Chapter -5

Insertion sequence mediated genomic rearrangement in colony morphology variants of *Acidithiobacillus ferrooxidans*

6.1 INTRODUCTION:

Insertion sequence (IS) elements are small (<2.5kb), phenotypically cryptic segment of DNA with a simple genetic organization being capable of insertion on multiple sites in a target molecule (Mahillon and Chandler, 1998). They were initially identified as moving agents present in the plasmids that are capable of generating mutations as a result of their translocation. They often remain in cluster forming "islands" that promote the integration and excision of plasmid (Bukhari *et al.*, 1977). Very often two copies of certain ISs flanking a DNA segment are able to act in concert, rendering the intervening region mobile, being referred to as composite of compound transposons (Berg and Howe, 1989).

Over 500 different ISs have been identified till date. They can have pronounced and spectacular effects in shaping and reshuffling of the bacterial genome. Recent studies have shown such IS activity, which include the massive IS expansion during the emergence of some pathogenic bacterial species and the intimate involvement of ISs in assembling genes into complex plasmid structures. However, a global understanding of their impact on bacterial genomes requires detailed investigation and knowledge of their distribution across the eubacterial and archaeal kingdoms, understanding their partition between chromosomes and extra-chromosomal elements (e.g. plasmids and viruses), the factors which influence this, and appreciation of the different transposition mechanisms in action, the target preferences and the host factors that influence transposition. In addition, defective (non-autonomous) elements, which can be complemented by related active elements in the same cell, that also contribute to the evolution of genome organization are often overlooked in genome annotations (Siguier *et al.*, 2006).

6.1.1 Structure of Insertion Sequence

Organization

Structurally IS elements are small and compact DNA segments containing no other genes than those involved in their mobility. These include the active DNA sequences that define the ends of the element together with an enzyme, transposase (Tpase), which renders them the ability to move from one location to another. The Tpase is generally encoded by one or perhaps two open reading frames (ORFs) and covers nearly the entire length of the element.

Terminal inverted repeats

In the majority of the IS elements the sequence is terminated with a short terminal inverted repeat sequences (IR) between 10 and 40 bp. The IRs can be divided into two functional domains viz. (i) the one which is positioned within IR and is involved in Tpase binding and (ii) the terminal 2 or 3 bp which is involved in the cleavages and strand transfer reaction leading to transposition of the element (Derbyshire *et al.*, 1990; Makris *et al.*, 1988; Zerbib *et al.*, 1990). Beside these, indigenous promoters are often located partially within the IR sequence upstream of the Tpase gene, called IRL which may provide a mechanism for autoregulation of Tpase synthesis by Tpase binding.

Direct target repeats

Another typical feature of IS element is the presence of a short directly repeated sequences (DRs) at the flanking regions of the IS. Generation of staggered cuts at the target site actually leads to the generation of such DRs. The length of DRs ranges between 2 and 14 bp, their lengths being characteristics for a particular element. However sometimes there may be a total lack of the DRs because of the homologous inter-or intramolecular recombination between two IS elements, each with a different DR (Galas and Chandler, 1989; Turlan and Chandler, 1995).

6.1.2 Influence of IS element in the expression of neighboring genes

Certain IS elements have also been found that have deleterious effect on the expression of adjacent gene(s). When spontaneous Gal$^-$ mutations in *gal* operon of *Escherichia coli* K12 were examined, many were found to arise through the insertion of extra DNA into the bacterial genome (Starlinger and Saedler, 1976; Starlinger, 1980). When these insertions were examined, it was found that the same sequence occurred time and again in independent isolates which were recA mutants. In other words the same DNA sequence was repeated several times in the genome. Here, normal homologous recombination was ruled out as there is the absence of recA enzyme. Evolution of Gal$^-$ mutants was due to the disruption of the open reading frame of *gal* due to IS element that became incorporated within the gene. IS elements were found to be inserted into a gene in either of the two possible orientations. IS insertion into the *gal* operon, giving Gal$^-$ phenotype, are often very much polar (degree of polarity more than classical chain

termination mutations in the same region) (Jordan *et al.,* 1967). At least in case of IS*1* (in both orientations) and IS*2* (in orientation I), the polarity is due to *rho*-dependent transcriptional termination signals in the insertion elements (Besemer and Herpers, 1977). Gal^+ phenotype can be restored by internal deletion of IS*1* that removes the transcriptional termination site.

Many IS elements, however, have been found to activate the expression of neighboring genes. It has been shown that some IS elements such as IS*1*, IS2 etc. possess outwardly directed -35 promoter hexamer located in the terminal IRs. Therefore when it is placed at a correct distance from a resident -10 hexamer by transposition, new promoters capable of driving the expression of neighboring genes can be created (Galas and Chandler, 1989). In addition several inwardly directed -10 hexamers have also been detected in certain IRs. Therefore when two ends of such elements are juxtaposed in tandem, a combination of the -10 with a -35 hexamer resident in the neighboring right end can lead to the generation of a relatively strong promoter (Reimmann *et al.,* 1989). When either of IS*1* (in both orientations) or IS*2* (in orientation I) of *E. coli* K-12 is integrated it has been found to reduce the expression of operon. However when IS2 is integrated in orientation II it influences the genes lying downstream to express constitutively at a level three times higher than normal (Nevers and Saedler, 1977). This has been shown to be due to the generation of a strong promoter which can act as the transcription initiation point for the downstream gene. This promoter, however, may have other consequences when IS*2* is integrated in orientation I, when transcription from the IS*2* promoter may run head on into an RNA polymerase coming from the existing promoter of the operon, thus contributing to the inhibitory effect of IS*2* in orientation I. The two antipodal effects of IS*2* suggests that it can act as a simple switch (Saedler *et al.,* 1974). On the other hand when IS*1* is integrated it promotes deletion of the neighbouring DNA sequences at high frequency (10^{-4}) leaving IS*1* intact (Reif and Saedler, 1975; Peterson *et al.,* 1975).

Insertion elements have also been found to be responsible for the 'replicon fusion' whereby two different plasmids containing IS elements present in the same host may fuse together which is independent of the host recombination system, occurring also in *recA* mutants (Ohtsubo *et al.,* 1980). It has been found that only the terminal repeats of IS elements and not the gene in between are required for the transposition. In case of IS*1* any gene could be mobilized to a new transposon by enclosing it between two copies of IS*1* (Calos and Miller, 1980; Rosner and Guyer, 1980).

Insertion sequences which are generally present in multiple copies in a genome can also serve as the regions of homology for the host's recombination system (Saedler and Heiβ, 1973). The best known example is the integration of F plasmid into the *E. coli* chromosome to give rise to Hfr strains. This recombination between plasmid and the chromosome seems to be due to the presence of several copies of insertion sequences in the plasmid and in the chromosome that results into the homologous recombination (Davison *et al.*, 1975). However unlike other IS mediated recombination this recombination has been found to be *recA* dependent (Cullum and Broda, 1979). On the other hand when a gene is flanked by two copies of insertion sequence in the same orientation, an unequal crossing-over between sister copies of this region can result in tandem duplications of the gene, a process called amplification (Rownd *et al.*, 1978; Schneider *et al.*, 2000). On the other hand IS elements can also promote deletion of the DNA sequence but at a lower frequency (Farabaugh *et al.*, 1978).However it must be realized that the IS-mediated genome rearrangement is dependent on the selection system used.

6.1.3 Role of IS element in evolution of bacteria

It is impossible to predict the outcome of evolution because of the two variables, the environmental condition and the genotype of an organism (de Visser *et al.*, 2004). However the use of microorganism, because of their rapid generation, large population, ease of experimental control, and wealth of genetic and physiological information, in evolution experiments have offered a promising approach for the empirical exploration of the evolutionary role of these variables (Elena and Lenski, 2003). The information regarding the changes in the phenotypes due to change in the environmental condition can be obtained from these microorganisms (Imhof and Schlotterer, 2001; Rozen *et al.*, 2002; Orr, 2003). These rare mutations are the ones that actually direct the evolution. The identification of these mutations that are beneficial to the organism, at the molecular level would throw some lights in understanding of the evolution (Rainey *et al.*, 2000; Travisano, 2001; Otto, 2002).

Insertion elements in bacterial genome can play a pivotal role in the spontaneous mutation (de Visser *et al.*, 2004). Because of their ability of insertion into the coding region of the gene they can lead to the disruption of a gene or may increase the expression if the element is carrying a promoter that can promote the adjacent gene (Schneider and

Lenski, 2004). These IS elements have been shown to play an important role in the host adaptation to the changes in the environment (Wery *et al.*, 2002; Bongers *et al.*, 2003). Additionally recombination between two homologous IS elements can generate chromosomal inversions or deletions (Schneider *et al.*, 2000) contributing significantly to spontaneous mutation in bacteria (Hall, 1999, Kitamara *et al.*, 1995).

From an evolutionary perspective two hypotheses have been proposed to explain as to how IS elements persist in genomes. i) IS elements are genome parasites that, on balance, harm their host, primarily by the increased rate of deleterious mutations that they cause. However they are maintained by replicative transposition to additional genomic structures like plasmids, along with the horizontal transfer into other lineages (Charlesworth *et al.*, 1994). ii) IS elements also generate occasional beneficial mutations and are important for the adaptive evolution of their hosts, and are maintained by selection for some useful mutations that they produce sometimes (Blot, 1994). Several evolution experiments have substantiated the effects in the phenotypes of the bacteria brought about by IS-mediation.

In the longest-running evolution experiment that involved 12 populations of *E. coli* B, whose populations were propagated for more than 20,000 generations by daily transfer in minimal medium supplemented with glucose, showed the competitive fitness of the bacteria to be increased by 70% as shown by their ancestor. Molecular characterization revealed that all the essential mutations detected were IS-mediated events, which involved insertions, deletions and inversions (Cooper and Lenski, 2000, Lenski, 2004, Lenski *et al.*, 1991). Another experiment utilized restriction fragment length polymorphism (RFLP) analyses using IS element probe in *E. coli* which had been stored in a stab for 30 years. Substantial genetic diversity was revealed in this experiment: among the 118 clones analyzed, 68 different hybridization patterns were detected, and it was suggested that IS elements are important sources of genetic instability in the stab-culture conditions (Naas *et al.*, 1994). It was also shown that the rate of IS-mediated genetic change is faster during starvation than during normal growth of the organism. Therefore, IS elements provide a useful marker for the determination of intra-strain polymorphisms and to quantify genome evolution (Papadopoulos *et al.*, 1999). IS-mediated changes in the genome of *E. coli* have also been found in the population grown for a prolonged period in the stationary phase in Luria Bertani medium. The survivors were found to have acquired mutations that conferred increased fitness during the stationary phase (Finkel and Kolter, 1999). The

above findings led to the discovery of a novel mechanism of evolution (Zinser *et al.*, 2003). It is interesting to note the suggestions of some workers who believed that the effect of stress such as starvation, increased oxygen levels, or any adverse condition activate IS elements (Naas *et al.*, 1994; Mahillon and Chandler, 1998; Hall, 1999) thereby helping their host to adapt to these adverse conditions. Thus, IS elements can even be viewed as adaptive mutator genes, that mediate mutations at increased rates (Chao *et al.*, 1983; Moxon *et al.*, 1994; Cooper *et al.*, 2001).

6.1.4 Insertion Sequences in *A ferrooxidans genome*

The total genome sequence for *Acidithiobacillus ferrooxidans* type strain ATCC 23270 and *A. ferrooxidans* ATCC 53993 have recently been completed at TIGR (The Institute for Genomic Research now called J. Craig Venter Institute) and at US DOE Joint Genome Institute respectively. The size of the genome has been determined to be 3.0 Mbp and 2.9 Mbp for *A. ferrooxidans* ATCC 23270 and ATCC 53993 strains respectively. They differ in their shape, while the former is rod in shape the latter shows a spiral shape (Valdes *et al.*, 2008).

Forty one IS elements have been identified of which thirty one have been classified as belonging to nine families according to the scheme proposed by Mahillona and Chandler (Siguier *et al.*, 2006). In *A. ferrooxidans* ATCC 23270 the largest group of *A. ferrooxidans* IS elements are IS*Afe3* and IS*Afe4* which belong to the IS110 and IS3 families respectively. In the bacterial genome they remain in 8 and 3 copies respectively. However all IS elements are not found in all the strains of *A ferrooxidans* e.g. IS*Afe1* which is associated with phenotypic switching in *A. ferrooxidans* (Holmes *et al.*, 2001) was not detected in the type strain *A. ferrooxidans* 23270 (Valdes *et al.*,2008).

Earlier, Yates and Holmes (1987) discovered two families of repeated sequences of 1.3 and 1.4 kb in sizes that were repeated 20 to 30 times per chromosome covering approximately 6% of the chromosome. One of the representatives of the family, IS*T2*, was sequenced and found to be 1408 bp long. It contained two imperfectly conserved, 25 bp inverted repeats at either end of the IS element and two 9 bp target site duplications immediately adjacent to the inverted repeats and three ORFs, the same feature that were used to identify IS elements in *Shigella sonnei, E. coli and Halobacterium volcanii* (Yates *et al.*, 1988). It was also reported that IS*T2* does not have rigid target-site specificity and

therefore can move to any place of the genome and its transposition does not involve recombination (Cadiz *et al.*, 1994). Another repeated element IS*T1* (now IS*Afe*1) about 1.3 kb in size was found to contain 26 bp imperfectly paired inverted repeats similar to the ISL3 family and 5 bp target duplications and an ORF that potentially codes for the transposase gene along with -35 and -10 region which are separated by 18bp space. Out of the six tested four of the target sites showed strong preference to AT-rich region for IS*Afe*1 (Zhao and Holmes, 1997; Holmes *et al.*, 2001). It has been suggested that the number of copies of IS*Afe1* in the genome is low compared to other IS elements. Southern hybridization experiments using IS*T2* and IS*T1* probes, showed different DNA banding in different strains of *A. ferrooxidans* after several generation of propagation. However there was high consistency as far as the distribution of IS*Afe*1 is concerned (Holmes *et al.*, 2001). In 1997 Chakraborty *et al.*, discovered another repeated element from *A. ferrooxidans* viz. IS*T445* in addition to IS*T2* which was distributed within the *A. ferrooxidans* genome at a frequency of 10 to 20 copies. IS*T445* was found to be 1219 bp long containing 8 bp terminal repeats which could be further extended to 23 or 48 bp with 9 and 26 mismatches respectively and three ORFs along with the promoter and sequence typical of the Shine-Dalgarno sequence. The promoter of the sequence was found to lie within the left inverted repeats. The DNA sequence of IS*T445* was found to show a significant homology with IS*AE1* of *Alcaligenes eutrophus.* They also found that the distribution of these elements were different in different strains substantiating the view that they are also mobile within the *A. ferrooxidans* genome (Chakraborty *et al.*, 2002).

Changes in the distribution of IS-elements can be related with the changes in the phenotype of *A. ferrooxidans*. Schrader and Holmes (1988) found the large, spreading, highly motile colonies in the medium containing thiosulfate and ferrous iron which lacked the ability to oxidize iron but retained the capacity to oxidize thiosulfate or tetrathionate. They suggested that the specific changes in the position of the mobile elements were associated with the formation of the large-colony variant. Beside that, it was also observed that in the wild type revertants specific changes in IS elements distribution returned to wild-type patterns suggesting its role in the phenotypic switching (Holmes and Haq 1989). In 2001 Holmes *et al.*, found that the phenotypic switching is correlated with the high frequency insertion and excision of IS*Afe1* into and out of the *resB* gene which encodes a cytochrome *c*-type maturation protein. They proposed that insertion of IS*Afe1* into *resB* gene inactivated the capacity of ResB to satisfactorily mature a *c*-type cytochrome

resulting into the loss of the ability to oxidize iron but not sulfur (Cabrejos *et al.*, 1999). Chakraborty *et al.* (2002) also found large spreading variants which were unlike Schrader and Holmes (1988) and were found to be able to oxidize reduced sulfur compounds as well as ferrous iron. This showed that loss of the ability to oxidize ferrous iron can not be the only reason for the generation of large-spreading variants.

The present work tries to explore the genes that might have been affected by the movement of the IS element using a novel methodology and PCR technology. A novel PCR-based strategy has been employed to look into altered distribution, caused by a single event of excision or insertion of the IS elements, in the genome of the mutants. The intervening DNA sequence(s) between two successive pre-existing IS elements (Inter-IS DNA sequence) in the genome could be successfully amplified by using combination of outwardly directed primers designed from the termini of IS*T2* and IS*T445* (Fig 5.1). Comparing the pattern of inter-IS amplicons of the wild and the mutant strains to identify the polymorphic bands enables to document an insertion or excision event of the IS. These primers were used in various combinations so as to get different banding profiles of flanked genomic DNA regions. Under the circumstance of already preoccupied/apparently fixed location of insertion sequences in a given genome of *A. ferrooxidans*, the amplification of DNA regions between two nearby IS elements (subject to the ability of the Taq polymerase to amplify IS flanked DNA regions of varied lengths, aborting many of the inter IS regions of relatively larger fragment, under a given PCR condition), could be amplified.

If under a compelling environmental condition the mobility of an IS element causes a beneficial mutation to allow certain cells to out compete others in forming colonies, genomic DNA of mutant cell lines is supposed to manifest certain changes in the location of IS elements keeping others in the same genomic location identical to the wild type. On comparing the amplicon profile of inter-IS DNA regions of the mutant with the mother strain one may expect appearance of new band(s) or disappearance of band(s) due to fresh insertion(s) or deletion(s) of insertion element(s), in the vicinity of an already existing 'lone' IS element (incapable of generating an amplicon under the existing PCR condition) to generate a new inter-IS region or a band that had suffered a deletion event of one of them or both. Whatever may be the case, the researcher will be in a state to locate polymorphic band(s). The sequence information from the polymorphic band would enable to locate the gene sequence(s) that has been affected by the mobility of the IS element.

6.2 MATERIALS AND METHODS

6.2.1 Chromosomal DNA preparation

Reagents:

i) 0.15 (M) NaCl-0.1 (M) EDTA (pH 8.0) solution:

ii) 3 (M) Na-acetate.

iii) TE buffer: (10X): 100mM Tris-Cl (pH 8.0) and 10mM EDTA (pH 8.0).

iv) Proteinase K: 20mg/ml in sterile 50mM Tris-Cl (pH 8.0), 1.5 mM Ca-acetate; stored at -20°C.

Procedure:

Genomic DNA preparation was made following modification of Marmur's procedure (Marmur, 1961) as described by Yates and Holmes (1987). Cells were grown in elemental sulfur medium up to late log phase and filtered through Whatman filter paper no.1 to remove un-dissolved sulfur particles. Cells were harvested from 200 ml of the culture by centrifuging at 10000 rpm for 10 min at 4°C. The cell pellet was washed with distilled water and re-suspended in 2 ml of 0.15 M NaCl-0.1 M EDTA (pH 8.0) and the concentrated solution was frozen at -20°C for at least 4 hrs. The frozen cells were rapidly thawed at 55°C and lysozyme (1.5 mg/ml) was added. The mixture was then incubated at 37°C for 15 min followed by the addition of SDS to the final concentration of 1% (w/v). The mixture was incubated at 55°C for 1 hour. Proteinase K was then added (5mg/ml) and the incubation was continued at 55°C until the solution became clear. The cell lysate was extracted with phenol, phenol:chloroform and then with chloroform at 4°C. Finally the aqueous phase was collected in a beaker placed in ice. 1/10 volume of 3M Na-acetate and double volume of chilled ethanol was then added. The DNA was spooled with a sterile bent glass rod. The DNA obtained was air dried and suspended in TE buffer (10:1).

6.2.2 Plasmid DNA preparation

Plasmid DNA was obtained by alkaline lysis (Brinboim and Doly, 1979).

6.2.2.1 Minipreparation of plasmid DNA

Reagents:

i) Solution I: 25 mM Tris-HCl (pH 8.0), 50 mM EDTA, 1% glucose (w/v).

ii) Solution II: 0.2 N NaOH, 1% SDS (w/v).

iii) Solution III: 5M Acetic aid and 5 M K-acetate mixed in ration of 2:1. pH adjusted to 4.8 with either solution.

*Solution I and Solution III were stored at 4°C. Solution II was freshly prepared.

iv) Ethanol

v) TE buffer

vi) RNase A: 10mg/ml in 0.01 M Na-acetate. The solution was heated to 100°C for 15 minutes and allowed to cool slowly to room temperature. pH was then adjusted by adding 0.1 volume of 1M Tris-Cl (pH 7.4). The solution was stored at -20°C.

Procedure:

Cells were harvested in 1.5 ml microfuge tube by centrifugation at 8000 rpm for 5 min at 4°C. The bacterial pellet was made as dry as possible and suspended in 100 µl of ice-cold Solution I by vigorous vortexing followed by the addition of 200 µl of freshly prepared Solution II.

The contents were mixed by inverting the tube rapidly for five times and stored in ice for 6 to 8 minutes.

150 µl of ice-cold Solution III was added and mixed by inverting the tube several times. The mixture was then stored on ice for 3 to 5 minutes.

The tube was centrifuged at 10000 rpm for 15 min at 4°C and the supernatant was transferred to a fresh tube.

An equal volume of phenol:chloroform was added and mixed by vortexing. The tube was then centrifuged at 10000 rpm for 2 minutes at 4°C. The aqueous phase of the mixture was transferred to a fresh tube.

Plasmid DNA was recovered by precipitation by adding double volume of ethanol at room temperature. The solution was mixed by vortexing and allowed to stand for 2 minutes at room temperature.

The precipitated DNA was collected by centrifugation at 12000 rpm for 5 minutes at 4°C in a microfuge tube. The supernatant was removed by gentle aspiration and the tube was allowed to stand in an inverted position on a paper towel at room temperature to allow all the liquid to drain away.

DNA was then dissolved in 50 µl TE (10:1) buffer containing DNase-free pancreatic RNase (20µg/ml) followed by incubation at 37°C for 1 hour. The solution was extracted with phenol, phenol:chloroform, chloroform and finally DNA was precipitated with ethanol. DNA pellet was washed with 70% ethanol, air dried and re-suspended in TE (10:1) buffer.

6.2.2.2 Maxipreparation of plasmid DNA

Reagents:

i) Solution I: 25 mM Tris-HCl (pH 8.0), 50 mM EDTA, 1% glucose (w/v).

ii) Solution II: 0.2 N NaOH, 1% SDS (w/v).

iii) Solution III: 5M Acetic aid and 5 M K-acetate mixed in ratio of 2:1. pH adjusted to 4.8 with either solution.

*Solution I and Solution III stored at 4°C. Solution II was freshly prepared.

iv) Antibiotic for plasmid selection.

v) Chloramphenicol (34mg/ml ethanol).

vi) Ethanol.

vii) Isopropanol

viii) STE buffer {10mM Tris-Cl (pH 8.0), 0.1 M NaCl and 1mM EDTA (pH 8.0)}

ix) TE buffer

x) Lysozyme: 10 mg/ml in 10mM Tris-Cl (pH 8.0)

xi) RNase A: 10mg/ml in 0.01 M Na-acetate. The solution was heated to 100°C for 15 minutes and allowed to cool slowly to room temperature. pH was then adjusted by adding 0.1 volume of 1M Tris-Cl (pH 7.4). The solution was stored at -20°C.

Procedure:

30 ml of the rich medium (Luria-Bertani broth) was inoculated with a single colony and incubated till late log phase (OD_{600} = 0.6). About 25 ml of culture was transferred to 500 ml of the medium containing the appropriate antibiotic in a 2 L flask. The culture was incubated with vigorous shaking to an OD_{600} of ~0.4. For amplifiable plasmid 2.5 ml of chloramphenicol was added and further incubated to 12-16 hrs at 37°C with constant shaking.

Cells were harvested by centrifugation at 5000 rpm for 15 minutes at 4°C and made as dry as possible. Cell pellet was re-suspended in 200ml STE buffer and cells were collected by centrifugation as described above. Cells were then frozen at -20°C for 3 hrs followed by thawing at room temperature.

Bacterial cells were re-suspended in 18 ml of Solution I and 2 ml of freshly prepared 10mg/ml lysozyme solution. After that 40 ml of freshly prepared Solution II was added. The contents were mixed thoroughly by gently inverting the tubes several times. The bottle was then incubated at room temperature for 5 to 10 minutes.20 ml of ice-cold Solution III was then added followed by mixing well by swirling the bottle several times and stored on ice for 10 minutes.

The bacterial lysate was then centrifuged at 11000 rpm for 30 minutes at 4°C. The supernatant was collected in a fresh tube and 0.6 volume of isopropanol was added and

allowed to stand in room temperature for 10 minutes. The precipitated DNA was recovered by centrifugation at 8000 rpm for 15 minutes at room temperature. Supernatant was decanted and the DNA pellet was rinsed with 70% ethanol. The pellet was dried and re-suspended in 3 ml of TE (pH 8.0) containing 20µl RNase A (10 mg/ml). The mixture was incubated at 37°C for 1 hour.

The solution was extracted with phenol, phenol:chloroform, chloroform and finally DNA was precipitated with ethanol. DNA pellet was rinsed with 70% ethanol, air dried and re-suspended in 1.5 ml TE (10:1) (pH 8.0).

6.2.3 Restriction digestion

Restriction enzymes were purchased from Bangalore Genei, India. Reaction was conducted in a small polypropylene tube containing 10 µl of the total mixture containing 1X buffer, 1 µg DNA and 1unit of the enzyme following the manufacturer's instruction. The mixture was incubated at 37°C for 12hrs. The DNA was ethanol precipitated and re-suspended in TE (10:0.1) and stored at -20°C.

6.2.4 Purification of DNA

DNA was purified by separating in the agarose gel electrophoresis and eluting it from low melting agarose gel.

6.2.4.1 Agarose gel electrophoresis

Reagents:

i) Agarose (SRL)

ii) 1 X TAE buffer: 40mM Tris-acetate, 1mM EDTA.

It was prepared by diluting the stock solution (50 X TAE). 50 X TAE buffer contained 242 g of Tris base (SRL), 57.1 ml glacial acetic acid and 100 ml of 0.5 M EDTA (pH 8.0).

iii) Ethidium bromide: Stock solution was prepared by dissolving 1 g of Ethidium bromide in 100 ml water. Required volume was added to TAE buffer with the final concentration of 1μg ml^{-1}.

For the separation of large and medium sized DNA fragments, 0.8% agarose gel electrophoresis was performed, whereas for smaller sized fragments 1.5% agarose gel electrophoresis was performed following the method of Sambrook *et al.,* (1989). Electrophoresis was performed in a horizontal tank containing 1 X TAE buffer using ~50 mV of electricity.

6.2.4.2 Elution from low melting agarose gel

1.5% low melting agarose purchased from SRL Company, India was used to isolate small amount of DNA of about 1μg or less. The gel was run for a required period of time under cold condition. After completion of the run, required DNA band was sliced with the help of sterile scalpel and taken in an eppendorf tube. The gel was melted by keeping in the water bath kept at 65°C for 5 minutes to which an equal volume 1X TE was added. The mixture was freeze-thawed thrice and the agarose particles were then removed by phenol extraction. Finally DNA was precipitated in isopropanol and was re-suspended in 0.1 X TE buffer and stored at -20°C.

6.2.5 Amplification of DNA using PCR (Polymerase Chain Reaction)

PCR amplification was performed using 'PCR Amplification Kit' (GENEI, India), in 50 μl reaction volume, following instructions provided by the supplier. Each 50 μl PCR mix contained; 3 μl of 10mM dNTP mix, 5 μl of 10X buffer containing 15 mM $MgCl_2$, 12.5 pmol of each forward and reverse primer, 10 ng of template DNA and 1U *Taq* DNA Polymerase. Taq DNA polymerase was added after adding all the ingredients of the mixture. The PCR was done in a GenAmp PCR system (Applied Biosystems).

Purity of DNA was tested with the help of *UV* spectrophotometer by determining the OD values at 260 and 280 nm (ratio of the two readings for pure DNA was considered to be 2.0), while concentration was determined by using the standard that the concentration of DNA will be 10 times the OD reading at 260nm.

6.2.5.1 Primers used

A. To amplify the 16S rDNA a set of eubacterial universal primers 27f and 1492r (Lane, 1991) were used. Sequence of the primers used:

```
27 f:      5' AGAGTTTGATCCTGGCTCAG 3'
1492r:     5' TACGGTTACCTTGTTACGACTT 3'
```

The program set to amplify 16S rDNA was: 5 min of initial denaturation at 94°C, 30 sec of denaturation at 94°C, 30 sec for annealing at 58°C, and 1.5 min at 72°C for extension, for a total of 35 cycles.

B. To amplify IS*T445* and IS*T2* repetitive elements of *A. ferrooxidans* a set of primers were designed using the conserved terminal inverted repeats of the insertion sequences. For this total DNA sequences of IS*T445* and IS*T2* were retrieved from the Gene Bank and the primers having almost same melting temperature [T_m = {2(A+T) + 4(G+C)} °C] values were designed as shown below. Inwardly directed primers are colored blue and outwardly directed primers are colored red. The symbols '<' and '>' shows the orientation of the primers when they are annealed with template DNA. The inverted repeats are underlined. The primer sequences and their melting temperatures are given in Table 5.1.

B (1) Complete IS*T445* sequence showing Primer sequences used in the amplification of IS*T445* and inter IS polymorphic DNA regions:

(RP3)
<<<<<<<<<<<<<<<<<<<
ccgagaagacgcctaactca
(A_1)
5'
ggctcttctgcggattgagtgggtaggggtagaatgctgtggagcagttgggtcccgaataggagttag
>>>>>>>>>>>>>>>>>>

ctgatggtcccagaagatctgttttctctcgcgttagggttggtaccgccggtggttggtggacggaatg

tgaccttccaagtggaggagaagcgcctggatctgcacatcaactttcccaagggcagtcgctttgcttg

cccgtatgtggtgaggagtgtccggtacatgcaccctggaactataccggcggcacatggtttatttcag

catgaagtcctatcccacgccgctgaccctggtcggagatgccagtattgacggtagcccgctggtcggc

gagccgacacgctcctgtgcgggtgattgaccattatgtgcccgaagcccgtaccagggtggatatggcc

aaatgtccatgcccggttcggcgtcgaattgagaaccaagcagtcggcgcggtcatgactacaatcacgc

tcttcgtggatctgaatgcccggcgactcttgttccgtacgcccggcaaggatgccaagacctttgagaa

attctccgaagatctacaggcccatggtggtagcgcggaaggcatcaaccgaattgtgagcatggccctc

tcgccggccttccagaaaggggcttgccgaagcacctgcccaatgctgagatcaccttcgaatcgttttc

acctcatgaagctcgttcaacgaggccgtggatgccgtgcgcaaggggaagtccttcacccagccaaat

ctcaaaaagacccgctggctctggctcaagaatgattgtaacctcaaggccaggcagaaagaaaagctgc

aggaattattaaaagaccagaacctgaaaacggcacaggcctaccagtttcgctgacttttcaggcattc

ttcacagtcaagaatcgcaccagggcgccaaccctgctcaaggcctggatggaaaacgccaaggacagcg

gaatgccgcctatggtcaaggtcgcctacaccattccattgaacaattgggatggccgttgcttccgctg

gttcgaaaagccagattaccaatggaattcttggaaggcttcaacagcctcaatcaatccgccaaggcca

aggcccggggttaccgaccacaagaattttatcaacatggcctacctgatcctgggtaagctggatctca

ggctacccacttgaaaatgacgaagagcc 3′ (RP4)

actttttactgcttctcgg (A_2)

B 2) Complete IS*T2* sequence showing Primer sequences used in the amplification of IS*T2* and inter IS polymorphic DNA regions:

(RP1)

ctcgatatcagtttagacca

(G_3)

5′

gagctatagtcaaatctggtgtatgagcattttctgacatagatcaggcggcggtctgttgatcatccgg

cagggtgaggttagccggtatgccatcgataaatttcacccctgcaaaaagctgcaggactttctctgga

gcataaatcccaatccagcgtttttcagcctgctgcagcatcttgaaacccaatcccagaaagctgctgc

gtgatacacagttcttcgtccgggtagtccggtggcgtaccgtagcgaaggtggactcaatggcattggt

ggtccggatatgccgccagtgttcggccggaaaatcatagaaagcgagcaattcagcccgatcttttcc

agtttagccacagccttgggatactttgcctggtaattgcgcacaaacacgtccagtgctttttccgcag

cctggcgattggcggccatccagatttcctgcaacgctgccttggctttgctctgttgggctttgggaag

ttcgttgaggatgttggcggtcttatgcacccagcagcgttgctgaccagtttcgggataggcttcatcc

agtgcggcccaaaaccccatggcaccatccccaatggcgagcaaaggggccgtctccagaccgcgcgcct

gcaggtcacgcaggatctccagccaggaagctttggattcgcgcaggccgtcactgaccatgaccagctc

tttcttgccctctgccgtcacgccgataatcaccagcaggcagatacgcggatcctctccgcacggaggt

tcgtataaataccgtctacccaccagtaagcatagcgctttcctgtaggagcggcgttgccaatgggcat

actcttgcgcccactccgccttgagacgtcccaacaccgcaggcgaaagtcccttggcctcatcacccag

caaaatggaaagggcttcctgcatgtggccggaagacaccccatgcagatagagccaaggtactgtagcg

gctaccgttcgtgatttgcgtacatacggaggcgccagtaccgaattgaatttgatccccgatcctgacg

gtcccgcaccttgggtactttgacgggtaccggacccagagcggtcatgatctcgcgctccggcagatgc

ccattacgcacgaccgcctgacgcccatcaaccatccgcaccgtcgcaaattcttccagcaataccgcca

cctctgcctctatggcctgctcaatgcagagtgccgcgcgtcgaagtatcccttcaatgcccagacccaa

ctctcccatcccacctgctattacggtattcttttccacggcgtactccttatgttgctctttgagtcgg
>>>>>>>>>>>>
aaacccttttgtagcaacagtacgccacctcattcaagccagttgtaacgcgcccatac**accactttcgag**
<<<<<<<<<<
gtgaaagctc

>>>>>>>>
catagctc 3' (RP2)
<<<<<<<<
gtatcgag (G_4)

6.2.5.1 PCR Program

The PCR programs set for the amplification of IS*T445* and *IST2* were as follows:

IS*T445*: 30 sec at 94°C for initial denaturation, 30 sec at 50°C for annealing and 3 min at 68°C for the extension, for a total of 30 cycles.

IS*T2*: 30 sec at 94°C for initial denaturation, 30 sec at 48°C for annealing and 3 min at 68°C for the extension, for a total of 30 cycles.

Although sixteen different combinations of outwardly directed primers could be made (Fig 5.1) for the amplification of the inter IS regions of the genome, large differences in the T_m values (Table 5.1) allowed to use only two combinations, *i.e.*, i) RP1 and RP4, ii) RP2 and RP3. The PCR program set for the amplification of inter IS regions were as follows:

i) RP1 and RP4: 30 sec at 94°C for initial denaturation, 30 sec at 52°C for annealing and 3 min at 72°C for the extension, for a total of 30 cycles.

ii) RP2 and RP3: 30 sec at 94°C for initial denaturation, 30 sec at 55°C for annealing and 3 min at 72°C for the extension, for a total of 30 cycles.

6.2.6 Cloning of PCR products

6.2.6.1 Chemicals and reagents

All the chemicals used in the experiment were purchased from SRL Fine Chemicals (SRL India Ltd.). De-ionized double distilled water was used for the preparation of reagents. The reagents were either filter-sterilized or autoclaved, wherever required, before use and preserved in glass containers.

6.2.6.2 Elution of the PCR products from low melting point agarose

Approximately 100 µl of PCR products were directly precipitated and dehydrated in ethanol and were suspended in 25 µl TE (10mM Tris-Cl : 1 mM EDTA, pH 7.5). 20 µl of the suspended DNA was loaded in an 'Low Melting Point Agarose' gel and were subjected to electrophoresis at 40–45 mV for at least 8 –10 h. Setting and running of the gel was performed at 4°C in 1X TAE following techniques stated in 'Basic Methods in Molecular Biology' (Davis *et al.* 1987). The required band (s) were sliced and taken in a 2 ml microcentrifuge tube, and were melted at 65 ^{0}C for 5 min. To the melted agarose equal volume of TAE buffer was added and agarose particles were removed by repeated phenol extraction. Finally DNA was precipitated, suspended in TE (10mM Tris-Cl: 0.1 mM EDTA, pH 7.5) and quantified.

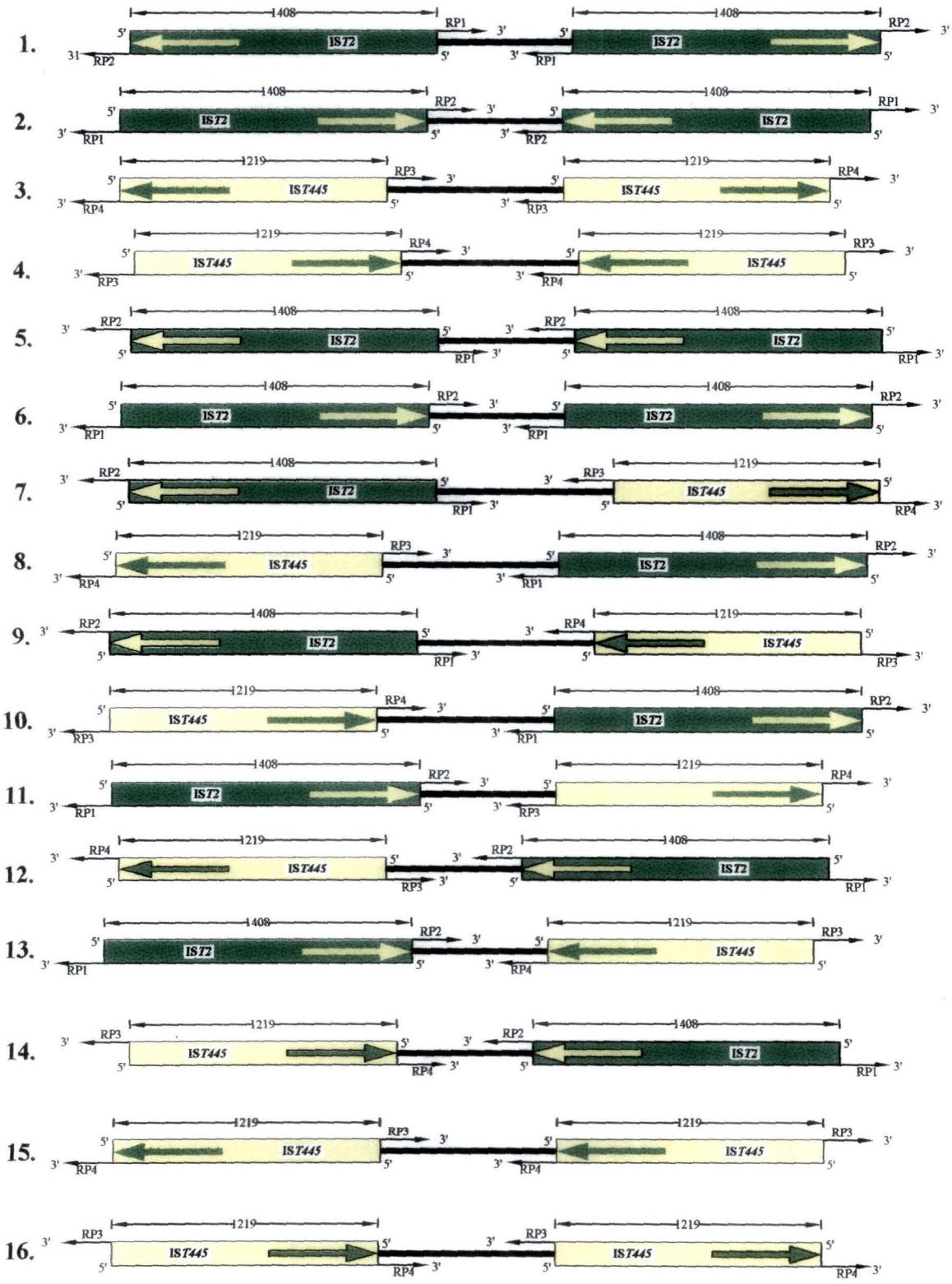

Fig 5.1: Schematic represantation of possible combinations of outwardly directed primers. Small arrows indicate the direction of primers, bigger arrows indicate the orientation of Insertion Sequence elements. The **Bold Line** represents the amplifiable inter IS region permitted by Taq polymerase. RP1 and RP4 primers combination was used to amplify the inter IS region as shown in **9 and 10 while RP2 and RP3 primers combination was used to** amplify the region as shown in **11 and 12**. Other combinations were not possible because of the large differences in their *Tm* values.

Table 5.1: Name and sequences of primers along with their T_m and GC% values.

Name	Sequence	T_m (°C)	GC %
A_1	GGCTCTTCTGCGGATTGA	56	55.6
A_2	GGCTCTTCGTCATTTTCA	52	44
G_3	GAGCTATAGTCAAATCTG	50	38.9
G_4	GAGCTATGCTCGAAAGTG	54	50
RP1	ACCAGATTTGACTATAGCTC	56	40
RP2	ACCACTTTCGAGCATAGCTC	60	50
RP3	ACTCAATCCGCAGAAGAGCC	62	55
RP4	CTTGAAAATGACGAAGAGCC	58	45

6.2.6.3 Cloning of the PCR product

The pGEM–T Easy Vector System II, purchased from Promega Corporation, Madison, USA, was used for the cloning of PCR products following manufacturer's instructions.

6.2.6.4 Ligation

The vector (25 ng) and the insert DNA were taken in the molar ratio of 1:2 and were suspended in 4µl sterile double distilled water. 5 µl of 2X rapid ligation buffer and 1 µl of T4 DNA ligase were added and thoroughly mixed by vortexing and centrifugation. The mixture was kept at 4°C for at least 16 h, heated at 60°C for 10 min and 5-7 µl was used to transform competent *E. coli* cells.

6.2.7 Transformation of recombinant plasmids

6.2.7.1 Preparation of fresh competent *E. coli* cells using Calcium Chloride and Transformation

A single colony from a plate freshly grown for 16-20 hours at 37°C was taken and transferred into 100 ml of LB broth in a 1 L flask. The culture was incubated for approximately 3 hours or more at 37°C with vigorous shaking to attain the viable cell number of 10^8 cells ml^{-1}. The cells were transferred aseptically to a sterile, ice-cold 50 ml polypropylene tube. The culture was cooled to 0°C by storing the tube on ice for 10 minutes. The cells were then recovered by centrifugation at 5000 rpm for 10 minutes at 4°C. The medium was decanted from the cell pellet. The pellet was re-suspended in 10 ml

of ice cold 0.1M $CaCl_2$ and stored on ice for 30 minutes. The cells were recovered by centrifugation at 4000 rpm for 10 minutes at 4°C. The fluid was decanted from the cell pellet, and the tube was placed in an inverted position for 1 minute to drain all the traces of fluid. The pellet was re-suspended in 2 ml of ice cold 0.1M $CaCl_2$ for each 50 ml of original culture. Using a chilled, sterile pipette tip, 200 µl of each suspension of the competent cells was transferred to a sterile microfuge tube. DNA (not more than 50 ng in a volume of 10 µl) was added to each tube. Gentle swirling of the microfuge tubes was done to mix the contents. The tubes were stored on ice for 30 minutes. The tubes were then transferred to a rack placed on a circulating water bath (42°C) for exactly 90 seconds followed by immediate transferring to an ice bath for chilling for 1-2 minutes. 800 µl of Luria-Bertani (LB) medium was added to each tube and incubated for 45 minutes in a water bath set at 37°C to allow the bacteria to recover and to express antibiotic resistance marker encoded for the plasmid. Appropriate volume (upto 200 µl per 9 cm petriplates containing 15 ml medium each) of competent cells was transferred onto LB agar plate containing appropriate antibiotic. A sterile bent glass rod was used to spread the transformed cells over the surface of the agar plate. The plates were left at room temperature until the liquid had been absorbed. Finally plates were inverted and incubated at 37°C for 12-16 hours for the appearance of colonies.

6.2.7.2 Identification of bacterial colonies that contain recombinant plasmids by α-complementation method

To sterile Luria–Bartani agar plate containing 50 µg ml^{-1} ampicillin, 0.5 mM isopropylthiogalactoside (IPTG) and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (80µg $µl^{-1}$), transformed competent cells were plated as described earlier. It was possible to recognize colonies that carried putative recombinant plasmids by blue-white screening of the colonies. Insertion of foreign DNA into the polycloning site of plasmid vectors leads to the incapability of complementation. Bacteria carrying recombinant colonies therefore produce white colonies. Re-circularized vectors containing bacteria could utilize chromogenic substrate X-Gal and thus form blue colonies. Recombinant-plasmids were isolated and screened by alkaline lysis method (Brinboim and Doly, 1979). Purified clones were checked by *Eco*RI restriction digestion followed by agarose gel electrophoresis.

6.2.8 DNA sequencing

The recombinant plasmids were directly used for sequencing of the inserts using primers for T7 and SP6 promoters. Nucleotide sequencing was performed with the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kits (Perkin-Elmer) using specific primer and the reaction was analyzed in an 'AB1 PRISM 377 DNA Sequencer'.

6.2.8 Computer analysis of the sequence

6.2.8.1 Restriction analysis

For the determination of restriction map of the sequence, the software package NEB CUTTER v. 2.0 of New England Biolabs Inc. was used from the website http://tools.neb.com/NEBcutter2.

6.2.8.2 Determination of open reading frame

The possible open reading frame was obtained from the sequence by using the NCBI tool for ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/).

6.2.8.3 Similarity Search

For similarity search studies of nucleic acid and amino acid sequences the BLAST N and BLAST P programs were used (Altschul *et al.*, 1997) from the website http://www.ncbi.nlm.nih.

6.2.8.4 Conserve domain search (CDD)

The Domain analysis of the amino acid sequence of the inserts was done using the program of Marchler-Bauer and Bryant (2004) from the website http://www.ncbi.nlm.nih.

6.2.8.5 Structure and function analysis of protein

For determination of probable secondary and three-dimensional structure of the peptides, the software packages, 'PredictProtein', 'SWISS-Pdb Viewer 3.7 (SP5)' and 'POV-Ray' from http://www.embl-heigelberg.de, http://www.expasy.org/spdbv/, and

http://www.povray.org/povlegal.html respectively were used. The function of the peptides was analyzed using the ProtFun 2.1 software of the website http://www.cbs.dtu.dk.

6.2.9 Southern Hybridization

DIG High prime DNA labeling and detection kit', purchased from Roche Diagnostics GmbH, Penzberg, Germany, was used for the experiments following manufacturer's instructions.

6.2.9.1 Preparation of probes

PCR amplification: For the preparation of probes the primers G_3 and G_4 were used to amplify the IS*T2* fragment from the NCIB 8455 strain obtained from Bose Institute, Kolkata, India. The parameters for the PCR amplification were kept same as described earlier. Elution of the PCR products from low melting point agarose was performed as described earlier.

6.2.9.2 Labeling of the PCR amplicons

The random primed DNA labeling (non-radioactive) DIG- (Digoxigenin-dUTP) DNA labeling kit (Roche diagnostics, Germany) was used for the labeling reaction following manufacturer's instructions. Approximately 1.0 µg of template DNA was taken in 16 µl of sterile double-distilled water; denatured in a boiling water bath and immediately transferred to an ice bath. To the denatured DNA 4 µl of 'DIG-High Prime' was added, thoroughly mixed and incubated overnight at 37°C. On the next day 2 µl of 0.2 M EDTA was added to the mixture to stop the reaction.

6.2.9.3 Transfer of DNA to the nylon membrane

In an agarose gel (1%) the *Eco*RI restricted genomic DNA of *Acidithiobacillus ferrooxidans* strains DK1, positive control NCIB 8455 and negative control (*Escherichia coli* XL1) were loaded in separate lanes. The gel was electrophoresed at 50 mV for at least 5h. The electrophoresed DNA products were then transferred to a positively charged nylon membrane by capillary transfer method (Sambrook *et al.*, 1989) after depurination (0.25 N HCl, for 15 min); denaturation (1.5 M Nacl, 0.5 mM NaOH, for 20 min); and

neutralization (1.0 M Tris-Cl, 2.0 M NaCl, pH 5.0, for 25 min). After transfer, the DNA on the membrane was fixed using a UV-crosslinker (Stratagene, USA).

6.2.9.4 Pre-hybridization

The membrane was placed in a plastic hybridization bag and was allowed to wet in 10 ml of preheated (40°C) DIG-Easy Hyb solution. The hybridization bag was properly sealed with care not allowing the trapping of any bubbles. The sealed bag was agitated in a water bath at 40°C for 30 min.

6.2.9.5 Hybridization

The pre-hybridization solution was forced out from the bag and 10 ml of fresh DIG-Easy Hyb solution was introduced again. The DIG-labeled DNA probe was denatured in boiling water bath for 5 min and was immediately placed on ice bath. The denatured probe was then introduced in to the hybridization bag and was mixed thoroughly taking care to avoid foaming. The bag was immediately placed into a water bath at 65°C and agitated for 15-16 h.

6.2.9.6 Post hybridization wash

After hybridization the membrane was placed in a plastic container and washed twice for 5 min in 50 ml Solution-A (2X SSC, 0.1% SDS) at 25°C under constant agitation. Stringent washes were also done twice for 15 min (2 X 15 min) in solution-B (0.5 X SSC, 0.1 % SDS) at 50°C.

6.2.9.7 Blocking of membranes

After hybridization and stringent washes, the membrane was rinsed in 'washing buffer' (0.1 M Maleic acid; 0.15 M NaCl; pH 7.5; 0.3% Tween 20). Then the membrane was incubated for 30 min in 100 ml 1X 'Blocking solution' (prepared by diluting 10X Blocking reagent with wash buffer). The incubated membrane was dipped into 20 ml 'Antibody solution' (10 µl of Anti-Digoxigenin-AP was added in 50 ml Blocking solution) for 30 min and then washed twice in 10 ml wash buffer for 15 min to remove the unbound antibody.

6.2.9.8 Detection of hybridization by color development

After blocking, the membrane was transferred to a glass tray and equilibrated with 20 ml detection buffer (0.1 M Tris-Cl, 0.1 M NaCl, pH 9.5) for 5 min. The membrane was then placed in a fresh glass tray and 10 ml of freshly prepared 'color-substrate solution' (200 µl NBT/BCIP stock solution to 10 ml detection buffer) was poured on it. The membrane was preserved in dark and the reaction was stopped after 16 h by transferring the membrane in water.

6.3 RESULTS

6.3.1 16S rRNA gene amplification

About 1.5 kb 16S rDNA of the isolate *Acidithiobacillus ferrooxidans* strain DK6.1 was amplified and checked in an agarose gel comparing with a molecular size marker (Figure 5.2). The PCR product was purified and cloned in pGEM T-easy vector (Fig 5.3) prior to sequencing. The recombinant plasmid containing 1.5 kb PCR product of DK6.1 was named as pAR1.1.

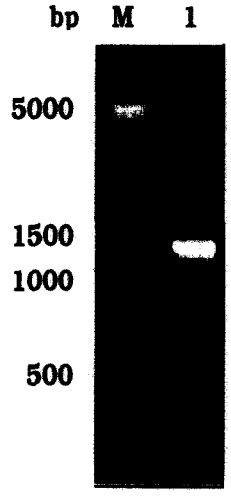

Fig 5.2: Lane 1, amplified 16S rDNA from *A. ferrooxidans* DK6.1; and lane M, 500 bp DNA ladder. corresponding sizes of the fragment is given.

The recombinant plasmids were transformed into the competent cells of *E. coli* XL1 Blue and the clones containing recombinant plasmids were detected using blue-white

screening method in X-gal containing plates. White cells were taken and minipreparation of plasmid was carried out to check the presence of insert in the plasmid. Once the presence of desired insert in the plasmid was confirmed by restriction digestion, maxipreparation of plasmid was carried out. And the sequencing of the inserts of the recombinant plasmids was done using SP6 and T7 primers.

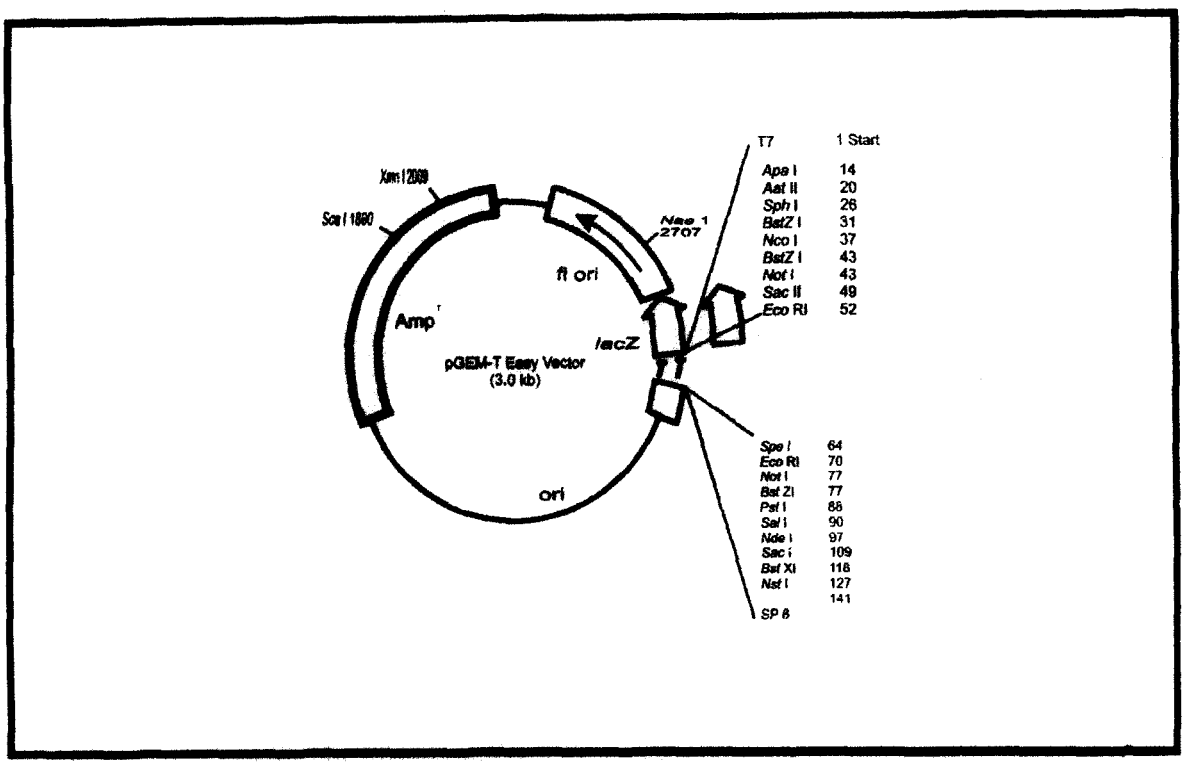

Fig 5.3: Diagrammatic representation of pGEM-T easy vector.

The 16S rRNA gene sequences of the isolates *Acidithiobacillus ferrooxidans* DK6.1 was deposited in EMBL nucleotide database under accession no. FJ602383. Blanstn result of the 16SrRNA and assigning the strains from Garubathan AMD to *Acidithiobacillus ferrooxidans* has already been described in chapter 3.

6.3.2 Determination of presence of IS*T445* and IS*T2* sequences in Garubathan strains

Using the primers A_1-A_2 and G_3-G_4, IS*T445* and IS*T2* fragments from DNA of strains DK1, GBVI, and CMO II along with the positive control *A. ferrooxidans* NCIB 8455 (obtained from Bose Institute, Kolkata, India) were amplified (Fig 5.4). The sizes of both fragments of IS*T445* and IS*T2* generated from Garubathan strains were found to be similar to that of the positive control NCIB 8455 that were around 1.2 kb and 1.4 kb

respectively. Since the nucleotide sequences of these fragments are already known (Yates *et al.*, 1988; Chakraborty *et al.*, 1997) the undertaken research work included the preparation of restriction maps using NEBcutter v2.0 (Fig 5.5A and 5.5 B). Southern hybridization was performed to detect the presence of IS*T2* fragment in one of the strains (DK1) isolated from Garubathan along with the positive control NCIB 8455 and negative control *E. coli* XL1 Blue. Hybridization was conducted using the IS*T2* fragment isolated from NCIB 8455 as probe labeled with Digoxigenin-dUTP. The hybridization reaction was conducted by incubating the IS*T2* probe with *Eco*RI digested genomic DNA at 65°C overnight. The presence of a multiple copies of IS*T2* in DK1 as well NCIB 8455 strains but no hybridization signal with *E. coli* XL1 Blue DNA was detected (Fig 5.6).

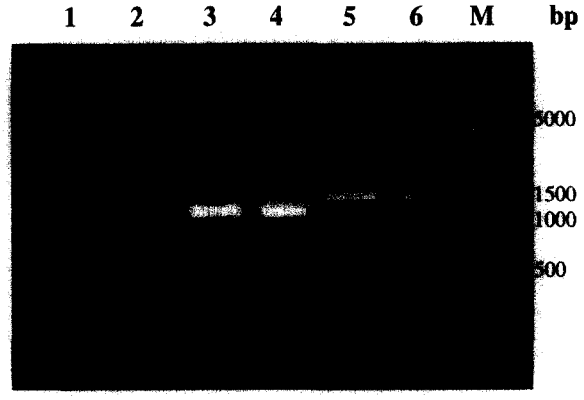

Fig 5.4: IS*T445* and IS*T2*amplification of different strains of *Acidithiobacillus ferrooxidans* strains from Garubathan AMD.
Lane M, 500 bp DNA ladder; corresponding sizes of the fragment is given; Lanes **1** to **4**: IS*T*445 fragments generated from the strains DK1, **GBVI, CMO II** and NCIB 8455 respectively. Lane **5** and **6**: IS*T2* fragments generated from DK1 and NCIB 8455 respectively.

6.3.2.1 Restriction fragment analysis of amplicon generated with IS*T445* specific internal primers

Since the the restriction map of IS*T445* is available, the amplicon generated from DK1 and NCIB8455 DNA (using IS*T445* specific primers) were digested with *PstI* enzyme and analyzed by agarose gel electrophoresis. In case of NCIB8455, the DNA fragments so generated were of the sizes corresponding to the sizes predicted from NEBCutter v 2.0 analysis. However in case of DK1, *PstI* digested DNA showed three

fragments of DNA. The sizes of the two DNA fragments out of three were consistent with that of the DNA fragments produced from *Pst*I digested NCIB8455 DNA (Fig 5.7). This result suggested that in case of DK1 in addition to IS*T445* another unknown DNA (~ 1.1 kb) fragment got co-amplified, which did not bear any *Pst*I site.

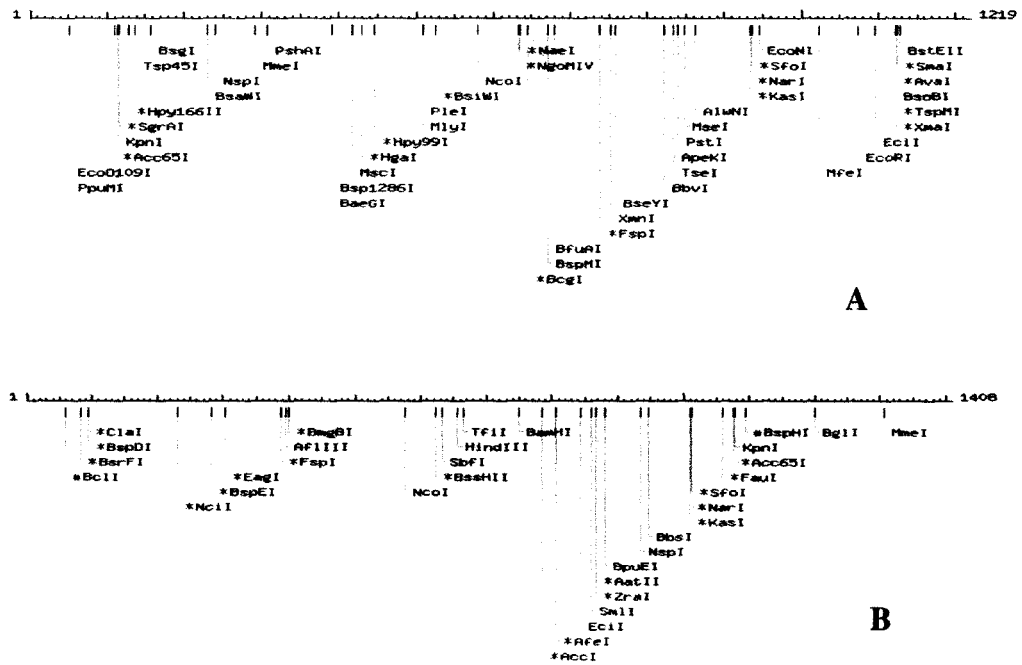

Fig 5.5: Restriction map of **A**, IS*T445* and **B**, IS*T2* prepared by using NEBCutter v 2.0.

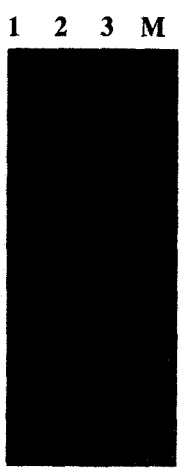

Fig 5.6: Distribution of IS*T2* IS element in the genome of *A. ferrooxidans* strains. Genomic DNA was digested with *Eco*RI. Lanes: **1**, *E. coli,* **2**, DK1, and **3**, NCIB 8455. M, is the marker DNA;sizes of the DNA fragments are given.

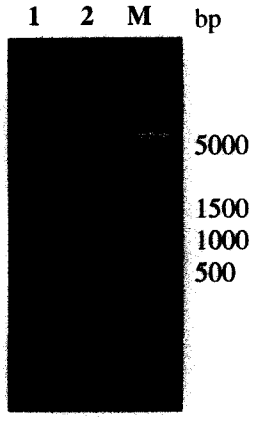

Fig 5.7: *Pst*I digestion of the DNA fragments generated by IS*T445* primers. 1, DK1, 2, NCIB8455, M, 500 bp DNA ladder

6.3.2.2 Cloning and sequence analysis of the unknown amplicon generated with IS*T445* specific internal primers

The 1.1 kb amplicon having no *Pst*I restriction site was cloned in pGEM-T easy vector and the recombinant plasmid was named as pAR2. The plasmid was transformed into the competent cells of *E. coli* XL1 Blue. After selecting the transformants via blue-white screening method, plasmid was prepared and the nucleotide sequence of the insert was determined. The complete nucleotide sequence of the amplicon containing 1081 bases was obtained (GQ 305308) using SP6 and T7 sequencing primers. Interestingly this sequence was found to bear inverted repeats at left and right termini identical to terminal repeats of IS*T*445 (Fig 5.8). The restriction map of pAR2 sequence also has been presented here showing no *Pst*I site (Fig 5.9).

6.3.2.3 Analysis of pAR2 sequence

BlastN was performed using NCBI tool which revealed that the sequence is 98.8% homologous to the glutamyl-tRNA reductase of *Acidithiobacillus ferrooxidans* 23270 (Fig 5.10). However sequence did not show any homology with the first 9 bases and the last 10 bases of pAR2 nucleotide sequence within which the actual inverted terminal repeat is present, indicating that the gene may be intervened by the inverted repeats of IS*T445*.

```
(A1)>>>>>>>>>>>>>>
ggctcttctgcggattgaggctcatccttcggcgtcactgagacgaaagagaatatccagtgccgcaac
cagactttcgttggtcgcgtcctgacaaggctggcgaagggtcgcgatggggtcgtgaagaactttatt
catcaaggctttggagaaggcttccagaaccgcacgagggtcctgcccctgatccagataatggctaaa
acgccgtagctcctcctggcgacgtccctccacatgatcacgcaggcgacggatggcgggcaccacgtc
cagactctcccgccattgctggaactcgccgacctcatcggcaatgatcagctccgccgccgacgccgc
ctcgcggcgggcgcgcatgcccgcctgggcgatatcgttcagatcgtccagcgtatagagaaaacactg
cgcaacaccttccacctccggggcgatatcccggggtacggcaagatccaccagcatcaggtcacccccg
cgcccgccgtgccatgactgcggagatcgtttcccgcgtcactatgggcagcaggctcgcggtgcaact
gaccaccacatcggcatcgtgcaacagatgggggatggcctccagcgcatgggagtcgcccgtaaattt
ttcggcgagttgctgaccgcgttccgcactgcggttggctacggcgaatctctccaccccatgctcgcg
cagatgggtcgccaccagctcgatggtatccccggcaccgatcaacagcaccgatttgccttcgagact
gcccagcaactgcttggccaaacatacggcggcataggcgacactgaccggggccgaaccaatggcggt
ctccgagcgcacccgctttgccacgcggaaggcccagtgcagcaggcgattcaataccggcccggcggc
accgttatccgccgccgcctgataggcatccttgacctgaccgagaatctgcggctctccgatgatcat
ggaatccagcccgcaggccacgcgaaacaggtggcgtacggcttccgcatcgctcgaatggtaaatatg
cccatccagcaatcgcgggtccatgccgtgaaaatgacgaagagcc
                          <acttttactgcttctcgg<          (A2)
```

Fig 5.8: Nucleotide sequence of the DNA insert of pAR2. Inverted repeats are underlined and primer sequenced is highlighted in bold letters. '<' and '>' mark the direction of primers

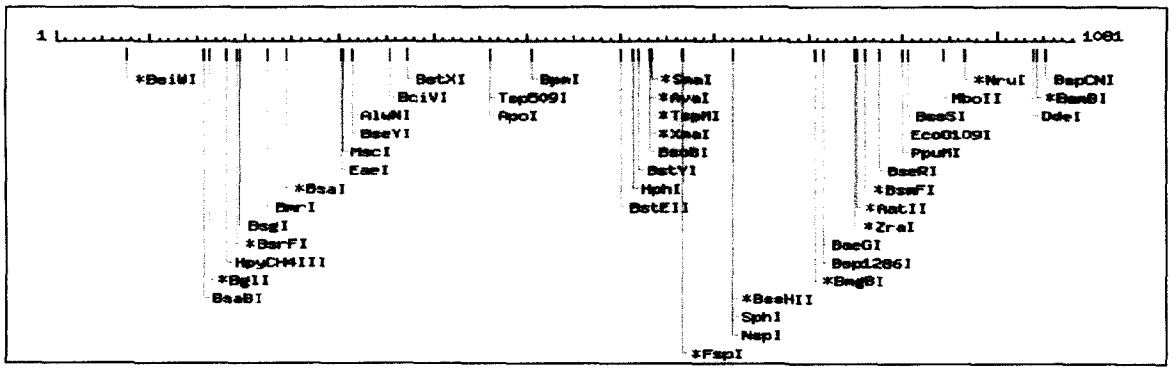

Fig 5.9: Restriction map of pAR2 insert using NEBCutter v2.0

```
PAR2    10        GCGGATTGAGGCTCATCCTTCGGCGTCACTGAGACGAAAGAGAATATCCAGTGCCGCAAC  69
                  ||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Aferr   2456312   GCGGACTGAGGCTCATCCTTCGGCGTCACTGAGACGAAAGAGAATATCCAGTGCCGCGAC  2456371
PAR2    70        CAGACTTTCGTTGGTCGCGTCCTGACAAGGCTGGCGAAGGGTCGCGATGGGGTCGTGAAG  129
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr   2456372   CAGACTTTCGTTGGTCGCGTCCTGACAAGGCTGGCGAAGGGTCGCGATGGGGTCGTGAAG  2456431
PAR2    130       AACTTTATTCATCAAGGCTTTGGAGAAGGCTTCCAGAACCGCACGAGGGTCCTGCCCCTG  189
                  |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Aferr   2456432   AACTTTATTCATCAAGGCCTTGGAGAAGGCTTCCAGAACCGCACGAGGGTCCTGCCCCTG  2456491
PAR2    190       ATCCAGATAATGGCTAAAACGCCGTAGCTCCTCCTGGCGACGTCCCTCCACATGATCACG  249
                  |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Aferr   2456492   ATCCAGATAATGGCTAAAACGCCGTAGCTCCTCCTGGCGTCGTCCCTCCACATGATCACG  2456551
PAR2    250       CAGGCGACGGATGGCGGGCACCACGTCCAGACTCTCCCGCCATTGCTGGAACTCGCCGAC  309
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr   2456552   CAGGCGACGGATGGCGGGCACCACGTCCAGACTCTCCCGCCATTGCTGGAACTCGCCGAC  2456611
PAR2    310       CTCATCGGCAATGATCAGCTCCGCCGCCGACGCCGCCTCGCGGCGGGCGCGCATGCCCGC  369
                  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr   2456612   CTCCTCGGCAATGATCAGCTCCGCCGCCGACGCCGCCTCGCGGCGGGCGCGCATGCCCGC  2456671

PAR2    370       CTGGGCGATATCGTTCAGATCGTCCAGCGTATAGAGAAAACACTGCGCAACACCTTCCAC  429
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr   2456672   CTGGGCGATATCGTTCAGATCGTCCAGCGTATAGAGAAAACACTGCGCAACACCTTCCAC  2456731
```

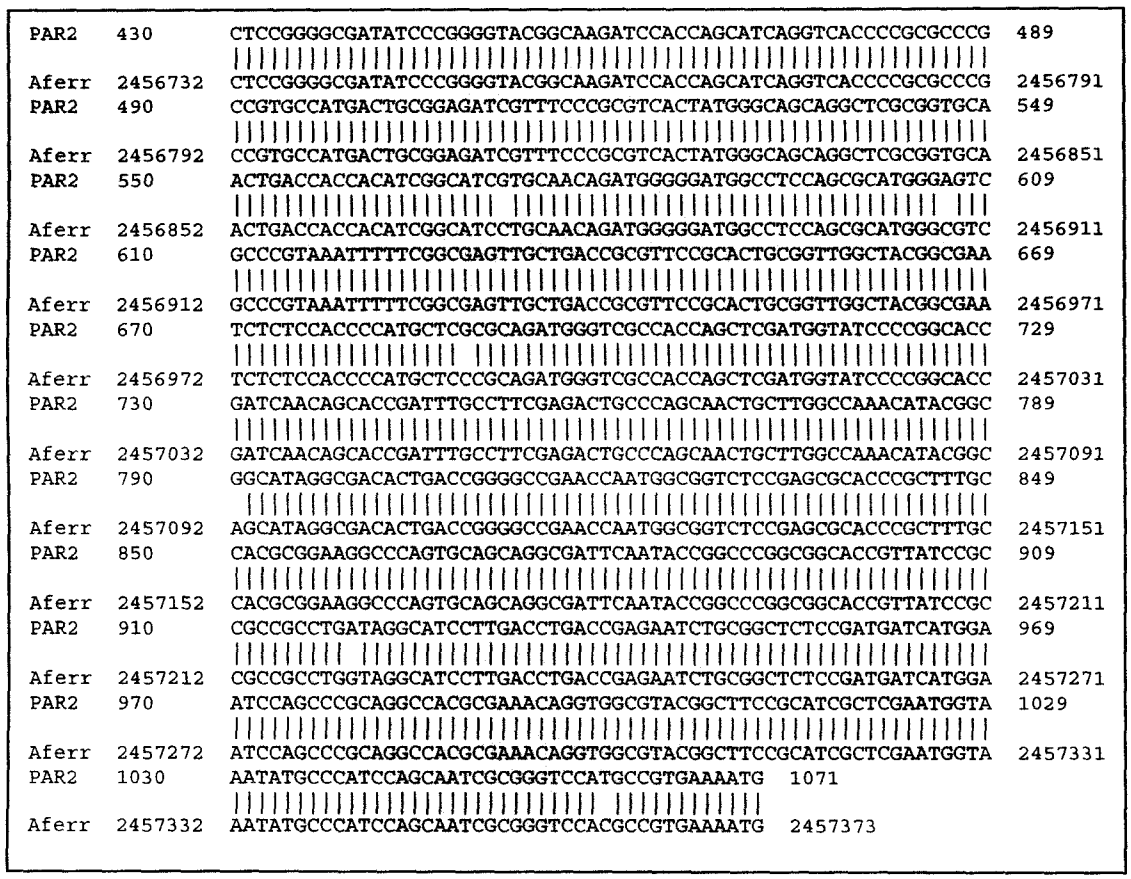

Fig 5.10: BlastN result of pAR2 nucleotide sequence showing 98.8% homology with the '+' strand of *A. ferrooxidans* 23270.

Further sequence analysis revealed that -2 frame of the sequence contained an open reading frame (ORF) of 345 amino acids starting with the ATG (Methionine codon) as the initiation codon and TGA as terminating codon located at 1059th and 24th nucleotide respectively (Fig 5.11). Pairwise alignment study revealed that the sequence is 98.6% homologous to the amino acid sequence of *A. ferrooxidans* glutamyl tRNA reductase gene retrieved from the website http://www.ncbi.nlm.nih.gov/protein/YP_002220789.1 (Fig 5.12). The same study was also made with the representative molecule of glutamyl tRNA reductase (1gpj_a) of *Methanoparus kandleri* (Moser *et al.*, 2001) which was already crystallized and whose 3D structure has been determined. The result showed a significant similarity of 36% homology with 0.7% of gap frequency (Fig 5.13). Although the sequence showed very high similarity with the glutamyl tRNA reductase gene of *A. ferrooxidans* pairwise alignment revealed that the pAR2 sequence is truncated at the N-terminal end by 91 amino acids. Similarly when compared with 1gpj_a it was found that 63 amino acid missing from the N terminal end. However in both the cases the C-terminal end showed no deletion in the sequence.

```
1059 atggacccgcgattgctggatgggcatatttaccattcgagcgat
     M  D  P  R  L  L  D  G  H  I  Y  H  S  S  D
1014 gcggaagccgtacgccacctgtttcgcgtggcctgcgggctggat
     A  E  A  V  R  H  L  F  R  V  A  C  G  L  D
 969 tccatgatcatcggagagccgcagattctcggtcaggtcaaggat
     S  M  I  I  G  E  P  Q  I  L  G  Q  V  K  D
 924 gcctatcaggcggcggcggataacggtgccgccgggccggtattg
     A  Y  Q  A  A  A  D  N  G  A  A  G  P  V  L
 879 aatcgcctgctgcactgggccttccgcgtggcaaagcgggtgcgc
     N  R  L  L  H  W  A  F  R  V  A  K  R  V  R
 834 tcggagaccgccattggttcggccccggtcagtgtcgcctatgcc
     S  E  T  A  I  G  S  A  P  V  S  V  A  Y  A
 789 gccgtatgtttggccaagcagttgctgggcagtctcgaaggcaaa
     A  V  C  L  A  K  Q  L  L  G  S  L  E  G  K
 744 tcggtgctgttgatcggtgccggggataccatcgagctggtggcg
     S  V  L  L  I  G  A  G  D  T  I  E  L  V  A
 699 acccatctgcgcgagcatggggtggagagattcgccgtagccaac
     T  H  L  R  E  H  G  V  E  R  F  A  V  A  N
 654 cgcagtgcggaacgcggtcagcaactcgccgaaaaatttacgggc
     R  S  A  E  R  G  Q  Q  L  A  E  K  F  T  G
 609 gactcccatgcgctggaggccatcccccatctgttgcacgatgcc
     D  S  H  A  L  E  A  I  P  H  L  L  H  D  A
 564 gatgtggtggtcagttgcaccgcgagcctgctgcccatagtgacg
     D  V  V  V  S  C  T  A  S  L  L  P  I  V  T
 519 cgggaaacgatctccgcagtcatggcacggcgggcgcggggtgac
     R  E  T  I  S  A  V  M  A  R  R  A  R  G  D
 474 ctgatgctggtggatcttgccgtaccccgggatatcgccccggag
     L  M  L  V  D  L  A  V  P  R  D  I  A  P  E
 429 gtggaaggtgttgcgcagtgttttctctatacgctggacgatctg
     V  E  G  V  A  Q  C  F  L  Y  T  L  D  D  L
 384 aacgatatcgcccaggcgggcatgcgcgcccgccgcgaggcggcg
     N  D  I  A  Q  A  G  M  R  A  R  R  E  A  A
 339 tcggcggcggagctgatcattgccgatgaggtcggcgagttccag
     S  A  A  E  L  I  I  A  D  E  V  G  E  F  Q
 294 caatggcgggagagtctggacgtggtgcccgccatccgtcgcctg
     Q  W  R  E  S  L  D  V  V  P  A  I  R  R  L
 249 cgtgatcatgtggagggacgtcgccaggaggagctacggcgtttt
     R  D  H  V  E  G  R  R  Q  E  E  L  R  R  F
 204 agccattatctggatcaggggcaggaccctcgtgcggttctggaa
     S  H  Y  L  D  Q  G  Q  D  P  R  A  V  L  E
 159 gccttctccaaagccttgatgaataaagttcttcacgaccccatc
     A  F  S  K  A  L  M  N  K  V  L  H  D  P  I
 114 gcgacccttcgccagccttgtcaggacgcgaccaacgaaagtctg
     A  T  L  R  Q  P  C  Q  D  A  T  N  E  S  L
  69 gttgcggcactggatattctctttcgtctcagtgacgccgaagga
     V  A  A  L  D  I  L  F  R  L  S  D  A  E  G
  24 tga 22
     *
```

Fig 5.11: ORF obtained from the -2 reading frame of the sequence of pAR2.

```
pAR2,          2 LRHFHGMDPRLLDGHIYHSSDAEAVRHLFRVACGLDSMIIGEPQILGQVKDAYQAAADNG
Afer232,      93 LCHFHGVDPRLLDGHIYHSSDAEAVRHLFRVACGLDSMIIGEPQILGQVKDAYQAAADNG
                 * **** ************************************************

pAR2,         62 AAGPVLNRLLHWAFRVAKRVRSETAIGSAPVSVAYAAVCLAKQLLGSLEGKSVLLIGAGD
Afer232,     153 AAGPVLNRLLHWAFRVAKRVRSETAIGSAPVSVAYAAVCLAKQLLGSLEGKSVLLIGAGD
                 ************************************************************

pAR2,        122 TIELVATHLREHGVERFAVANRSAERGQQLAEKFTGDSHALEAIPHLLHDADVVVSCTAS
Afer232,     213 TIELVATHLREHGVERFAVANRSAERGQQLAEKFTGDAHALEAIPHLLQDADVVVSCTAS
                 ************************************ ********** ***********

pAR2,        182 LLPIVTRETISAVMARRARGDLMLVDLAVPRDIAPEVEGVAQCFLYTLDDLNDIAQAGMR
Afer232,     273 LLPIVTRETISAVMARRARGDLMLVDLAVPRDIAPEVEGVAQCFLYTLDDLNDIAQAGMR
                 ************************************************************

pAR2,        242 ARREAASAAELIIADEVGEFQQWRESLDVVPAIRRLRDHVEGRRQEELRRFSHYLDQGQD
Afer232,     333 ARREAASAAELIIAEEVGEFQQWRESLDVVPAIRRLRDHVEGRRQEELRRFSHYLDQGQD
                 ************** ********************************************

pAR2,        302 PRAVLEAFSKALMNKVLHDPIATLRQPCQDATNESLVAALDILFRLSDAEG
Afer232,     393 PRAVLEAFSKALMNKVLHDPIATLRQPCQDATNESLVAALDILFRLSDAEG
                 ***************************************************
```

Fig 5.12: Pairwise alignment of the amino acid sequence of pAR2 and the glutamyl tRNA reductase of *A. ferrooxidans* 23270.

```
pAR2,          16 AEAVRHLFRVACGLDSMIIGEPQILGQVKDAYQAAADNGAAGPVLNRLLHWAFRVAKRVR
igpj_a,        79 SEAVRHLFRVASGLESMMVGEQEILRQVKKAYDRAARLGTLDEALKIVFRRAINLGKRAR
                   ********** ** **  **  ** *** **  **  *     *      *    ** *

pAR2,          76 SETAIGSAPVSVAYAAVCLAKQLLGSLEGKSVLLIGAGDTIELVATHLREHGVERFAVAN
igpj_a,       139 EETRISEGAVSIGSAAVELAERELGSLHDKTVLVVGAGEMGKTVAKSLVDRGVRAVLVAN
                   ** *    **   *** **   ****  * **  ***     **  *   **    ***

pAR2,         136 RSAERGQQLAEKFTGDSHALEAIPHLLHDADVVVSCTASLLPIVTRETISAVMARR-ARG
igpj_a,       199 RTYERAVELARDLGGEAVRFDELVDHLARSDVVVSATAAPHPVIHVDDVREALRKRDRRS
                   *  **   **    *            *   ***** **   *             *  *

pAR2,         195 DLMLVDLAVPRDIAPEVEGVAQCFLYTLDDLNDIAQAGMRARREAASAAELIIADEVGEF
igpj_a,       259 PILIIDIANPRDVEEGVENIEDVEVRTIDDLRVIARENLERRRKEIPKVEKLIEEELSTV
                        * * ***    **        * ***  **      **      *  *  *

pAR2,         255 QQWRESLDVVPAIRRLRDHVEGRRQEELRRFSHYLDQGQDPRAVLEAFSKALMNKVLH
igpj_a,       319 EEELEKLKERRLVADVAKSLHEIKDRELERALRRLKTG-DPENVLQDFAEAYTKRLIN
                     *                      ** *   *  * ** **  *  *
```

Fig 5.13: Pairwise alignment of the amino acid sequence of pAR2 with 1gpj_a molecule showing 36% homology with 0.7% gap frequency.

When BlastP was performed with the help of the NCBI tool given in the website http://www.ncbi.nlm.nih.gov/blast/Blast.cgi the result showed the sequence to have the putative conserved domain, where the first 230 amino acids sequence showed the homology with the NADP-binding domain of glutamyl tRNA reductase which belongs to

NADB_Rossmann superfamily whereas the sequence starting from 246 to 344 homology with dimerization domain of glutamyl tRNA reductase (Fig 5.14).

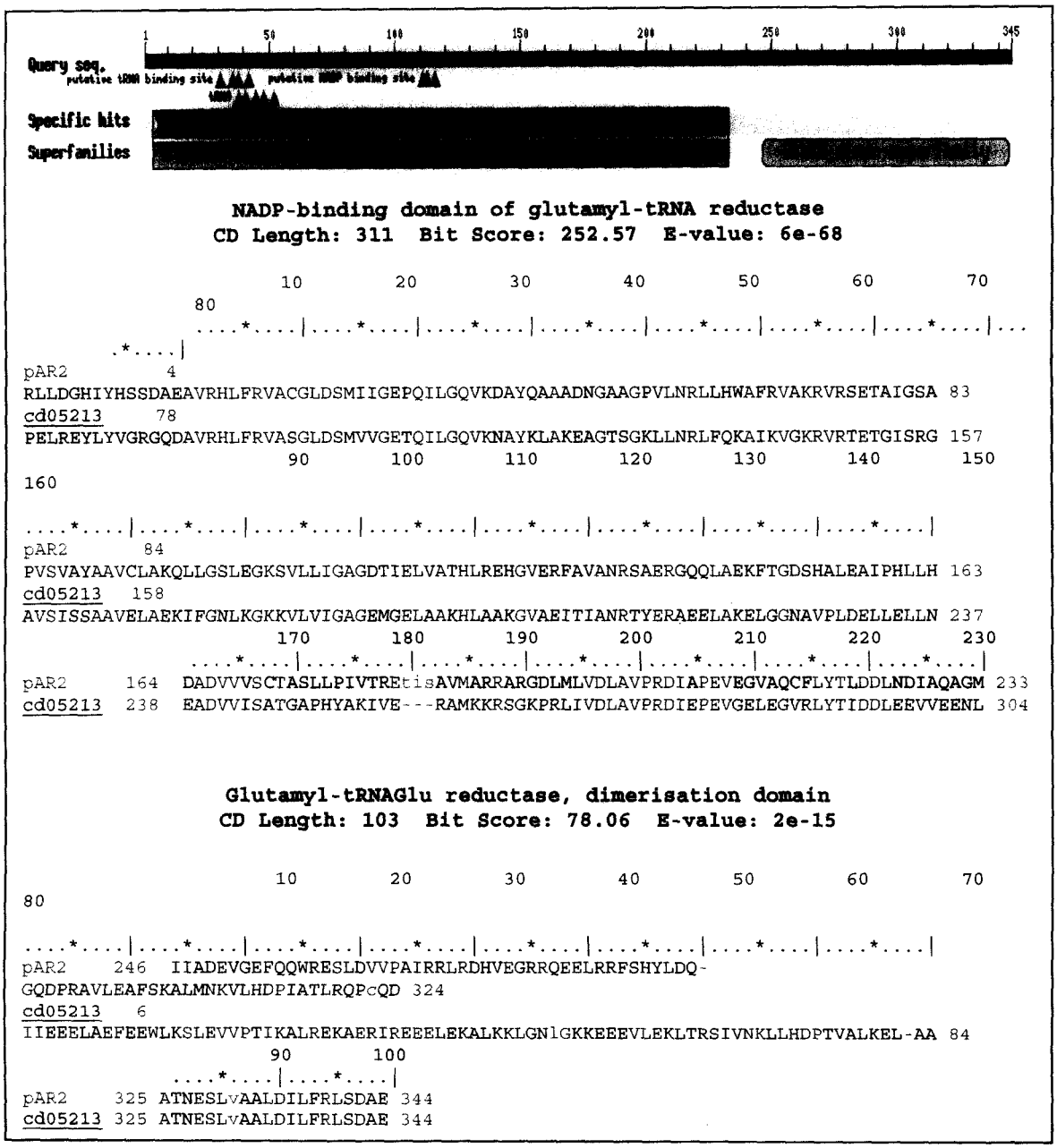

Fig 5.14: The amino acid sequence of pAR2 showing the conserved domain sequence of glutamyl tRNA reductase.

Although the ORF typically began with the ATG codon but no typical Shine-Dalgarnho sequence AGGAGG was found at upstream of the initiation codon. It is likely to infer that the sequence is truncated at N-terminal point *i.e.* the ATG codon which is being referred here as the initiation codon, is actually not the initiation codon but only an

intermediate codon. The above results have already substantiated the probability of truncation at N-terminal end. However with the help of the software package SWISS-pdb Viewer v3.7 the three dimensional structures of 1gpj_a as well as pAR2 could be created and compared (Fig 5.15 and 5.16). The structure shows very high similarity between these two. Like 1gpj_a which remains in dimer form the dimer form of pAR2 also could be generated (Fig 5.17). Glutamyl-tRNA reductase molecule from *Methanoparus kandleri* has been extensively studied. The molecule has been detected to remain in dimer form containing two monomers each having a length of ~125A° and together they span a width of ~165A°. Each monomer contains three distinct linearly arranged domains. Domain I is the N-terminal domain and consists of two sub-domains. Sub-domain I (1-77 residues) contains four β-sheets and 3 α-helices which are arranged in βαββααβ-motif typical to the RNA binding protein. Sub-domain II consists of three anti-parallel α-helices. Together these two sub-domains form a distorted anti-parallel helix bundle. Domain II (149-285 residues) is composed of central six stranded β-sheets linked to the Domain I by a short linker residues (142-148). Domain two has the classical nucleotide-binding fold. Domain III(355-360) is composed of three α-helices and is the dimerization domain. It is linked to the domain II by a 'spinal' α-helix (286-352) comprising of 18 α-helical turns (Moser *et al.*, 2001). Even after seeing very high similarity with the representative molecule, the functionality of the truncated glutamyl-tRNA reductase is yet to be tested following expression coupled to structure-function studies.

Fig 5.15: 3D structure of **A**, 1GPJA; **B,** pAR2. In case of 1GPJA, the first 66 amiino acids have been deleted from N-terminal to fit with pAR2

The nucleotide sequence of pAR2 showed almost 100% similarity with the sequence of glutamyl-tRNA of *A. ferrooxidans* except at the terminal inverted repeats (Fig 5.10), this truncated nature of pAR2 may be the consequence of integration of the inverted sequence in the gene.

Fig 5.16: 3D structure of **A**, 1GPJA; **B**, pAR2. All the amino acids have been accounted.

6.3.2.4 Restriction fragment analysis of amplicon generated with IS*T2* specific internal primers

Just like IS*T445* when IS*T2* was PCR amplified from the Garubathan strains DNA fragment size of around 1.4kb could be amplified (Fig 5.4). The DNA fragment was analyzed by digesting the fragment with *Hin*dIII enzyme, the recognition site for which is present in IS*T2* sequence at 658 nucleotide position. The result showed that in case of DK1 just like in case of IS*T445*, three DNA fragments were found while in case of control DNA sample NCIB 8455 two DNA fragments corresponding to size of amplified fragment were generated (Fig 5.18). This led to conclude that during the amplification of IS*T2* another DNA fragment also got amplified.

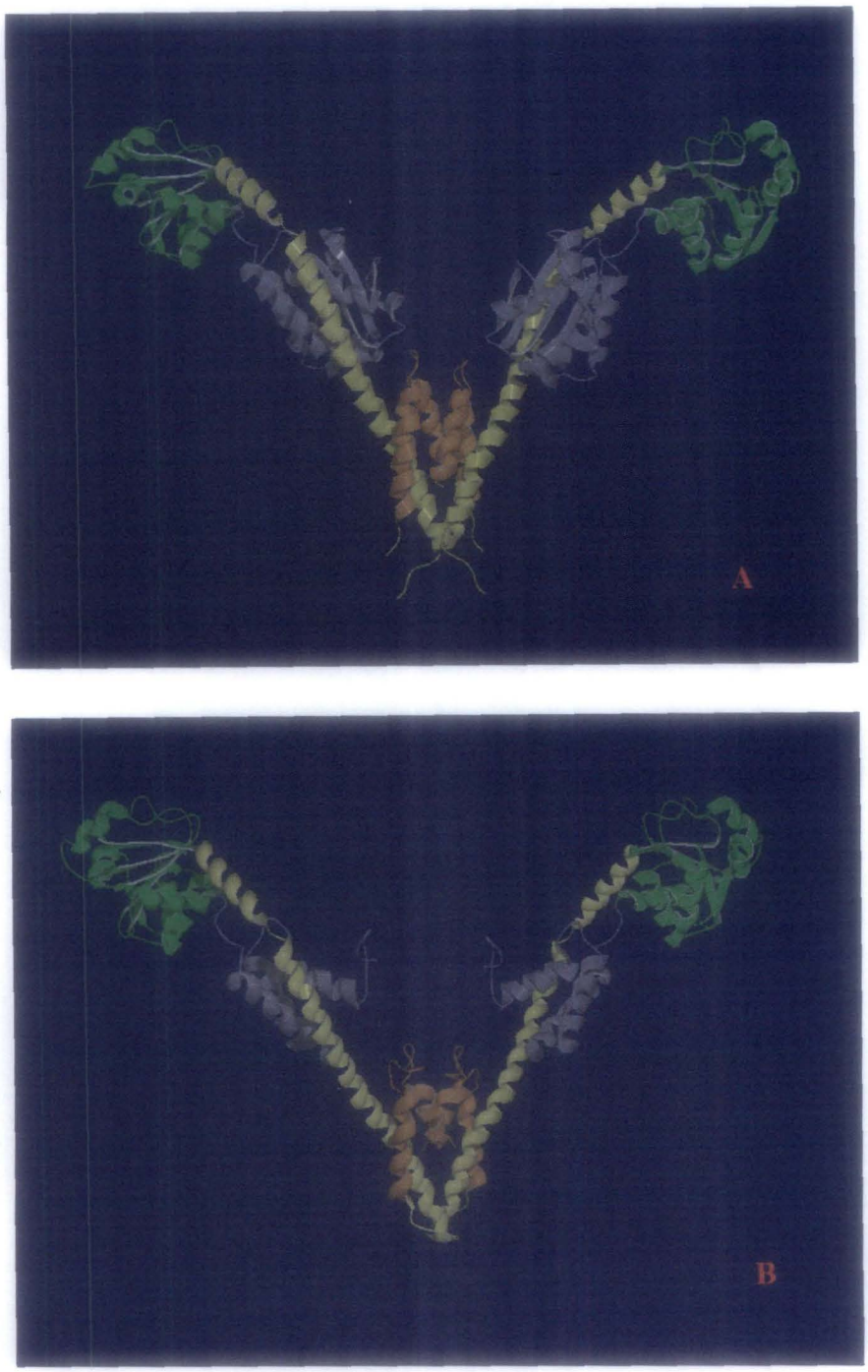

Fig 5.17: Dimer structure of glutamyl tRNA . **A**, 1GPJA; **B**, pAR2

6.3.2.5 Cloning of unknown PCR amplicon generated with IS*T2* internal primers and construction pAR4 clone

The amplicon of DK1 generated with IS*T2* internal primers (which did not bear the *Hin*dIII site) was cloned in pGEM-T easy vector and the recombinant plasmid was screened, selected and named as pAR4. Minipreparaton followed by maxipreparation of the plasmid was carried out and sequencing of the insert was carried out using T7 and SP6

primers. Two sequences (part A 568bp and part B 581bp) obtained from both ends were presented; complete sequence was not done with internal primer walking (Fig 5.19 A and B). Interestingly here also the two ends of the sequence had identical inverted repeats of IS*T2*.

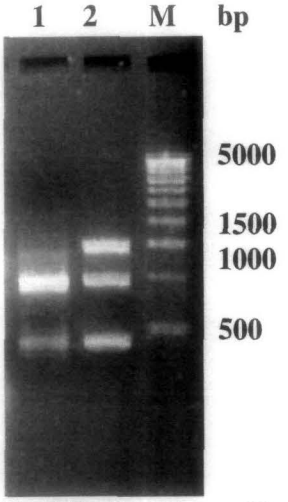

Fig 5.18: *Hind*III digestion of Lane **1**, NCIB8455 and lane **2**, DK1. **M**, 500 bp DNA ladder

(G3)

GAGCTATAGTCAAATCTGGCATGGTGGCAGGGCGTGTCATCCTGATCTCAATAATGGGGCAGTGCCCGTGGCGCCCAGCGCCATCGCC
ATTGCCGGTGAAGAAGTACATACCCCCAAAGGCAAGCAGACATACACCACCCCGCGTGCTCTGGCTGACGACGAAATCCCCGGGATTG
TCGAAGGCTTTGCCCAGGCTGCCCGCAATGCCCGGTTGGCCGGTTTTGACCTGGTAGAAGTGCATGGCGCCAACGGTTATTTGCTGGA
TGAATTTTTGCGGGATGGTGCCAATCATCGCCAGGGACCCTATGGTGGAGATCGGGAAAATCGCGCCCGTTTTCTTTTTGAAGTTCTG
GCGGCAGTCTGCCAGGCGGTCGATGCAGATCGTGTCGGAGTGCGCCTTTCGCCACTGAACAGTTTTAACGATATGCGGGATAGTGATC
CCGTTGGTTTGCTATCCTGGCTTGCTGGTCGACTGAATGATTTCGGATTGGCCTATTGCATCTGATGCGCGCAGATTTTTATGGACGT
CAGACCGGAGATGTCCTGGCACCAGCCCGGAAACATTTTC (568)

[A]

(G4)

GAGCTATGCTCGAAAGTGGAGGCGTGCAGTGGATGACCGCAGGGCGGGGCGTGATTCACTCGGAAATTCCGGAACAGGAAAAAGGCGT
CATGGAGGGATTTCAGCTCTGGCTCAACCTGCCGAGCCAGGACAAGATGGCACCGCCGTGGTATCGCGATTTCTCTGCCGCAGATTTA
CCGCGCTTTGTGACTCAGGAAGGAGTCGTCGTCACGGTAATTGCCGGAGAGAGTCATGGAGTCATTGGCGCGGTAACGCGGGAGAGCA
CAAATCCGCTCTATCTGGACCTGCAATTGCCGGAAGGCGCACATTTCAAACAGCAACTGCCTGCTGATTACAACGCCTTTCTTTATGC
ATATCGCGGTACCCTGGAAATCGTCGGCACACCTGTGTCTGTCCAGAGCATGGCCATACTCGAGAATGGCCCGGAAACCGACGGTGTG
GACATCAAGGCGCTTACGAGTGCACGCTTACTGCTGATTGCGGGCCAGCCGTTGCATGAGCCAATTGTCCAGTATGGACCATTTGTCA
TGAATACTCGGGAGGAAATATACCAGGCTTTTGAGGATATGAAGGAAGGCCGG (581)

[B]

Fig 5.19: Nucleotide sequence of pAR4; **A**, sequence done with SP6 primer; **B**, sequence done with T7 primer. Primer sequence is highlighted by bold letters and inverted repeats are underlined.

6.3.2.6 Analysis of the sequence of pAR4 clone

BlastN was conducted with both the sequences. The nucleotide sequence (part B) that was determined using T7 primer (Fig 5.20) showed 98% similarity with *A. ferrooxidans* ATCC 23270 genome sequence. The other part of the sequence (part A) showed 67% homology with NADH:flavin oxidoreductase/flavin oxidase gene of *Magnetococcus sp.* but no homology with *A. ferrooxidans* DNA (Fig 5.21).

```
PAR4   7         TGCTCGAAAGTGGAGGCGTGCAGTGGATGACCGCAGGGCGGGGCGTGATTCACTCGGAAA  66
                 |||||||||| |||||||||| |||||||||||||||| ||||||||||||||||||||
Aferr  2904995   TGCTCGAAAGCGGAGGCGTGCAATGGATGACCGCAGGGCGAGGCGTGATTCACTCGGAAA  2905054

PAR4   67        TTCCGGAACAGGAAAAAGGCGTCATGGAGGGATTTCAGCTCTGGCTCAACCTGCCGAGCC  126
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr  2905055   TTCCGGAACAGGAAAAAGGCGTCATGGAGGGATTTCAGCTCTGGCTCAACCTGCCGAGCC  2905114

PAR4   127       AGGACAAGATGGCACCGCCGTGGTATCGCGATTTCTCTGCCGCAGATTTACCGCGCTTTG  186
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr  2905115   AGGACAAGATGGCACCGCCGTGGTATCGCGATTTCTCTGCCGCAGATTTACCGCGCTTTG  2905174

PAR4   187       TGACTCAGGAAGGAGTCGTCGTCACGGTAATTGCCGGAGAGAGTCATGGAGTCATTGGCG  246
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Aferr  23270     TGACTCAGGAAGGAGTCGTCGTCACGGTAATTGCCGGAGAGAGTCATGGAGTCACTGGCG  2905234

PAR4   247       CGGTAACGCGGGAGAGCACAAATCCGCTCTATCTGGACCTGCAATTGCCGGAAGGCGCAC  306
                 |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Aferr  23270     CGGTAACGCGGGAGAGCACAAATCCACTCTATCTGGACCTGCAATTGCCGGAAGGCGCAC  2905294

PAR4   307       ATTTCAAACAGCAACTGCCTGCTGATTACAACGCCTTTCTTTATGCATATCGCGGTACCC  366
                 |||||||||||||||||| |||||||||||||||||||||||||||||||| ||||
Aferr  23270     ATTTCAAACAGCAACTGCCGGCTGATTACAACGCCTTTCTTTATGCATATCGCGGAACCC  2905354

PAR4   367       TGGAAATCGTCGGCACACCTGTGTCTGTCCAGAGCATGGCCATACTCGAGAATGGCCCGG  426
                 |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Aferr  23270     TGGAAATCGTCGGCACACCTGTGTCTGTCCAGCGCATGGCCATACTCGAGAATGGCCCGG  2905414

PAR4   427       AAACCGACGGTGTGGACATCAAGGCGCTTACGAGTGCACGCTTACTGCTGATTGCGGGCC  486
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr  23270     AAACCGACGGTGTGGACATCAAGGCGCTTACGAGTGCACGCTTACTGCTGATTGCGGGCC  2905474

PAR4   487       AGCCGTTGCATGAGCCAATTGTCCAGTATGGACCATTTGTCATGAATACTCGGGAGGAAA  546
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr  23270     AGCCGTTGCATGAGCCAATTGTCCAGTATGGACCATTTGTCATGAATACTCGGGAGGAAA  2905534

PAR4   547       TATACCAGGCTTTTGAGGATATGAAGGAAGGCCGG  581
                 |||||||||||||||||||||||||||||||||||
Aferr  23270     TATACCAGGCTTTTGAGGATATGAAGGAAGGCCGG  2905569
```

Fig 5.20: BlastN result of part B nucleotide sequence of pAR4 showing 98% homology with the *A.ferrooxidans* 23270 nucleotide sequence.

Three ORFs 192, 109 and 60 amino acids were obtained from part B sequence of pAR4 insert (Fig 5.22 A, B and C) from +3, +2 and -2 reading frame respectively.. The

ORF containing a sequence of 192 amino acids showed 98% homology with Pirin domain protein of *A.* ferrooxidans 53993. Beside that the presence of conserved domains for

Cupin_2 and Pirin_C superfamily were also found (Fig 5.23). The result showed that the first 36 amino acids have the homology with the C-terminal part of the Cupin_C while the last 101 amino acids starting from amino acid 91 showed the containment of Pirin_C superfamily. BlastP result of 109 long amino acid chain did not show any significant homology with any of the non-redundant GenBank CDS while 60 amino acids long ORF showed 33% homology with the ATP-dependent ligase of *Syenococcus* sp.

```
PAR4  134     GACGACGAAATCCCCGGTATTGTCGAAGGCTTTGCCCAGGCTGCCCGCAATGCCCGGTTG  193
              ||  ||||||  |  ||  |  |||||  |  ||  |||||||  ||  ||||||  |||||||
Mag   958330  GATGACGAACTGCCGGCCATTGTAAATGGGTTTGCCCGTGCGGCCCGCCATGCCCGCGCA  958389

PAR4  194     GCCGGTTTTGACCTGGTATAAGTGCATGGCGCCAACGGTTATTTGCTGGATGAATTTTTG  253
              |||||  |||||    ||||  |  ||||||||  |||||  ||  ||   ||||  |||  |||||  ||
Mag   958390  GCCGGGTTTGATGGGGTAGAGGTGCATGGGGCCAATGGCTACCTGCTCGATCAATTTCTG  958449

PAR4  254     CGGGATGGTGCCAATCATCCCCAGGGACCCTATGGTGGAGATCG--GGAAAATCGCGCCC  311
              ||||||||    |||  ||||  |    ||  |||||||||  |  |||   |||  ||  |||||||
Mag   958450  CGGGATGGCAGCAACCATCGCAGCGGTGCCTATGGTG--GCTCGGTGGAGAACCGCGCCC  958507

PAR4  312     GTTTTCTTTTTGAGGTTCTGGCGGCAGTCTGCCAGGCGATCGATGCA--GATCGTGTCGG  369
              ||||  |  ||  |||||  |  |  |||  ||||  ||  ||     |  |||  ||  ||  ||  ||
Mag   958508  GTTTGTTGTTGGAGGTGATTGAGGCGGTCTCCCTAGC--CTGGGGCAGTGAGCGGGTGGG  958565

PAR4  370     AGTGCGCCTTTCGCCACTGAACAGTTTTAACGATATGCGGTATAGTGATCCCGTTGGTTT  429
              ||||  |  ||  ||  ||  |||||  ||||||||||||     |  ||  |||||  |  ||  ||
Mag   958566  GCTGCGGATCTCTCCCCTTAACAGCTTTAACGATATGATCGACAGCGATCCTATGGGGTT  958625

PAR4  430     GCTATCCTGGCTTGCTGGTCGACTGAATGATTTCGGATTGGCCTATTTGCATCTGATGCG  489
              |      |||||  ||    |||||  ||  |  |    ||||||||||  ||||  ||||||||
Mag   958626  GGCGGAGTGGCTGGCCAAGCGACTAAACGCCTATCCATTGGCCTATCTGCACCTGATGCG  958685

PAR4  490     -CGCAGATTTTTATGGACGTCGGACCGGAGATGTCCTGGCACCAGCCCGGAAACATTTTC  548
               |||  ||  ||||  ||  |  |     ||||||||  |  |  |  |  |||  |  ||  |  |
Mag   958686  TCGC-GACTTTTTGGGTCAACAACAGGGAGATGTGGTTCCCCTGGTACGGCAGCACTATA  958744

PAR4  549     AGGGCCCCTTGATTGTCAACATGGGTTA  576
              |||||  ||  ||||  |  ||  ||||||||
Mag   958745  AGGGCACCCTGATCGGTAATATGGGTTA  958772
```

Fig 5.21: BlastN result of part A nucleotide sequence of pAR2 showing 67% homology with *Magnetococcus* sp. MC-1 (Mag).

From part A nucleotide sequence of pAR4, three ORFs of 122, 70 and 38 amino acids were found from -1, +3 and -1 reading frame respectively (Fig 5.24). BlastP of 122 long amino acids sequence showed the highest homology of 25% only with transcriptional regulator of Ara family protein of *Stigmatella aurantiaca* while that of 70 amino acids sequence showed the highest homology of 41% with the NADH dehydrogenase subunit I

of *Streptomyces ghanaensis* ATCC 14672. On the other hand 38 amino acid sequence did not show any significant homology with any of the non-redundant GenBank CDS.

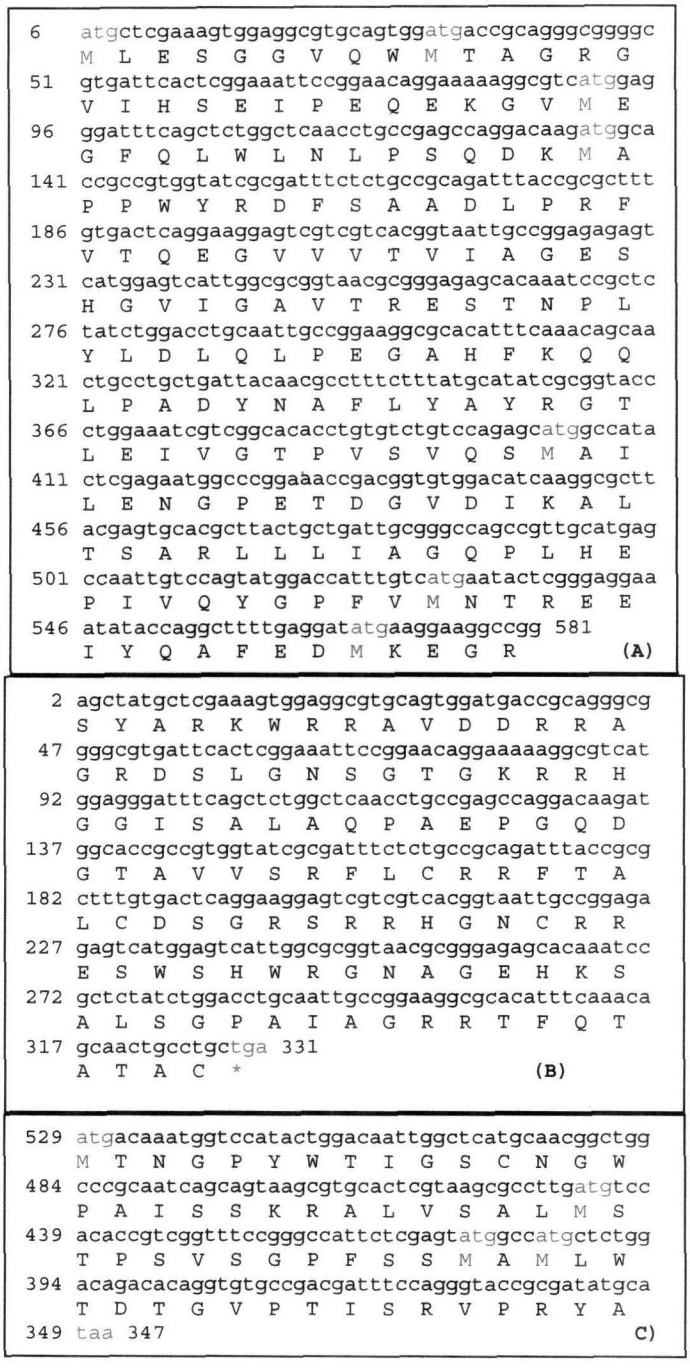

Fig 5.22: ORF obtained from part B of the pAR4 insert. **A,** 192 amino acids long at +3 frame; **B,** 109 amino acids long at +2 frame; and **C,** 60 amino acids at -2 reading frame.

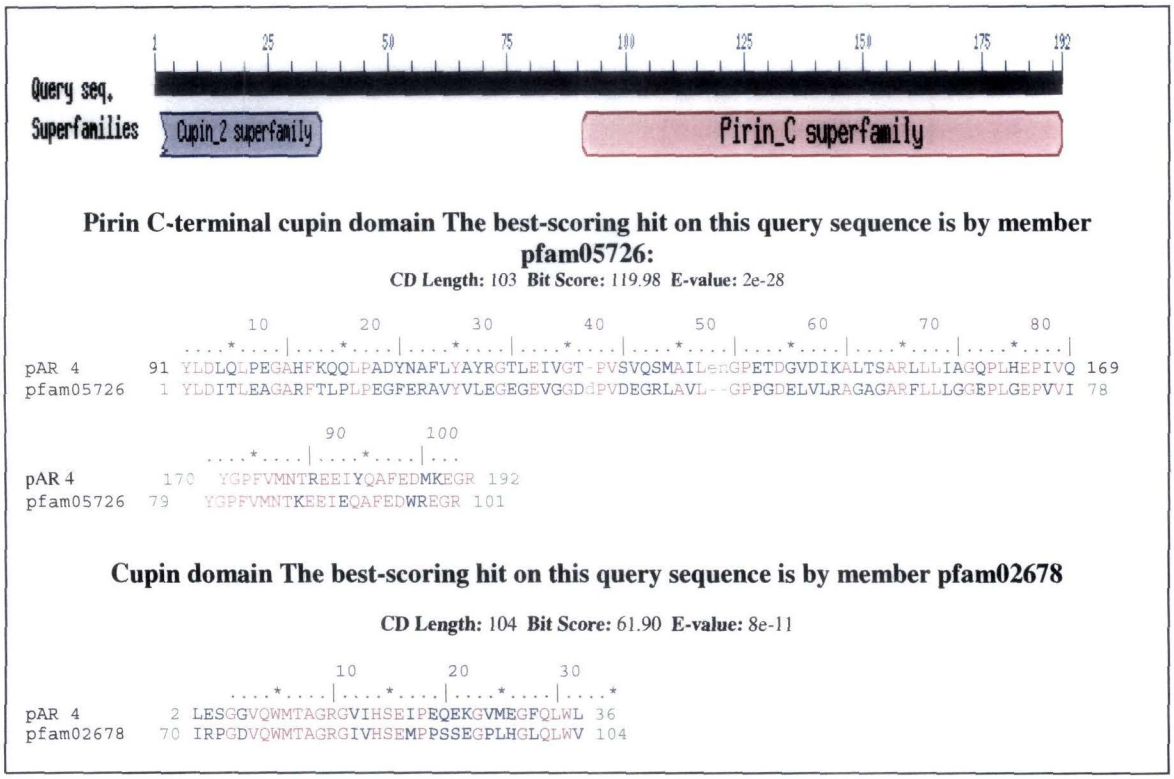

Fig 5.23: Amino acid sequence of partB of pAR4 showing conserved domain sequence of Pirin_C superfamily.

6.3.3 Ampification of inter-IS sequence using outwardly directed primers of IS*T445* and IS*T2*:

The idea behind the designing of outwardly directed primers sequences to amplify the inter IS regions has already been explained in the last two paragraphs of the introduction. The primers were named as RP1, RP2, RP3 and RP4 that will move in the direction opposite to G_3, G_4, A_1, and A_2 respectively (Fig 5.1 and Table 5.1). Using the combinations of primers having more or less same T_m value, PCR amplifications were conducted for different strains (both wild and colony morphology variants) and the profiles of amplified DNA bands specific to a particular strain were obtained. Comparison of 'fingerprint of amplified fragments' (FAF) of the wild with their colony morphology variants confirmed their mother-daughter identity. However FAFs derived from different mother strains of *A. ferrooxidans* (DK1, GBVI, CMO II, CMI I) established their distinctiveness. Moreover, presence of few amplified fragments of same size present in the profile of all the strains tested indicated the conservation of few inter-IS fragment which in evolution suffered minimum IS-mediated rearrangements.

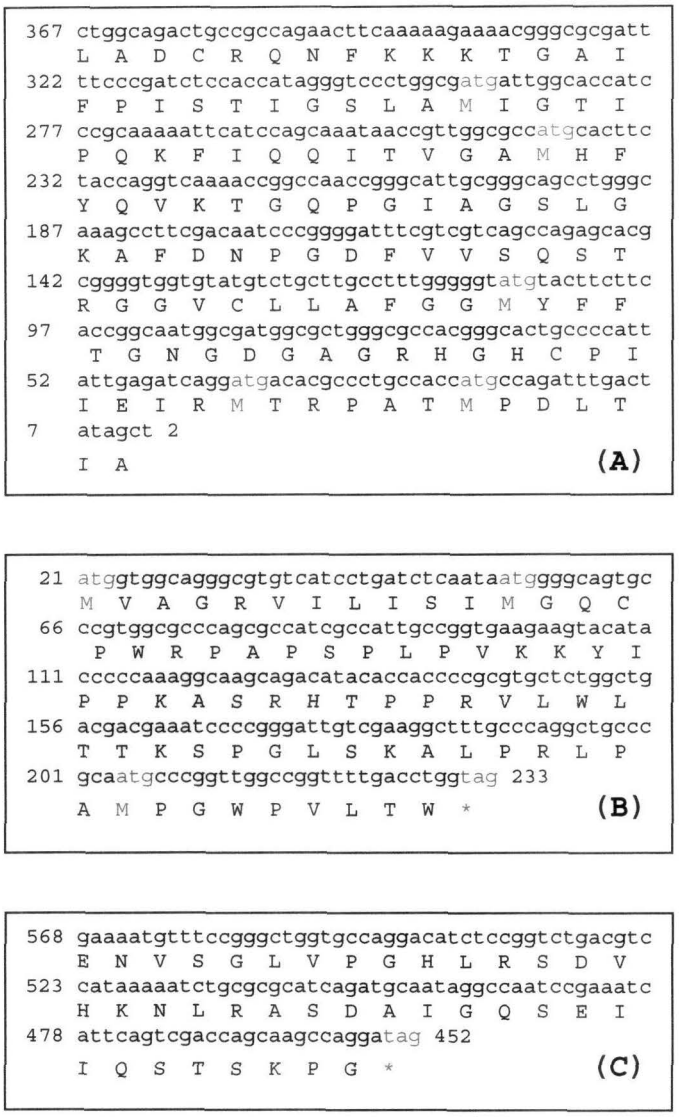

Fig 5.24: ORFs obtained from partA nucleotide sequence of pAR4 at: **A**, -1; **B**, +3, and **C**, -1 reading frame respectively.

Theoretically there are sixteen different possible combinations of primers (RP1-RP1, RP2-RP2, RP3-RP3, RP4-RP4, RP1-RP2, RP2-RP1, RP1-RP3, RP3-RP1, RP1-RP4, RP4-RP1, RP2-RP3, RP3-RP2, RP2-RP4, RP4-RP2, RP4-RP3 and RP3-RP4) (Fig 5.1). In practice, two combinations were optimized which yielded consistent results. The T_m values of different outwardly directed primers are summarized in the Table 5.1. These calculated T_m values have allowed selecting two combinations of primers i.e. RP1-RP4 and RP2-RP3. In other theoretical possible combinations the chances of nonspecific amplification of DNA would be augmented. To cite an example, in case of RP1 and RP3 combination of primers, maintaining the annealing temperature at 58°C (4°C below the T_m of RP3) will not allow RP1 to bind with the template DNA because the T_m of RP1 is still 2°C below the annealing temperature. On the other hand if the temperature is lowered down to 54°C (2°C

below the T_m of RP1), RP3 can bind nonspecifically to the template DNA because now the annealing temperature is 8°C below its T_m value. Hence only two combinations of primers (RP1-RP4 and RP2-RP3) were used.

6.3.3.1 Inter IS amplification using RP2 and RP3 primers and the analysis of the DNA fragments

A total of eight strains (four wild and their respective colony morphology variants) viz. DK1, DK1S1, GBVI, GBVIS2, CMO II, CMOS1, CMI I and CMIS2 (symbol 'S' placed after mother strain names denotes spreading variant) were used in this experiment. PCR was conducted at 55°C annealing temperature. Once the PCR amplification was accomplished, the products were analyzed by running the agarose gel electrophoresis. The gel contained 2% agarose and the electrophoresis was performed at 50mV of electricity for about one hour. The separated DNA bands in the gel were observed and documented. With the help of the marker DNA molecule the size of each DNA fragment was manually determined and tabulated (Fig 5.25 and Table 5.2).

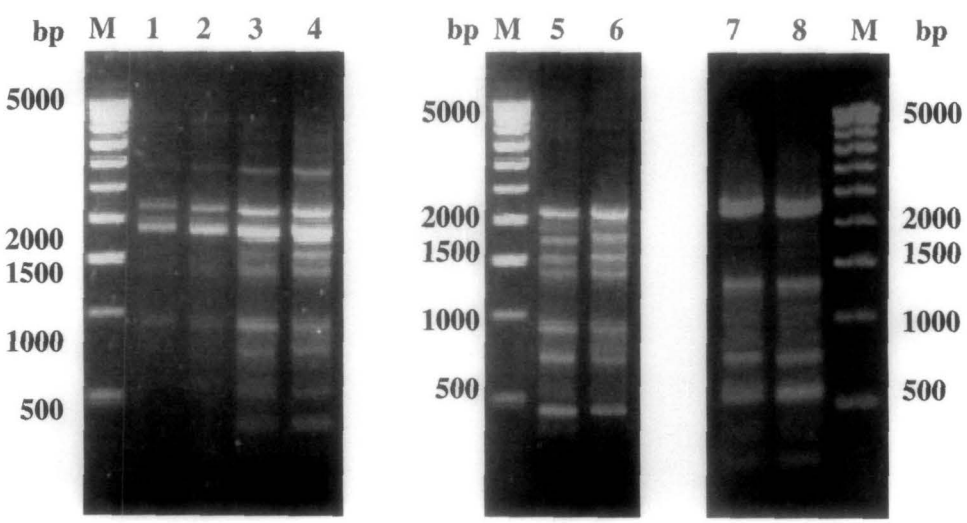

Fig 5.25: PCR amplification of inter-IS regions of different strains of *A. ferrooxidans* using RP2-RP3 combination of primers.
Lanes: M, 500 bp DNA ladder; **1**, DK1; **2**, DK1S1; **3**, CMO II; **4**, CMOS1; **5**, CMI I; **6**, CMIS2; **7**, GBVI; **8**, GBVIS2.

Table 5.2: Analysis of inter-IS DNA fragments (Kb) generated from wild and colony morphology variants strains of *A. ferrooxidans* using RP2-RP3 primers

	STRAINS							
Size (Kb)	DK1	DK1S1	GBVI	GBVIS2	CMO	CMOS2	CMI	CMIS2
2.8	√	√	-	-	√	√	√	√
2.1	√	√	√	√	√	√	√	√
1.8	√	√	-	-	√	√	√	√
1.5	-	-	-	-	√	√	√	√
1.4	-	-	-	-	√	√	√	√
1.1	-	-	√	√	-	-	-	-
0.9	-	-	-	-	√	√	√	√
0.8	-	-	-	-	√	√	√	√
0.7	-	-	√	√			-	-
0.5	-	-	√	√	√	√	-	-
0.4	-	-			√	√	√	√
0.3	-	-	√	√	-	-	-	-

'√' & '-' denote the presence and absence of the DNA fragments
DK1, GBVI, CMO and CMI are the wild type strains while DK1S1, GBVIS2, CMOS2 and CMIS2 are their colony morphology variants respectively.

Analysis of these amplified DNA fragments have revealed that some DNA band(s) are highly conserved among the strains. e.g. a DNA band of 2.1 kb has been found in all the strains tested irrespective of their origin (Table 5.2). There were eight DNA bands (2.8 kb, 2.1 kb, 1.8 kb, 1.5 kb, 1.4 kb 0.9 kb, 0.8 kb and 0.4 kb) present in all CM strains (CMO,CMI, CMOS1, and CMIS2) indicating that these inter-IS DNA fragments are highly conserved among the strains of coal mine. An additional amplicon of 0.5 kb present in both mother (CMO) and its variant (CMOS2) reveals the genotypic difference between CMO and CMI strains. However, the number of such conserved DNA bands was found to be lower in case of DK (DK1 and DK1S1) and GB (GBVI and GBVIS2) strains. FOFs so obtained with wild type and their colony morphology variants displayed high Dice coefficient. Hence it can be inferred that the bands are strain specific and can be efficiently used for typing *A. ferrooxidans* strains.

The common DNA bands existing in the strains can be considered as possible ancestral fragments and it may be assumed that the genomic reorganization of IS elements in different strains had originated from one or few common ancestral DNA sequences residing in a particular DNA locus at a time 0. Let n_A be the number of ancestral copies which is common in the strains tested, and n_{XI} be the number of additional DNA fragment(s) that have been generated by the movement of IS elements in strain X with the

total (N_X) of $n_A + n_{X1}$. Similarly for strain Y, N_Y will be equal to $n_A + n_{Y1}$. The number of identical amplified DNA band shared by these two strains will be equal to n_{XY}. Assuming that all the same sized DNA fragments generated by outwardly directed primers are those that remained unchanged from the common ancestor, the expected proportion of fragments that remain unchanged in both strains were calculated by the formula, $D = 2n_{XY} / (N_X+N_Y)$, which is known as Dice coefficient of similarity (Goldstein *et al.*, 1995; Nei and Li, 1979). Highest D value was obtained for the pairs between the wild type and their respective colony morphology variant/mutants (e.g between DK1 and DK1S1 or CMO and CMOS2 etc.) while the lowest D value was obtained for the pair between CMI and GB strains (Table 5.3). This indicated that GB strains are most distantly related with CMI strains compared to the others. On the other hand very low D value for these DNA bands obtained from these strains indirectly indicated the differential movement of the IS elements in the strains of *A. ferrooxidans.*

Table 5.3: Dice similarity coefficient (D) of different sample pairs

	DK1	DK1S1	GBVI	GBVIS2	CMO	CMOS2	CMI	CMIS2
DK1	1							
DK1S1	1	1						
GBVI	0.25	0.25	1					
GBVIS2	0.25	0.25	1	1				
CMO	0.17	0.17	0.29	0.29	1			
CMOS2	0.17	0.17	0.29	0.29	1	1		
CMI	0.18	0.18	0.15	0.15	0.88	0.88	1	
CMIS2	0.18	0.18	0.15	0.15	0.88	0.88	1	1

As was expected, the amplicons generated with RP2 and RP3 combination of primers amplifying DNA region flanked by IS*T2* on one side and IS*T445* on the other in a particular orientation, showed the different patterns of distribution of IS elements in different strains isolated from different locations. However, no unique band in the pairs between the wild and their colony morphology variant was found. Absence of polymorphic band(s) in the array of amlicons generated with primer pairs (RP2 and RP3) for the wild type and variants can not be taken for granted that rearrangement in the genome of CMVs due to movement of either of the two or both ISs did not occur at all. Under the set of PCR reaction condition, the chances of amplification of larger inter-IS regions have diminished blurring the prospect of identifying unique bands.

Therefore in the search of the unique band(s) another combination or primers was used in the next experiment.

6.3.3.2 Inter IS amplification using RP1 and RP4 primers and the analysis of the DNA fragments

For the primer combinations of RP1, and RP4, the annealing temperature was kept at 52°C. Similar to the generation of amplicons with RP2 and RP3 combination, a profile of amplicons (FAFs) were obtained from different strains of *A. ferrooxidans* (Fig 5.26). The sizes of the different fragments were determined and tabulated (Table 5.4). Unlike in RP2-RP3 combination, no DNA band of identical sizes could be identified. Dice coefficients for all the pairs were measured and it was found that the strains isolated from same location revealed greater similarities (Table 5.5). However, in case of DK1 and DK1S1, D value was not found to be 1.0 due to the occurrence of a unique band of 0.4 kb in DK1S1 (which was absent in its parent strain DK1) (Fig 5.26).

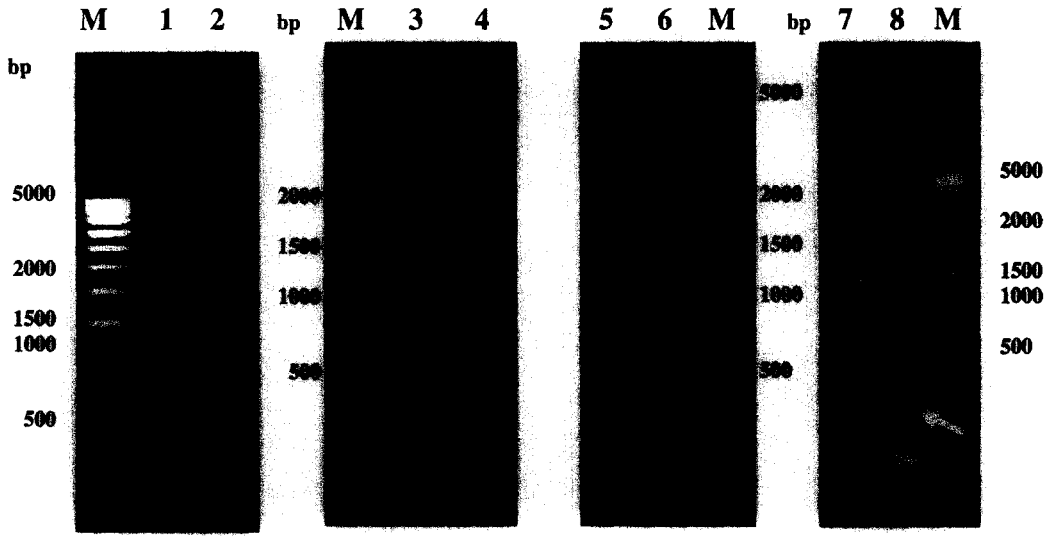

Fig 5.26: PCR amplification of inter-IS regions of different strains of *A. ferrooxidans* using RP1-RP4 combination of primers. Lanes: **M**, 500 bp DNA ladder; **1**, CMO II; **2**, CMOS2; **3**, GBVI; **4**, GBVIS2; **5**, CMI I; **6**, CMIS2; **7**, DK1; **8**, DK1S1.Unique band is shown by an arrow .

Amplification of unique DNA fragment of about 0.4 kb was repeatedly observed in DK1S1 but not in other colony morphology variants. The PCR product was then run in the low melting agarose gel electrophoresis and the DNA of the unique band was eluted from the gel. The DNA was then cloned in pGEM-T easy vector as described in the material and methods. The recombinant plasmid was named pAR1. Cloning was followed by the screening and sequencing.

Table 5.4: Analysis of inter-IS DNA fragments (Kb) generated from wild and colony morphology variants strains of *A. ferrooxidans* using RP1-RP4 primers

	STRAINS							
Size (Kb)	DK1	DK1S1	GBVI	GBVIS2	CMO	CMOS2	CMI	CMIS2
2.8	-	-	-	-	-	-	√	√
2.3	-	-	√	√	√	√	-	-
2.0	-	-	-	-	-	-	√	√
1.7	-	-	√	√	√	√	-	-
1.3	-	-	-	-	√	√	-	-
1.2	√	√	-	-	-	-	-	-
1.1	-	-	-	-	-	-	√	√
1.0	-	-	√	√	-	-	√	√
0.8	√	√	-	-	-	-	-	-
0.7	-	-	√	√	-	-	√	√
0.5	√	√	√	√	-	-	√	√
0.4	-	√	-	-	-	-	-	-

'√' & '-' denote the presence and absence of the DNA fragments

DK1, GBVI, CMO and CMI are the wild type strains while DK1S1, GBVIS2, CMOS2 and CMIS2 are their colony morphology variants respectively.

Table 5.5: Dice similarity coefficient (D) of different sample pairs

	DK1	DK1S1	GBVI	GBVIS2	CMO	CMOS2	CMI	CMIS2
DK1	1							
DK1S1	0.85	1						
GBVI	0.25	0.22	1					
GBVIS2	0.25	0.22	1	1				
CMO	0	0	0.5	0.5	1			
CMOS2	0	0	0.5	0.5	1	1		
CMI	0.22	0.2	0.54	0.54	0	0	1	
CMIS2	0.22	0.2	0.54	0.54	0	0	1	1

6.3.4 Analysis of pAR1 sequence

A sequence of 420 bp was obtained from DNA fragment cloned in pAR1 which contained inverted repeats of IS*T445* and IS*T2* at the terminal ends (Fig 5.27). The sequence has been deposited in the Genbank (Accession number of GQ305311). BlastN result showed 95% identity with the portion of *A. ferrooxidans* ATCC 53993 genome sequence, which featured the natural resistance-associated macrophage protein of *A. ferrooxidans* (Fig 5.28). However on careful analysis it was revealed that the first 20 bp and the last 20 bp sequences which were actually the part of inverted repeats of IS elements (designed as outwardly directed primers) followed by 4 bases ought to have been

target site duplicated direct repeat, indicated the disruption of the intact ORF due to transpositional event of IS*T2* and IS*T445*. Further analysis with BlastX (translated nucleotide sequence query versus protein sequence) showed 97% identities with natural-resistance macrophage resistance protein of *Acidithiobacillus ferrooxidans* ATCC 53993 (Fig 5.29).

```
5'
(RP1)
>>>>>>>>>>>>>>>>>>>>
ACCAGATTTGACTATAGCTCATCCTGCACATCCTGCGGCAAAGCAGCCTGGGCAGCATCGGGACCGCCCTGTTTGC
CCTGGGCCTCGTCGAGGCTGGGCTGATAGCGGCCATCGCCATTACCGCCAGCACTTCCTGGGCGGTGGGCGAGGCG
CTCAAACTGCCCCGCAGCCTGAATCTCAAACCGCAACGCGCACTGCCTTTTTATCTGTCCGGCATACTCAGCGCCG
CCATGGCCGCCATGGTGGTCCTGATTCCTCATATCCCCTTGGGCTTCCTGAATCTTACCGTCCAGGTGATTGCCTC
CATTTTCATGCCCGCCGCCATGCTATTCCTGCTCATGCTGCTCAATGACCGGGACATCATGGGAAGCTATATCAAC
CGGCCATGGCAGAATTGGTTGGCTCTTCGTCATTTTCAAG (RP4)
                      <<<<<<<<<<<<<<<<<<<< 3'
```

Fig 5.27: Nucleotide sequence of pAR1 containing the inverted repeats assigned to IS*T445* and IS*T2*. The primer sequences are highlighted by bold letters and the inverted repeats are underlined (5' end contains inverted repeats of IS*T2* and 3' end contains inverted repeats of IS*T445*.

```
PAR1         20      CATCCTGCACATCCTGCGGCAAAGCAGCCTGGGCAGCATCGGGACCGCCCTGTTTGCCCT  79
                     ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Aferr53993 147502    CATCCTGCGCATCCTGCGGCAAAGCAGCCTGGGCAGCATCGGGACCGCCCTGTTTGCCCT  147443

PAR1         80      GGGCCTCGTCGAGGCTGGGCTGATAGCGGCCATCGCCATTACCGCCAGCACTTCCTGGGC  139
                     ||||||||||||||| ||||||||  ||||||||||||||||||||||||||||||||||||
Aferr53993 147442    GGGCCTCGTCGAGGCAGGGCTGATTGCGGCCATCGCCATTACCGCCAGCACTTCCTGGGC  147383

PAR1         140     GGTGGGCGAGGCGCTCAAACTGCCCCGCAGCCTGAATCTCAAACCGCAACGCGCACTGCC  199
                     ||||||||||||||||||||||||||||||||||||||||||| |||||  |||||||||  ||||
Aferr53993 147382    GGTGGGCGAGGCGCTCAAACTGCCCCGCAGCCTGAATCTTAAACCACAACGCGCATTGCC  147323

PAR1         200     TTTTTATCTGTCCGGCATACTCAGCGCCGCCATGGCCGCCATGGTGGTCCTGATTCCTCA  259
                      |||||||||||||||||||||||||||||  ||||||||||||||||||  ||||||||  ||
Aferr53993 147322    CTTTTATCTGTCCGGCATACTCAGCGCCGGCATGGCCGCCATGGTGGTGCTGATTCCCCA  147263

PAR1         260     TATCCCCTTGGGCTTCCTGAATCTTACCGTCCAGGTGATTGCCTCCATTTTCATGCCCGC  319
                      ||||||  ||||  ||||||||  ||  ||||||||||||||||||||||||||||||||
Aferr53993 147262    CATCCCCCTGGGGTTCCTGAACCTGACCGTCCAGGTGATTGCCTCCATTTTCATGCCCGC  147203

PAR1       320       CGCCATGCTATTCCTGCTCATGCTGCTCAATGACCGGGACATCATGGGAAGCTATATCAA  379
                     |||||||||  ||  ||||||||||||||||||||||||||||||||||||||||||||||
Aferr53993 147202    CGCCATGCTGTTTCTGCTCATGCTGCTCAATGACCGGGACATCATGGGAAGCTATATCAA  147143

PAR1         380     CCGGCCATGGCAGAATT  396
                     |||||||||||||||||
Aferr53993 147142    TCGGCCATGGCAGAATT  147126
```

Fig 5.28: BlastN result of pAR1 showing 95 % identity with *A. ferrooxidans* ATCC 53993 genome sequence.

```
PAR1      36  RQSSLGSIGTALFALGLVEAGLiaaiaitastsWAVGEALKLPRSLNLKPQRALPFYLSG  215
              RQSSLGSIGTALFALGLVEAGLIAAIAITASTSWAVGEALKLPRSLNLKPQRALPFYLSG
A ferr   311  RQSSLGSIGTALFALGLVEAGLIAAIAITASTSWAVGEALKLPRSLNLKPQRALPFYLSG  370

PAR1     216  ILSAAMAAMVVLIPHIPLGFLNLTVQVIASIfmpaamlfllmllNDRDIMGSYINRPWQN  395
              ILSA MAAMVVLIPHIPLGFLNLTVQVIASIFMPAAMLFLLMLLNDRDIMGSYINRPWQN
A ferr   371  ILSAGMAAMVVLIPHIPLGFLNLTVQVIASIFMPAAMLFLLMLLNDRDIMGSYINRPWQN  430

PAR1     396  WLA  404
              + A
A ferr   431  YAA  433
```

Fig 5.29: BlastX result of pAR1 showing 97 % homology with the natural-resistance macrophage protein of *A. ferrooxidans* ATCC 53993.

Using ORF finder tool three ORFs coding for 132, 116 and 82 amino acids from +3, +2 and -2 frames for the sequence were obtained (Fig 5.30). The ORF containing 132 amino acids sequence showed 96% homology with the natural resistance-associated macrophage protein (Nramp) of *Acidithiobacillus ferrooxidans* ATCC 53993. The sequence was found to contain conserved domain for MntH transporter protein belonging to the Nramp superfamily (Fig 5.31). Other ORFs did not show any significant homology with any of the non-redundant GenBank CDS.

As the total genome sequence of *A. ferrooxidans* is now available in GenBank, total gene sequence coding for MntH protein was retrieved and analyzed. MntH gene of *A. ferrooxidans* was found to contain a sequence of 1371 bp in the '-' strand of *A. ferrooxidans* genome starting from 148416 to 147046. In the first step of the analysis, pair-wise alignment between the *A. ferrooxidans* MntH and cloned amplicon sequence was performed. The aligned sequence of *A. ferrooxidans* ATCC 53993 genome sequence with amplicon sequence from DK1S1 showed a certain degrees of homology in few bases with the inverted repeats of ISs in the terminal region (Fig 5.32). This result indicated the possibility of genomic rearrangement by insertion of IS due to recombimational event disrupting the MntH gene. Non specific annealing of primers leading to amplification of that DNA region has been ruled out by the PCR condition which would not allow the binding of primers with the DNA sequences having around 55% homology with two gaps in between. In addition to that, the same fragment of DNA under the same PCR condition could not be amplified from DK1 strain. Hence, the non-specific amplification of inter IS region was ruled out. Therefore the insertion of two ISs in the genome is very likely; or the presence of such homologous sequences could have led to the homologous recombination

of MntH gene sequence with the IS elements resulting into the integration of the IS elements into the gene.

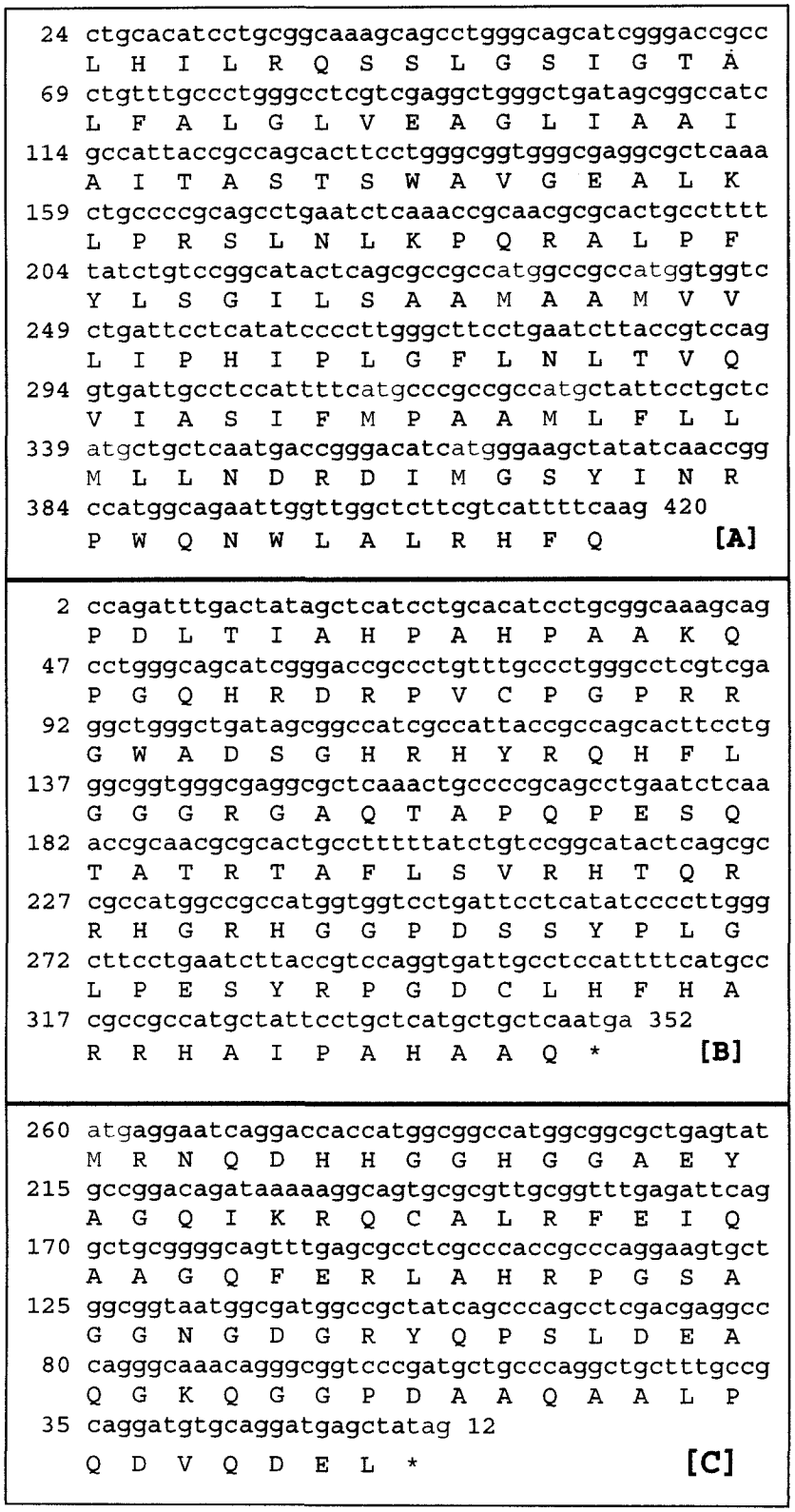

Fig 5.30: ORFs of pAR1 sequence obtained from **A**, +3 frame; **B**, +2 frame and **C**, -2 frame containing 132, 116 and 82 amino acids respectively.

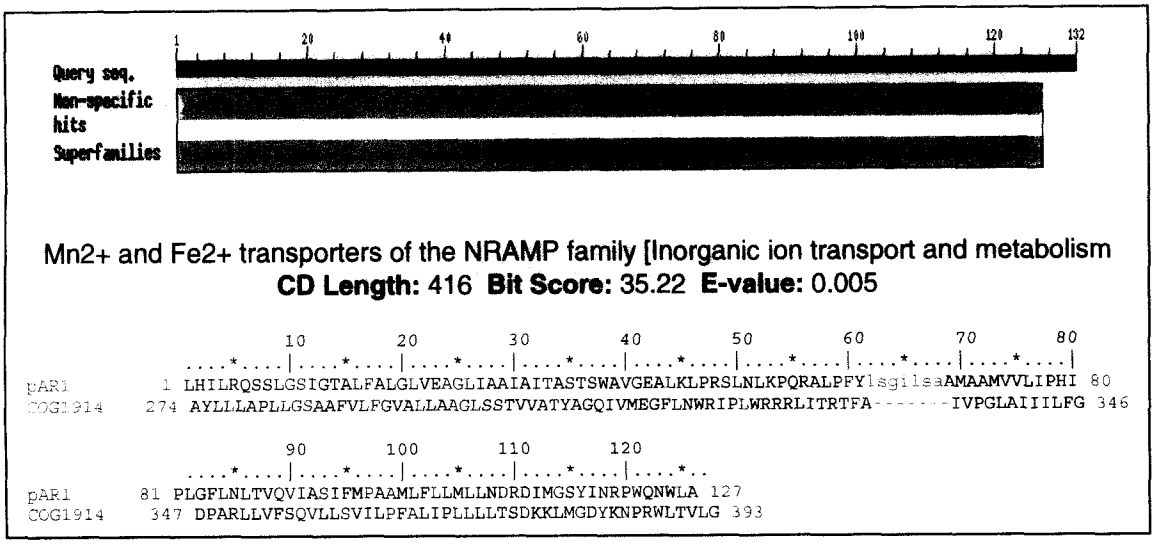

Fig 5.31: pAR1 sequence showing the conserved sequence for MntH of the Nramp family protein.

```
pAR1,          3 CAGATTTGA-CTATAGCTCATCCTGCACATCCTGCGGCAAAGCAGCCTGGGCAGCATCGG
Afer53,      898 CAACTTTGATATACAGGCCATCCTGCGCATCCTGCGGCAAAGCAGCCTGGGCAGCATCGG
                 **  *****  ** **  ******** **********************************

pAR1,         62 GACCGCCCTGTTTGCCCTGGGCCTCGTCGAGGCTGGGCTGATAGCGGCCATCGCCATTAC
Afer53,      958 GACCGCCCTGTTTGCCCTGGGCCTCGTCGAGGCAGGGCTGATTGCGGCCATCGCCATTAC
                 ********************************** ******** *****************

pAR1,        122 CGCCAGCACTTCCTGGGCGGTGGGCGAGGCGCTCAAACTGCCCCGCAGCCTGAATCTCAA
Afer53,     1018 CGCCAGCACTTCCTGGGCGGTGGGCGAGGCGCTCAAACTGCCCCGCAGCCTGAATCTTAA
                 ************************************************************ **

pAR1,        182 ACCGCAACGCGCACTGCCTTTTTATCTGTCCGGCATACTCAGCGCCGCCATGGCCGCCAT
Afer53,     1078 ACCACAACGCGCATTGCCCTTTTATCTGTCCGGCATACTCAGCGCCGGCATGGCCGCCAT
                 *** ********* **** ***************************** ************

pAR1,        242 GGTGGTCCTGATTCCTCATATCCCCTTGGGCTTCCTGAATCTTACCGTCCAGGTGATTGC
Afer53,     1138 GGTGGTGCTGATTCCCCACATCCCCCTGGGGTTCCTGAACCTGACCGTCCAGGTGATTGC
                 ****** ******** ** ****** **** ******** ** *****************

pAR1,        302 CTCCATTTTCATGCCCGCCGCCATGCTATTCCTGCTCATGCTGCTCAATGACCGGGACAT
Afer53,     1198 CTCCATTTTCATGCCCGCCGCCATGCTGTTTCTGCTCATGCTGCTCAATGACCGGGACAT
                 *************************** ** ******************************

pAR1,        362 CATGGGAAGCTATATCAACCGGCCATGGCAGAATTGGTTGGCTCTTCGTCATTTTC 417
Afer53,     1258 CATGGGAAGCTATATCAATCGGCCATGGCAGAATTACGCGGCA-TTCGCCATCGTC
                 ****************** ***************    ***  **** ***  **
```

Fig 5.32: Pairwise alignment of the pAR1 sequence with MntH gene of *A. ferrooxidans* ATCC 53993 showing 91.8% similarity. The primer sequences designed from inverted repeats are highlighted in bold letters.

MntH gene in the genome of *A. ferrooxidans* ATCC 53993 was thoroughly analyzed (Fig 5.33) and superimposed with inter-IS genomic sequence to locate the

position of insertion and predict the effect of such insertions. The promoter sequence containing both -10 and -35 regions along with the Shine-Dalgarno sequence could be detected in the upstream of the sequence. Interestingly, while looking for the transcription terminators at the downstream of the MntH gene, it was found that immediately after the stop codon of MntH gene, another initiation codon followed after a gap of 11 bp. The second gene of 198 bp encodes for a hypothetical protein of *A. ferrooxidans*. The third gene sequence starts after 17 bp from the stop codon of the second gene with an initiation codon. The third gene has been found to contain 702 bp long DNA sequence encoding for the MgtC protein responsible for the transport of Mg^{2+} ion. After the stop codon of MgtC gene a gap of 30 bp was encountered before the initiation codon of the last gene of this operon which coded for carbohydrate-sensitive Porin molecule. After the porin gene there is an intergeneric gap of 695 base pairs. The probable sequences for the hairpin formation that may lead to the transcription termination could be predicted at the downstream of the porin gene. This indicated that the MntH is the first gene to be transcribed of the novel transport operon. An opportunity to name this operon as MnMgCarpo operon (MntH, MgtC, Carbohydrate-sensitive porin) has been taken.

148802
agccgattca tctgatgtgc ccggggaatc gggccgcgtc caagtccctt gggcggattg
gatctgtcgt atcgggtgac cctctggatg gagccggtcg atttttctt gacgcgcttc
tgcgggatgg tgtaaaagct cgacatcatt ttggcgttag gaggcagcga ggcttggaca
gtcaccctag taggccccat gggggcaact cagggccctg gatttggggc ggcttggtca
agtaacgctt gcgccgccgt ctggtcctcg ggcggtgtga accatttttc cctattcaca
tgtgggttcc cggccttgct tctggcgaga cgggcggggt ggttgtgtcg aagtcgcgct
-35
ccattgttca ttGtggagga tgcgttatgt cggcaatacc ggaaacccct gcggcaccca
-10 TI SD IC 1
ccgaggcgga tctcctgcag catctgcccg aggatgcgcg tcagcgcatc caggatcgct
ggcgggttta ccaactgcgc catcatccct cctggtggcg gcgactcctg ctgtttttga
gtctgatcgg gccgggcatc ctcgtcatga tcgccgacaa cgatgccggc ggcgtcatca
cctacgccca gacaggggca atgtacggta tcggtttttt tatcccttc ctgttgctca
tgatcccggt ggcctatgtg gtgcaggaaa tgaccgtgcg tcttggtgca gtgacccatc
gcggtcacgc cgagatgatc tgggggcgtt atggcgcctt ttggggcgtg ttttccctgc
tcgatctcac cgtggccaac atcctgaccc tagtgaccga gttcatcggc atccgggtgg
gcatgagcat cttttccgtg ccgccgatca tttccgtacc ggtcgcttgg gtattcgtcg
cttccatcat gatctttttg cattacagca cctgggaacg attggcgctg tggatcgccg
ccggcaacat cattttgtc cccctggcca ttgccgcgca tccggactgg agccgggtcg
tggccgccgt tggcaactgg catattcccg ccggcgccgc cctgggtgcc tttacctatg
tcatcctggc caacctgggc accaccgtcg ccccctggat gctgtttttc gagcagtcct
ccgtcgtcga caagggcctg acggtcaatg acatccccag cggtcaggcg gacacggcca
ttggcagcct ctccatgggg atcatcgcca ttgccatcgt cgtgctgacg ggaacgctgg
tctatggaca cgcggacgcg gtcaactttg atatacaggc catcctgcgc atcctgcggc
aaagcagcct gggcagcatc gggaccgccc tgtttgccct gggcctcgtc gaggcagggc

tgattgcggc catcgccatt accgccagca cttcctgggc ggtgggcgag gcgctcaaac
tgccccgcag cctgaatctt aaaccacaac gcgcattgcc cttttatctg tccggcatac
tcagcgccgg catggccgcc atggtggtgc tgattcccca catccccctg gggttcctga
acctgaccgt ccaggtgatt gcctccattt tcatgcccgc cgccatgctg tttctgctca
tgctgctcaa tgaccgggac atcatgggaa gctatatcaa tcggccatgg cagaattacg
cggcattcgc catcgtcgga tttctcatcc tggccaacgg gctttatggc tataccgtgg
ttttcccttc ctcc**tga**gga tccttcc**atg** caccacaaag aagactttca tgcggcacca

SC 1 **IC 2**

acccaagatt ttctggcgtt tttcgaggag gaaacccatc tggaacccca acccctcacc
cgcgccccgg ttcggggatg gatccaggtc gccttctggg gcctgcgcat ctacatcgtg
gtgatgctgg tcatggtggc catcggcttc gcgcgcggga ttcat**tga**cc gggggatatc

SC 2

cgtc**atg**acg ccacaggtgc ccaacctcat tctcactgcg gaagacttgc tcctcgcctt

IC 3

tgtcctgggt gccctcattg gtattgagcg gcaatggcgt caacggactg cgggcctgcg
aaccaatacg ctggtggcgg tgggcgctgc catcttcgtg gacctggccc tgcgcatggg
cggatcggca gaagccatcc gggtagtcgc ttacgtcgtt tcgggtgtgg gctttttggg
cgcgggaacc atcatgaagg aaggggccaa tgtgcgcggc ctcaataccg ccgcgacctt
gtggggctcc gcggccgtcg gcgcctgcgc cggggcggag caatggctgc cggcgatcct
ggcggcgggc ttcgtgatca gtgccaacac cctcttgcgc ccagtggtga accgcatcaa
ccgggcgcca ctgagcgaga ccggcaccga actcaattat acattctctc tcatcgtgcc
ccgcggccag gaagcggcac tcctggatat cctcagcgaa aaactggaaa cagccggtta
tcccgtacgc gatttgcgca tgcacccctt cggacagagc gaaacggaaa tgcatgcgat
ccttttgccc agcagcgtcg atggcgaggt gctggatcgc ctgtccgcgc agttcagcca
gatgccccag gtgcgccagg ccttctggaa ccagagcgcc ggtgag**tga**c cggaggcaac

SC 3

aacccgagga gcaccaga**at** **g**aatcgattc ctgtcgtgcc gcaggttggc ctgctgggcc

IC 4

acgggcttgc tcatggtcct tccggcggtt tcggcgtccg ctgctccaac ggggattccg
gcagcgaccc acgcatcatt gtcctcccta cccctgtcgg ccgccttcct gcccgcctat
caacagtggt tcgatgggct agcctgggcg gcccatatcg acacggagac ggtatcgaat
ctggtcggcg gcatcgcccc gggaacggtc gcgatcctgg tgggtcacgc cggcttttcc
tatcacaccg cccgcgcagg tctctggacg ggcggcaagt tcacgctgaa cgtgctcggc
atcgagagcg ggaatccgga ggatagggta ggggatgtcc aaggcgtcag caatctgtat
gcgcgcagcc agtttcggct ctatcgtctg acgttccggc agcacctcgg accagccacc
ctgcgtcttg gactgatgag cgccaatgat tactttgacg tcaccgggca agcggcccaa
ctcatcaaca gctccttcgg catcaccccc accctgagcg gcaatatacc ggggatggcg
acctatccct actcggggct gggcgtcatg acggcggtcc atggtggtgc atgggaggcc
aaggcgggct ggtatcaggg tgaccccatt caccccttct ccgatccctt cggtcagggc
gatatggcgt tggcggaact cgattaccgc aacacgaagc aggagcgaac ctatgtcctc
aaactcggtg cctggcacta ccagcaacgc gcgatctaca acgccgtatt gggaccgcgc
accagcggcc tgtacgccgt ggccgaatac gccgaaccac ttccgcgcgg ccgcaggctc
ggcctcttcc tgcaggcggg ttccaccccg cagagcgtca accttattcc ttactatctg
ggggccggct tccgggtcaa ggggttgtta cccgggcgcg gggcggatgt attcagcatg
gggatgacgc gggcctggtt gcgcaacgca ccgtatccgg cggagaccgc ctgtgaaatc
acctatgccg cccaagtggt gccccacatc ttcttacaac cggatcttca atgggtcgtt
catcccggcg gtctctaccc caacgccctg gtggcgggcc tgagcttgca tgtgcagctt
ttt**tga**tgac gt***gcccttcc aggaaagggc*** tgctcgtaat ggcaggacaa tgtgacagcc

SC 4 ***Probable transcription terminator***

gcattgcgat catgtgaact agaaatggaa ttggcgcctg tgtggggagg gaacagggtc
accctcggcg tagcccgttg ttgcatgaac aagtgagtaa cgtggacgtg tccctggtgc
ccgagccaga atccgcaaaa gacgctcacg gctattacca tgagtatctt cagggaaatg
cgaaagcggc gcagcacgta ccacagcccc gatcaggaag gccggtggcg atcctttgcg
gattgctgcc accagtagac ggctttgtgg gtattctggg gtacgccctc tccgaagttg
taggcaaagg cgaggttgaa cgcgccctgg gcattcctgg gcattcccgg ccaaagctgc
cttacgctta cggtaccaga ccgcggcttt tttccagtga gtgtggtcag ccaaattgat

ggtcctctgg tcaacttatt tgaggggcag acggcggcaa taaggccacg cgacccggtt
tcgtgaacag gtggcgtggc tatggactga gcagatcctc acaacggagg atcgtcgtca
gtttgttggc gccttttta acaaccactt gttcatcaga actcaggcag agcttccgca
actcttcccg cacggtattc catcagttga cgagagacga tc**atg**cggcc cctgcgggaa
IC 5
gtaattccgt cttaatccag ctaagtgtat cgcggacgct taccgagaaa tcgcgcgcca
tccaacccag ttcgcgttcc gcacgggcgc tcgtaaagtg cgcctcgcga tggccgtact
ccttgacgaa cgcacag 144046

Fig 5.33: *A. ferrooxidans* ATCC 53993 gene showing the operon of four genes (MnMgCarpo operon). The sequence in pink color shows the region which is homologous to the pAR1 sequence. Probable transcriptional terminator region is underlined and italicized **TI**: Transcription initiation point; **SD**: Shine-Dalgarno sequence; **IC:** Initiation codon; **SC:** Stop codon.

MntH gene spans a length of 1371 bp in the genome of *A. ferrooxidans* ATCC 53993 (Fig 5.33). Critical analysis of the MntH gene revealed that the positions of the insertion of IS*T2* and IS*T445* elements to be 914 bp and 1291 respectively from the initiation codon of the gene (Fig 5.34 and Fig 5.35). It was also found that the insertion of IS*T2* at the 914 bp position results into the generation of a new non sense codon at 960 bp from the initiation codon (Fig 5.36) suggesting that the insertion of IS*T2* element causes the disruption of MntH gene and probably the MnMgCarpo operon.

ttgtgtcgaagtcgcgctccattgttcattGtggaggatgcgtt**atg**tcggcaataccggaaaccccctgcgg
Promoter **SD** **Initiation codon of MntH**
cacccaccgaggcggatctcctgcagcatctgcccgaggatgcgcgtcagcgcatccaggatcgctggcgggt
ttaccaactgcgccatcatccctcctggtggcggcgactcctgctgtttttgagtctgatcgggccgggcatc
ctcgtcatgatcgccgacaacgatgccggcggcgtcatcacctacgcccagacaggggcaatgtacggtatcg
gtttttttatccccttcctgttgctcatgatcccggtggcctatgtggtgcaggaaatgaccgtgcgtcttgg
tgcagtgacccatcgcggtcacgccgagatgatctgggggcgttatggcgccttttggggcgtgttttccctg
ctcgatctcaccgtggccaacatcctgaccctagtgaccgagttcatcggcatccgggtgggcatgagcatct
tttccgtgccgccgatcatttccgtaccggtcgcttgggtattcgtcgcttccatcatgatctttttgcatta
cagcacctgggaacgattggcgctgtggatcgccgccggcaacatcatttttgtccccctggccattgccgcg
catccggactggagccgggtcgtggccgccgttggcaactggcatattcccgccggcgccgccctgggtgcct
ttacctatgtcatcctggccaacctgggcaccaccgtcgccccctggatgctgtttttcgagcagtcctccgt
cgtcgacaagggcctgacggtcaatgacatccccagcggtcaggcggacacggccattggcagcctctccatg
gggatcatcgccattgccatcgtcgtgctgacgggaacgctggtctatggacacgcggacgcggtcaactttg
atatacaggc**gagctatgctcgaaagtg**gtgtatgggcgcgttacaactggct**tga**atgaggtggcgtactgt
LIR of IST2 **Newly generated stop codon**
tgctacaaagggtttccgactcaaagagcaacataaggagtacgccgtggaaaagaataccgtaatagcaggt
gggatgggagagttgggtctgggcattgaagggatacttcgacgcgcggcactctgcattgagcaggccatag
aggcagaggtggcggtattgctggaagaatttgcgacggtgcggatggttgatgggcgtcaggcggtcgtgcg

taatgggcatctgccggagcgcgagatcatgaccgctctgggtccggtacccgtcaaagtacccaaggtgcgg
gaccgtcaggatcggggatcaaattcaattcggtactggcgcctccgtatgtacgcaaatcacgaacggtagc
cgctacagtaccttggctctatctgcatggggtgtcttccggccacatgcaggaagccctttccattttgctg
ggtgatgaggccaagggactttcgcctgcggtgttgggacgtctcaaggcggagtgggcgcaagagtatgccc
attggcaacgccgctcctacaggaaagcgctatgcttactggtgggtagacggtatttatacgaacctccgtg
cggagaggatccgcgtatctgcctgctggtgattatcggcgtgacggcagagggcaagaaagagctggtcatg
gtcagtgacggcctgcgcgaatccaaagcttcctggctggagatcctgcgtgacctgcaggcgcgcggtctgg
agacggccccttgctcgccattggggatggtgccatggggttttgggccgcactggatgaagcctatcccga
aactggtcagcaacgctgctgggtgcataagaccgccaacatcctcaacgaacttcccaaagcccaacagagc
aaagccaaggcagcgttgcaggaaatctggatggccgccaatcgccaggctgcggaaaaagcactggacgtgt
ttgtgcgcaattaccaggcaaagtatcccaaggctgtggctaaactggaaaaagatcgggctgaattgctcgc
tttctatgattttccggccgaacactggcggcatatccggaccaccaatgccattgagtccaccttcgctacg
gtacgccaccggactacccggacgaagaactgtgtatcacgcagcagctttctgggattgggtttcaagatgc
tgcagcaggctgaaaaacgctggattgggatttatgctccagagaaagtcctgcagctttttgcaggggtgaa
atttatcgatggcataccggctaacctcaccctgccggatgatcaacagaccgccgcctgatctatgtcagaa
aatgctcatac**accagatttgactatagctcC**atcctgcgcatcctgcggcaaagcagcctgggcagcatcg
RIR of IST2
ggaccgccctgtttgccctgggcctcgtcgaggcagggctgattgcggccatcgccattaccgccagcacttc
ctgggcggtgggcgaggcgctcaaactgccccgcagcctgaatcttaaaccacaacgcgcattgccctttat
ctgtccggcatactcagcgccggcatggccgccatggtggtgctgattccccacatccccctggggttcctga
acctgaccgtccaggtgattgcctccattttcatgcccgccgccatgctgtttctgctcatgctgctcaatga
ccgggacatcatgggaagctatatcaatcggccatggcagaatt<u>GGTT</u>**ggctcttcgtcattttcaag**tgg
TD LIR of IST445
gtagcctgagatccagcttacccaggatcaggtaggccatgttgataaaattcttgtggtcggtaaccccggg
ccttggccttggcggattgattgaggctgttgaagccttccaagaattccattggtaatctggcttttcgaac
cagcggaagcaacggccatcccaattgttcaatggaatggtgtaggcgaccttgaccataggcggcattccgc
tgtccttggcgttttccatccaggccttgagcagggttggcgccctggtgcgattcttgactgtgaagaatgc
ctgaaaagtcagcgaaactggtaggcctgtgccgttttcaggttctggtcttttaataattcctgcagctttt
ctttctgcctggccttgaggttacaatcattcttgagccagagccagcgggtctttttgagatttggctgggt
gaaggacttccccttgcgcacggcatccacggcctcgttgaacgagcttcatgaggtgaaaacgattcgaag
gtgatctcagcattgggcaggtgcttcggcaagcccctttctggaaggccggcgagagggccatgctcacaat
tcggttgatgccttccgcgctaccaccatgggcctgtagatcttcggagaatttctcaaaggtcttggcatcc
ttgccgggcgtacggaacaagagtcgccgggcattcagatccacgaagagcgtgattgtagtcatgaccgcgc
cgactgcttggttctcaattcgacgccgaaccgggcatggacatttggccatatccaccctggtacgggcttc
gggcacataatggtcaatcacccgcacaggagcgtgtcggctcgccgaccagcgggctaccgtcaatactggc
atctccgaccagggtcagcggcgtgggataggacttcatgctgaaataaaccatgtgccgccggtatagttcc
agggtgcatgtaccggacactcctcaccacatacgggcaagcaaagcgactgcccttgggaaagttgatgtgc
agatccaggcgcttctcctccacttggaaggtcacattccgtccaccaaccaccggcggtaccaaccctaacg
cgagagaaaacagatcttctgggaccatcagctaactcctattcgggacccaactgctccacagcattctacc
ccctacccc**actcaatccgcagaagagcc**acgcggcattcgccatcgtcggatttctcatcctggccaacgggc
RIR of IST445
tttatggctataccgtggttttcccttcctcc**<u>tga</u>**
Stop codon of MntH

Fig 5.34: MntH gene sequence showing disruption by the integration of IS*T445* and IS*T2* elements within it. Promoter region, initiation codon and stop codon are underlined. Sequence in blue color is the superimposed region of pAR1.RIR: Right inverted repeat; LIR: Left inverted repeat; SD: Shine-Dalgarno sequence; TD: Probable target duplication.

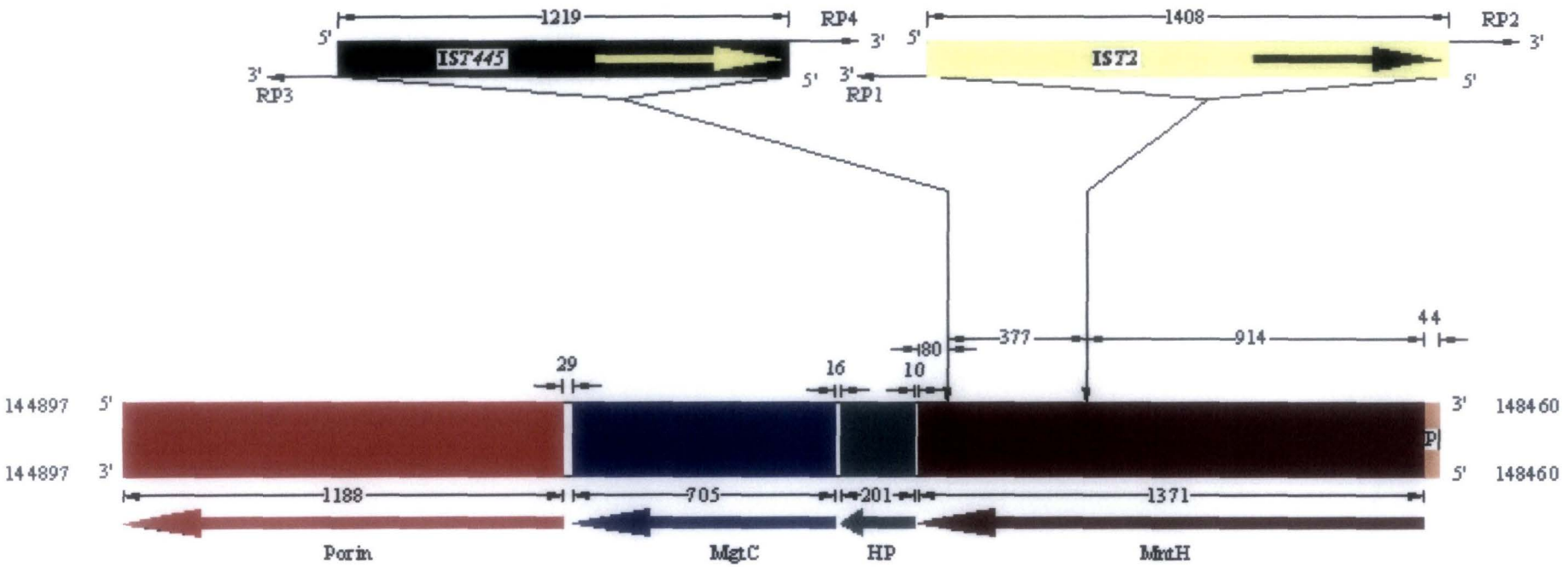

Fig 5.35: Diagrammatic representation of the positions of insertion of IS*445* and IS*2* in *A. ferrooxidans* operon containing MntH, Hypothetical protein (HP), MgtC and Carbohydrate sensitive-Porin molecule genes lying in the *A. ferrooxidans* ATCC 53993 genome spanning 144897-148460 region. Bigger arrows show the orientations of genes and small arrows show the direction of primers (RP1, RP2, Rp3 and RP4). P stands for Promoter. The numerical digits represents the size of the genes in bp.

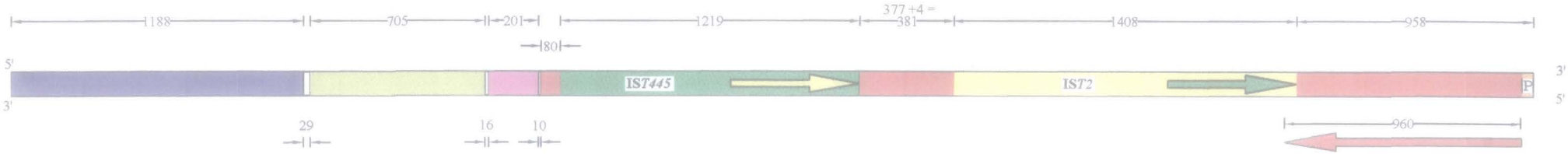

Fig 5.36: Schematic representation of the truncated MntH gene structure (shown in red colour) formed due to the insertion of IS *T2* and IS*T445* elements and premature termination of MntH gene due to generation of a new stop codon at 960 bp position.

6.4 DISCUSSION

From the above results it was found that *A. ferrooxidans* showed the presence of repeated elements (IS*T*445 and IS*T*2) in the genome (Fig 5.4 and 5.6). The molecular characterization of the organisms was also determined.

In addition to IS*T*445 and IS*T*2 two other genes viz. Glutamyl tRNA reductase and Pirin was amplified by the primers designed to amplify IS elements. Almost a total sequence of glutamyl tRNA reductase gene (pAR2) was obtained which showed the conserved domains of NADB_Rossmann superfamily during BlastP analysis. Glutamyl tRNA reductase catalyzes the conversion of glutamyl-tRNA to glutamate-1-semialdehyde, initiating the synthesis of tetrapyrrole. The sequence has been analyzed and compared with the reference molecule obtained from SwissProt protein database. The reference molecule 1GPJA isolated from *Methanoparus kandleri* (Moser *et al.*, 2001) showed a high similarity with the glutamyl tRNA reductase obtained from Garubathan strain. This was further substantiated by the 3D structures obtained with the help of software Spdb-Viewer and POV-Ray which showed the two structures to be highly comparable to one another except that the pAR2 sequence is truncated at its N-terminal end (Fig 5.12 and Fig 5.13). The point of interest here is that the pAR2 sequence was amplified with the internal primers (A_1 and $A_{2)}$ which were designed from the conserved inverted repeat sequences of IS*T445*. Therefore the two ends of the pAR2 sequence had the exact sequences as that of the inverted repeats of IS*T445* (Fig 5.8). Nevertheless the translated sequence yielded a protein that showed 98.8% identity with glutamyl tRNA reductase gene of *A. ferrooxidans* and 37% identity with 1gpj_a from *Methanoparus kandleri*. Again by using G_3 and G_4 internal primers, in addition to IS*T*2 (which was desired) another DNA segment was amplified, of which one part of the sequence was found to code for the Pirin gene of *A. ferrooxidans* while other part of the sequence did not match with any gene sequences available in the GenBank.

IS elements in bacterial genomes can be a significant source of spontaneous mutation. By inserting into coding regions they can disrupt a gene, while an increase in gene expression may occur when an element carrying an outwardly directed promoter inserts upstream of a gene (Schneider and Lenski, 2004). Different copies of the same IS element may also form the substrate for recombination and result in chromosomal rearrangements such as deletions, duplications or inversions (Gray, 2000). The role of IS

elements in evolution has not been clearly elucidated. A school has viewed them as parasites in the genome having potential deleterious effect on their host, while the opposing view emphasizes beneficial effects and their improvement in host adaptation (Richle *et al.*, 2001). IS elements can ever be seen as adaptive mutator genes, since the mutations they mediate can occur at increased rates particularly during adverse conditions such as starvation or oxygen stress (Naas *et al.*, 1994). These studies often involved simple homogenous environments with conditions that either allowed regular growth or caused starvation. There has been studies that compared IS- mediated mutations under growth and starvation conditions (Hall, 1999).

In the present work, it was sought to detect and characterize IS-mediated mutations in populations of *Acidothiobacillus ferrooxidans* DK1 that are adapting to thiosulphate-agar growing condition in the laboratory. *A. ferrooxidans* obtains energy either from the oxidation of ferrous ions to ferric or reduced- sulfur compounds to sulfuric acid. Most isolates of *A. ferrooxidans* have remarkably modest nutritional requirements. Aeration of a sample of iron pyrite in acidified water is sufficient to support growth at the expense of pyrite. The pyrite provides energy source and **low pH** provides the growth environment. When attempts were taken to grow *A. ferrooxidans* on solid medium hurdles appeared in providing the acidified growth environment. Agar gets hydrolyzed in presence of acid, and hydrolyzed carbon compounds are toxic to *A. ferrooxidans*. Hence the heterotrophs come into picture to rescue and help colonize autotrophic *A. ferrooxidans* in the milieu created by the acidophilic heterotroph on the agar surface. It has been shown in this study that thiosulfate inhibits the growth of acidophilic heterotroph *Acidiphilium*; therefore, presumably the heterotrophic rescuers in the putatively pure culture of *A. ferrooxidans* fails to provide the growth environment free of toxic organics (produced by acid hydrolysis of agar) in thiosulfate agar medium for growth and manifestation of *A. ferrooxidans* colonies. The condition that prevails in thiosulfate agar is a serious stress to the population of *A. ferrooxidans* seeded on it.

This work has attempted to reason and look into the role of ISs in enabling a fraction of the population to emerge out successfully in colonizing on thiosulfate. Variations in IS copy number and position among related *A. ferrooxidans* suggested that the elements are mobile. The basic interest of this work was to investigate the effect of stress due to the presence of thiosulfate in acidic agar medium, as earlier authors have

suggested that some adverse conditions activate IS elements (Schrader and Holmes, 1988; Chakraborty *et al.*, 2002), thereby perhaps helping their host to adapt to those adverse conditions. Basically, stressful conditions can enhance the transposition rate of IS elements, but also the selection of IS-mediated mutations, and prevailing approaches can not distinguish between the possibilities. In this study, we have adopted a novel strategy in fishing out the rearranged DNA segment in the genome that appears in the variants due to activity of IS elements. Bypassing the classical southern hybridization to locate additional IS bands followed by cloning and sequence of the IS flanked DNA region to know the mutations in a particular gene, we looked in for inter IS DNA regions i.e. the DNA regions spanned by two IS elements. In a given genome, before any fresh insertion or deletion events, the inter IS genomic regions should be fixed; on movement of the ISs, there will be either a formation of a new inter IS region {because of the earlier location of an IS and insertion of another copy in its vicinity (left or right) or insertion of two new IS copies in an unaffected region }or loss of an existing inter IS region {due to deletional event of one of the existing IS which bracketed the inter IS region}. The PCR strategy of using outwardly directed primers from IS*T2* and IS*T445* in pairs enabled to identify one such novel inter IS band in the colony morphology variant of DK1 which appeared due to movement of IS within the host cell.

In the post-genomic, gene identification and genome organization analysis allow initial predictions to be made as to whether genes are linked in possible functional units. This information can then be used to reconstruct metabolic pathways "in silico" and to begin to unravel the often complicated and multilevel regulation of cellular functions. Missing enzymes and potential bottlenecks can be identified and central control modes of regulation can be pinpointed. The present study has opened up another opportunity to map IS insertion mutations. We traced IS-mediated mutations in *A. ferrooxidans* DK1 and the unique inter-IS genomic region of DKIS1 (colony morphology variant of DK1 strain) was cloned and sequenced which matched with part of the Mn^{2+} transporter gene. This chapter described a simple method of mapping IS-mediated mutations, which identified a DNA region essential for transport of Manganese, Magnesium and carbohydrate. The Natural Resistance Associated Macrophage Protein (N RAMP) class of Mn^{2+} transporter gene of the "Mn Mg Carpo Operon" (as we named it) got disrupted by insertion of IS*T2* and IS*T 445*.

The NRAMP class of Mn^{2+} transporter also depends on membrane potential and not on ATP (Saier, 2000). It is found virtually in all bacterial types, and also in yeast, plants and animals.It is involved in highly competitive battle over Mn^{2+} by bacteria living alone in external environments or in mixed populations with other microbes, or within plant or animal cells. These are members of a large class of membrane transporters called 'ABC' transporters, but at other times (and for heterotrophic microbes) for uptake of amino acids or sugars (Saier, 2000). Due to insertional mutagenic effect in the first gene of the Mn Mg Carpo Operon, the downstream genes would also be affected due to polar effect. The third cistron of the operon codes for carbohydrate porin. Proteins called 'porins' are present in the outer membrane of gram negative bacteria, and these proteins function as channels for the entrance and exit of hydrophilic low molecular weight substances. Several porins have now been identified, and both specific and non-specific classes are known. Some porins are highly specific because they contain a specific binding site for one or more substances. Porins are transmembrane proteins and associate to form small membrane holes (~ 1nm in diameter). Apparently a mechanism exists for opening and closing the pores. The carbohydrate-selective porin Opr B family includes for example the *Pseudomonas aeruginosa* Porin B, a substrate-selective channel for a variety of different sugars. This protein may facilitate the diffusion of variety of diverse compounds, but is restrictive to carbohydrates and does facilitate glucose fusion across the outer membrane (Wylie and Worobee, 1995). Using interposon mutagenesis, strains of *P. aeruginosa* were generated which lack or over express the substrate-selective OprB porin of that species.A marked decrease or increase in the initial uptake of glucose by those strains verified the role of OprB in facilitating the diffusion of glucose across the outer membrane of *P. aeruginosa*. It was also demonstrated that the loss or over expression of OprB had a similar effect on the uptake of three other sugars- mannitol, glycerol and fructose to support the growth of bacteria. In our case, if the expression of the carbohydrate-porin of *A. ferrooxidans* is severely affected due to polar mutation the entry of carbohydrates will be severely affected.

To an obligate autotrophic organism like *A. ferrooxidans* this would not severely affect the growth; but will offer a selective advantage in an environment like thiosulfate-agar (acidified) medium; where entry of agar hydrolyzing carbohydrates (toxic to *A. ferrooxidans*) would be prevented. The colony morphology variant DKS1 will therefore, spread in response to the gradient of thiosulfate for enhanced mobilization of thiosulfate

into the cell for deriving energy. In an earlier study it was reported that spreading variants of *A. ferrooxidans* demonstrated greater chemotactic activity toward thiosulfate (Chakraborty *et al.*, 2002). There may be cross talk operating between chemotoxins and cation transport. Under the condition of the disruption of Mn^{2+}gene (as in DKS1), the cells will be driven towards higher gradient of ions for facilitating diffusion of cations through less specific cation transport system. The genotypic characterization of the colony morphology variant was the most important objective that the chapter has attempted to address.

6.5 CONCLUSION

1. *Acidithiobacillus ferrooxidans* strains isolated from AMD samples of Garubathan from Darjeeling Himalayas produced unique fingerprints of amplified fragments (FAFs).The technique of using outwardly directed primers of insertion sequences in pair-wise combination to amplify inter IS DNA fragments produced unique strain specific profile. This profile can be efficiently used for typing *A. ferrooxidans* strains/ strain identification.

2. In addition to the amplification of IS*T2* and IS*T445* with internal primer specific to each element, two other DNA fragments that co-amplified were cloned, sequenced and analyzed using different bioinformatics tools. The one that co-amplified with IS*T445* was found to be a truncated glutamyl-tRNA reductase gene at N-terminal end (as inferred after being compared with the intact sequence derived from whole genome sequence). The 3D model of the protein was constructed and compared with the ideal model. *A. ferrooxidans* protein was found to be highly similar to the representative molecule of glutamyl-tRNA reductase. Another DNA fragment that co-amplified with IS*T2* was cloned, sequenced and analyzed. A part of the sequence was homologous to Pirin molecule of *Acidithiobacillus ferrooxidans.*

3. Amplification of inter-IS regions using a novel PCR based technique (using outwardly directed primers of IS*T*445 and IS*T*2) could be successfully reproduced in tracking genomic regions affected by the movement of ISs under cultural stress condition.

4. When inter IS DNA banding profiles between the wild type (DK1) and its colony morphology variant (DK1S1) was analyzed, one unique polymorphic inter-IS band appeared in DK1S1. It was sequenced and analyzed using various bioinformatics tools and *in-silico* superimposed on *A. ferrooxidans* while genome sequence to map mutations due to IS insertions in an operon (which was named as MnMgCarpo operon).

5. The possible effect of the disruption of MntH gene of the MnMgCarpo operon leading to the advantage of the colony morphology variant, DK1S1, on thiosulfate agar medium was discussed in details

6.6 REFERENCES

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

Berg, D. E. and Howe, M. M. (ed.) (1989). Mobile DNA. *American Society for Microbiology*, Washington, D C.

Besemer, J. and Herpers, M. (1977). Suppression of polarity of insertion mutations within the *gal* operon of *Escherichia coli. Mol. Gen. Genet.* **151**: 295-304.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic. Acid. Res.* **7**: 1513-1523.

Blot, M. (1994). Transposable elements and adaptation of host bactria. *Genetica.* **93**:5-12.

Bongers, R. S., Hoefnagel, M. H. N., Starrenburg, M. J. C., Siemerink, M. A. J., Arends, J. G. A., *et al.*, (2003). IS*981*-mediated adaptive evolution recovers lactate production by IdhB transcription activation in a lactate dehydrogenase-deficient strain of *Lactococcus lactis. J. Bacteriol.* **185**: 4499-4507.

Bukhari, A. I., Shapiro, J. A. and Adhya, S. L. (1977): DNA insertion elements, plasmids, and episomes. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY.

Cabrejos, M. E., Zhao, H. L., Guacucano, M., Bueno, S., Levican, G., Garcia, E., Jedlicki, E. and Holmes, D. S. (1999). IST1 insertional inactivation of the *resB* gene implications for phenotypic switching in *Thiobacillus ferrooxidans. FEMS Microbiol. Lett.* **175**: 223-229.

Cadiz, R., Gaete, L., Jedlicki, E., Yates, J., Holmes, D. S. and Orellana, O. (1994). Transposition of IST2 in *Thiobacillus ferrooxidans. Mol. Microbiol.* **12**: 165-170.

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

Chakraborty, R., Deb, C., Lohia, A. and Roy, P. (1997). Cloning and characterization of high-copy-number novel insertion sequence from chemolithotrophic *Thiobacillus ferrooxidans. Plasmid* **38**: 129-134.

Chakraborty, R., Singh, A., Lahiri, C., Deb, C. and Roy, P. (2002). Colony morphology mutants of chemolithotrophic *Acidithiobacillus ferrooxidans* are associated with altered genomic distribution of family 1 repetitive DNA sequence. *Curr. Sci.* **8**: 1009-1014.

Chandler, M. and Mahillon, J. (2002). Insertion sequences revisited. *In* N.I. Craig, R Craigie, M Gellert, A M Lambowitz (edn.), *Mobile DNA II*, American Society of Microbiology, Washington D C.

Chao, L., Vargas, C., Spear, B. B. and Cox, E. C. (1983). Transposable elements as mutator genes in evolution. *Nature* **303**: 633-635.

Charlseworth, B., Sniegowsky, P. and Stephen, W. (1994). The evolutionary dynamics of repeatitive DNA in eukaryotes. *Nature* **371**: 215-220.

Cooper, V. S. and Lenski, R. E. (2000). The population genetics of ecological specialization in evolving *E. coli* population. *Nature* **407**: 736-739.

Cooper, V. S., Schneider, D, Blot, M. and Lenski, R. E. (2001). Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* **183**: 2834-2841.

Cullum, J. and Broda, P. (1979). Chromosome transfer and Hfr formation of F in *rec*$^+$ and *recA* strains of *Escherichia coli* K12. *Plasmid* **2**: 358-365.

Davis, L. G., D-Dibner, M. and Battey, J. F. (1987). 'Basic Methods in Molecular Biology'. Elsevier, New York, Amsterdam, London.

Davison, N., Deonier, R. C., Hu, S. and Ohtsubo, E. (1975). Electron microscope heteroduplex studis of sequence relations among plasmids of *Escherichia coli*. The DNA sequence organization of F and F primes and the sequences involved in Hfr formation. pp. 56-65. *In* Schlessinger (ed.). *Microbiology*. American Society of Microbiology.

de Visser, J. A. G. M., Akkermans, A. D .L., Hoekstra, R. F. and de Vos. W. M. (2004). Insertion-sequence-mediated mutations isolated during adaptation to growth and starvation in *Lactococcus lactis*. *Genetics* **168**: 1145-1157.

Derbyshire, K. M., Kramer, M. and Grindley, N. D. (1990): Role of instability in the cis action of the insertion sequence IS903 transposase. *Proc. Natl. Acad. Sci.*, USA. **87**: 4048-4052.

Elena, S. and Lenski, R. F. (2003). Evolution experiments withmicroorganisms: the dynamics and genetic bases of adaptation. *Natl. Rev. Genet.* **4**: 457-469.

Farabaugh, P. J., Schmeissner, U., Hofer, M. and Miller, J. H. (1978). Genetic studies of the *lac* repressor.VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J. *Mol. Biol.* **126**: 847-863.

Finkel. S. E. and Kolter, R. (1999). Evolution of microbial diversity during prolonged starvation. *Proc. Natl. Acad. Sci. USA*, **96**: 9721-9726.

Galas, D. J. and Chandler, M. (1989). Bacterial insertion sequences. p109-162. *In* D E Berg and MM Howe (ed.). *Mobile DNA*. American Society for Microbiology, Washington D C.

Goldstein, R., Sun, R., Jiang, R., Sajjan, U., Forstner, J. F. and Campenelli, C. (1995). Structurally variant classes of pilus appendages fibers coexpressed from *Burkholdria (Pseudomonas) cepacia. J. Bacteriol.* **177**: 1039-1052.

Gray, Y. H. M. (2000). It takes two transposons to tango. *Trends Genet.* **16**: 461-468.

Hall, B. G. (1999). Spectra of spontaneous growth-dependent and adaptive mutations at *ebgR. J. Bacteriol.* **181**: 1149-1155.

Holmes, D. S. and Haq, R. Ul. (1989). Adaptation of *Thiobacillus ferrooxidans* for industrial applications, p 115-127. *In* J. Salley, R G L McCready, and P L Wichlacz (ed.), *Biohydrometallurgy-1989*. CANMET, Ottawa, Canada.

Holmes, D. S., Zhao, H. L., Levican, G., Ratouchniak, J., Bonnefoy, V.,Varela, P. and Jedlicki, E. (2001). ISAfe1, an ISL3 family insertion sequence from *Acidithiobacillyus ferrooxidans* ATCC 19859. *J. Bacteriol.* **183**(14): 4323-4329.

Imhof, M. and Schlotterer, C. (2001). Fitness efeects of advantageous mutations in evolving *Escherichia coli* population. *Proc. Natl. Acad. Sci., USA.* **98**: 181-187.

Jordan, E., Saedler, H. and Starling, P. (1967). Strong polar mutations in the transferase gene of the galactose operon in *Escherichia coi. Mol.Gen. Genet.* **100**: 296-306.

Kitamura, K., Torii, Y., Matsuoka, C. and Yamamoto, K. (1995). DNA sequence changes in mutation in th *tonB* gene on the chromosome of *Escherichia coli*K-12: Insertion elements dominate the mutational spectra. *Jpn. J. Genet.* **70**: 35-46.

Lane, D. J. (1991). 16S/23S rRNA sequencing. pp. 115-175. *In* (ed.) *by* E. Stackebrandt & M. Goodfellow. *Nucleic Acid Techniques in Bacterial Systematics*. Chichester: Wiley.

Lenski, R. E. (2004). Phenotype and genomic evolution during a 20,000-generation esperiment with thebactrium *Escherichia coli. Plant Breed Rev.* **24**: 225-265.

Lenski, R. E., Rose, M. R., Simpson, S. C. and Tadler, S. C. (1991). Long-term experimental evolution in *Escherichia coli.* I. Adaptation and divergence during 2000 generations. *Am. Nat.* **138**: 1315-1341.

Mahillon, J. and Chandler, M. (1998): Insertion Sequences. Microbiol. *Mol. Biol. Rev.* **62**: 725-774.

Makris, J. C., Nordmann, P. L. and Reznikoff, W. S. (1988): Mutational analysis of insertion sequence 50(IS50) and transposon 5 (Tn5) ends. *Proc. Natl. Acad. Sci. USA*, **85**: 2224-2228.

Marchler-Bauer, A. and Bryant, S. H. (2004). "CD-search : protein domain annotations on the fly". *Nucleic. Acids. Res.* **32** : W327-W331.

Marmur, J. (1961). A procedure for isolation of deoxyribonucleic acid from microorganism. *J. Mol. Biol.* **3**: 208-218

Moser, J., Schubert, W. T., Beier, V., Bringemeier, I., Jahn, D. and Heinz, D. W. (2001). V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. *EMBO J.* **20**: 6583-6590.

Moxon, E. R., Rainey, P. B., Nowak, M. A. and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci inpathogenic bacteria. *Curr. Biol.* **4**: 24-33.

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

Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.*, USA. **76**: 5269-5273.

Nevers, P. and Saedler, H. (1977). Transoposable genetic elements as agents of instability and chromosomal rearrangements. *Nature* **268**: 109-115.

Ohtsubo, E., Zenilman, M. and Ohtsubo, H. (1980). Plasmids containing insertion elements are potential transposons. *Proc. Natl. Acad. Sci. USA.* **77**: 750-754.

Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics* **163**: 1519-1526.

Otto, S. P. (2002). Evolving beyond point mutations. *Trends Ecol. Evol.* **17**: 110.

Papadopoulos, D, Schneider, D., Meier-Eiss, J. Arber, W., Lenski, R. E. and Blot, M. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci.* USA, **96**: 3807-3812.

Peterson, P. A., Ghosal, D., Sommer, H. and Saedler, H. (1979). Development of a system useful for studying formation of unstable alleles of IS2. *Mol. Gen. Genet.* **173**: 15-21.

Rainey, P. B., Buckling, A., Kassen, R. and Travisano, M. (2000). The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends Ecol. Evol.* **15**: 243-247.

Reif, H. J. and Saedler, H. (1975). IS*1* is involved in deletion formation in the *gal* region of *E. coli* K-12. *Mol. Gen. Genet.* **137**: 17-28.

Reimmann, C., Moore, R., Little, S., Savioz, A., Willets, N. S. and Haas, D. (1989). Genetic structure, function and regulation of the transposable elements IS21. *Mol. Gen. Genet.* **215**: 416-424.

Richle, M. M., Bennet, A. F. and Long, A.D. (2001). Genetic architecture of thermal adaptation in *Escherichia coli. Proc. Natl. Acad. Sci.USA.* **98**: 525-530.

Rosner, J. L. and Guyer, M. S. (1980). Transpositin of IS*1*-lambda-B10-IS*1*- from a bacteriophage lambda derivative carrying the IS*1*-*cat*-IS*1* transposon (Tn9). *Mol. Gen. Genet.* **178**: 111-120.

Rownd, R., Miki, T., Appelbaum, E. R., Miller, J. R., Finkelstein, M. and Bartock, C. R. (1978). Dissociation, amplification, and reassociation of composite R-plasmid DNA. pp. 138-146. *In* (ed.) D Schlessinger, *Microbiology,* American Society for Microbioogy, USA.

Rozen, D. E., De Visser, J. A. G. M. and Gerrish, P. J. (2002). Fitness effects of fixed beneficial mutations in microbial populations. *Curr. Biol.* **12**: 1040-1045.

Saier, M H> Jr. (2000). A functional-phylogenetic system for transmembrane solute transfer. *Microbiol. Mol. Biol. Rev.* **64**: 354-411.

Saedler, H. and Heiβ, B. (1973). Multiple copies of the insertion – DNA sequences IS*1* and IS*2* in the chromosome of *Escherichia coli* K 12. *Mol. Gen. Genet.* **122:** 267-277.

Saedler, H., Reif, H. J., Hu, S. and Davidson, N. (1974). IS2, a genetic element for turn-off and turn-on of gene activity in *Escherichia coli*, *Mol. Gen. Genet.* **132**: 265-289.

Sambrook, J., Fritsch, E. F. and. Maniatis, T. (1989): "Molecular Cloning: A Laboratory Manual", *2nd edition,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schneider, D. and Lenski, R. E. (2004). Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res. Microbiol.* **155**: 319-327.

Schneider, D., Duperchy, E., Coursange, E., Lenski, R. E. and Blot, M. (2000). Long term experimental evolution in *Escherichia coli*IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics.* **156**: 477-488.

Schrader, J. A. and Holmes, D. S. (1988). Phenotypic switching of *Thiobacillus ferrooxidans. J. Bacteriol.* **170**(9): 3915-3923.

Siguier, P. Filee, J. and Chandler, M. (2006). Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* **9**(5): 526-531.

Starlinger, P. (1980). IS elements and transposons. *Plasmid.* **3**: 241-259.

Starlinger, P. and Saedler, H. (1976). IS-elements in microorganisms. *Curr. top. Microbiol. Immunol.* **75**: 111-152.

Travisano, M. (2001). Evolution: towards a genetical theory of adaptation. *Curr. Biol.* **11**: R440-R442.

Turlan, C. and Chandler, M. (1995). IS1-mediated intramolecular rearrangements: formation of excised transposon cirlcles and replicative deletions. *EMBO J.* **14**: 5410-5421.

Valdes, J., Pedroso, I., Quatrini, R., Dodson, R. J., Tettelin, H., Blake, (II) R., Eisen, J. A. and Holmes, D. S. (2008). *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics.* **9**: 597.

Wery, J., Hidayat, B., Kieboom, J. and de Bont, J. A. M. (2002). An insertion sequence prepares *Pseudomonas putida* S12 for severe solvent stress. *J. Biol. Chem.* **276**: 5700-5706.

Wylie, J. L and Worobee, E. A. (1995). The OprB porin plays a central role in carbohydrate uptake in *Pseudomonas aeruginosa. J. Bacteriol.* **177**: 3021-3026.

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). IS*T2*: An insertion sequence from *Thiobacillus ferrooxidans. Proc. Natl. Acad. Sci. USA*, **85**: 7284-7287.

Zerbib, D., Prentki, P., Gamas, P., Freund, E., Galas, D. J. and Chandler, M. (1990). Functional organization of the end of IS1: specific binding site for an IS1- encoded protein. *Mol Microbiol.* **4**:1477-1486.

Zhao, H. and Holmes, D. S. (1993). Insertion sequence IST1 and associated phenotypic switching in *Thiobacillus ferrooxidans*. p. 667-671. *In* A. E. Torma, M. L. Apel, and C. L. Brierly (ed.), *Biohydrometalurgical technologies, vol. 2*. TMS Press, Warrendale, Pa.

Zinser, E. R., Schneider, D., Blot, M. and Kolter, R. (2003). Bacterial evolution through the selective loss of beneficial genes: Trade-offs in expression involving two loci. *Genetics* **164**: 1271-1277.