Biochem. **201**:175-182.

Gottesman, M. M. and Rosner, J. L. 1975. Acquisition of a determinant of chloramphenicol resistance by coliphage lambda. Proc. Natl. Acad. Sci USA. **72**:5041-5045.

Grunstein, M., and Hogness, D. S. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci USA. **72**:3961-3965.

Hauer, B. and Shapiro, J. A. 1984. Control of Tn*7* transposition. Mol. Gen. Genet. **194**:149-158.

Hedges, R. W., and Jacob, A. E. 1974. Transposition of ampicillin resistance from RP4 to other replicons. Mol. Gen. Genet. **132**:31-40.

Heffron, F., Reubens, C, and Falkow, S. 1975. Translocation of a plasmid DNA sequence which mediates ampicillin resistance: molecular nature and specificity of insertion. Proc. Natl. Acad. Sci USA. **72**:3623-3627.

Heinikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351-369.

Hermolin, J., Gallant, J. Fillingame, R. H. 1983. Topology, organization and function of the psi subunit in the F_0 sector of the H+ -ATPase of *E.coli*. J. Biol. Chem. **258**:14550-14555.

Hermolin, J and Fillingame, R. H. 1989. Assembly of F_0 sector of *Escherichia coli* H^+ ATP synthase. Interdepenndence of subunit insertion into the membrane. J. Biol. Chem. **270**:2815-2817.

Hernalsteens, J. P., van Montagu, M., Holsters, Zambryski, P., Depicker, A., De Beuckeleer, M., Engler, G., Lemmers, M., Willmitzer and Schell J. M. 1980. The interaction of Agrobacterium T1-plasmid DNA and plant cells. Proc Soc Lond [Biol] **210**:351-365.

Hirsch, H., Saedler, H. and Starlinger, P. 1972. Insertion mutations in the control region of the galactose operon of *E.coli*: Physical characterization of the mutants. Mol. Gen. Genet. **115**:266-276.

Holmes D. S. and R. Ul Haq. 1990. Adaptation of *Thiobacillus ferrooxidans* for industrial applications, p115-127. *In* J. Salley, R. G. L. McCready, and P. L. Wichlacz (ed.), Biohydrometallurgy-1989. CANMET, Ottawa, Canada.

Höltje, J. V., and U. Schwarz. 1985. Biosynthesis and growth of the murein sacculus, p 77-119. *In* N. Nanninga (ed.), Molecular cytology of *Escherichia coli*. Academic Press, Inc., New York.

Holzenburg, A., Jones, P. C., Franklin, T., Padi, T., Heimburg, D., Marsh, Findlay, J. B. and Finbow, M. E. 1993. Evidence for a common structure for a class of membrane channels. Eur. J. Biochem. **213**:21-30.

Hoppe, J. and Sebald, W. 1986. Topological studies suggest that the pathway of the protons through F_0 is provided by amino acid residues accessible from the lipid phase, Biochimie. (Paris) **68**:427-434.

Hu, S., Otsubo, H., Davidson, N. and Saedler, H. 1975. Electron microscope heteroduplex studies of sequence relations among bacterial plasmids: identification and mapping of the insertion sequences *IS1* and *IS2* in F and R plasmids. J. Bacteriol. **22**:762-775.

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

Ish-Horowicz, D. and Burke J.F. 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989-2998.

Johnston, B., Clennell, A. M. and Rawlings, D. E. 1995. Structure and function of Tn5467, a Tn*21*-like transposon located on the *T.ferrooxidans* broad host range plasmid pTF-FC2. Appl. Envir. Micro. **61**:4223-4229.

Kahmann, R. and Kamp, D. 1979. Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature **280**:247-250.

Kahn, K. and Schaefer, M. R. 1995. Characterization of transposon 5469 from cyanobacterium *Fremyella diplosiphon*. J. Bacteriol, **177**:7026-7032.

Karavaiko, G. I., Golovacheva, R. S., Pivovarova, T. A., Tzaplina, I. A. and Vartanjan, N. S. 1988. Thermophilic bacteria of the genus *Sulfobacillus*, page 29-41. In Biohydrometallurgy-87. Norris, P. R. and Kell, D. P. (Eds.) Science and Technology Letters, Kew, Surrey, UK.

Kennedy, J. F., Barker, S. A. and Humpphreys, J. D. 1976. Microbial cells living immobilised on metal hydroxides. Nature **261**:242-244.

Koonin E. V. and Rudd, K. E. 1993. SpoU protein of *Escherichia coli* belongs to a new family of putative rRNA methylases. Nucleic Acids Res. **21**:5519.

Kopecko, J. and Cohen, N. 1975. site specific *rec*A independent recombination between bacterial plasmids: involvement of palindromes at the recomnation loci. Proc. Natl. Acad. Sci USA. **72**:1373-1377.

Kornfield, R. 1967. Studies on L-glutamine D-fructose 6- phosphate amidotransferase. J. Biol. Chem. **242**:3135-3141.

Krumholz, L. R., Esser, U. and Simoni, R.D. 1989. Nucleotide sequence of the *unc* operon of *Vibrio alginolyticus*. Nucleic Acids Res. **17**:7993-7994.

Kubo, K. and Craig, N. 1990. Bacterial transposon Tn*7* utilizes two classes of target sites. J. Bacteriol. **172**:2774-2778.

Kucharczyk N., Denisot, M. A. Le Goffic, F. and Badet, B. 1990. Glucosamine-6-phosphate synthase from *E.coli*: determination of the mechanism of inactivation by N^3 fumaroyl-L-2-3 diaminoproprionic derivatives. Biochemistry **29**:3668-3676.

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

Lane, D.J., A. P. Harrison, Jnr., D. Stahl, B. Pace, S.J. Giovannoni, G.J. Olsen, and N. R. Pace. 1992. Evolutionary relationship among sulphur- and iron-oxidizing eubacteria. J. Bacteriol. **174**:267-278.

Lee, C. H., Bhagwhat, A. and Heffron, F. 1983. Identification of a transposon Tn*3* sequence required for transposition immunity. Natl. Acad. Sci. USA. **80**:6765-6769.

Lewis, M. J., Chang, J. A. and Somoni, R. D. 1990. A topological analysis of subunit *a* from *Escherichia coli* F_1F_0-ATP synthase predicts eight transmembrane segments, J. Biol. Chem. **265**:10541-10550.

Lichtenstein, C. P. and Brenner, S. 1982. Unique insertion site of Tn*7* in the *E.coli* chromosome. Nature (London) **297**:601-603.

Lichtenstein, C. P. and Brenner, S. 1981. Site-specific properties of transposition to the *E.coli* chromosome. Mol. Gen. Genet. **183**:380-387.

Liu, X., Petersson, S. and Sandstrom, A. Mesophilic versus moderate thermophilic bioleaching. Biohydrometallurgy Technologies. Vol I. A. E. Torma, J. E. Wey, and Lakshmanan, V. I. (ed.) pp 29-38.

Lizama, H. M. and Sankey, B. M. 1993. Oxidation of H_2S by *Thiobacillus thiooxidans* is inhibited by substrate and methane. Biohydrometallurgy Technologies. Vol II. A. E. Torma, J. E. Wey, and C. L. Brierley (ed). pp 339-364.

Lundgren, D. G. and Silver, M. 1980. Ore leaching by bacteria. Ann. Rev. Microbiol. **34**:263-268.

Makino, K., Amemura, M., Shinagawa, H., Kobayashi, A. and Nakata, A. 1985. Sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. J. Mol. Biol. **184**:231-240.

Makino, K., Shinagawa, H., Amemura, M., Kimura, S., Nakata, A. and Ishihama, A. 1988. Regulation of the phosphate regulon of *E.coli*. Activation of *pstS* transcription by the PhoB protein in vitro. J. Mol. Biol. **203**:85-95.

Makino, K., Shinagawa, H., Amemura, M., Yamada, M. and Nakata, A. 1990. Signal transduction in the phosphate regulon of the *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. J. Mol. Biol. **210**:551-559.

Malamy, M. H. 1966. Frameshift mutations in the lactose operon of *Escherichia coli*. Cold Spring Harb Symp Quant Biol. **31**:189-201.

Malamy, M. H. 1972. Electron microscopy of polar insertions in the lac operon of *Escherichia coli*. Mol. Gen. Genet. **119**:207-222.

Malamy, M. H. and Bennett, R. L. 1970. Arsenate resistance mutants of *E.coli* and phosphate transport. Biochem. Biophys. Res. Commun. **40**:496-503.

Maniatis T., Fritsch E.F., Sambrook J. 1982 Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. New York.

McKnight G. L., S. L. Mudri, S.H. Mathewest, R. R.Traxinger, S. Marshall, P. O. Sheppard, and P. J. O'Hara. 1990. Molecular cloning, cDNA synthesis, and bacterial expression of human glutamine:fructose-6-phosphate amidotransferase. J. Biol. Chem. **267**:25208-25212.

Medveczky, N. and H. Rosenburg. 1970. The phosphate-binding protein of *Escherichia coli*. Biochim. Biophys. Acta. **211**: 158-168.

Mei, B. and Zalkin, H. 1989. A Cysteine-Histidine-Aspartate catalytic triad is involved in Glutamide Amide Transfer Function in *pur*F-type Glutamine amodotransferases. J. Biol. Chem. **264**:16613-16619.

Mengin-Lecreulx, D. and J. van Heijenoort. 1993. Identification of *glmU* gene encoding N-acetylglucosamine 1-phosphate uridyltransferase in *Escherichia coli*. J. Bacteriol. **175**:6150-6157.

Michaelis, G., Douglass, S., Tsai, M. J. and Criddle M. J. 1970. Mitochondrial DNA and suppressiveness of petite mutants in *Saccharomyces cerevisiae*. Biochem. Genet. **5**:487-495.

Michaelis, G., Saedler, H., Venkov, P. and Starlinger, P. 1969. Two insertions in the galactose operon having different sizes but homologous DNA sequences. Mol. Gen. Genet. **104**:371-377.

Miller, J. H. and Calos M. P. 1980. Transposable elements. Cell **20**:579-595.

Miller, M. J. Oldenburg, M. and Fillingame, R. H. 1990. The essential carboxyl group of in subunit *c* of the F_1F_0 ATP synthase can be moved and H(+)-translocating function retained. Proc. Natl. Acad. Sci USA. **87**:4900-4904

Mitcell, P. 1966. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. Biol. Rev. **41**:455-502 (1966).

Mizuuchi, K. 1984. Mechanism of transposition of bacteriophage Mu. Polarity of the strand transfer reaction at the initiation of the transposon. Cell **39**:395-404.

Mustin, C., Ph. de Donato and Berthelin, J. 1993. Surface oxidized species, a key factor in the study of bioleaching processes. Biohydrometallurgical Technologies. In A. E. Torma, J. E. Wey and V. I. Lakshmanan ed. 1993. Vol 1. pp. 175-184.

Nakata, A., Ishino, Y., Shinagawa, H., Makino, K. and Amemura, M. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *E.coli* and identification of the gene product. J. Bacteriol. **169**:5429-5433.

Ochi, K. 1989. The tsr gene-coding pIJ702 prevents thiopeptin from inhibiting ppGpp synthesis in *Streptomyces lividans*. FEMS Microbiol Lett. **52**:219-223.

Ogasawara, N. and Yoshikawa, H. 1992. Genes and their organization in the replicon of the bacterial chromosome. Mol. Microbiol. **6**(5):629-634.

Ohtsubo, E., Hu, S., Davidson, N. and Saedler H. 1975. Electron microscope heteroduplex studies of sequence relations among bacterial plasmids: identification and mapping of the insertion sequences IS1 and IS2 in F and R plasmids. J. Bacteriol. **122**:746-775.

Olson, G. J. 1994. Microbial oxidation of gold ores and gold bioleaching. FEMS Microbiol. Lett. **119**:1-6.

Orle, K. A. and Craig, N. L. 1991. Identification of transposition prroteins encoded by the bacterial transposon Tn*7* [corrected and republished article originally printed in Gene 1990, Nov 30: **96**(1):1-7] Gene **104**:25-31.

Ouellette, M., and Roy, P. Homology of ORFs from Tn2603 and from R46 to site specific recombinases. 1987. Nucl. Acids Res. **15**:10055-10059.

Park, J. T. 1987. Murein synthesis, p663-671. *In* F.C. Neidhardt, J. L. Ingraham, K. B. Louw, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biol, vol. 1. American Society for Microbiology, Washington, D.C.

Perlin, D. S. and Senior, A. E. 1985. Functional effect and cross-reactivity of antibody to purified subunit b (uncF protein) of *Escherichia coli* proton-ATPase. Arch. Biochem. Biophys. **236**:603-611.

Plumbridge, J. A., O. Cochet, Souza, J. M., Altramirano, M. M., Calcagno, M. L. and Badet, B. 1993. Coordinated regulation of amino sugar-synthesizing and -degrading enzymes in *Escherichia coli* K-12. J. Bacteriol. **175**(16):4951-4956.

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

Qadri, M. I., Flores, C. C., Davis, J. and Lichtenstein, C. 1989. Genetic analysis of *att*Tn*7*, the transposon Tn*7* attachment site in *Escherichia coli* using a novel M13-based transduction assay. J. Mol Biol. **207**:85-98.

Radstrom, P., Skold, O., Swedberg, J., Roy, P. H. and Sundstrom, L. 1994. Tn5090 of plasmid R751 which carries integron is related to Tn*7*, Mu and retroelements. J. Bacteriol **176**:3257-3268.

Raetz, C. R. H. 1987. Structure and biosynthesis of lipid A in *Escherichia coli*, pp. 498-503. In F. C. Niedhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington., D.C.

Rao, N. N. and Torriani, A. 1990. Molecular aspects of phosphate transport in *Escherichia coli*. Mol Microbiol. **4**(7):1083-1090.

Rawlings, D.E., D. R. Woods and N.P. Mjoli. 1991. The cloning and structre of genes from the autotrophic biomining bacterium *Thiobacillus ferrooxidans*, p. 215-237. In. P.J. Greenaway (ed.). Advances in gene technology, vol. 2. JAI Press, London.

Richet, E and O. Raibaud. 1989. MalT, the regulatory protein of the *Escherichia coli* maltose system is an ATP-dependent transcriptional activator. EMBO, J. **8**:981-987.

Rogers, M., Ekaterrinaki, N., Nimmo, E. and Sherratt, D. 1986. Analysis of Tn*7* transposition. Mol. Gen. Genet. **205**:550-556.

Ronson, C. W., Nixon, B. T., Albright, L. M. anf Ausubel, F. M. 1987. *Rhizobium meliloti ntrA (rpoN)* gene is required for diverse metabolic functions. J. Bacteriol. **169**:2424-2431.

Rosenburg, H., Gerdes, R. G., and Chegwidden, K. 1977. Two systems for the uptake of phosphate in *E.coli*. J.Bacteriol. **131**:505-511.

Rosenburg, H. 1987. *In* ion transport in Prokaryotes. Rosen, B. P. and Silver, S. (eds). New York: Academic Press, pp. 205-248.

Ross, D. G., Swan, J. and N. Klecker. 1979. Physical structures of Tn*10*-promoted deletions and inversions: role of 1400 base pairs inverted repetitions. Cell **16**:721-731.

Saedler, H. and Starlinger, P. 1967. O degree mutation in the galactose operon in *E.coli*: II. Physiological characterization. Mol. Gen. Genet. **100**: 190-202.

Saedler, H., Jordan, E. and Starlinger P. 1968. OO and strong polar mutations in the gal operon are insertions. Mol. Genet. Genet. **102**:353-363.

Sambrook, J., Fritsch F. and Maniatis T. 1989. Molecular cloning: a laboratory manual; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sand, W., Gehrke, T., Hallmann, R., Rohde, K., Sobotke, B. and Wintzien, S. 1993. In-situ Bioleaching of metal sulphides: The importance of *Leptospirillum ferrooxidans*. Biohydromemetallurgical Technologies. VolI. pp 15-25.

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

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

Schneider, E and Altendorf, K. 1987. Bacterial adenosine 5'-triphosphate synthase (F_1F_0): purification and reconstitution of F_o complexes and biochemical and functional characterization of their subunits, Microbiol. Rev. **51**:477-497.

Schrader, J. A. and D. S. Holmes. 1988. Phenotypic switching of *Thiobacillus ferrooxidans*. J. Bacteriol. **170**:3915-3023.

Scordilis, G. E., Ree, H. and Lassie, T. G. 1987. Identification of transposable elements which activates gene expression in *Pseudomonas cepacia*. J. Bacteriol. **169**:8-13.

Sekine, Y., Eisaki, N. and Ohtsubo, E. 1996. Identification and characterization of the linear 1S3 molecules generated by staggered breaks. J. Biol. Chem. **271**:197-202.

Senior, A. E. 1990. The proton-translocating ATPase of *Escherichia coli*. Annu. Rev. Biophys. Chem. **19**:4-41.

Shapiro, J. A., and Adhya, S. L. 1969. The galactose operon of *E.coli* K-12. II. A deletion analysis of the operon structure and polarity. Genetics **62**:249-264.

Shapiro, J. A. 1969. Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. J. Mol. Biol. **40**:93-105.

Silverman, M. P. 1967. Mechanism of bacterial pyrite oxidation. J. Bacteriol. **94**:1046-1051.

Simonsen, C. C., Ellson, Y and Levinson, A. D. 1983. Identification of the type1 trimethoprim resistance reductase specified by the *E.coli* R-plasmid R43: comparison with procaryotic and eucaryotic dihydrofolate reductase. J. Bacteriol. **155**:1001-1008.

Smith, G. and Jones, P. 1986. Tn*7* transposition: a multigene process. Identification of a regulatory gene product. Nucl. Acids Res. **14**:7915-7927.

Sprague, G.F., Jr, Bell, R. M. and Cronan, J. E. 1975. A mutant of *E.coli* auxotrophic for organic phosphates: evidence for two defects in inorganic phosphate transport. Mol. Gen. Genet. **143**:71-77.

Starlinger, P. and Michaelis, G. 1968. Suppression and the sequential appearance of the galactose enzymes in a transferase amber mutant of *Escherichia coli*. Mol. Gen. Genet. **102**:367-369.

Starlinger, P. and Saedler, H. 1977. IS elements in microorganisms. Curr. Top Microbiol. Immunol **75**:111-152.

Steffens, K., Schneider, E., Herkernhoff, B., Schmid, R. and Altendorf, K. 1987a. F_0 portion of *Escherichia coli* ATP synthase Further resolution of trypsin-generated from subunit *b*, J. Biol. Chem. **262**:5866-5869.

Steffens, K., Di Giola, A., Deckers-hebestreit, G. and Altendorf, K. 1987b. The structural and functional relationship of ATP synthases (F_0F_1) from *E.coii* and the thermophilic bacterium PS3, J. Biol. Chem **262**:6334-6338.

Strominger, J. L., and M. S. Smith. 1959. Uridine diphosphoacetylglucosamine pyrophosphorylase. J. Biol. Chem. **234**:1822-1827.

Sundstrom, L. and O. Skold, 1990. The *dhfr*I trimethoprim resistance gene of Tn*7* can be found at specific sites in other genetic surroundings. Antimicrob Agents Chemother **34**:642-650.

Sundstrom, L., Roy, P. H. and Skold O. 1991. Site-specific insertion of three structural gene cassettes in Tn*7*. J. Bacteriol. **173**:3025-3028.

Surin, B. P., and Downie, J.A. 1988. Characterization of the *Rhizobium leguminosarum* genes nodLMN involved in efficient host-specific nodulation. Mol. Microbiol. **2**:173-183.

Surin, B. P., Dixon, N. E. and Rosenburg, H. 1986. Purification of PhoU protein, a negative regulator of the pho regulon of *E.coli* K-12. J. Bacteriol. **168**:631-635.

Surin, B. P., D. A. Jans, A. L. Fimmel, D.C. Shaw, G.B. Cox, and H. Rosenburg. 1984. Structural gene for the phosphate-repressible phosphate binding protein of *Escherichia coli* has its own promoter: complete nucleotide sequence of the *phoS* gene. J. Bacteriol. **157**:772-778.

Suzuki, I., Chang, C. W. and Takeuchi, T. L. 1994. Oxidation of organic sulphur compounds by Thiobacilli, ACS Symp. Series **550**:60-67.

Takeyama M., Hsu, S. Y., Noumi, T., Maeda., Ishibashi S. and Futai, M. 1988. Beta subunit of *E.coli* F_1-ATPase. An amino acid replacement within a conserved sequence G-X-X-X-X-G-K-T/S) of the nucleotide binding proteins. FEBS lett. **218**:222-226.

Thomson, J.A., M. Hendson, and R. M. Magnes. 1981. Mutagenesis by insertion of drug resistance transposon Tn*7* into a vibrio species. J. Bacteriol. **157**:772-778.

Tichy, R., Grotenhuis, J. T. C., Janssen, A., van Houten, R., Rulkens, W. H. and Lettinga, G. 1993. Application of the sulphur cycle for bioremediation of soils polluted with heavy metals. *In* Proc. Int. Conf. Contaminated Soil '93, ed. F. Arendt, G. J. Annokée, R. Bosman and W. J. van der Brink. Kluwer Academic Publishers, Dordrecht, pp 1461-1462.

Tiedge, H., Schäfer, G. and Mayer, F. 1983. An electron microscopic approach to the quaternary structure of mitochondrial F_1-ATPase, Eur. J. Biochem. **132**:37-45.

Tietze, E., Brevet, J. and Tschape, H. 1987. Relationship among the streptothricin resistance transposons Tn*1825* and Tn*1826* and the trimethoprim resistance transposon Tn*7*. Plasmid **18**:246-249.

Tso, J. Y., Zalkin, H., van Cleemput, M., Yanofsky, C., and Smith, J.M. 1982. Nucleotide sequence of *Escherichia coli pur*F and deduced amino acid sequence of glutamine phosphribosylpyrophosphate amidotransferase. J. Biol. Chem. **257**:3525-3531.

Tsuprun, V. L., Mesyanzhinova, I. V., Koslov, I. A. and Orlova, E.V. 1984. Structure of ATP-synthase studied by electron microscopy and image processing, FEBS Lett. **167**:285-289.

Turner, P., Barber, C. and Daniels, M. 1984. Behaviour of the transposons Tn5 and Tn*7* in *Xanthomonas campestris pv. campestris*. Mol. Gen. Genet. **193**:153-157.

Ullrich, J. and van Putten, P. M. 1995. Identification of the Gonococcal *glmU* gene encoding the enzyme N-acetylglucosamine 1-Phosphate Uridyltransferase involved in the synthesis of UDP-GlcNAc. J. of Bact. **177**:6902-6909.

Vaara, M. 1992. Eight bacterial proteins includin]g UDP-N-acetylglucosamine acyltransferase (LpxA) and three vother transferases of *Escherichia coli*, consist of a six-residue periodicity theme. FEMS Microbiol. Lett. **97**:249-254.

van der Steen, J. J. D., Doddema, H. J. and de Jong, G. 1992. Uitloging van zware metalen uit afvalstromen met behulp van thiobacilli (Removal of heavy metals from waste streams using thiobacilli). Ministry of Housing, Physical Planning, and Enviromental (VROM), Directoraat-Generaal Milieubeheer, Rapport 1992/10, The Hague.

Vermoote, P. 1988. Thesis. Université Paris VI, Paris, France.

Vignais, P. V., Lunard, J., Issartel, J. P. and Dupuis, A. 1985. Interaction between the oligomycin sensitivity protein (OSCP) and beef heart mitochondrial F_1-ATPase. 2. Identification of the interacting F_1 subunits by cross-linking. Biochem J. **24**:734-739.

Vik, S. B., and Dao, N.N. 1992. Prediction of transmembrane topology of F_0 proteins from *E.coli* F_1F_0 ATP synthase using variational and hydrophobic moment analyses, Biochim. Biophys. Acta **1140**:199-207.

von Meyenburg, K., B. B. Jφrgensen, J. Nielsen, and F. G. Hansen 1982. Promoters of the *atp* operon coding for the membrane-bound ATP synthase of *E.coli* mapped by Tn*10* insertion mutations. Mol. Gen. Genet. **188**:240-248.

von Meyenberg, K., and F. G. Hansen. 1980. The origin of replication, *oriC* of the *E.coli* chromosome: genes near *oriC* and construction of *oriC* deletion mutants. ICN-UCLA Symp. Mol.Cell.Biol. **19**:137-159.

Waddell, C. S and Craig, N. L. 1989. Tn**7** transposition: recognition of the *att*Tn*7* target sequence. Proc. Natl. Acad. Sci. USA **86**:3958-3962.

Waddell, C. S. and Craig, N. L. 1988. Tn*7* transposition: two transposition pathways directed by five Tn*7*-encoded genes. Genes Develop. **2**:137-147.

Walker, J. E., Gay, N. J., Saraste, M. and Eberle, A. N. 1984. DNA sequence around the *Escherichia coli unc* operon. Completion of the sequence of a 17 kilobase segment containing *asnA, oriC, unc, glmS* and *phoS*. Biocem. J. **224**:799-815.

Walker, J. E., and Collinson, I. R. 1994. the role of the stalk in the coupling mechanism of F_1F_0-ATPase. FEBS Lett. **346**:39-43.

Walker, J. E., Gay, N. J. and Tybulewicz, V. L. 1986. Insertion of transposon Tn*7* into the *Escherichia coli glmS* transcriptional terminator. Biochem J. **243**:111-117.

Wanner, B. L and McSharry, R. 1982. Phosphate-controlled gene expression in *Escherichia coli* K-12 using Mudl-directed *lacZ* fusions. J. Mol. Biol. **158**:347-363.

Weng, M., and Zalkin, H. 1987. Structural role for a conserved region in the CTP synthetase glutamide amide transfer domain. J. Bacteriol. **169**:3023-3028.

Wheatcroft, R. and S. Laberge. 1991. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence IS*Rm3*: similarity between the putative transposase encoded by IS*Rm3* and those encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IS*T2*. J. Bacteriol. **173**:2530-2538.

Whitby, M. C. and Lloyd, R. G. 1995. Branch migration of three-strand recombination intermediates by RecG, a possible pathway for securing exchanges initiated by 3'-tailed duplex DNA. EMBO J. **14**:3303-3310.

Wilkens, S. and Capaldi, R. A. 1994. Assymmetry and structural changes in ECF_1 examined by cryoelectronmicroscopy. Biol. Chem. Hoppe-Seyler **375**:43-51.

Willsky, G. R. and Malamy, M. H. 1974. The loss of the PhoS periplasmic protein leads to a change in the specificity of a constitutive inorganic phosphate transport system in *Escherichia coli*. Biochem. Biophys. Res. Commun. **60**:226-233.

Willsky, G. R. and Malamy, M. H. 1980. Characterization of two genetically separable inorganic phosphate transport systems in *Escherichia coli*. J. Bacteriol. **144**:356-365.

Winterburn, P. J. and and Phelps, C. F. 1973. in The enzyme of glutamine metabolism (Prusiner, S. and Stadtman, E. R., eds.), pp. 343-363, Academic Press, New York.

Winterburn, P. J., and Phelps, C. F. 1971. Studies on the control of hexosamine biosynthesis by glucosamine synthetase. Biochem. J. **121**:711-720.

Wolf-Watz, H. and Norquist, A. 1979. Deoxyribonucleic acid and outer membrane protein: binding to outer membrane protein involves a specific protein. J. Bacteriol. **140**:43-49.

Wolf-Watz, H. and Masters, M. 1979. Deoxyribonucleic acid and outer membrane: strains diploi for *oriC* region show elevated levels of deoxyribonucleic acid-binding protein and evidence for specific binding of the *oriC* region to outer membrane. J. Bacteriol. **140**:50-58.

Wolf-Watz, H. 1984. Affinity of two different regions of the chromosome to the outer membrane of *E.coli*. J. Bacteriol. **157**:968-970.

Wu, H. C., and T.C. Wu. 1971. Isolation and characterization of a glucosamine-requiring mutant of *Escherichia coli* K-12 defective in glucoamine-6-phosphate synthetase. J. Bacteriol. **105**:455-466.

Yamada, M., Makino, K., Shinagawa, H. and Nakata, A. 1990. Regulation of the phosphate regulon of *Escherichia coli*: properties of the *pho*R deletion mutants and subcellular localization of the PhoR protein. Mol. Gen. Genet. **220**:366-372.

Yates, J. R. and Holmes, D. S. 1987. Two families of repeated DNA sequences in *Thiobacillus ferrooxidans*. J. Bacteriol.
169:1861-1870.

Yates, J. R. Cunningham R. P. and Holmes, D. S. 1988. IST2: an insertion sequence from *Thiobacillus ferrooxidans*. Proc Natl Acad. Sci **85**:7284-7287.

Youvan, D. C., Elder, J. T., Sandlin, D. E., Zsebo, K., Alder, D. P., Panopoulos, N. J., Marrs, B. L. and Hearst, J. E. 1982. R-prime site-directed transposon Tn7 mutagenesis of the photosynthetic apparatus in *Rhodopseudomonas capsulata*. J. Mol. Biol. **162**:17-41.

Zalkin, H. and Mei, B. 1990. Amino terminal deletions define a glutamide transfer domain in glutamine phosphoribosylpyrophosphate amido transferase and other *pur*F-type amidotransferases. J. Bacteriol. **172**:3512-3514

Zalkin, H. and Weng, M. L. 1987. Structural role for a conserved region in the CTP synthetase glutamide amide transfer domain. J. Bacteriol. **169** (7):3023-3028.

Zhang, Y., and Fillingame, R. H. 1994. Essential aspartate in subunit *c* of F_1F_0 ATP synthase. Effect of position 61 substitutions in helix-2 on function of Asp^{24} in helix-1. J. Biol. Chem. **269**:5473-5479.